ANALYSIS OF THE HUMAN MICROTUBULE INTERACTOME USING MULTIDIMENSIONAL CHROMATOGRAPHY AND TANDEM MASS SPECTROMETRY

A dissertation presented

by

CHAO GONG

to

The Department of Chemistry at Brown University

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Chemistry

Brown University
Providence, Rhode Island

May 2013

This dissertation by Chao Gong is accepted in its present form

by the Department of Chemistry as satisfying the

dissertation requirement for the degree of Doctor of Philosophy.

Date _____ _____

Professor Carthene R. Bazemore-Walker, Advisor

Recommended to the Graduate Council

Date _____ _____

Professor Shouheng Sun, Reader

Date _____ _____

Professor Arthur R. Salomon, Reader

Approved by the Graduate Council

Date _____ _____

Professor Peter M. Weber,
Dean of the Graduate School

Curriculum Vitae

Chao Gong was born in Nanjing, Jiangsu Province of China on January 5th of 1982. He went to Nanjing University for undergraduate study since 2000 and graduated with a B.Sc. in Chemistry in 2004. Chao joined Brown University in August, 2006 and worked in Prof. Carthene Bazemore-Walker's group for his Ph.D. in Chemistry. His research interest is in the fields of proteomics, including LC-MS method development and quantitative analysis of proteins/peptides, especially for the microtubule associated proteome.

Publications:

1. **Gong, C.**; Bazemore-Walker, C. R. "High Resolution RP/RP-MS/MS for Analysis of the Human Microtubule Interactome," to be submitted.
2. **Gong, C.**; Bazemore-Walker, C. R. "Proteomic Characterization of Human Microtubule-Associated Interactome," in preparation for submission.
3. **Gong, C.**; Bazemore-Walker, C.R. "Comprehensive characterization of tubulin isotypes and identification of novel post-translational modifications," in preparation for submission.
4. **Gong, C.**; Bazemore-Walker, C. R. "Quantitative Proteomics Using Reductive Isotopic Diethylation and 2D RP-RP LC-MS/MS: Application to the Microtubule-

Associated Proteome". in preparation for submission.

5. Ellisor, M. B., Hogan, M. J., **Gong, C.**; Bazemore-Walker, C. R. "Quantitative Proteomics of MCF-7 Breast Cancer Cells," to be submitted.
6. Miao, JW., Zhou, J., Song, GH., Fan, YN., **Gong, C.** "Chromium Oxide supported on mesoporous SBA-15 as catalysts for oxidative dehydrogenation of propane by CO2," Chin. J. Inorg. Chem., 2005, 21(10), 1541-1545.

ACKNOWLEDGEMENTS

There are many people I would like to express my sincere appreciation for their support and help during my Ph.D. study. First of all, I would like to thank my advisor, Prof. Carthene Bazemore-Walker. I appreciate all her contributions of time, ideas, and teaching to me about how to think, talk and perform experiments like a scientist. She is a great mentor to provide insightful guidance and large freedom to explore my own research ideas. I am also thankful for the rigorous scholarship attitude she interprets.

Thanks also go to my committee members, Prof. Shouheng Sun and Prof. Arthur Solomon. Prof. Sun is a great scientist and kind person who had helped me several times during my graduate study. Prof. Solomon had given valuable suggestion and instrument support during the collaboration with my lab.

Many thanks to all Bazemore-Walker group members, especially to Yuan Cao and Hongbo Gu who had discussed with me and we were helping each other with individual projects. In addition, I would like to thank past members of Solomon group, Kebing Yu, Lulu Cao and Vinh Nguyen, for their patience and efforts on training me.

Finally I would like to thank my significant other, Jingyu Liu, who is always by my side sharing the vicissitudes of life. It’s your involvement and support that can make me achieve further success. I must thank my dear parents, for their understanding, endless patience and unconditional love to me.

Abstract of “Analysis of The Human Microtubule Interactome by Multidimensional Chromatography and Tandem Mass Spectrometry”, By Chao Gong, Ph.D., Brown University, May 2013.

Recent developments in mass spectrometer technology has increased the sensitivity and speed to an unprecedented level, along with the continuing improvements of high performance liquid chromatography (HPLC); they are combined as the most powerful tool for analysis of proteins, which are usually at low concentration levels in complicated biological samples. Often the challenge of shotgun proteomic analysis is related to the high complexity of the sample and limited instrument duty-cycle, which requires high resolution separation of components within the sample prior to mass spectrometer analysis. Aside from qualitative information of protein identifications from proteomic study, quantitative information is of increasing demand for the biological understanding of cellular functions or potential biomarkers.

We seek to investigate two dimensional reversed phase HPLC (2D-RP-HPLC) to provide high efficiency separation of peptides for proteomic study of complex biological systems. The utility of this method is illustrated for the characterization of the human microtubule interacting proteome, which is essential for the understanding of anti-tumor drug effects. In order to quantify expression changes of microtubule interacting proteins after drug treatment of cancer cells, we developed reductive diethylation strategy to

isotopically label peptide amine groups for the differentiation of control samples and drug-treated samples.

Our results from the study of 2D-RP-HPLC peptide separation demonstrate its superior resolving power and selective separation mechanism for the comprehensive characterization of the microtubule interactome. Moreover, the reductive diethylation strategy exhibits high efficiency for stable isotopic labeling of peptides and no retention shift behavior of labeled peptide pairs even after 2D-RP-HPLC separation. Combining these two techniques provides an inexpensive and high-throughput workflow for quantitative proteomics. Finally, we identified a few protein candidates from the microtubule interactome that were up-regulated or down-regulated with response to a cancer therapeutic agent using this method.

Table of Contents

List of Figures

List of Tables

Chapter 1

GENERAL INTRODUCTION

1.1 Mass Spectrometry Analysis for Quantitative Proteomics

Recent advances in the study of biological systems benefit from the complete DNA sequencing of many genomes, including human. However, biological functions inside living creatures are usually fulfilled at proteins' level rather than genes. Recent studies suggest that there is no good correlation between mRNA expression and protein quantity [1, 2]. In addition, gene mutation and post-translational modifications (PTMs) would both affect the translated protein for its downstream cellular activity. "Proteomics" is the study of the entire complement of proteins expressed in a specific state of an organism or a cell population [3]. Thus, the object of proteomics is always a protein mixture containing hundreds of proteins from an isolated biological complex or even more than 10,000 different proteins from mammalian cell cultures. Furthermore, the quantity of different proteins within the study varies based on their expression levels. The ultimate goal of proteomics is to identify and quantify all proteins, as well as their PTMs on the corresponding amino acid sites.

With the development of mass spectrometry (MS) instrumentation, MS-based proteomics has become the most powerful tool for systems biology owing to the superior sensitivity and fast duty-cycle for high throughput analysis. General shotgun proteomics uses mass spectrometry to analyze peptides derived from enzymatic digestion of the protein mixture. After obtaining mass to charge (m/z) information of the peptide ions in the MS spectrum, collision induced dissociation (CID) is used to break up the peptide backbone at different sites between amino acids (shown in Figure 1.1). The m/z values of

fragment ions are recorded in the resulting MS/MS spectra. Recorded data are processed by software searching against protein databases to match experimental MS/MS spectra with theoretical MS/MS spectra of a particular peptide sequence derived from a parent protein in the database. This is called a spectra peptide match (SPM). In this way, the existence of peptides and corresponding proteins in the analyzed sample are determined in shotgun proteomics.

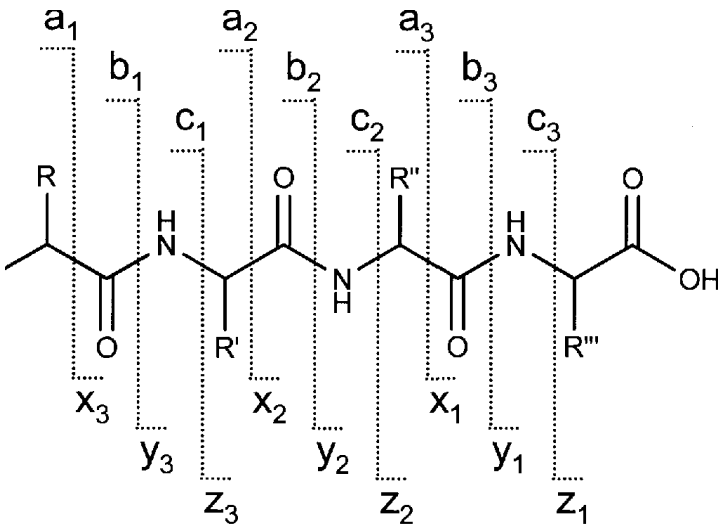


Figure 1.1. Backbone fragment of peptide ions by CID and nomenclature of fragment ions [4]. Peptide fragment ions after CID are labeled consecutively from the original amino terminus. Charged N-terminal fragments are categorized as a, b, c ions and charged C-terminal fragments are categorized as x, y, z ions. The a_m, b_m, c_m and x_m, y_m, z_m ions are defined according to the bond it breaks and m represents the number of amino acids these fragment ions contain.

Mass spectrometers usually consist of the following parts: the ionization source, the mass analyzer, and the data processing electronics. Major mass analyzers includes: ion trap (IT), orbitrap, ion cyclotron resonance (ICR), quadrupoles (Q), and time-of-flight

(TOF). Performance comparison of mass spectrometer instruments utilized in latter chapters is summarized in Table 1.1.

Table 1.1. Performance comparison of some mass spectrometer instruments [5].

Instrument	Resolution	Mass accuracy	Sensitivity	Dynamic range	Scan rate
LIT (LTQ)	2000	100 ppm	Femtomole	1e4	Fast
LTQ- Orbitrap	100,000	2 ppm	Femtomole	1e4	Moderate
LTQ-FTICR	500,000	<2 ppm	Femtomole	1e4	Slow, slow
Q-TOF	10,000	2–5 ppm	Attomole	1e6	Moderate, fast

LIT (LTQ from Thermo) stands for linear ion trap mass analyzer; FTICR stands for Fourier transform ion cyclotron resonance mass analyzer; Q-TOF stands for quadrupole time-of-flight hybrid mass spectrometer.

Most of the early studies in proteomics emphasized identifying as many proteins as possible. However, to better understand the function of a protein, especially in case it is a potential biomarker that would be differentially expressed in normal or diseased cells, quantitative protein profiling is essential for proteomics. For protein quantification, the requirement for data quality, in terms of quantifiable information content, is higher than that necessary for protein identification only. In general, due to the dynamic range of protein expressions within biological samples and instrument limitations from mass spectrometric technology, higher abundance proteins have a better chance to be identified

and further quantified while lower abundance proteins would be a challenge even for identification (shown in Figure 1.2).

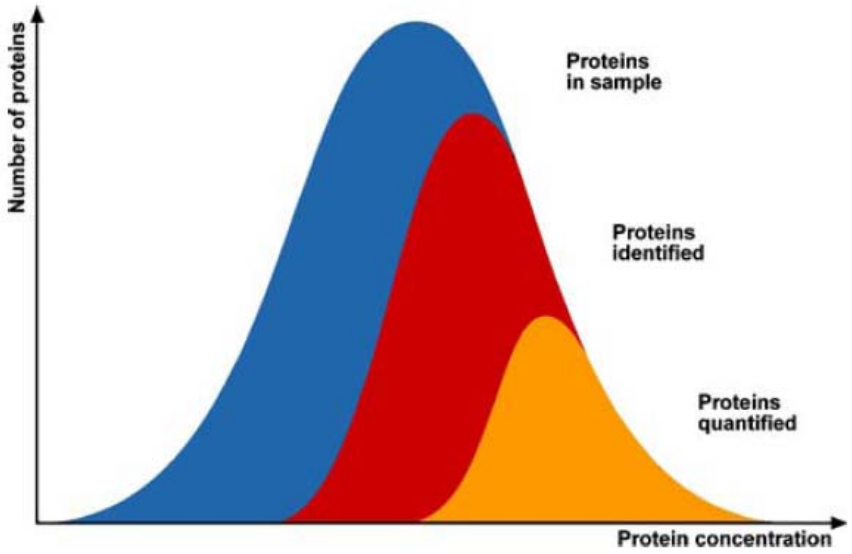


Figure 1.2. Schematic representation of the fraction of a proteome that can be identified or quantified by mass spectrometry based proteomics [6]. Cellular proteins span a wide range of expression and only a fraction of all these proteins in a sample are typically identified by current mass spectrometric technologies. Due to limited data quality, only a fraction of all identified proteins can also be reliably quantified.

1.2 Methodologies for Quantitative Proteomics

1.2.1 Spectral Counting

This label-free quantitation method is based on counting the number of spectrum peptide matches (SPMs) associated with a specific protein. The assumption of this methodology is that the more abundant a protein is within the sample, the more abundant the peptides derived from this protein would be in the digested mixture and the more

likely these peptides will be selected for MS/MS analysis. Spectral counting has been shown to have a linear dynamic range of 2 orders of magnitude for six protein markers in the background of yeast digest [7]. Further study using yeast culture in media of different nitrogen isotopes (^{14}N and ^{15}N) and with incorporation of statistical analysis, the result showed proteome-wide model-based assessment of differential expression by spectral counting [8].

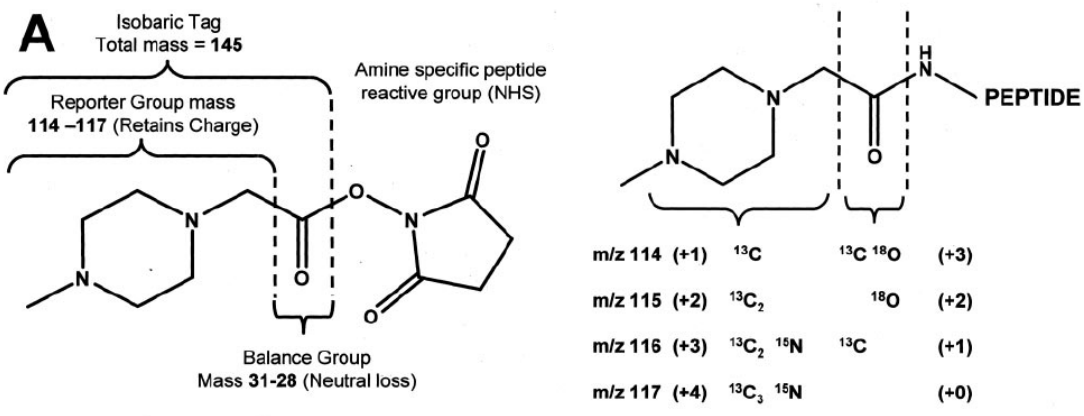
One disadvantage to spectral counting is that high-abundance proteins can mask low-abundance proteins [9] and it assumes that the linearity of response is the same for every protein. However, differences in terms of ionization efficiency among peptides derived from different proteins would render spectral counting method a biased result. In order to circumvent this issue, absolute protein expression profiling (APEX) [10, 11] was developed to incorporate machine learning techniques for accounting differences in peptide physical properties with the purpose of improving spectral counting accuracy. In a study on yeast and *Escherichia coli* using APEX, the dynamic range was found to be 2–4 orders of magnitude and the results were in agreement with Western blot and 2D gel studies [11].

The other disadvantage associated with label-free spectral counting is the limitation for sample multiplexing since each sample has to be analyzed individually. This would not only increase sample handling and instrument running time, but also require highly reproducible experimental factors (such as sample loading, multidimensional peptide separation, electrospray ionization (ESI) performance, etc) for analyzing complicated

samples. For these reasons, the label-free spectral counting method is generally considered inferior in quantification accuracy compared to stable isotope labeling method [12].

1.2.2 Isobaric Labeling

Unlike the label-free method, isobaric labeling attaches a mass tag to the reactive amine groups of a peptide (ie. N-terminal and lysine side chain). Currently, the iTRAQ (isobaric tag for relative and absolute quantification) labels [13] from ABSciex and the TMT (tandem mass tag) labels [14] from Thermo Fisher are commercially available tagging technologies. The mechanism of isobaric labeling is illustrated for iTRAQ reagents in Figure 1.3, TMT labels use similar principles. In general, the introduced mass tags have the same mass, and are composed of a reporter group (114 to 117 Da), a balance group (31 to 28 Da) and a amino-reactive group. During the MS scan, peptides with same sequence but coming from differentially labeled samples would emerge as a



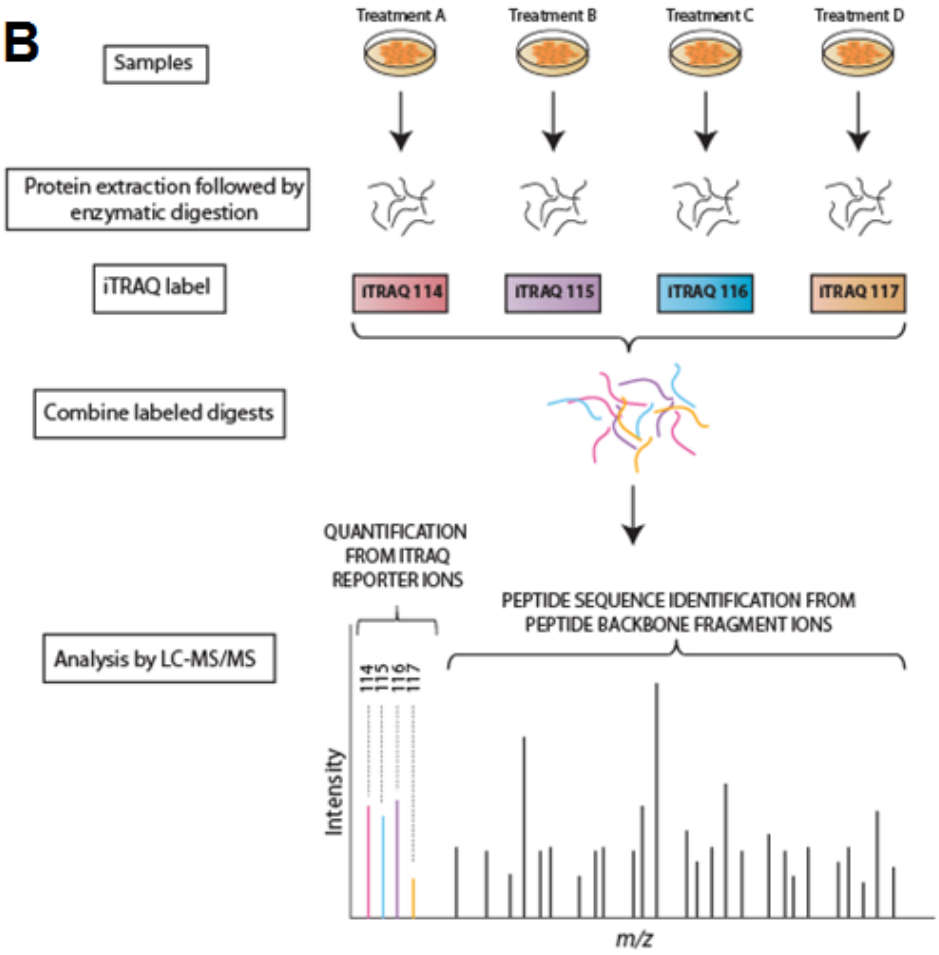


Figure 1.3. Illustration of iTRAQ isobaric labeling strategy [13]. (A) Structure and components of iTRAQ isobaric tag. (B) Schematic experiment flow of isobaric quantification using iTRAQ (4-plex) labeling tags.The iTRAQ technology utilizes isobaric reagents to label the primary amines of peptides. For 4-plex iTRAQ reagents, each consist of an N-methyl piperazine reporter group with mass=114, 115, 116, or 117, a balance group with mass=31, 30, 29, or 28, and an N-hydroxy succinimide ester group that is reactive with the primary amines of peptides (A). The function of balance group in each of the iTRAQ reagents is to make the same peptides from different samples to have isobaric (identical) mass after labeling. The key steps in iTRAQ quantification experiment are shown in schematic workflow (B). Samples for quantification are prepared under various treatment conditions and followed by cell lysis to extract proteins. Each protein sample is then digested using enzyme (such as trypsin) to generate proteolytic peptides. Each peptide digest is labeled with a specific iTRAQ reagent and then the labeled digests are combined into one sample mixture. The combined peptide mixture is analyzed by LC-MS/MS for both identification and quantification. The relative quantity of a peptide among the treated samples is determined by comparing the intensities of reporter ion signals in the MS/MS scan.

single MS peak. During CID fragmentation, the reporter group would be released from labeled peptides thus allowing quantitative comparison between different samples based on reporter ion intensity. With iTRAQ or TMT tags, the labeling is done at the peptide level; and theoretically, peptides from a tryptic digest should all be labeled.

The obvious advantage of isobaric labeling is the ability to multiplex up to eight samples using iTRAQ 8-plex kits in parallel within a single LC-MS/MS analysis to compare the quantity change of a specific protein with different treatments. However, these labeling reagents are expensive and are not stable during long-term storage. In addition, the labeling protocol and follow-up sample cleaning procedure before MS analysis are complicated and should be conducted with extreme care.

1.2.3 Stable Isotope Labeling

The difference between isobaric labeling and stable isotope labeling for quantitative proteomics is that the former introduces isobaric tags of equal mass and quantifies during MS/MS mode while the latter labels peptides with heavy stable isotopes (hydrogen and deuterium, ^{12}C and ^{13}C, ^{14}N and ^{15}N, ^{16}O and ^{18}O) among different samples and quantifies under MS mode. The concept of stable isotope labeling is based on the assumption that the abundance of a peptide is reflected by its extracted ion chromatogram (XIC) as shown in Figure 1.4, thus comparing the XIC of light-labeled and heavy-labeled forms of peptide would provide relative quantification information.

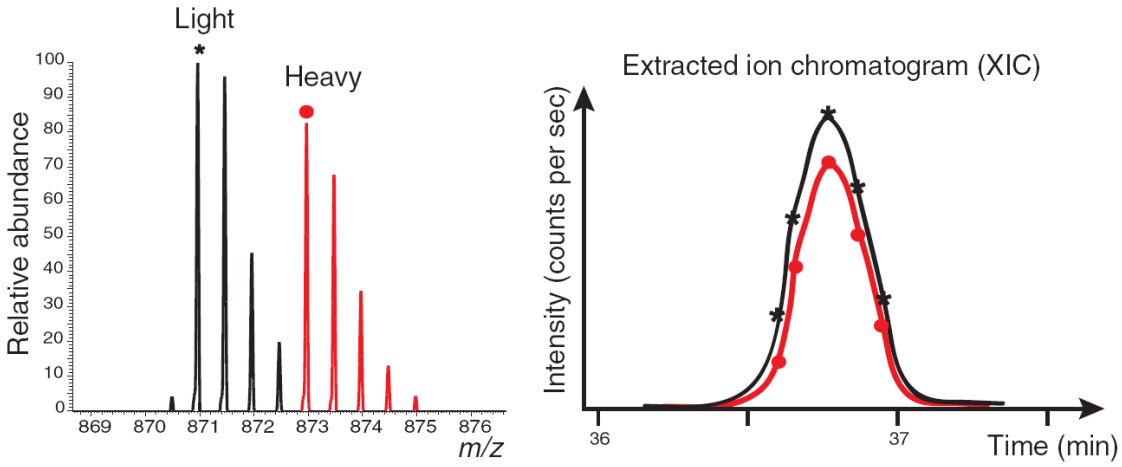


Figure 1.4. Extracting quantitative information from XIC of mass spectra data [15]. During a typical LC-MS/MS analysis, the MS signal (in the left panel) for a particular peptide ion is sampled several times within its elution profile, forming the XIC (in the right panel). When a doubly charged peptide is labeled with isotopic reagents with mass shift of 4Da, their m/z difference would be 2Da in the MS mode, thus allowing quantification comparison between the light-labeled and heavy-labeled forms based on corresponding XIC area.

Stable isotope labeling can be mainly classified into three categories by the ways it introduces mass shift into peptides: 1. metabolical; 2. chemical; 3. enzymatical. Each method has its strong points and weakness and is explained below.

i) Metabolical labeling

The earliest possible point for introducing a stable isotope signature into peptides is to metabolically label all proteins during cell growth and division. The most popular metabolical labeling method called stable isotope labeling by amino acids in cell culture (SILAC) approach, was introduced by Mann and co-workers in 2002 [16]. The principle of SILAC method [17] is illustrated in Figure 1.5.

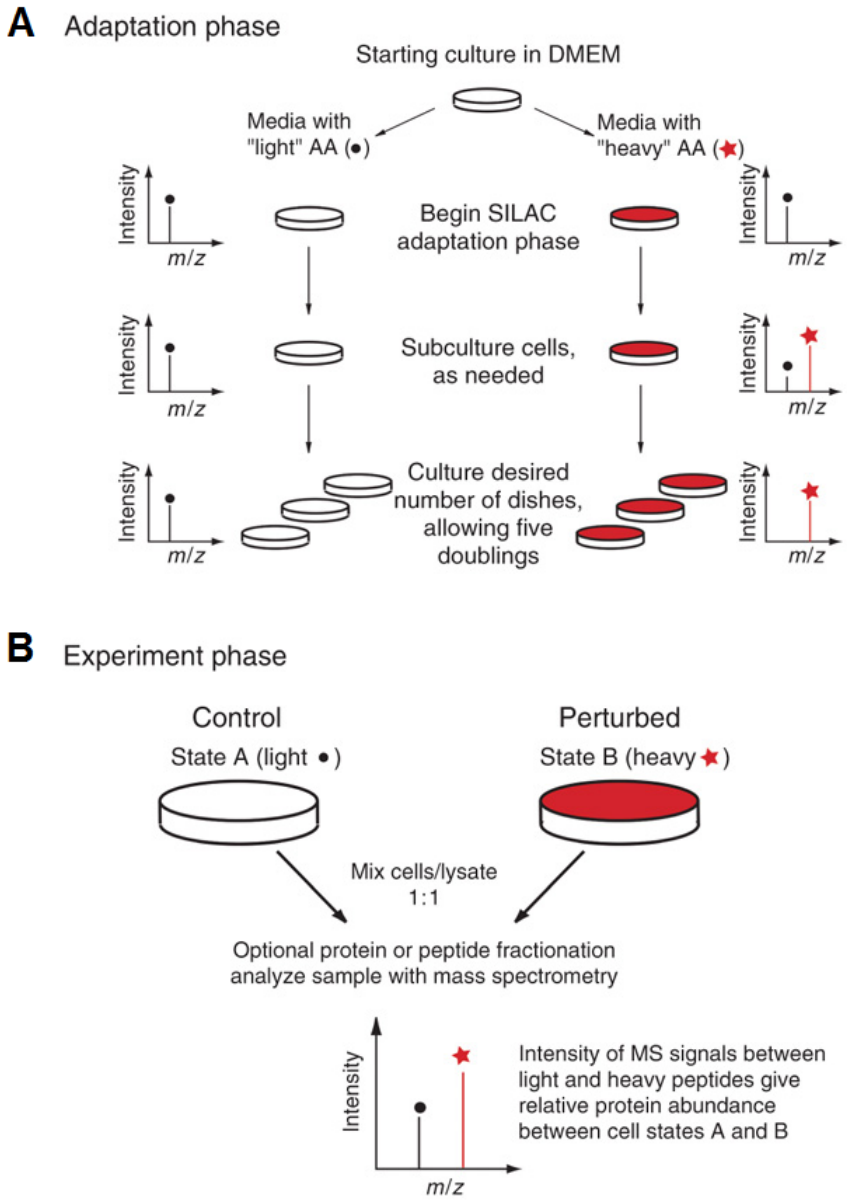


Figure 1.5. Schematic overview of SILAC experiment flow [17]. A typical SILAC experiment basically consists of an adaptation phase (**A**) and an experiment phase (**B**). During the adaptation phase, cells are grown in SILAC media with light and heavy isotopes until the heavy cells have fully incorporated the heavy amino acids (red star). This would allow two SILAC cell pools to be fully distinguishable by precursor MS of light and heavy SILAC peptides (black dot and red star respectively), and then can be mixed and analyzed as a single sample. In the experiment phase, the two cell populations are differentially treated to induce changes in the proteome. The samples are then mixed together for digestion to peptides, and subsequently analyzed by MS for protein quantification between different proteomes.

The advantage of SILAC is that it's straightforward with no additional sample handling step after cell culture. However, it's mostly applicable to cultured cells and caution should be taken to prevent possible amino acid conversion (arginine to proline) during cell culture in SILAC mediums [18].

ii) Chemical labeling

In the pioneering work of Gygi *et al*., an isotope-coded affinity tag (ICAT) approach [19] was developed in which cysteine residues were specifically derivatized by biotinylated iodoacetamide (IAA) with a heavy form (deuterated) and a light form (hydrogen). Using this reagent, both isotope labeling and affinity purification on cysteine-containing peptides were achieved together. However, since cysteine-containing peptides only constitute a small percentage of total peptides derived from tryptic digestion, this method is not practical for large-scale quantitative proteomics with the majority of protein identification based on lysine and arginine peptides.

Other chemical labeling methods involve esterification of carboxyl groups [20] and acetylation of free amines [21-23]. The limitation of esterification is due to efficiency and esters would be partially hydrolyzed under acidic conditions of online RP-HPLC-MS/MS analysis [24]. In terms of acetylation, charge state reduction would reduce peptide ionization and decrease its chance to be selected for MS/MS fragmentation [25].

One method of particular interest is dimethylation of amine groups from n-termini and lysine side chain of peptides. This method was originally developed by Hsu *et al.* [26]

and has been explored further [27-34]. This method is promising in terms of high labeling efficiency and uncompromised ionization of the labeled peptides. However, the limited mass shift of methyl groups between ^{12}C and ^{13}C isotopes would require deuterium substitution to increase mass difference between heavy and light labeled peptides. This would result in a retention time shift of peptide pairs during RP-HPLC elution as has been noted before [24].

iii) Enzymatic labeling

A hydroxyl group from water could be introduced into the carboxyl group formed during amide bond hydrolysis. Thus if the proteolysis is carried out in $H_2{}^{18}O$ environment, all peptides will be labeled with ^{18}O except the peptide originating from the C-terminus of the protein. Mirgorodskaya *et al.* performed enzymatic digestion in the presence of ^{18}O and ^{16}O water [35] and compared peptide and protein relative quantification by the introduced isotopic labels. This approach has been utilized to code peptides for studies of C-terminal characterization [36], de-novo sequencing [37], and protein-protein interactions [38]. However, this method requires special sample handling to reduce sample loss and minimize back exchange of ^{18}O labeled peptides into ^{16}O form [39].

1.3 Conclusion

In conclusion, several methods have been developed for the analysis of complex proteome or sub-proteome using quantitative mass spectrometry and each has its strong points as well as limitations. In addition, further improvements to experimental strategies

are needed particularly for the quantitative analysis of post-translational modifications (PTMs). At the current stage, the field is still under development to generate quantitative proteomic data for comprehensive investigation of a biological phenomenon. With the exponential increase in data volume and complexity from quantitative proteomics studies, appropriate statistical approaches need to be developed for meaningful interpretations of the results in the biological context.

References

1. Gygi, S.P., et al., *Correlation between protein and mRNA abundance in yeast.* Molecular and Cellular Biology, 1999. **19**(3): p. 1720-1730.
2. Anderson, L. and J. Seilhamer, *A comparison of selected mRNA and protein abundances in human liver.* Electrophoresis, 1997. **18**(3-4): p. 533-537.
3. Wilkins, M.R., et al., *From proteins to proteomes: Large scale protein identification by two-dimensional electrophoresis and amino acid analysis.* Bio-Technology, 1996. **14**(1): p. 61-65.
4. Steen, H. and M. Mann, *The ABC's (and XYZ's) of peptide sequencing.* Nature Reviews Molecular Cell Biology, 2004. **5**(9): p. 699-711.
5. Yates, J.R., C.I. Ruse, and A. Nakorchevsky, *Proteomics by Mass Spectrometry: Approaches, Advances, and Applications.* Annual Review of Biomedical Engineering, 2009. **11**: p. 49-79.
6. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: a critical review.* Analytical and Bioanalytical Chemistry, 2007. **389**(4): p. 1017-1031.
7. Liu, H.B., R.G. Sadygov, and J.R. Yates, *A model for random sampling and estimation of relative protein abundance in shotgun proteomics.* Analytical Chemistry, 2004. **76**(14): p. 4193-4201.
8. Choi, H., D. Fermin, and A.I. Nesvizhskii, *Significance Analysis of Spectral Count Data in Label-free Shotgun Proteomics.* Molecular & Cellular Proteomics, 2008. **7**(12): p. 2373-2385.
9. Chen, C.H., *Review of a current role of mass spectrometry for proteome research.* Analytica Chimica Acta, 2008. **624**(1): p. 16-36.
10. Braisted, J.C., et al., *The APEX Quantitative Proteomics Tool: Generating protein quantitation estimates from LC-MS/MS proteomics results.* BMC Bioinformatics, 2008. **9**: p. -.
11. Lu, P., et al., *Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation.* Nature Biotechnology, 2007. **25**(1): p. 117-124.
12. Simpson, K.L., A.D. Whetton, and C. Dive, *Quantitative mass spectrometry-based techniques for clinical use: Biomarker identification and quantification.* Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences, 2009. **877**(13): p. 1240-1249.
13. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.* Molecular & Cellular Proteomics, 2004. **3**(12): p. 1154-1169.
14. Thompson, A., et al., *Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS.* Analytical Chemistry, 2003. **75**(8): p. 1895-1904.

15. Ong, S.E. and M. Mann, *Mass spectrometry-based proteomics turns quantitative.* Nature Chemical Biology, 2005. **1**(5): p. 252-262.
16. Ong, S.E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.* Molecular & Cellular Proteomics, 2002. **1**(5): p. 376-386.
17. Ong, S.E. and M. Mann, *A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC).* Nature Protocols, 2006. **1**(6): p. 2650-2660.
18. Bendall, S.C., et al., *Prevention of amino acid conversion in SILAC experiments with embryonic stem cells.* Mol Cell Proteomics, 2008. **7**(9): p. 1587-97.
19. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nature Biotechnology, 1999. **17**(10): p. 994-999.
20. Goodlett, D.R., et al., *Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation.* Rapid Communications in Mass Spectrometry, 2001. **15**(14): p. 1214-1221.
21. Regnier, F.E., et al., *Comparative proteomics based on stable isotope labeling and affinity selection.* Journal of Mass Spectrometry, 2002. **37**(2): p. 133-145.
22. Che, F.Y. and L.D. Fricker, *Quantitation of neuropeptides in Cpe(fat)/Cpe(fat) mice using differential isotopic tags and mass spectrometry.* Analytical Chemistry, 2002. **74**(13): p. 3190-3198.
23. Ji, J.Y., et al., *Strategy for qualitative and quantitative analysis in proteomics based on signature peptides.* Journal of Chromatography B, 2000. **745**(1): p. 197-210.
24. Julka, S. and F. Regnier, *Quantification in proteomics through stable isotope coding: A review.* Journal of Proteome Research, 2004. **3**(3): p. 350-363.
25. Barrios-Llerena, M.E., et al., *The use of a novel quantitation strategy based on Reductive Isotopic Di-Ethylation (RIDE) to evaluate the effect of glufosinate on the unicellular algae Ostreococcus tauri.* Journal of Proteomics, 2011. **74**(12): p. 2798-2809.
26. Hsu, J.L., et al., *Stable-isotope dimethyl labeling for quantitative proteomics.* Analytical Chemistry, 2003. **75**(24): p. 6843-6852.
27. Boersema, P.J., et al., *Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics.* Nature Protocols, 2009. **4**(4): p. 484-494.
28. Boersema, P.J., et al., *Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates.* Proteomics, 2008. **8**(22): p. 4624-4632.
29. Guo, K., C.J. Ji, and L. Li, *Stable-isotope dimethylation labeling combined with LC-ESI MS for quantification of amine-containing metabolites in biological samples.* Analytical Chemistry, 2007. **79**(22): p. 8631-8638.
30. Hsu, J.L., S.Y. Huang, and S.H. Chen, *Dimethyl multiplexed labeling combined with microcolumn separation and MS analysis for time course study in proteomics.* Electrophoresis, 2006. **27**(18): p. 3652-3660.
31. Ji, C.J., et al., *Identification and quantification of differentially expressed proteins in E-cadherin deficient SCC9 cells and SCC9 transfectants expressing E-cadherin*

by dimethyl isotope labeling, LC-MALDI MS and MS/MS. Journal of Proteome Research, 2005. **4**(4): p. 1419-1426.
32. Ji, C.J. and L. Li, *Quantitative proteome analysis using differential stable isotopic labeling and microbore LC-MALDI MS and MS/MS.* Journal of Proteome Research, 2005. **4**(3): p. 734-742.
33. Ji, C.J., N. Guo, and L. Li, *Differential dimethyl labeling of N-termini of peptides after guanidination for proteome analysis.* Journal of Proteome Research, 2005. **4**(6): p. 2099-2108.
34. Hsu, J.L., et al., *Beyond quantitative proteomics: Signal enhancement of the a(1) ion as a mass tag for peptide sequencing using dimethyl labeling.* Journal of Proteome Research, 2005. **4**(1): p. 101-108.
35. Mirgorodskaya, O.A., et al., *Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using O-18-labeled internal standards.* Rapid Communications in Mass Spectrometry, 2000. **14**(14): p. 1226-1232.
36. Kosaka, T., T. Takazawa, and T. Nakamura, *Identification and C-terminal characterization of proteins from two-dimensional polyacrylamide gels by a combination of isotopic labeling and nanoelectrospray Fourier transform ion cyclotron resonance mass spectrometry.* Analytical Chemistry, 2000. **72**(6): p. 1179-1185.
37. Qin, J., C.J. Herring, and X.L. Zhang, *De novo peptide sequencing in an ion trap mass spectrometer with O-18 labeling.* Rapid Communications in Mass Spectrometry, 1998. **12**(5): p. 209-216.
38. Back, J.W., et al., *Identification of cross-linked peptides for protein interaction studies using mass spectrometry and O-18 labeling.* Analytical Chemistry, 2002. **74**(17): p. 4417-4422.
39. Sevinsky, J.R., et al., *Minimizing back exchange in O-18/O-16 quantitative proteomics experiments by incorporation of immobilized trypsin into the initial digestion step.* Analytical Chemistry, 2007. **79**(5): p. 2158-2162.

Chapter 2

COMPREHENSIVE TWO-DIMENSIONAL REVERSED PHASE LIQUID CHROMATOGRAPHY MASS SPECTROMETRY ANALYSIS OF THE HUMAN MICROTUBULE-INTERACTING PROTEOME

2.1 Introduction

Microtubules (MTs) are dynamic polymers known as major structural components of eukaryotic cytoskeletons and play a critical role in many cellular processes, including cell division, cell motility, intracellular trafficking, and cell shape maintenance [1]. The microtubule has a hollow cylindrical filament structure and is composed of α- and β-tubulin dimmers (its structural subunit), as shown in Figure 2.1. Each α- or β-tubulin monomer has about 450 amino acids and an approximately 55 kDa molecular weight. The microtubule undergoes rapid transitions between growth and shrinkage by exchange of free dimmers at its ends, this is called dynamic instability.

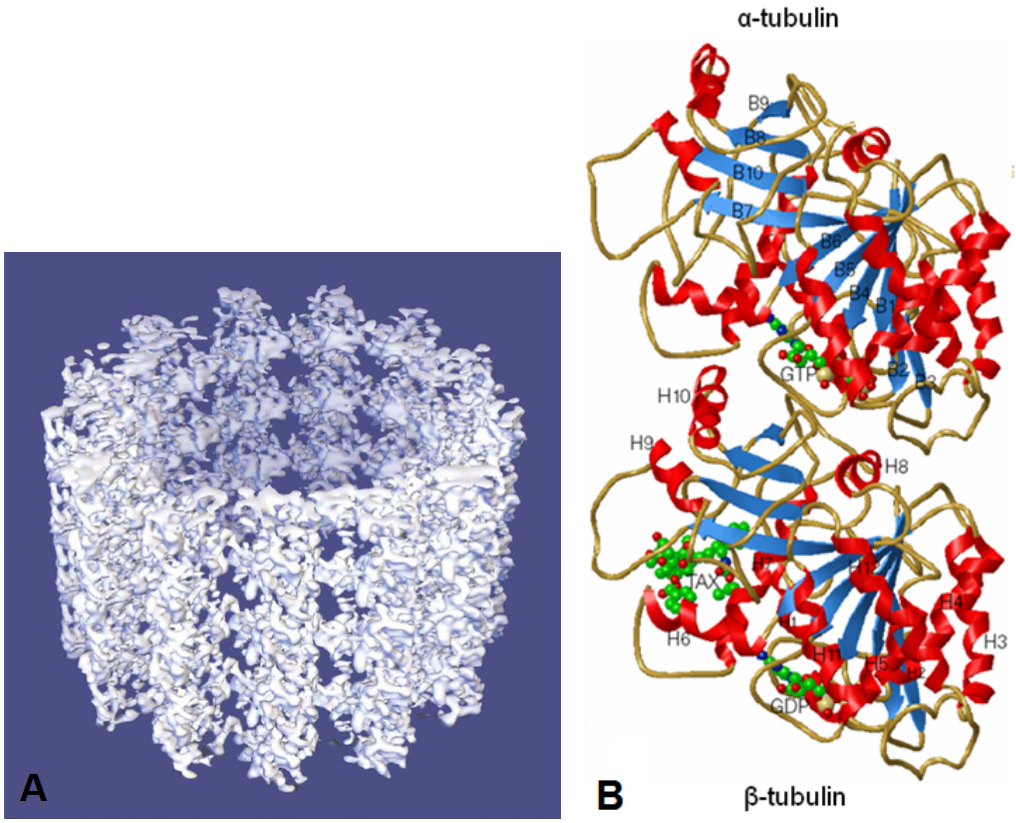


Figure 2.1. Three dimensional structure of microtubule and ribbon diagram of tubulin dimer. (A) Cryo-electron microscopy and image processing show microtubule as staggered

assembly of 13 protofilaments of tubulin heterodimers in the cylinder wall [2]. (B) Ribbon diagram of the tubulin dimer structure by electro crystallography shows α-tubulin on top and β-tubulin at bottom) [3]. Each α- and β-tubulin has a nucleotide binding site at the N-terminal domain (residues 1-206, B1 to B6, H1 to H5) and can bind a guanosine triphosphate (GTP) [3]. After GTP binding, the N-terminal domain of α-tubulin interacts with the Central domain (residues 207-384, B7 to B10, H6 to H10) of β-tubulin through longitudinal contacts to form tubulin dimer. GTP on α- tubulin do not hydrolyze while GTP on β-tubulin can hydrolyze to guanosine diphosphate (GDP).

Microtubule associated proteins (MAPs) are commonly considered as proteins that can bind to, and alter the dynamic of microtubules (MTs). As illustrated in Figure 2.1B, the C-terminal domain (residues 385-450, H11 to H12) of tubulin is on the outside layer of the microtubule cylinder wall, and is mainly involved in the binding of MAPs. MAPs regulate microtubule cytoskeleton and could be enriched by cycles of assembly/disassembly of MTs [4]. Due to their diversified functions, MAPs could be classified into three categories: 1. structural MAPs that stabilize and promote the assembly of MTs, represented by MAP4, TAU; 2. motor MAPs that move along MT for their cellular activities, such as Kinesin, Dynein and their relatives; 3. proteins that are not traditionally called MAPs but had been reported interaction with microtubules, such as heat shock proteins, glycolytic enzymes (e.g. GAPDH), kinases, etc [4]. Figure 2.2 displays structural MAPs example of MAP-2 protein [5] and motor MAPs example of kinesin and dynein [6]. During isolation, binding partners of MAPs would also be co-purified and overall, we can call this complex microtubule-binding proteins or microtubule interacting proteins.

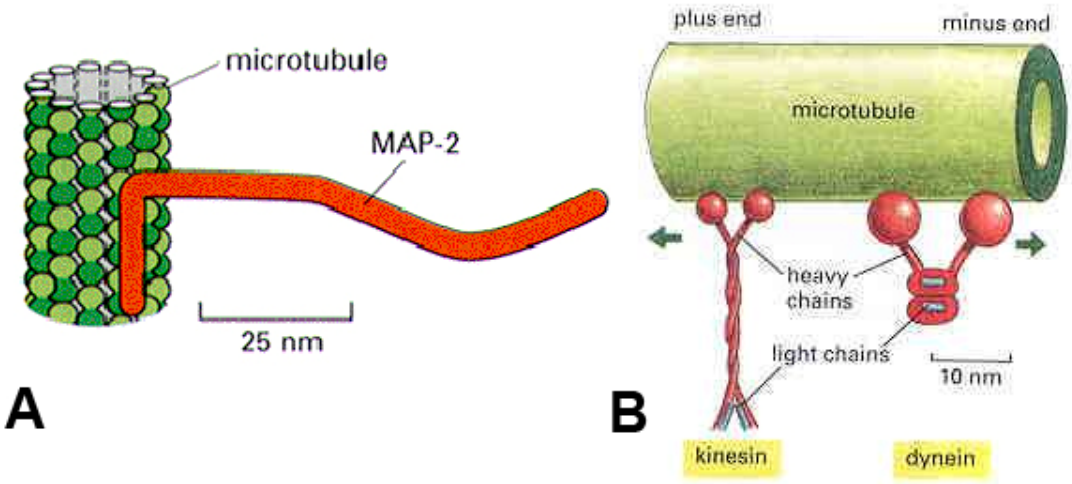


Figure 2.2. Illustrative microtubule associated protein (MAP) examples. (A) Structural MAP protein MAP-2 binding to microtubule outer layer [5]. (B) Movement of kinesin and cytoplasmic dynein along microtubule [6].

MAPs are important cancer chemotherapy targets because many effective anticancer drugs are also microtubule-targeting agents and may alter the activity or expression of MAPs through changing microtubule dynamic polymerization process [7]. Potential mechanisms between microtubule-targeting drugs and MAPs (both stabilizing and destabilizing MT) and their effects on MT dynamics are illustrated in Figure 2.3. Known members of the MAP family include protein products of oncogenes, tumor suppressors, and apoptosis regulators. This suggests that alteration of microtubule dynamics may be a critical event in either the initiation or progression of tumorgenesis [8]. Therefore, a large-scale analysis of MAPs must be conducted in order to understand, at the molecular level, how anticancer drugs mediate their cytotoxic activity on tumor cells.

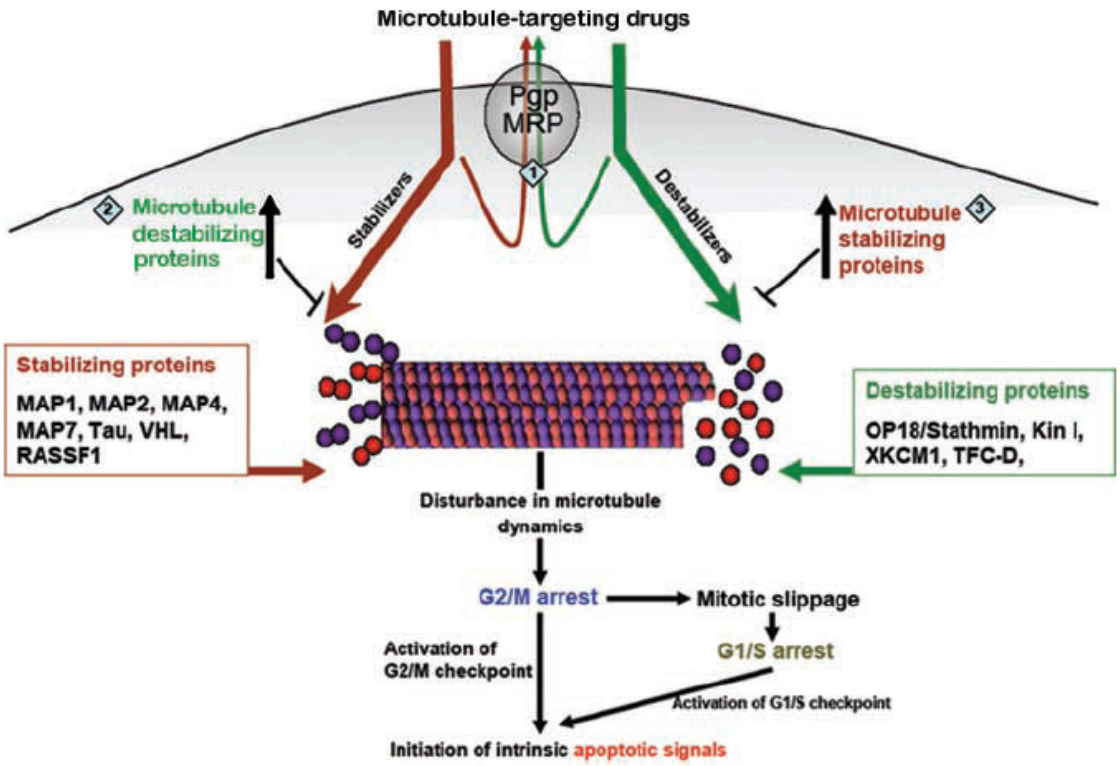


Figure 2.3. Effects of microtubule-targeting drugs on MAPs and microtubule dynamics in cancer cells [7]. Microtubule-targeting agents influence intracellular MAPs (MT stabilizers in red and MT destabilizers in green) and regulation of microtubule dynamics. Disturbance of microtubule dynamics can trigger cancer cell apoptosis by two pathways: arrest in G2/M phase or arrest at G1-S checkpoint of cell cycle.

One biochemical approach has been established to isolate this complex with the facilitation of microtubule-stabilizing agent, Taxol. This procedure involves cycles of polymerization of endogenous tubulin, co-sedimentation of microtubules and their bound proteins, and final release of microtubule interacting proteins by elevated ionic strength [9-11]. Due to the biological nature of the microtubule interaction network, the isolated protein complex mixture of this method contains both high abundance proteins such as heat shock protein family [12, 13] and other low abundance proteins which are difficult

to characterize by typical LC-MS/MS analysis. Owing to the biological interest of MAPs, people have all along been seeking to characterize this complex. A summary of previous proteomic studies on the analysis of microtubule-interacting proteins is listed in Table 2.1 in reverse chronological order. Recently, Patel et al. [14] chose first dimension RP-HPLC

Table 2.1. Proteomic Studies on Analysis of Microtubule-interacting proteins

S.NO	Isolation Procedure	Sample Type	# Proteins Identified	Instrument Used	Citation
1	Taxol precipitation	Bovine Brain	571	GeLC; CHIP 6330 LC/MSD Trap XCT Ultra ion trap	Kozielski et al. (2011)[15]
2	Taxol precipitation	Mouse Macrophage cell line	409	Protein RPLC Chip; 6340 ion trap	Patel et al. (2009)[14]
3	Taxol precipitation	Xenopus egg extracts	318	GeLC; LIT LTQ	Gache et al. (2010)[16]
4	Taxol precipitation	Rat Brain	391	MALDI-TOF; MALDI-TOF/TOF (ABI4700)	Sakamoto et al. (2008)[17]
5	Taxol precipitation	Drosophila Melanogaster embryos	250	1D and 2D-PAGE Q-TOF1	Hughes et al. (2008)[18]*
6	Tubulin affinity chromatography	Arabidopsis suspension cells	122	2DGE; Qstar Pulsar I Q-TOF	Chuong et al. (2004)[19]*
7	Taxol precipitation	Artemia franciscana embryos	55	2DGE; QTRAP	O'Connell et al. (2006)[20]
8	Immunoaffinity column	HeLa cells	14	GeLC; QTOF Ultima	Gache et al. (2005)[21]
9	Taxol precipitation	Xenopus Lavis egg extracts	41	1D PAGE ; MALDI-TOF; Qstar Pulsar I Q-TOF	Liska et al. (2004)[22]

S.NO represents study number. * Few standard MAPs identified.

to fractionate at the protein level, then performed online RP-HPLC separation for the digested peptides (both under low-pH conditions) to be analyzed by iontrap instrument, resulting in the identification of 409 proteins that bind directly or indirectly to microtubules from mouse macrophage cell line. Kozielski et al. [15] utilized SDS-PAGE method to separate MAPs isolated from bovine brain, after in-gel digestion and online LC-MS/MS analysis, they claimed 573 protein IDs (actually 571 because of a mistake). This result is the combination of proteins from tubulin-rich and tubulin-depleted extractions after searching both bovine and human databases. These high proteomic coverage results are not only contributed by the high throughput MS analysis, but also from the multidimensional fractionation strategy to reduce sample complexity.

Our research focused on the proteomic study of MAPs from human cancer cell line, in the hope of finding potential biomarker for cancer therapy and understanding the complicated microtubule interacting network. Considering the biological nature of MAPs, the isolated complex contains both high abundance proteins such as heat shock protein family [12, 13] and other low abundance proteins which would be difficult to characterize by typical LC-MS/MS. To circumvent this, we utilized the 2D-RP-HPLC peptide separation strategy which has not been explored in similar studies. This method was recently introduced by Gilar et al. by using two stages of reversed phase separation under high and low pH conditions, which provides high peak capacity in each RP dimension and greatest peptide separation orthogonality under pH 10 in the first dimension and pH 2.6 in the second dimension (comparable to 2D SCX-RP-HPLC setup) [23].

2.2 Materials and Methods

2.2.1 Materials

All chemicals including reagents for cell culture and isolation of microtubule-binding proteins were purchased from Sigma-Aldrich Inc. (St. Louis, MO) except fetal bovine serum (FBS) was from Atlanta Biologicals (Lawrenceville, GA) and protease inhibitor cocktail was from Roche Applied Science (Indianapolis, IN). NuPAGE Novex 4-12% Bis-Tris gel was obtained from Invitrogen (Grand Island, NY) while standard laboratory buffers were from Fisher Scientific (Pittsburgh, PA). Sequencing-grade modified trypsin was purchased from Promega (Madison, WI).

2.2.2 Cultivation of Human Jurkat T Cells

Jurkat T (E6-1) cell line was a gift of Prof. Arthur Salomon in Brown University which was originally obtained from the American Type Culture Collection (Rockville, MD). Cells were maintained at 37℃ supplied with 5% CO_2 and cultivated in suspension in RPMI-1640 medium containing 10% FBS and 1% L-Glutamine-Penicillin-Streptomycin stabilized solution with normal doubling time of 24 hours. Cell growth was monitored by counting population under a microscope.

2.2.3 Isolation of Microtubule-binding Proteins and Gel Electrophoresis

Microtubule-binding proteins were isolated from Jurkat cells using the taxol stabilization procedure [8-10] with some modifications. Briefly, a population of 5×10^8

of Jurkat cell was resuspended in 1mL MME buffer (0.1M 2-(N-morpholino) ethanesulfonic acid (MES), 1mM $MgCl_2$, 1mM EGTA, pH 6.9) containing protease inhibitor cocktail and 1mM DTT, followed by sonication to generate cell lysate. The resulting cell lysate was then centrifuged at 120,000g for 1 hour at 4 ℃ to separate cytosolic extract (supernatant) from DNA and cell debris (pellets). The cytosolic supernatant (containing soluble tubulin) was then transferred and added with 20μM Taxol and 1mM GTP, and incubated at 37℃ for 30 minutes to polymerize microtubules. This mixture was transferred on top of a 10% sucrose MME cushion containing 20μM Taxol, 1mM GTP and then centrifuged at 80,000 g for 30 min at 37℃ to precipitate microtubules along with its binding proteins as pellets. After centrifugation, the supernatant was discarded and microtubule pellets (with binding proteins) at tube bottom was carefully rinsed with MME buffer to remove any soluble cytoplasmic contamination. To dissociate microtubule-binding proteins from microtubules, the previous microtubule pellets were resuspended in 200μL MME buffer containing 20μM Taxol, 1mM GTP and 0.35M NaCl at 37℃ and later centrifuged at 80000 g for 30 minutes at 37℃. Microtubule-binding proteins were then released to the supernatant while remaining microtubule pellets stayed at the bottom of the tube.

Gel electrophoresis was conducted on NuPAGE Novex 4-12% Bis-Tris gel in MOPs buffer according to the vendor's instructions. Microtubule-binding proteins and all intermediate products from the isolation procedure were confirmed by Commassie Blue staining.

2.2.4 Preparation of Tryptic Digests

After dialyzing the final supernatant, 70μg of microtubule-binding proteins were obtained in 200μL 50mM ammonium bicarbonate buffer (pH 8.5). Proteins were then reduced by 3mM DTT for 1 hour at 55℃, alkylated by 6.3mM iodoacetamide for 45 minutes at room temperature, and reduced again by 3mM DTT under same condition. One μg of modified trypsin was added in the end to digest proteins at 37℃ for 14 hours.

2.2.5 High pH First Dimension RP-HPLC Separation

Tryptic digest of microtubule-binding proteins was injected into an Agilent 1200 HPLC system equipped with the Agilent ZORBAX 300Extend-C18 column (150 × 2.1mm i.d, 3.5μm) with temperature setting at 35℃. Mobile phase A was composed of 1% methanol, 20mM ammonia (pH 10.4) and mobile phase B was 90% acetontrile, 1% methanol, 20mM ammonia. Injected peptides were separated by a gradient elution of 0-70% B in 35 min with constant flow rate at 0.10 mL/min. UV absorbance was monitored at 214 nm and fractions were manually collected every minute. Typically these tryptic peptides were confined in 20 fractions from 20 to 40 of elution time. Each collected HPLC fraction (100μL) was centrifuged in the CentriVap Concentrator (LABCONCO, Kansas City, MO) to reduce sample volume to near dryness, and then reconstituted in 20μL 0.1% acetic acid.

2.2.6 Nanoflow RP-HPLC-MS/MS Analysis

For each concentrated HPLC fraction of peptides, a 2μL aliquot was taken and then diluted (1:10) by adding 18μL of 0.1% acetic acid. An Eskigent autosampler was used to load 5μL of the diluted sample onto a self-made capillary trapping column. The trapping column was made by packing an 4cm length reversed-phase C18 material (POROS 10 R2, 10μm particle, AB Sciex, Foster City, CA) into a 75μm ID fused-silica capillary column (Polymecro Technologies, Phoenix, AZ) with self-made frit on one end. Connected to the trapping column, was a self-made analytical column, which was constructed by pulling a 5μm tip from a 50μm ID fused-silica capillary column (Polymecro Technologies, Phoenix, AZ) and packing 7cm reverse-phase C18 material (Monitor 100A-SPHERICAL SILICA, 3μm particle, Column Engineering Inc, Downers Grove, IL) from the other end. Tempo nanoMDLC system was used to provide HPLC flow to the trapping column and its mobile phase A was composed of 2% acetontrile, 0.1% formic acid while mobile phase B was 98% acetontrile, 0.1% formic acid. After sample loading, peptides were eluted off the trapping column by a gradient elution of 0-30% B in 40 minutes, 30-60% B in 40 minutes, and 60-95% B in 20 minutes with constant flow rate at 80 nL/min and furthered ionized by electrospray ionization source.

The MS/MS analysis was conducted on QSTAR Elite hybrid quadrupole time-of-flight mass spectrometer (AB Sciex) with ionization voltage at 1800V and precursor ion scan from 250 to 2000m/z with 0.2 second accumulation time. For each precursor ion scan, the instrument would select the 4 most abundant ions (with +2 or higher charge state and at least 100 counts of intensity) to apply automatic collision energy, and

perform product ion scan from 70 to 2000m/z with maximum 2 second accumulation time after collision induced dissociation. The dynamic exclusion period was set at 20 seconds for recurring precursor ions.

2.2.7 Bioinformatic Analysis

The raw data file (*.wiff) from LC-MS/MS analysis of each collected HPLC fraction was searched both individually and jointly against International Protein Index (IPI) human database (v3.64, released September 24th, 2009) by using ProteinPilot^{TM} software version 2.0.1 (revision 67476) from AB Sciex. The features of ProteinPilot software is that, common modifications, substitutions, and cleavage events are already built-in functions in the software. After pooling all LC-MS/MS data designated for searching, ProteinPilot uses the Paragon^{TM} algorithm [24] to perform peptide identification, followed by the Pro Group^{TM} algorithm to perform a statistical analysis on the peptides found to determine the minimal set of confident protein identifications with redundancy removed. The search parameters for this work were set for thorough search of sample identification with digestion by trypsin, cystein alkylation by iodoacetamide, and ID focus on biological modifications.

In this work, the peptide identification threshold was set at 99% confidence level from Paragon algorithm [24] analysis with at least 8 amino acids and less than 50 ppm mass error to be considered valid. For protein identification, the grouped protein result from ProteinPilot searching was manually filtered to guarantee each protein has at least

two unique peptides identified with different amino acid sequences. The unique peptide used here means the peptide that is identified by MS/MS spectrum with 99% confidence and assigned to only one protein (or protein set) in the database after protein grouping by Pro Group algorithm. Three biological replicates were conducted independently and the final reported proteins were the proteins that were consistently identified in all 3 biological replicates and had at least 2 unique peptides.

The false positive detection rate (FDR) was estimated as previously described [25, 26] using the concatenated normal sequence + reverse sequence database method and based on the valid peptides after removing redundancy from the searching result. For each independent biological replicate, the FDR for identified peptides was less than 0.35% as calculated.

Four hundred and fifty-three microtubule-binding proteins were reported and protein identities were received in the form of IPI database accession numbers. To understand the biological functions of these identified proteins, they were classified according to their functions from Gene Ontology (GO) database [27, 28] supplemented by manual data mining of references. Uncharacterized proteins were those that have not been investigated previously or with indefinite functions.

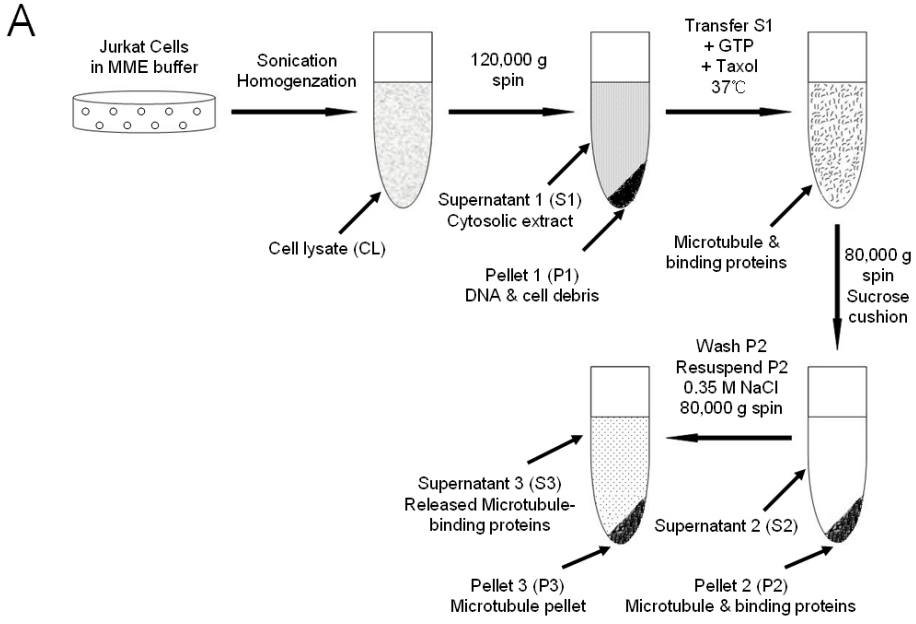
Known protein interactions among the identified proteins were created by Cytoscape software (version 2.6.3) with the BisoGenet [29] plugin (version 1.4). Interactions were identified using the INTACT (from European Bioinformatics Institute), MINT

(Molecular Interactions Database) and DIP (Database of Interacting Proteins) databases. An input of 466 genes (from the 453 identified proteins) resulted in 464 protein nodes and 452 interactions. After removing protein nodes that had no interactions with other proteins, the proposed microtubule interaction network with 247 nodes and 433 interactions was created.

2.3 Results

2.3.1 Isolation of the Biological Sample is Reproducible.

Microtubule-binding proteins were isolated from three independent cultures of approximately 5×10^8 Jurkat leukemia cells using a slightly modified version of the Taxol-based isolation procedure originally described by Vallee [9-11]. The isolation scheme is shown in Figure 2.4A. Each biological sample yielded >70-µg protein after



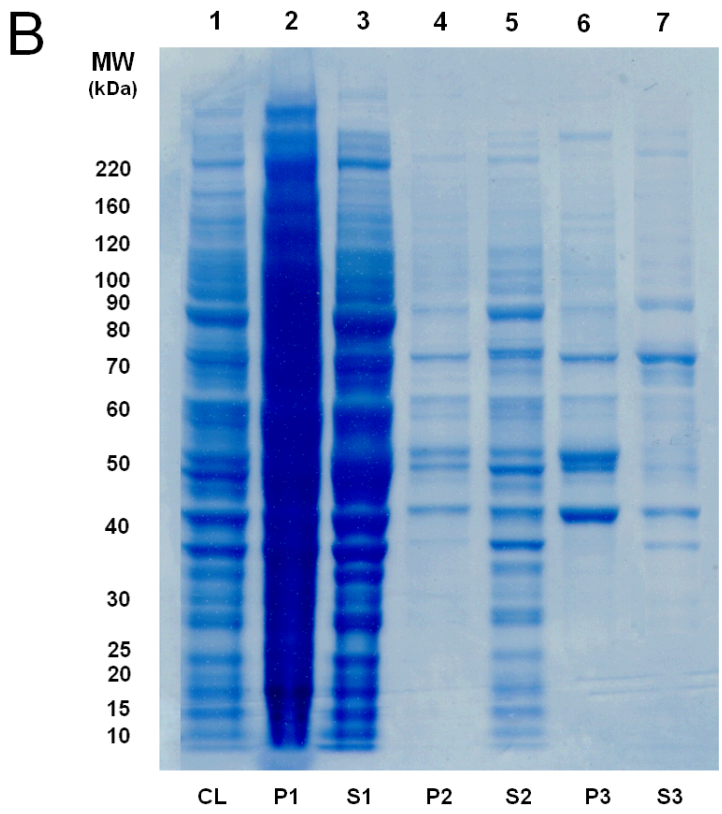


Figure 2.4. Isolation of microtubule-binding proteins from Jurkat cell lysate. (A) Schematic isolation procedure. (B) Image of isolated proteins after SDS-PAGE and coomassie blue staining. Proteins in lane 1 to 7 of gel image are from MAPs purification scheme: Cell Lysate (CL); Pellet I (P1), DNA and cell debris pellet; Supernatant (S1), cytosolic extract; Pellet II (P2), Taxol-stabilized microtubule pellet with MAPs on it; Supernatant II (S2); Pellet III (P3), Taxol-stabilized microtubule pellet after NaCl treatment; Supernatant III (S3), released MAPs.

dialysis; and when resolved by SDS-PAGE, gave a characteristic protein pattern (Lane S3 in Figure 2.4B).

The three biological samples were individually trypsinized and separately fractionated using a high-pH tolerant C18 column and mobile phases containing 20 mM ammonia, pH 10.4 [30]. The three replicate analyses showed good sample-to-sample

reproducibility as supported by the UV chromatograms for the first dimension separations except for the broad initial peak in replicates 1 and 2 due to non-retained contaminants (Figure 2.5). There is a slight change in the chromatographic profile of replicate 3 as compared to replicate 1 and 2, which could be a result of the natural variation in protein expression in asynchronous cells or due to column fouling. Peptides typically eluted within the 20-41 minute window and 1-min fractions were collected during that time. A total of 63 fractions (21 fractions per replicate) were analyzed by on-line low-pH LC-MS/MS.

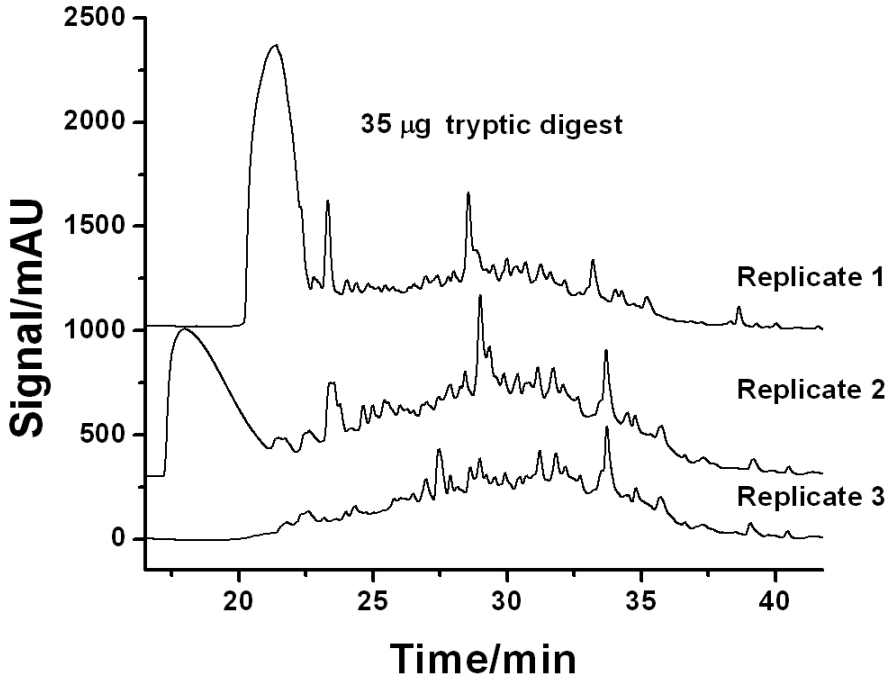
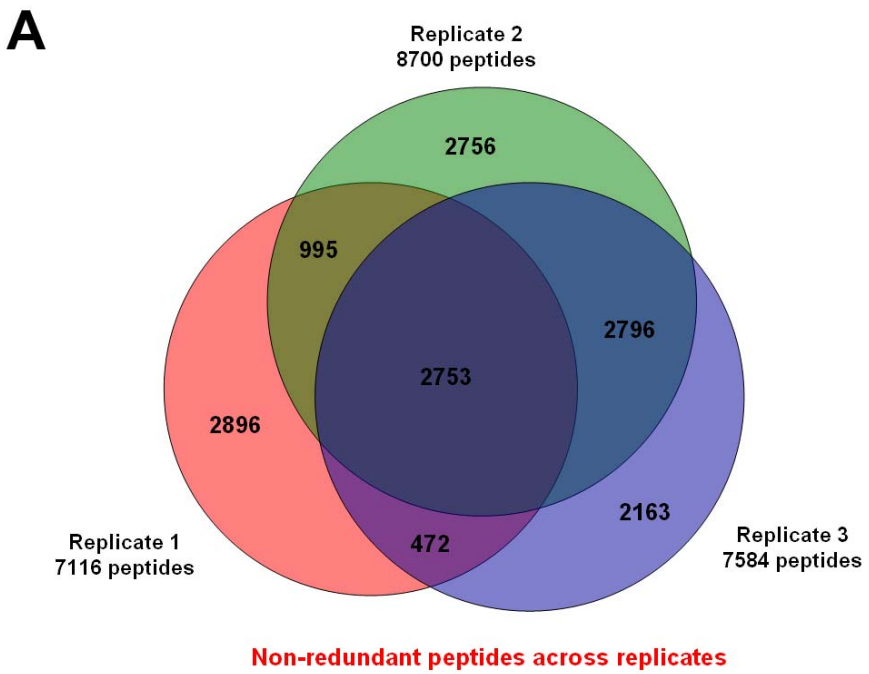


Figure 2.5. UV chromatograms of offline high-pH RP-HPLC fractionation of tryptic digests derived from microtubule-binding proteins isolated from Jurkat cell lysates by three biological replicates. Three biological replicates of microtubule-binding proteins were isolated from Jurkat leukemia cells and trypsinized. Thirty-five micrograms of each digest was injected onto an Agilent ZORBAX 300Extend-C18 column at 35°C. Peptides were separated at 0.10

mL/min using a gradient of 0-70% B in 35 min. Mobile phase A was composed of 1% methanol, 20 mM ammonia, pH 10.4 and mobile phase B was 90% acetontrile, 1% methanol, 20 mM ammonia. UV absorbance was monitored at 214 nm and fractions were collected every minute.

After offline high-pH RP-HPLC separation of the tryptic digests of isolated MAPs from Jurkat cell lysate, all fractions were analyzed by online low-pH RP-HPLC coupled with tandem mass spectrometry. The results from each high-pH-RP-MudPIT (multidimensional protein identification technology) experiment, in terms of peptide and protein identifications, are shown in Figure 2.6A and 2.6B respectively. Analysis of the three biological replicates resulted in the identification of 14,231 non-redundant peptides that represent 945 proteins. The presence of a protein in a given sample was considered



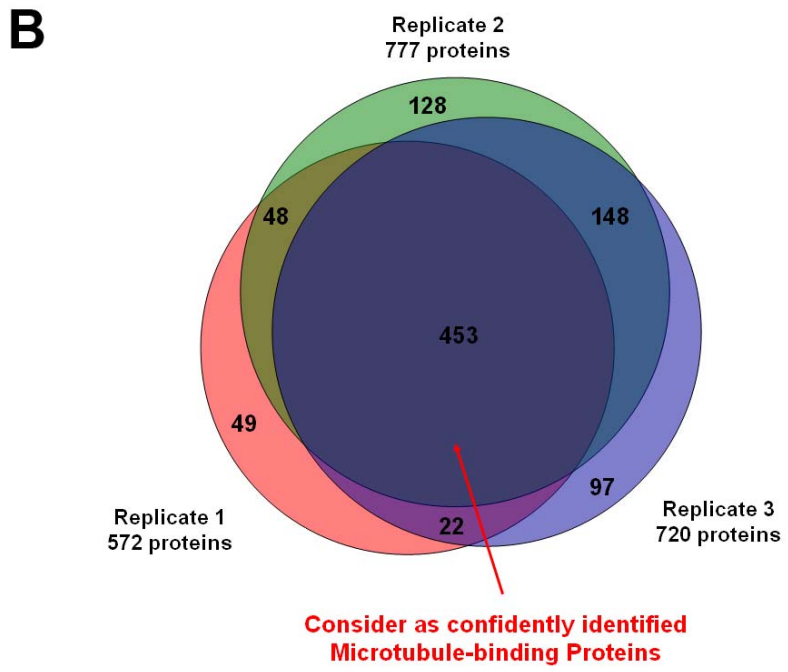


Figure 2.6. Venn diagrams summarizing peptides and proteins identified from MudPIT analysis of microtubule-binding proteins across three biological replicates. (A) Identified non-redundant peptides across replicates. (B) Identified microtubule-binding proteins across replicates. (Note, proteins shown in (B) had been filtered with at least two unique peptide identifications at the 99% confidence level. If the restriction is set to one 99% confidence unique peptide for each valid protein, there were 818, 1087, and 1070 proteins identified for each replicate, respectively.)

valid when at least two unique peptides (at the 99% confidence level from Paragon algorithm [24] analysis of ProteinPilot software) were detected for that protein. The peptide level FDR was estimated to be less than 0.35% for each replicate analysis. Four hundred fifty-three common proteins remain when we removed proteins found in only one or two replicates. These 453 common proteins represent the most biological-robust data set due to the application of very strict filtering criteria across the three biological replicates. The data presented in Figure 2.6B also indicates that identification at the

protein level is reproducible (varying from 58%-79%) even in light of sample-to-sample variability. However, this was not the case at the peptide level. Only about 1/3 of the non-redundant peptides from each replicate were found in all three replicate analyses most likely due to the stochastic nature of information-dependent proteomic analysis associated with LC-MS/MS shotgun proteomics.

2.3.2 Classification of 945 Identified Proteins based on Biological Functions and Comparison with Literature

As mentioned above, 945 high-confidence proteins in total were identified from all three biological replicates as members of the human microtubule interactome and this is by far the most comprehensive data set of microtubule-interacting proteins comparing to similar studies [14-22]. These proteins were grouped into eleven functional categories based on manual review of published literature and use of the Gene Ontology (GO) database [27, 28]. As shown in Figure 2.7, nearly 20% of the proteins sort into GO categories that describe expected functions of MT-binding proteins (i.e. cell division and growth/maintenance, cytoskeleton and cytoskeleton-associated, MT motors and motor-associated proteins, and non-motor MAPs). Tau-C, a well-appreciated non-motor MAP known for Alzheimer's disease, was identified in this study and has not previously been reported in similar studies for mammalian cell lines. Interestingly, the microtubule co-sedimentation procedure co-purified 276 metabolic enzymes in total from three replicates, suggesting the indirect involvement of microtubule and metabolic enzymes through

microtubule-binding proteins. The complete protein identification list and functional grouping categories are available in Supplemental Table 1 as excel sheet.

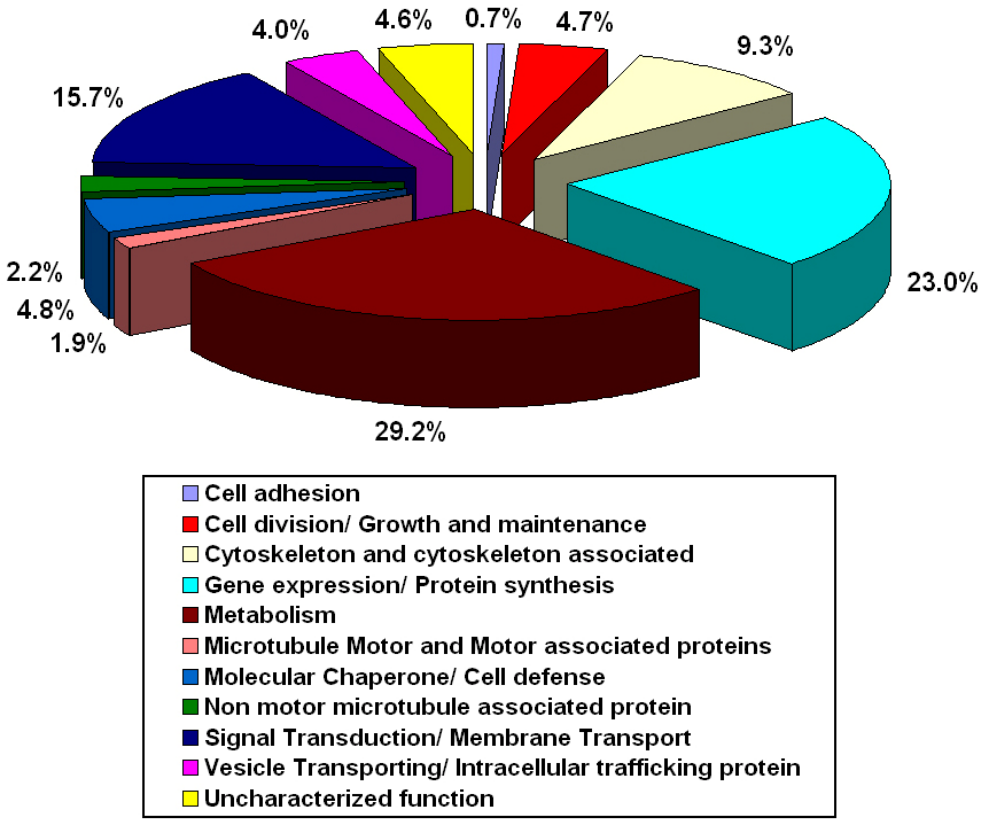


Figure 2.7. Grouping of identified proteins based on biological functions from GO. The 945 identified proteins (that bind directly or indirectly to microtubules) were classified into eleven categories according to Gene Ontology (GO) to illustrate their biological functions: 0.7% (7 proteins) were annotated as cell adhesion; 4.7% (44 proteins) as cell division/growth and maintenance; 9.3% (88 proteins) as cytoskeleton and cytoskeleton associated; 1.9% (18 proteins) as microtubule motors and motor associated proteins; 2.2% (21 proteins) as non-motor microtubule associated proteins; 23.0% (217 proteins) involved in gene expression/protein synthesis; 29.2% (276 proteins) involved in metabolism; 15.7% (148 proteins) involved in signal transduction/ membrane transport; 4.8% (45 proteins) as molecular chaperone/cell defense; 4.0% (38 proteins) involved in vesicle transporting/intracellular trafficking and 4.6% (43 proteins) have uncharacterized functions.

In Figure 2.8, we compared our protein identification list with results from two recent large-scale proteomic studies of the microtubule interacting proteins for mammalian brain tissue [15] or mammalian cell macrophage [14]. Our proteomic identifications overlap about 50% to 70% of each individual result, while the overlap between these two studies is about 1/3. In addition, we have the largest number in terms of protein identifications, which is about 65% to 130% increase from previous studies. The reason for such difference maybe contributed to the high resolving capability of 2D-RP-HPLC peptide

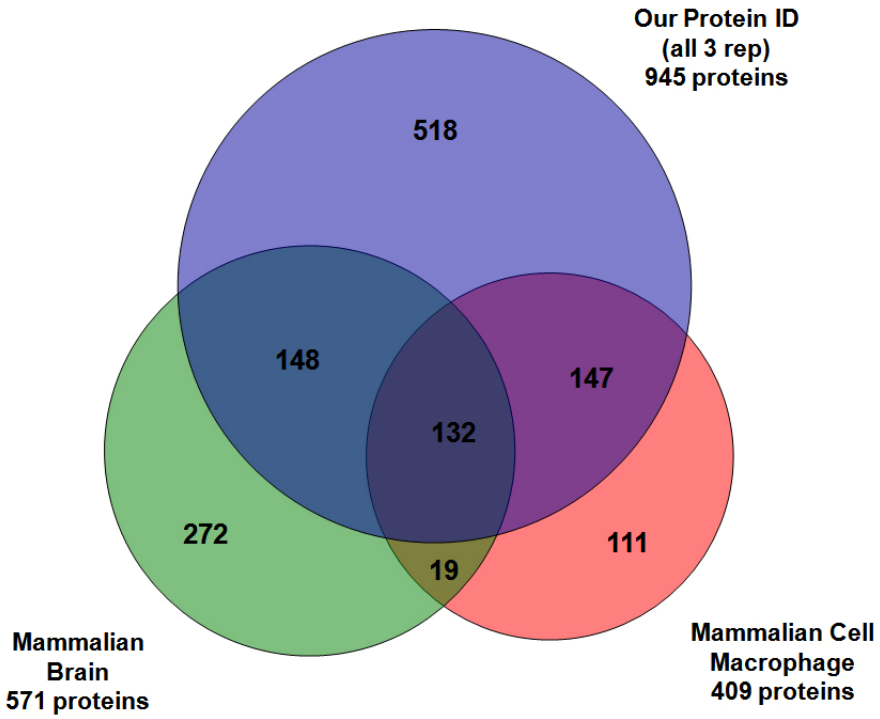


Figure 2.8. Comparison of identified proteins from our 2D-RP-HPLC peptide separation method with other proteomic studies of microtubule interacting proteins in mammalian brain tissue [15] or mammalian cell macrophage [14]. Protein accession numbers in the result of individual study are converted to their corresponding gene symbols and the comparison of protein identifications is based on the converted gene symbols.

separation method we utilized. In the mammalian cell macrophage study [14], isolated proteins were fractionated using RP-HPLC before digestion while in the mammalian brain tissue study [15]; proteins were separated by gel electrophoresis. In both studies, peptides after digestion were only separated by the online low-pH RP-HPLC prior to MS analysis.

2.3.3 Bioinformatic Analysis of Binary Protein Interactions for the 453 Reproducibly Identified Microtubule-binding Proteins and Quantification Analysis

As mentioned above, we have 453 common proteins that were reproducibly identified across three independent biological replicates after strict filtering criteria. To further understand this comprehensive data set of proteins that was observed because of their interactions with MTs, a protein network diagram was created to elucidate their underneath connection. Known protein interactions among the identified microtubule-interacting proteins were created by Cytoscape software (version 2.6.3) with the BisoGenet plugin (version 1.4). Interactions were identified using the INTACT (from European Bioinformatics Institute), MINT (Molecular Interactions Database) and DIP (Database of Interacting Proteins) databases. An input of 466 genes (from the 453 identified proteins) resulted in 464 protein nodes and 452 interactions. After removing protein nodes that had no interactions with other proteins, the proposed microtubule interaction network with 247 nodes and 433 interactions was illustrated in Figure 2.9.

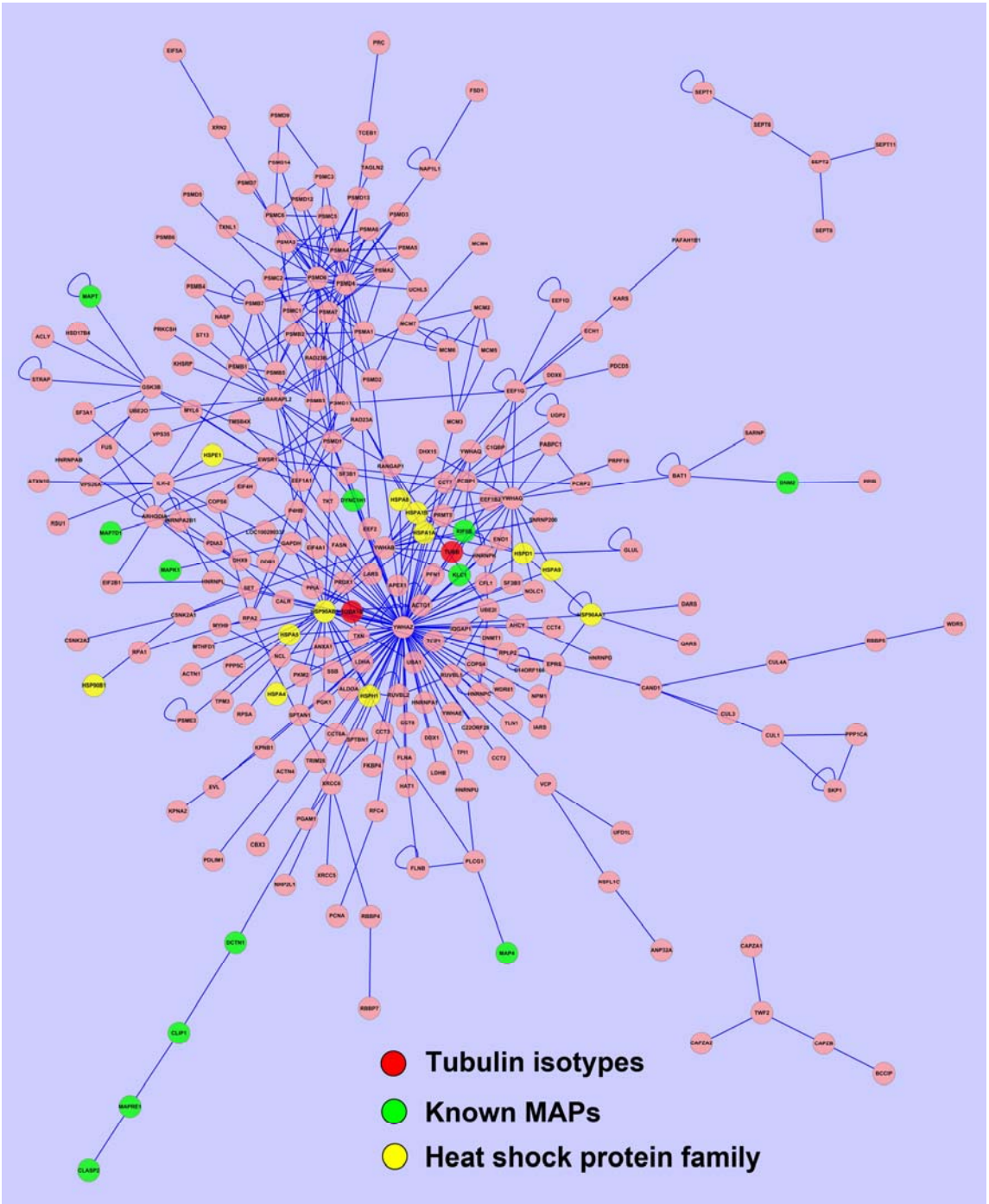


Figure 2.9. Protein interaction network of the 453 common proteins (bind directly or indirectly to microtubules) across three biological replicates. Each node in the diagram represents one protein gene symbol and each line between two nodes demonstrates that there is documented protein-protein interaction evidence for these two genes. There are a total of 247 nodes and 433 interactions supported by 860 evidence codes in the diagram.

Upon closer look at these protein-protein interactions, they are supported by 860 highly-confident low through-put experimental evidence codes from the databases (shown in Figure 2.10). So this microtubule interaction network is convincing and insightful in understanding its biological function and cellular process.

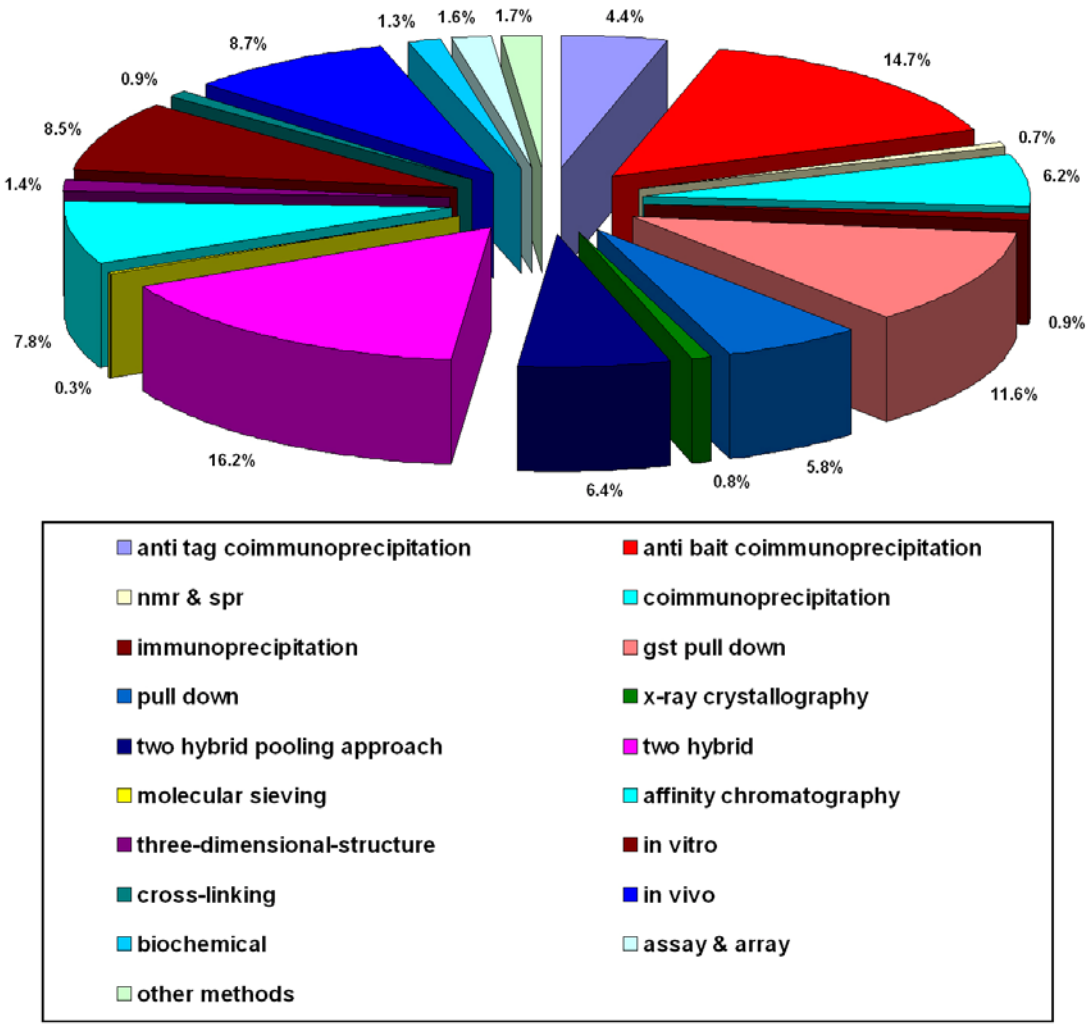


Figure 2.10. Summary of 860 evidence codes from protein interaction databases for the 433 protein-protein interactions.

Quantification information about the 453 common proteins was obtained by searching MS/MS spectrum using Mascot 2.2 (Matrix Science) and processing searching result with ProteoIQ 2.0 (Nusep) software for spectral counting of peptides with ion score threshold of 30. The combinational quantificational result of 453 common proteins in different biological function groups is illustrated in Figure 2.11. It's noticeable that for cytoskeleton and cytoskeleton associated proteins, although their protein IDs constitute

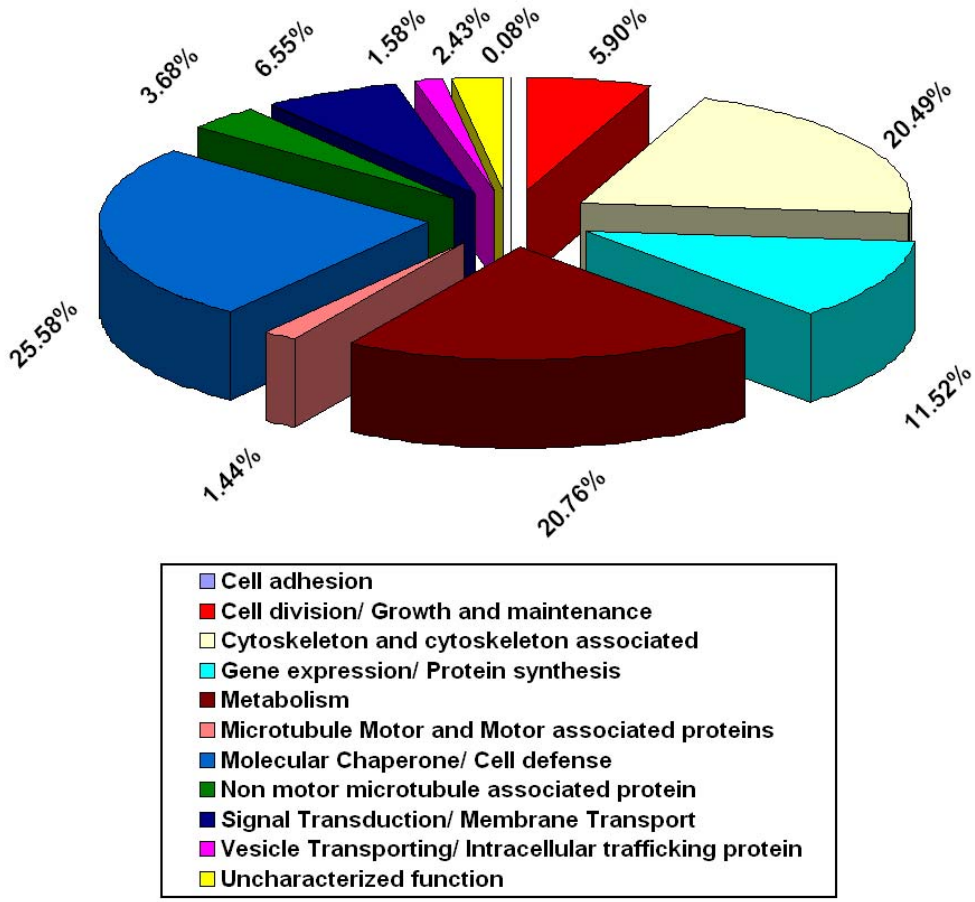


Figure 2.11. Protein group abundance for the 453 common proteins across biological replicates based on spectral counting quantitation.

11.3% of the total 453, their abundance is 20.5% of total spectral counts. Similar tendency is found for molecular chaperone/cell defense proteins with 6.4% of protein IDs and 25.6% of spectral counts. The former could be anticipated since MT is component of cytoskeleton and microtubule interacting proteins bind to MT, and the later could be contributed to the fact that the most abundant microtubule-interacting protein, heat shock protein family, falls into this category.

2.4 Discussion

In this work, we isolated microtubule-binding proteins from cancer cells using the Taxol-fascilated stabilization procedure [9-11], and this is the most-utilized approach in similar studies [14-22]. Due to the strong stabilizing effect of Taxol on microtubules during isolation procedure [31], the yield of microtubule-binding proteins without use of Taxol would be extremely low. In the study of microtubule interactome in *Drosophila* embryos, Hughes *et al.* reported a 200 fold less protein yield for the microtubule and binding proteins pellet in the control sample (with only GTP) compared to Taxol-treated sample [18]. For these reasons, proteins identified in our result should be predominantly true microtubule-binding proteins because a negative control experiment without usage of Taxol would generate negligible protein identifications. Another reason we did not include a negative control in our experiments is that, known MAPs such as MAP1, MAP2, and Tau can decrease the critical concentration of tubulin for polymerization *in vitro* and also bind to the reconstituted microtubules [32]. So a negative control without Taxol could still generate positive MAPs.

Our microtubule-binding protein result is mostly consistent with previous proteomic studies [14-22] about microtubule associated proteins in terms of the diversity and biological functions of identified proteins. However, due to the high resolving power of 2D-RP-HPLC peptide separation method utilized in our approach, we have achieved the largest comprehensive data set of microtubule-binding proteins up to the present. From the 945 total proteins we identified from three biological replicates, with comparison to two recent studies (Figure 2.8), more than 50% of them are exclusively from our result. In addition, we successfully identified some of the most abundant MAPs: dynein, dynactin, kinesin, stathmin, MAPRE1, MAP1, MAP4, MAP7, CLIP1, and EML4, which were also confirmed in the previous research about microtubule associated proteins during macrophage activation [14]. Other than that, we also identified some of the well know MAPs that had never been reported in similar proteomic studies for mammalian cell lines, to our best knowledge. These MAPs include Tau, TPX2, CLASP2, DNM2, DRG1, NUSAP1, and RAE1.

Identification of low abundance proteins in a complex biological extract is usually challenging due to the ionization competition of their peptides with co-eluting peptides from higher abundance proteins. Based on ProteinPilot result, if we consider proteins with only 2 or 3 unique peptide identifications as low abundance proteins, here's the low abundance protein identification summary. For replicate 1, 247 (43%) proteins out of 572 total protein identifications are low abundance proteins; in terms of replicate 2 and 3, low abundance proteins are 329 (42%) out of 777 total and 305 (42%) out of 720 total,

respectively. This result might be contributed to the high resolving power of 2D RP-HPLC separation of peptides so that peptides from low abundance proteins could be confidently identified.

2.5 Conclusion

We sought to utilize 2D-RP- HPLC for the large-scale analysis of the human MT interactome due to the expected benefits of this separation method versus SCX-RP [33, 34]. Microtubule-binding proteins represent a functionally diverse set of proteins with different physiochemical properties that are directly or indirectly modulated by the microtubule network. While less complex than a whole cell lysate, this multi-protein sample would still generate a tryptic digest complex enough to challenge the separation capability of multidimensional chromatography coupled to tandem mass spectrometry. With efficient and orthogonal separation of peptides by 2D-RP-HPLC, our result of identified microtubule-interacting proteins is the most comprehensive data set to date even after exceedingly stringent filtering threshold of high-confidence peptides and across replicates. In addition, our functional analysis and quantification of the data set, as well as the constructed microtubule interaction network, would help system-level understanding of its biological significance.

References

1. Calligaris, D., et al., *Microtubule targeting agents: from biophysics to proteomics.* Cellular and Molecular Life Sciences, 2010. **67**(7): p. 1089-1104.
2. Li, H.L., et al., *Microtubule structure at 8 angstrom Resolution.* Structure, 2002. **10**(10): p. 1317-1328.
3. Nogales, E., S.G. Wolf, and K.H. Downing, *Structure of the alpha beta tubulin dimer by electron crystallography.* Nature, 1998. **391**(6663): p. 199-203.
4. Mandelkow, E. and E.M. Mandelkow, *Microtubules and Microtubule-Associated Proteins.* Current Opinion in Cell Biology, 1995. **7**(1): p. 72-81.
5. Alberts, B., *Molecular biology of the cell*. 4th ed. 2002, New York: Garland Science. xxxiv, 1548 p.
6. Alberts, B., *Molecular biology of the cell*. 3rd ed. 1994, New York: Garland Pub. xliii, 1294, 67 p.
7. Bhat, K.M.R. and V. Setaluri, *Microtubule-associated proteins as targets in cancer chemotherapy.* Clinical Cancer Research, 2007. **13**(10): p. 2849-2854.
8. Martello, L.A., et al., *Elevated levels of microtubule destabilizing factors in a Taxol-resistant/dependent A549 cell line with an alpha-tubulin mutation.* Cancer Research, 2003. **63**(6): p. 1207-1213.
9. Vallee, R.B., *A Taxol-Dependent Procedure for the Isolation of Microtubules and Microtubule-Associated Proteins (Maps).* Journal of Cell Biology, 1982. **92**(2): p. 435-442.
10. Vallee, R.B. and C.A. Collins, *Purification of Microtubules and Microtubule-Associated Proteins from Sea-Urchin Eggs and Cultured-Mammalian-Cells Using Taxol, and Use of Exogenous Taxol-Stabilized Brain Microtubules for Purifying Microtubule-Associated Proteins.* Methods in Enzymology, 1986. **134**: p. 116-127.
11. Vallee, R.B., *Purification of Brain Microtubules and Microtubule-Associated Protein-1 Using Taxol.* Methods in Enzymology, 1986. **134**: p. 104-115.
12. Bluhm, W.F., et al., *Specific heat shock proteins protect microtubules during simulated ischemia in cardiac myocytes.* American Journal of Physiology-Heart and Circulatory Physiology, 1998. **44**(6): p. H2243-H2249.
13. Petrasek, J., et al., *Heat-shock protein 90 is associated with microtubules in tobacco cells.* Protoplasma, 1998. **202**(3-4): p. 161-174.
14. Patel, P.C., et al., *Proteomic Analysis of Microtubule-associated Proteins during Macrophage Activation.* Molecular & Cellular Proteomics, 2009. **8**(11): p. 2500-2514.
15. Kozielski, F., et al., *Proteome analysis of microtubule-associated proteins and their interacting partners from mammalian brain.* Amino Acids, 2011. **41**(2): p. 363-385.
16. Gache, V., et al., *Xenopus Meiotic Microtubule-Associated Interactome.* PLoS One, 2010. **5**(2): p. -.
17. Sakamoto, T., et al., *Mass spectrometric analysis of microtubule co-sedimented proteins from rat brain.* Genes to Cells, 2008. **13**(4): p. 295-312.

18. Hughes, J.R., et al., *A microtubule interactome: Complexes with roles in cell cycle and mitosis.* Plos Biology, 2008. **6**(4): p. 785-795.
19. Chuong, S.D.X., et al., *Large-scale identification of tubulin-binding proteins provides insight on subcellular trafficking, metabolic channeling, and signaling in plant cells.* Molecular & Cellular Proteomics, 2004. **3**(10): p. 970-983.
20. O'Connell, P.A., et al., *Characterization of the microtubule proteome during post-diapause development of Artemia franciscana.* Biochimica Et Biophysica Acta-Proteins and Proteomics, 2006. **1764**(5): p. 920-928.
21. Gache, V., et al., *Identification of proteins binding the native tubulin dimer.* Biochemical and Biophysical Research Communications, 2005. **327**(1): p. 35-42.
22. Liska, A.J., et al., *Homology-based functional proteomics by mass spectrometry: Application to the Xenopus microtubule-associated proteome.* Proteomics, 2004. **4**(9): p. 2707-2721.
23. Gilar, M., et al., *Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions.* Journal of Separation Science, 2005. **28**(14): p. 1694-1703.
24. Shilov, I.V., et al., *The paragon algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra.* Molecular & Cellular Proteomics, 2007. **6**(9): p. 1638-1655.
25. Tang, W.H., I.V. Shilov, and S.L. Seymour, *Nonlinear fitting method for determining local false discovery rates from decoy database searches.* Journal of Proteome Research, 2008. **7**(9): p. 3661-3667.
26. Elias, J.E. and S.P. Gygi, *Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.* Nature Methods, 2007. **4**(3): p. 207-214.
27. Ashburner, M., et al., *Gene Ontology: tool for the unification of biology.* Nature Genetics, 2000. **25**(1): p. 25-29.
28. Harris, M.A., et al., *The Gene Ontology (GO) database and informatics resource.* Nucleic Acids Research, 2004. **32**: p. D258-D261.
29. Martin, A., et al., *BisoGenet: a new tool for gene network building, visualization and analysis.* BMC Bioinformatics, 2010. **11**: p. -.
30. Liu, H.J., et al., *Peptide mapping with liquid chromatography using a basic mobile phase.* Journal of Chromatography A, 2008. **1210**(1): p. 76-83.
31. Collins, C.A. and R.B. Vallee, *Temperature-Dependent Reversible Assembly of Taxol-Treated Microtubules.* Journal of Cell Biology, 1987. **105**(6): p. 2847-2854.
32. Murofushi, H., et al., *Purification and Characterization of a 190-Kd Microtubule-Associated Protein from Bovine Adrenal-Cortex.* Journal of Cell Biology, 1986. **103**(5): p. 1911-1919.
33. Washburn, M.P., D. Wolters, and J.R. Yates, *Large-scale analysis of the yeast proteome by multidimensional protein identification technology.* Nature Biotechnology, 2001. **19**(3): p. 242-247.

34. Peng, J.M., et al., *Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: The yeast proteome.* Journal of Proteome Research, 2003. **2**(1): p. 43-50.

Chapter 3

TWO DIMENSIONAL REVERSED-PHASE LC-MS/MS FOR DETECTION OF ACIDIC PEPTIDES

3.1 Introduction

With the development of high throughput mass spectrometry and high efficiency nanoscale liquid chromatography (nanoLC), LC-MS/MS has become the preferred peptide sequence analysis approach for proteome-scale comprehensive characterization of biologically-interesting protein complexes. On such basis, biological replicates are required to avoid the downfalls associated with under-sampling for determination of convincing interacting proteins within the targeted protein complex [1]. To date, the hurdles that obstruct complete proteomic characterization of complicated biological extracts are mainly due to the technical capabilities of current mass spectrometry instruments in terms of detection limit and dynamic range. The other hindrance associated with LC-MS/MS analysis is that biological samples of moderate complexity contain thousands of tryptic peptides at least and no single LC separation method prior to electrospray is capable of resolving so many components in a typical LC gradient [2].

The success for comprehensive analysis of whole proteomes resides in the sample preparation step as well as later LC-MS/MS analysis step. Successful sample preparation involves reducing sample complexity and removing solvent contaminants for the following mass spectrometer analysis. With regard to online LC-MS/MS analysis, the major analytical constraint is competitive ionization of co-eluting peptides and instrument limitation of MS/MS duty cycle. To circumvent both problems, multidimensional fractionation has been explored to assist thorough analysis of biological protein complexes extracts as well as its proteolytic peptide mixtures [3]. For protein-scale

separation, 2D-PAGE [4] had been widely used prior to mass spectrometry analysis. For peptide-level analysis, strong cation-exchange (SCX) [5], reversed-phase (RP) chromatography [6], ion-pairing reversed-phase (IP-RP) chromatography [7], and isoelectric focusing (IEF) [8] have been utilized as first dimension separation methods. These are usually coupled with second dimension online RP-HPLC for shotgun proteomics. The success of multidimensional separation schemes is based on two major factors: separation orthogonality of different dimensions and separation efficiency of the system employed in each dimension.

Recently Gilar *et al.* introduced a multidimensional fractionation scheme by using two stages of reversed phase separation under high and low pH conditions [2]. This method provides high peak capacity in each RP dimension, and they found greatest peptide separation orthogonality (comparable to 2D SCX-RP-HPLC setup) under pH 10 in the first dimension and pH 2.6 in the second dimension [2, 9]. Other advantages of this method are the compatibility with MS analysis owning to salt-free mobile phases and convenient sample handling through first dimension offline fractionation. Follow-up study with similar pH condition setup but alternation of ion-pairing mobile phases [10] showed that the elution behavior of peptides under first dimension high-pH RP is strongly dependent on its hydrophobic property; the underlying mechanism of peptide separation with regard to its pI property of this 2D RP-HPLC method has not been thoroughly explored.

In this work, we sought to use this two dimensional reversed phase high performance liquid chromatography (2D-RP-HPLC) separation of peptides to achieve comprehensive analysis of microtubule interacting proteins in order to increase our understanding of how microtubule activity is regulated. This 2D-RP-HPLC method is composed of a first offline RP-HPLC fractionation of peptides under high-pH conditions and a subsequent second online RP-HPLC separation of first dimension fractions under low-pH conditions. The second dimension solvents are compatible with electrospray MS analysis.

Background about microtubules (MTs), microtubule-associated proteins (MAPs) and microbubule-interacting complex had been described in previous chapter **2.1**.

In 2004, Chuong *et al.* identified 122 microtubule-binding proteins extracted from plant cells using affinity chromatography with two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and quadrupole time-of-flight (Q-TOF) instrument [11]. In 2009, Patel *et al.* claimed 409 proteins binding directly or indirectly to microtubules from Taxol extract of mouse macrophage cell line using iontrap instrument [12]. They used first dimension RP-HPLC offline fractionation at the protein level, followed by online RP-HPLC separation at the peptide level, both under low-pH condition [12]. In 2011, Kozielski *et al.* reported 573 proteins (actually 571 because of a mistake) extracted from bovine brain by Taxol method and using GeLC fractionation method and LC/MSD Trap XCT Ultra ion trap instrument [13]. All of these studies utilized the Taxol approach for the isolation of microtubule-interacting complex but none of them used 2D-RP-HPLC

peptide separation strategy. At this point, information regarding human microtubule interactome is still limited.

For our work, we isolated microtubule-binding protein complex from human cancer cells using the well-established Taxol method (n= three biological replicates). Using these samples as a starting point, we investigated the approach of first dimension offline high-pH RP-HPLC peptide fractionation coupling with second dimension online low-pH RP-HPLC separation for peptide analysis. We evaluated the peptide separation efficiency (resolving power) by this 2D-RP-HPLC method, and assessed the reproducibility of this method on a biological complex sample basis. In addition, peptide separation orthogonality as well as peptide retention behavior of this approach was investigated. For each RP dimension, we found correlation of peptide retention with respect to its isoelectric point (pI) property and experimental pH condition setup, which has not been thoroughly studied. This confirms our hypothesis that peptide elution behavior under RP system is influenced by the combinational effect of hydrophobicity and pI property. With three biological replicates after stringent filter, we identified 945 proteins in total and 453 proteins in common within each biological replicate.

Our result demonstrates the improved protein identifications of microtubule interacting proteome comparing to similar studies [11-16] with other protein separation methods. This indicates the high peptide separation efficiency by 2D-RP-HPLC method and implies the efficacy of this method for comprehensive proteomic study. In addition, this method could potentially be applied for separation and detection of acidic peptides.

3.2 Materials and Methods

3.2.1 Materials

As described in previous chapter **2.2.1**.

3.2.2 Cultivation of Human Jurkat T Cells

As described in previous chapter **2.2.2**.

3.2.3 Isolation of Microtubule-Binding Proteins

As described in previous chapter **2.2.3**.

3.2.4 Preparation of Tryptic Digests

As described in previous chapter **2.2.4**.

3.2.5 High-pH First Dimension RP-HPLC Separation

As described in previous chapter **2.2.5**.

3.2.6 Nanoflow RP-HPLC-MS/MS Analysis

As described in previous chapter **2.2.6**.

3.2.7 Bioinformatic Analysis

Software searching of MS data for protein and peptide identifications is described in previous chapter **2.2.7**.

To determine the elution trend for peptides during the high-pH reversed-phase HPLC separation, average pI value and average GRAVY (Grand Average of Hydropathy) [17] value for identified non-redundant peptides (with no modifications) within each HPLC fraction were calculated. The reason for choosing peptides without modifications is the limitation of available software when performing such calculation. Calculation of pI values was conducted using a pI calculator developed by Gauci *et al.* [18] using ExPASy's pKa set for amino acids. Calculation of GRAVY values was through a web-based GRAVY calculator from *http://gravy.laborfrust.de/* [19].

3.3 Results

We sought to utilize RP/RP HPLC for the large-scale analysis of the human MT interactome due to the expected benefits of this separation method versus SCX-RP [3, 20]. Microtubule-binding proteins represent a functionally diverse set of proteins with different physiochemical properties that are directly or indirectly modulated by the microtubule network. While less complex than a whole cell lysate, this multi-protein sample would still generate a tryptic digest complex enough to challenge the separation capability of multidimensional chromatography coupled to tandem mass spectrometry. So using this sample, we were able to evaluate the performance of 2D-RP-HPLC in shotgun proteomic analysis as well as extend our knowledge concerning the composition of the human MT interactome.

3.3.1 Two Dimensional RP-HPLC-MS/MS Analysis Results in the Identification of 945 Microtubule-Binding Proteins.

Details about high-pH peptide fractionation (shown in Figure 2.5) and low-pH online MS/MS analysis had been described in previous chapter **2.3.1**. Analysis of the three biological replicates resulted in the identification of 14,231 non-redundant peptides that represent 945 proteins. Four hundred fifty-three common proteins remain when we removed proteins found in only one or two replicates. These 453 common proteins represent the most biological-robust data set due to the application of very strict filtering criteria across the three biological replicates.

3.3.2 High-pH RP-HPLC Fractionation is High Quality and Consistent.

Evaluating the elution behavior of all identified peptides allowed us to assess the quality of the first dimension separation. As shown in Figure 3.1, there are no gross irregularities in the distribution of identified peptides across the three replicates; and in general, peptide identifications smoothly increased with the most identifications clustered in fractions 26-35. Also, the average number of peptide identifications is relatively consistent across the three replicates. We next determined whether a given peptide was found in more than one high pH HPLC fraction based on ProteinPilot peptide identification results. As indicated in Figures 3.1 and 3.2, the resolving capability of the high pH separation is high. The majority (~71%) of the identified peptides was eluted in a single fraction and an additional 24% were eluted in just two adjacent fractions. Closer

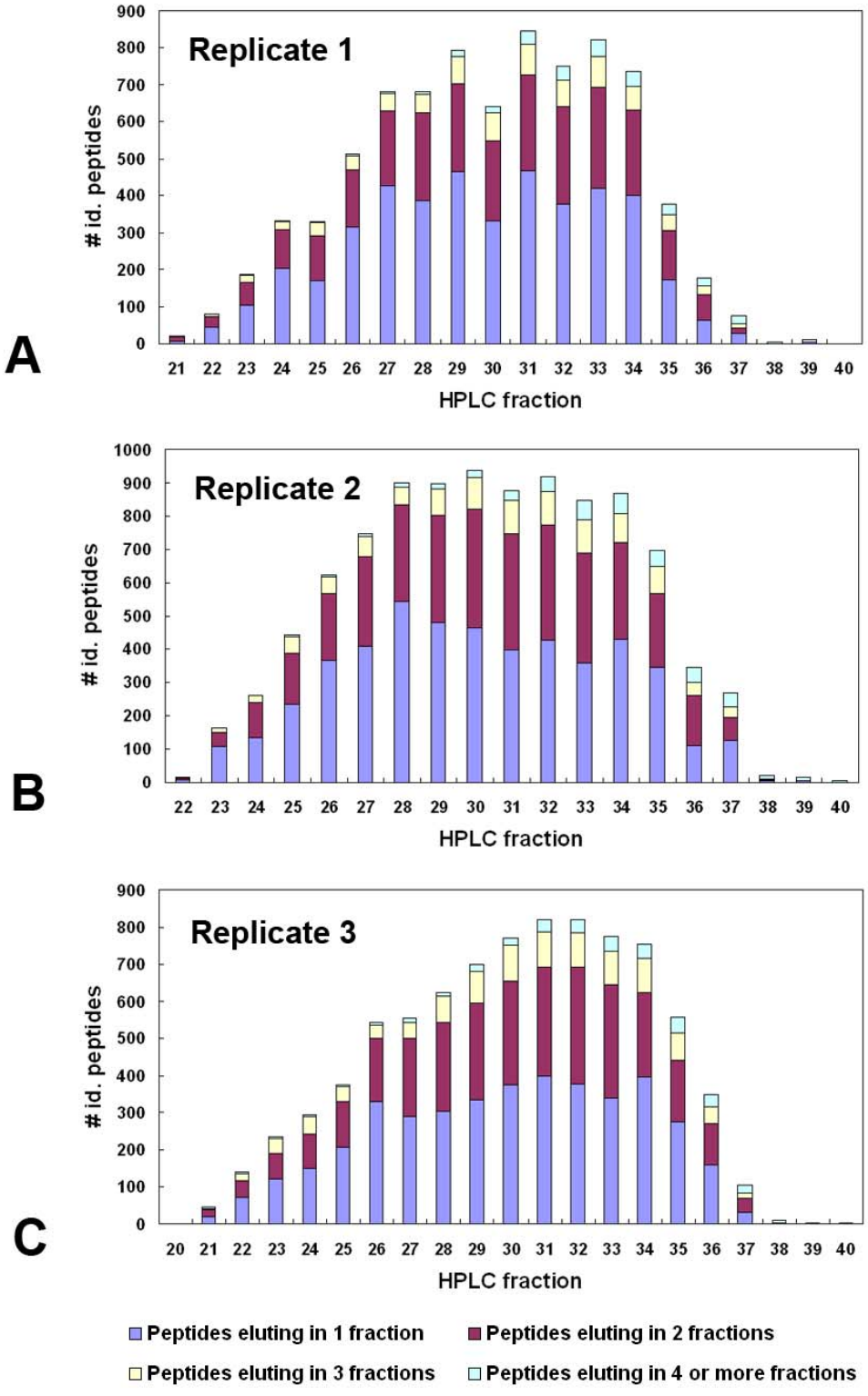


Figure 3.1. Distribution of identified peptides in the first dimension high-pH RP-HPLC separations.

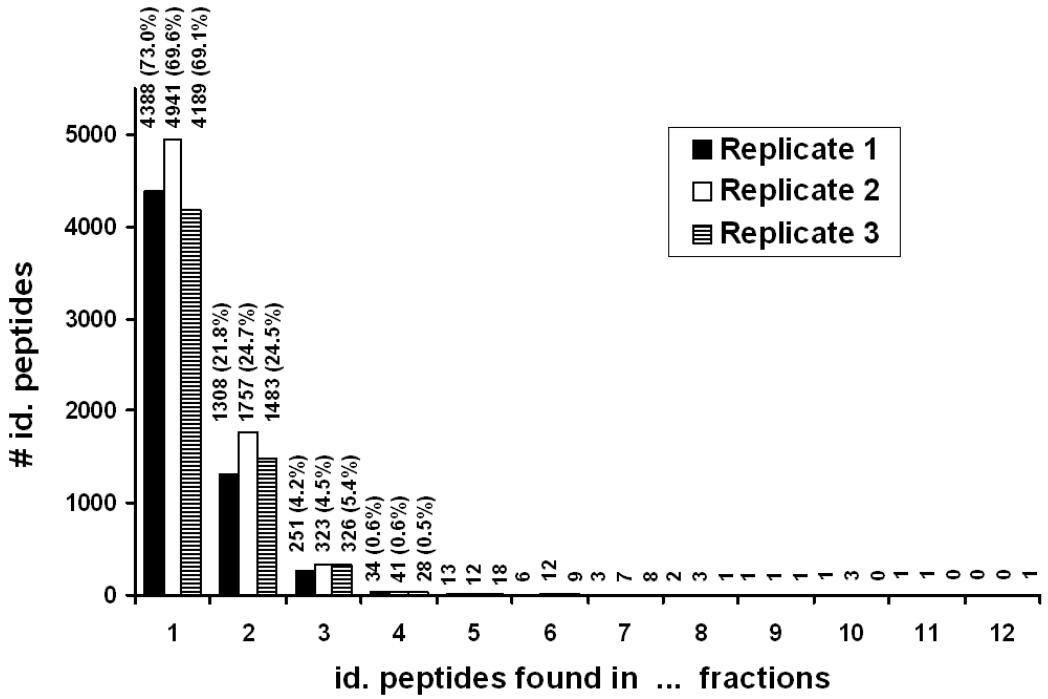


Figure 3.2. First dimension high-pH RP-HPLC peptide separation resolving power summary.

examination of peptide distribution revealed that only 485 peptides (~3% of total peptides identified) elute in one unique fraction and were found in all three replicates. One hundred sixty-seven of these (about 1/3) were consistently identified in the same exact fraction across three replicates. Nearly all of the remaining peptides were identified in adjacent fractions (±1) from replicate to replicate. Again, this is an anticipated result considering the irreproducibility of shotgun proteomic analysis using data-dependent control of MS/MS sampling. A final evaluation of five analytical characteristics (experimental precursor mass error, first dimension fraction number, second dimension retention time, ion counts, and peak width) for twenty of these peptide ions is presented

in Table 3.1. The twenty peptides were exclusively found in one unique fraction during the first dimension separation and were reproducibly identified in the same fraction in each independent replicate.

Table 3.1. Representative peptides from 2D-RP-HPLC separation.

Reproducibility of Run to Run analysis for Offline High-pH HPLC & Online Low-pH HPLC ESI-MS/MS approach

Peptide ID	pI	GRAVY	Offline High-pH Fraction	Precursor m/z	Mass Error (ppm)	Online Low-pH RT (min)	Intensity XIC area	Peak Width (min)
FSPDGELYASGSEDGTLR	3.92	-0.75	26	950.9297 ± 0.0042	3.93	50.55 ± 0.61	2427 ~ 11768	0.30 ~ 0.41
NASNTEKLTDQVMQNPR	6.07	-1.47	27	649.3196 ± 0.0061	6.18	42.63 ± 0.20	572 ~ 3433	0.20 ~ 0.32
KMNVSPDVNYEELAR	4.68	-0.91	27	588.9612 ± 0.0027	6.02	48.36 ± 0.24	3958 ~ 13559	0.34 ~ 0.45
MKQEILEEVVR	4.78	-0.35	28	458.5906 ± 0.0008	8.82	47.80 ± 0.47	3354 ~ 12410	0.27 ~ 0.43
AELRDDPIISTHLAK	5.38	-0.33	28	560.3084 ± 0.0021	7.63	44.57 ± 0.29	8321 ~ 13093	0.28 ~ 0.38
EIVLADVIDNDSWR	3.84	-0.05	29	822.9159 ± 0.0033	2.62	62.65 ± 0.25	1053 ~ 17705	0.40 ~ 0.55
VGNIEIKDLMVGDEASELR	4.18	-0.12	29	696.6981 ± 0.0040	7.42	56.64 ± 0.44	1362 ~ 6354	0.28 ~ 0.45
LAEMPADSGYPAYLGAR	4.37	-0.13	30	891.4349 ± 0.0037	5.47	51.12 ± 1.13	6567 ~ 76722	0.45 ~ 0.95
IMGIPEEEQMGLLR	4.25	-0.04	30	808.4160 ± 0.0049	5.85	57.03 ± 1.01	4939 ~ 38114	0.33 ~ 0.70
NELSGALTGLTR	6.00	-0.11	31	616.3340 ± 0.0037	3.71	50.26 ± 0.60	5090 ~ 8766	0.30 ~ 0.40
VTPPEGYEVVTVFPK	4.53	0.05	31	831.4368 ± 0.0043	7.36	54.77 ± 0.53	9458 ~ 12258	0.40 ~ 0.50
FLSDPQVHTVLVER	5.32	0.12	32	547.2987 ± 0.0024	2.95	47.50 ± 0.13	3486 ~ 14375	0.32 ~ 0.49
GSYGDLGGPIITTQVTIPK	5.83	0.08	32	959.0177 ± 0.0036	2.98	55.51 ± 0.10	2881 ~ 8238	0.30 ~ 0.36
ALGFPLERPK	8.79	-0.33	33	564.3316 ± 0.0022	2.57	47.29 ± 0.14	16338 ~ 33038	0.26 ~ 0.40
SSFYVNGLTLGGQK	8.31	-0.08	33	735.8819 ± 0.0004	1.42	52.91 ± 0.02	9127 ~ 19559	0.31 ~ 0.45
SKDDQVTVIGAGVTLHEA--LAAAELLKK	5.46	0.22	34	695.1463 ± 0.0074	10.76	64.44 ± 1.45	3756 ~ 6442	0.60 ~ 1.00
QIIQQNPSLLPALLQQI	5.52	0.32	35	959.0617 ± 0.0037	2.77	71.88 ± 1.97	3840 ~ 13094	0.35 ~ 0.48
SNEILTAIIQGMR	5.72	0.32	36	723.3985 ± 0.0027	8.11	66.90 ± 1.90	3960 ~ 5221	0.40 ~ 0.60
LVSSLTSGLLTIGDR	5.84	0.75	37	766.4353 ± 0.0064	6.53	59.80 ± 1.28	1746 ~ 16460	0.41 ~ 0.48
LISQIVSSITA	5.52	1.52	37	566.3292 ± 0.0034	9.58	60.37 ± 1.17	3538 ~ 12863	0.30 ~ 0.41

Twenty representative peptides that were only identified in one fraction and the same exact fraction during first dimension high-pH RP-HPLC fractionation, their calculated pI, GRAVY value, experimental precursor m/z and mass error, first dimension fraction number, second dimension online HPLC retention time, intensity of XIC area, and peak width were summarized.

3.3.3 2D RP-RP-HPLC-MS/MS Methodology is Reproducible and Orthogonal.

Further examination of the second dimension chromatographic elution time of representative peptide ions from Table 3.1 showed reproducible retention time for each peptide during the low-pH on-line RP-HPLC separation as well. Along with the results from the first dimension fractionation, this 2D RP-HPLC method provides robust separation of peptides before MS/MS analysis. However, the elution behaviors of peptides differ depending upon pH at which the separation is conducted, as illustrated in Figure 3.3 with low correlation value ($R^2 = 0.422$). This result suggests the orthogonality

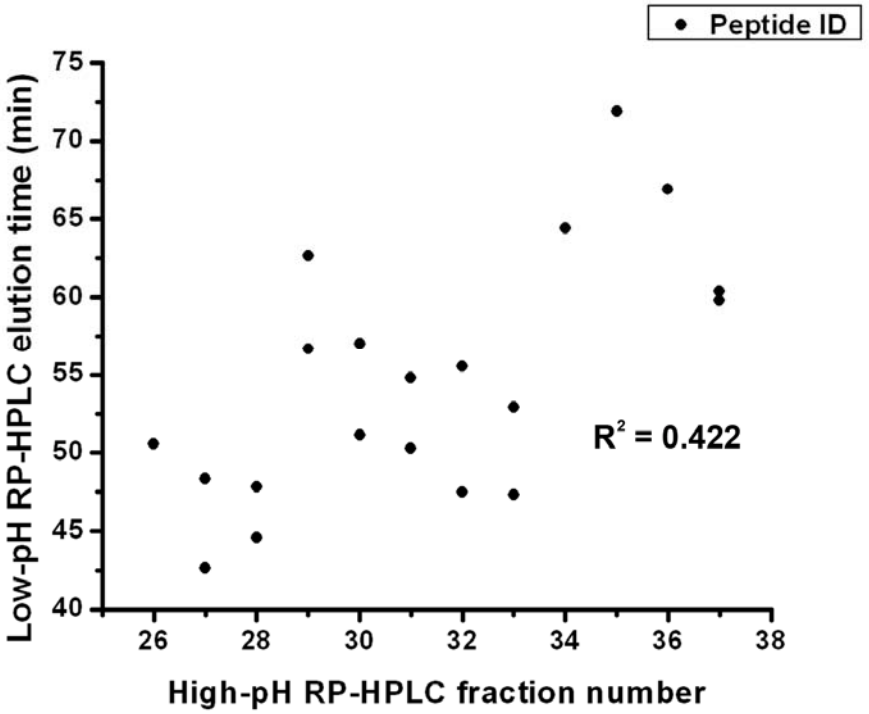


Figure 3.3. Orthogonality between each dimension retention times from 2D-RP-HPLC separation of representative peptides.

of peptide separation using this 2D RP-HPLC method with different pH setup, which is favorable for shotgun proteomic analysis in reducing sample complexity and resolving otherwise co-eluting peptides under just 1D RP-HPLC.

3.3.4 Selectivity of the 2D-RP-HPLC Peptide Separation is based upon both Hydrophobicity and pI Property.

Although RP-HPLC had been extensively studied and ubiquitously implemented in large-scale proteomic peptide analyse, the retention mechanism is not completely understood. Three models, namely the hydrophobic theory (solvophobic theory) [21], partition theory [22] and adsorption theory [22] had been proposed to describe the retention behavior, each with its strong points and weak points. In regards to peptides, their retention in RP-HPLC is most likely due to a combination of many different interactions simultaneously. Hydrogen-bonding interaction [23] and hydrophobic interaction [21] could be used to explain peptide retention behavior under RP-HPLC. The hydrogen bond is an interaction between two electronegative atoms, donor and acceptor, through an intermediate atom of hydrogen. The hydrogen donor group is composed of a highly electronegative atom and a hydrogen atom; the hydrogen acceptor group contains an electronegative atom with a lone pair of electrons [24]. The resulting hydrogen bond is a result of the partially positively charged H atom from the donor group being attracted to the lone electron pair from acceptor group [24]. The peptide backbone could not form hydrogen-bonding because the fully-bonded C atom is not electronegative and has no lone electron pairs.

To describe peptide retention in RP system, peptides were initially introduced into the column and adsorbed onto C18 packing materials due to hydrophobic interaction between C18 alkyl chains and peptide backbone. Meanwhile, strong hydrogen-bonding interactions exist between water molecules (from mobile phase) and polar groups of peptides (N-terminal, C-terminal, free amine side chain, etc). These hydrogen-bonding interactions would compete with existing hydrophobic interactions to elute a pepetide off the column. Typical peptide separation in RP system is achieved by gradient elution with increasing acetonitrile proportion in the mobile phase. Acetonitrile tends to associate with peptides and, upon forming layers of acetonitrile molecules between a peptide and packing materials, the hydrophobic interaction would be weakened. When the acetontrile concentration reaches a critical level, the competing hydrogen-bonding interactions prevails over the weakened hydrophobic interactions and peptides are eluted off the column. Because different peptides have varying hydrophobicities, their interaction strength with C18 packing material varies and more hydrophobic peptides would be eluted at higher acetonitrile percentage during the gradient.

Experimental pH conditions and peptide isoelectric point (pI), which is the characteristic pH for the neutral state peptide, could also influence the peptide retention in RP-HPLC system. If the peptide pI is significantly lower or higher than the mobile phase pH, it would be either positively or negatively charged, respectively. Water molecules tend to cluster around ions for the dissolution of an ionic substance. Similarly, peptides with a charge would be associated with water molecules. In such case,

hydrogen-bonding interaction between water molecules in the mobile phase and water molecules surrounding the peptides would cause the charged peptides to elute earlier than its neutral form. The greater the difference between peptide pI and experimental pH, the earlier such peptide would elute.

In order to confirm this hypothesis, we investigated peptide retention against grand average of hydropathy (GRAVY) [17] which is a measure of intrinsic peptide hydrophobicity and peptide pI value under both high-pH and low-pH conditions. The average GRAVY and pI values of the identified peptides found in each fraction (with at least 20 peptide IDs) were calculated and plotted as a function of first dimension fraction number (Figure 3.4). Tryptic peptides without any modifications were used to calculate the GRAVY and pI values. Under high-pH conditions, both average values show a positive correlation with time during the course of the first dimension gradient separation. However, GRAVY values exhibit a stronger positive correlation with increasing fraction number, which is an indication of retention time.

To further explore this phenomenon, the distribution of second dimension low-pH RP-HPLC retention times of identified peptides from three representative first dimension fractions (27, 31 and 34) were plotted against their GRAVY and pI values (Figure 3.5). Despite the wider elution time range of peptides in replicate 1 due to a mechanical problem with the LC at that time, the elution time ranges are consistent for replicate 2 and 3. The peptide elution trend under low-pH condition reluctantly showed a positive

correlation between GRAVY value and retention time. Interestingly, the trend of peptides eluting with decreasing pI value was much clearer under low-pH condition. These results

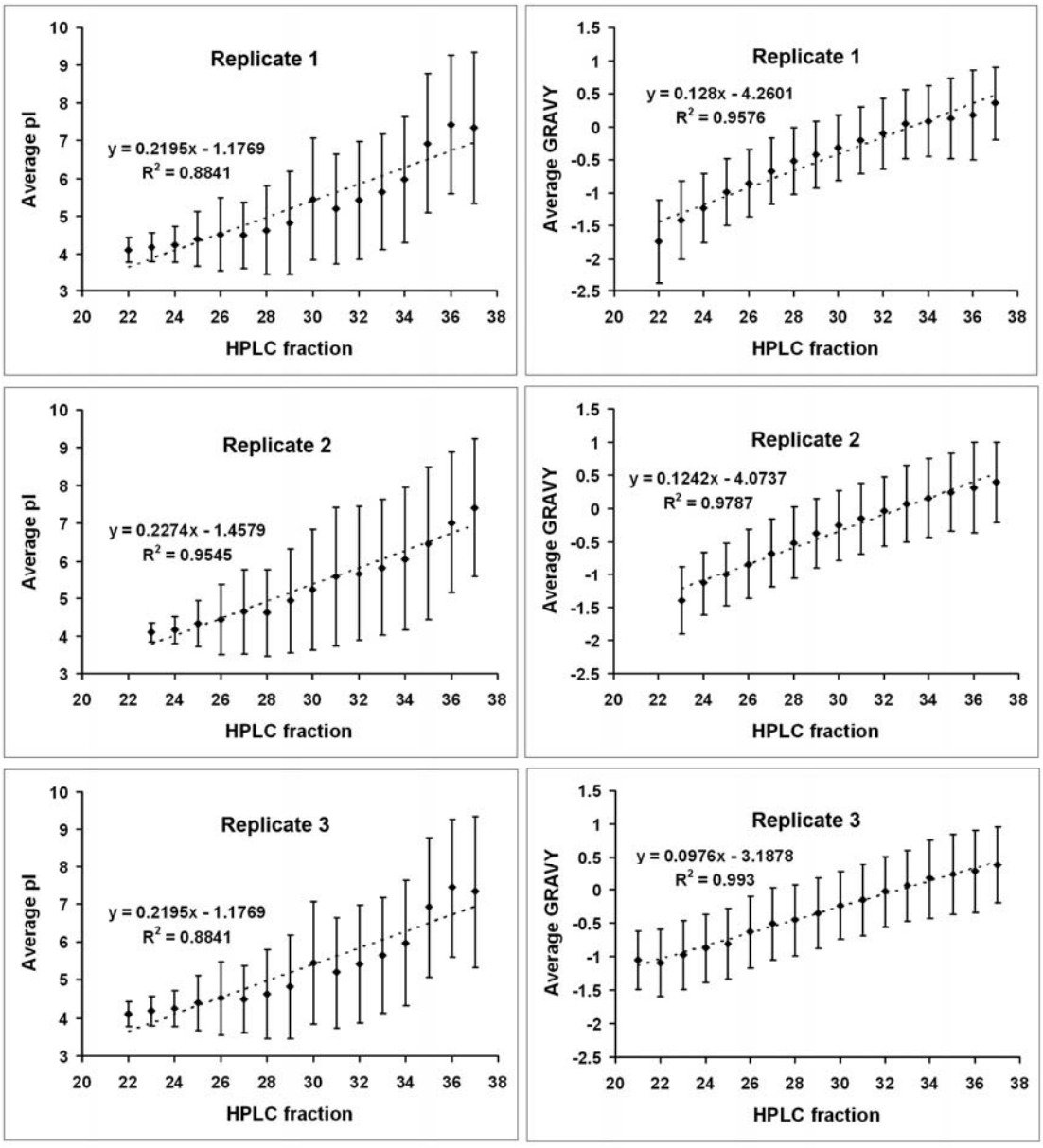


Figure 3.4. Correlation of peptide pI and GRAVY properties to first dimension high-pH RP-HPLC fraction number. For each first dimension high-pH RP-HPLC fraction, pI value and GRAVY value were calculated for identified peptides that have no modifications. The average pI and GRAVY for all peptides within each fraction were plotted against first dimension HPLC fraction number determined by retention time.

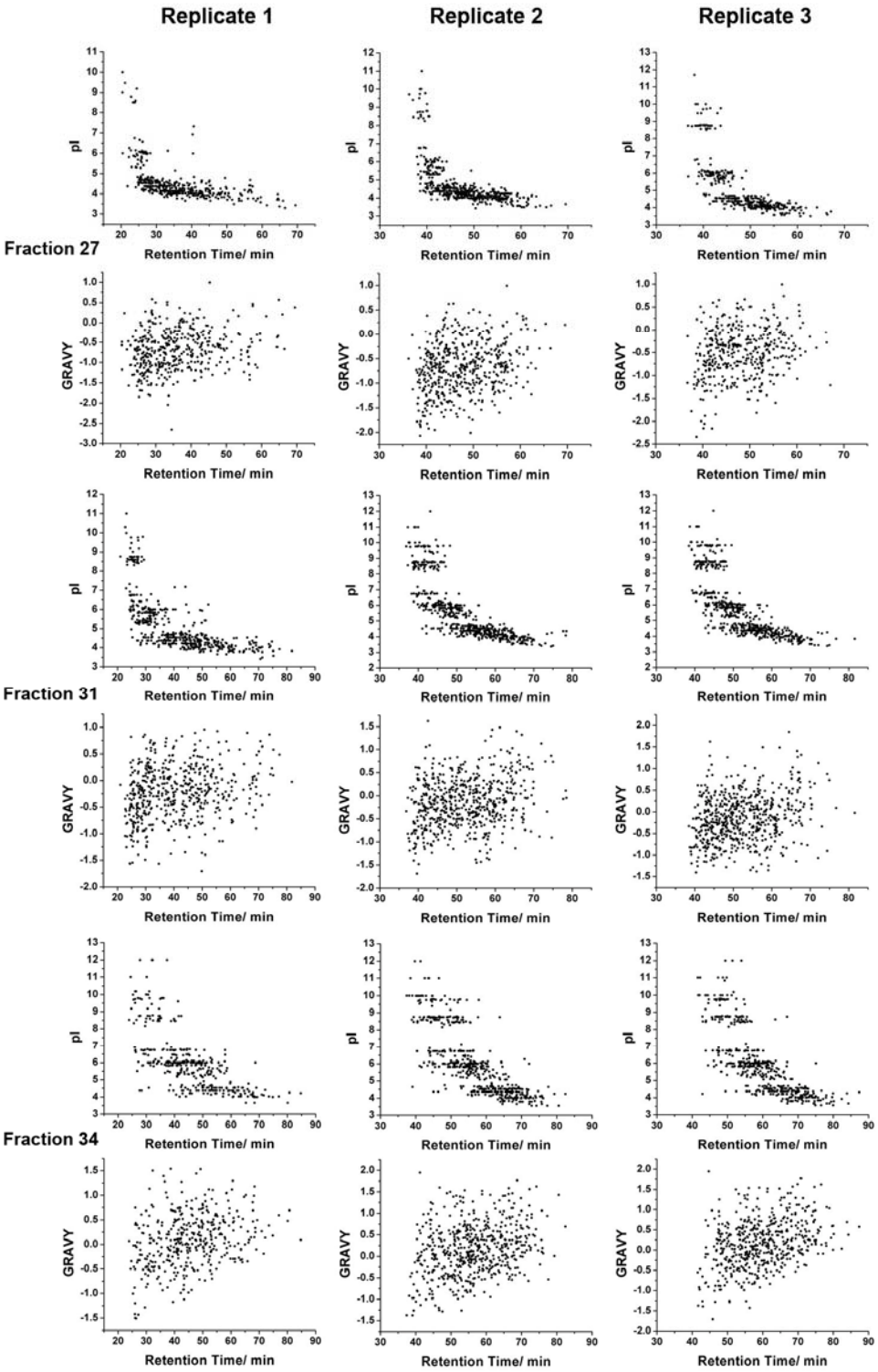


Figure 3.5. Correlation of peptide pI and GRAVY from first dimension fractions 27, 31 and 34 to peptide retention time during second dimension low-pH RP-HPLC separation.

indicate that peptide separation under RP-HPLC system is based upon both hydrophobicity and pI. While the understanding of peptide elution with increasing GRAVY value is straightforward, the influence of peptide pI value under specific pH conditions is consistent with our predication. The difference between the peptide pI and the pH of the separation is a measure of peptide chargeablity extent. Under the high-pH (10.4) condition, almost all peptides tend to be negatively charged and the chargeable extent of the peptide should increase with decreasing pI. So peptides with lower pI would carry more negative charge and they would be associated with more water molecules. Because of the hydrogen-bonding interaction of these water molecules with the water molecules in the mobile phase, these peptides would elute earlier under high-pH. On the contrary, under low-pH (2.8 for 0.1% formic acid) condition, almost all peptides tend to be positively charged and the chargeable extent increase with increasing pI. Therefore, this is the reason that peptides with higher pI elute earlier under low-pH condition.

3.3.5 Peptide Retention under Basic Conditions Increased with Decreasing Negative Charge.

We next consider how the first dimension fractionation was influenced by amino acid composition. Seven characteristics per peptide were considered: 1. number of polar amino acids; 2. number of nonpolar amino acids; 3. number of basic amino acids in terms of arginine (Arg), histidine (His) and lysine (Lys); 4. number of acidic amino acids in terms of aspartic acid (Asp) and glutamic acid (Glu); 5. peptide length in terms of amino acid number; 6. missed cleavage site by trypsin; 7. net charge under first dimension high-

pH conditions (pH 10.4). The average values of these characteristics were plotted according to the first dimension fraction number as shown in Figure 3.6. First, there is no apparent influence of peptide length nor missed cleavage sites on peptide retention under high pH since they are nearly constant across all fractions. However, peptide retention increased as the number of nonpolar amino acid increased. This is consistent with the hydrophobic property influence we examined earlier because polar amino acids are more hydrophilic while nonpolar amino acids are more hydrophobic. For the overall increasing hydrophobicity trend with first dimension RP-HPLC fractions, therefore, we would anticipate that early eluting peptides would contain more hydrophilic amino acids and correspondingly peptides that elute later should contain more hydrophobic amino acids. Lastly, during the first dimension separation, the overall charge of the peptides became less negative. This is due to the pH 10.4 separation conditions in the first dimension, peptide in the high pH solvent would become negatively charged and the peptides with lower pI would carry more negative charge. Net charge of the peptide would decrease its hydrophobic interaction with the reversed-phase packing material and increase hydrogen-bonding with the mobile phase. Thus causes the peptide to elute in early fractions. The contributions of net charge for a peptide would come from both its basic amino acids and acidic amino acids. Interestingly, the average number of basic amino acids per peptide across the first dimension fractions is relatively constant around 1.5. This phenomenon is most likely due to the use of trypsin to generate peptides with an arginine or lysine at the peptide's C-terminal. As a result, the decreasing trend of acidic amino acids per peptide

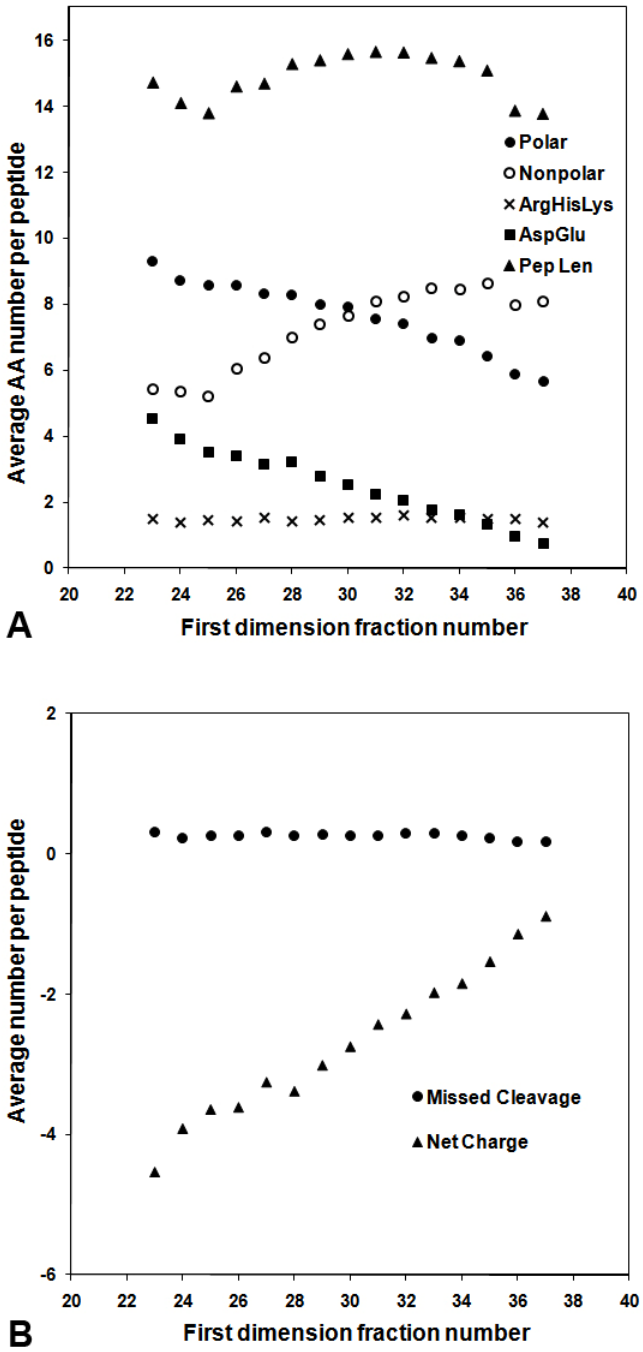


Figure 3.6. Peptide characteristics as a function of first dimension fraction number. Representative plot showing (A) Average value per peptide in terms of polar amino acids, nonpolar amino acids, basic amino acids of Arg, His and Lys, acidic amino acids of Asp and Glu, peptide length; (B) Average value per peptide in terms of missed cleavage sites and net charge under first dimension pH 10.4 condition.

across first dimension fractions illustrates its contribution to decreasing net charge (negative charge). The same trend of above described correlations is observed in all three replicate analyses.

3.4 Discussion

3.4.1 Increased Proteomic Identification of Microtubule Interactome via RP/RP-MS/MS.

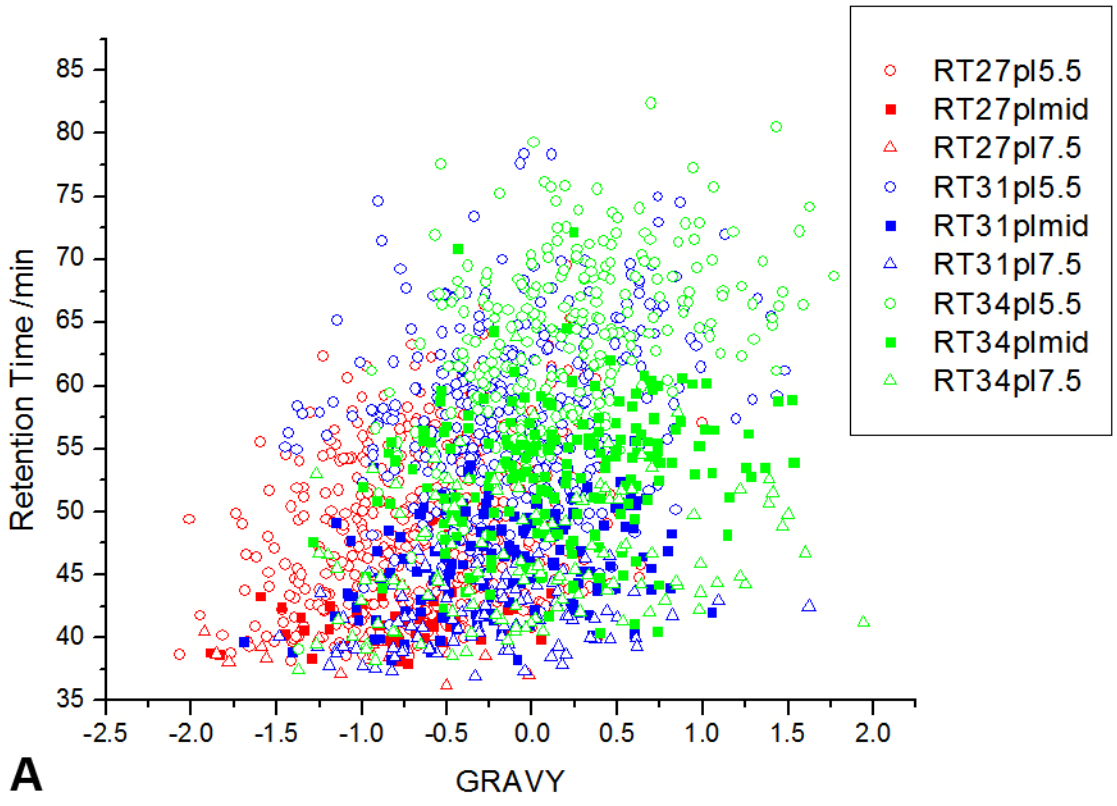
Our RP/RP-MS/MS analysis reveals that the human microtubule interactome contains 945 proteins in total and 453 in common after stringent filter of three biological replicate results. Compared to previous studies **[11-16]** that used different separation methods, our result also demonstrates increased proteomic identification for the same biological complex. This is mostly likely due to the high resolution, and orthogonal separation of peptides during each dimension of the 2D-RP-HPLC method. The overall result is an increase in the number of peptide identifications and protein identifications.

Identification of low abundance proteins in a complex biological extract is usually challenging due to the ionization competition of their peptides with co-eluting peptides from higher abundance proteins. From our protein identifications result, if we consider proteins with only 2 or 3 unique peptide identifications as low abundance proteins, then the following is observed. For replicate 1, 247 (43%) proteins out of 572 total protein identifications are low abundance proteins. For replicates 2 and 3, low abundance proteins are 329 (42%) out of 777 total and 305 (42%) out of 720 total, respectively.

These low abundance protein identifications occur because of the high resolving power of 2D RP-HPLC separation of peptides prior to MS/MS analysis.

3.4.2 Combined Influence of pI and Hydrophobicity on Peptide Retention during RP/RP Separation.

As shown in Figure 3.7A and 3.7B, we can readily discover that for peptides with similar second dimension retention times (indicated by y-axis value) but different colors (red, blue, green indicate first dimension fraction 27, 31, 34), the advantage of this 2D-RP-HPLC setup is that it resolves co-eluting peptides during low-pH RP by the first dimension high pH (10.4) fractionation. This increases their possibility of being sampled for MS/MS analysis. Moreover, from Figure 3.7A, we observed overall increased



A

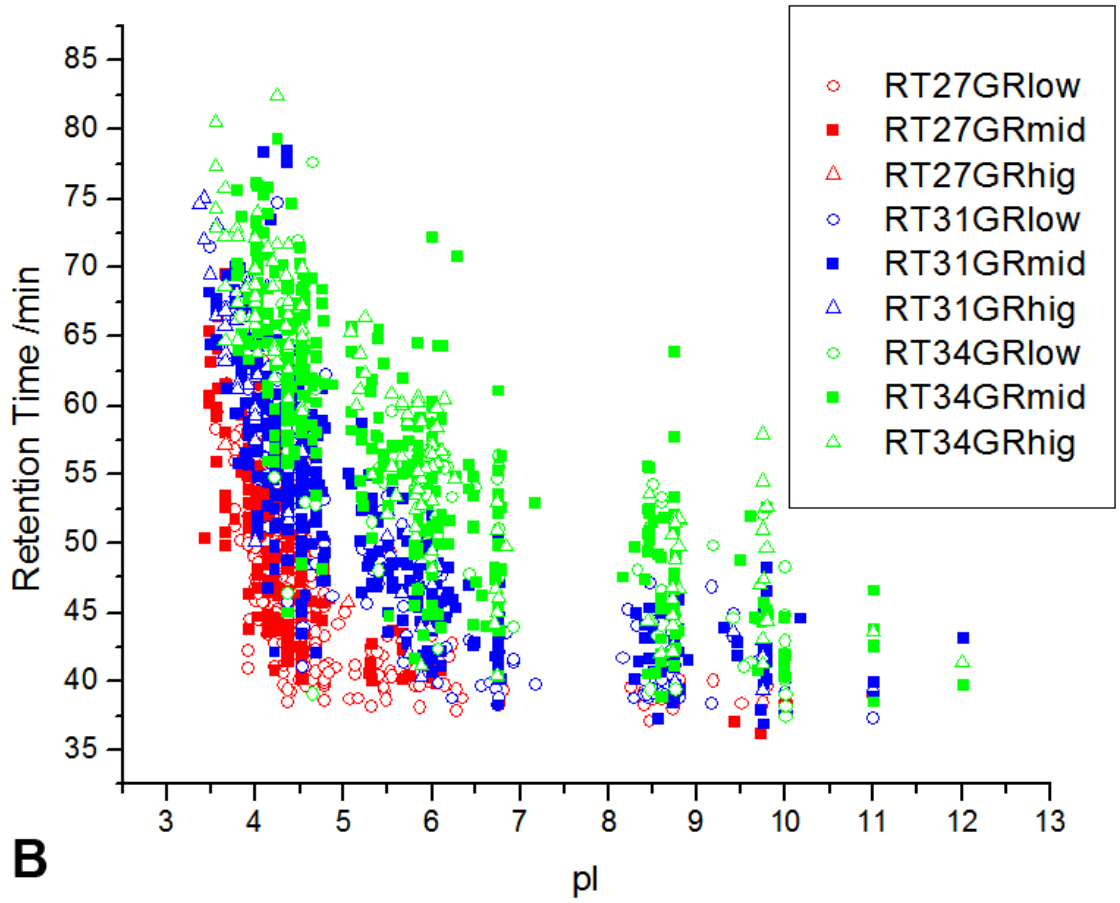


Figure 3.7. Combined influence of peptide pI and GRAVY on low-pH RP-HPLC retention time. (A) Second dimension retention time (y-axis) versus GRAVY of peptides (x-axis) from acidic to basic (different pI values). (B) Second dimension retention time (y-axis) versus pI of peptides (x-axis) from hydropholic to hydrophobic (different GRAVY values). Each dot in the plot represents one peptide ID. Peptides from first dimension fractions 27, 31, and 34 are indicated with different colors: red, blue and green. In figure (A), peptides are classified into 3 categories by pI values: <5.5 (acidic, circle symbol), 5.5 to 7.5 (neutral, solid square symbol), and >7.5 (basic, triangular symbol). In figure (B), peptides are classified into 3 categories by GRAVY values: <-0.5 (low, hydrophilic, circle symbol), -0.5 to 0.5 (medium, solid square symbol), >0.5 (high, hydrophobic, triangular symbol).

GRAVY from first dimension fractions 27 to 31, and to 34. Correspondingly, we also observed an overall increase in pI for these first dimension fractions. These phenomena

are consistent with peptide retention selectivity under first dimension high-pH conditions we described earlier.

In Figure 3.7, for the second dimension low-pH (pH 2.8) RP separation, pI influence (or GRAVY influence) on retention time of peptides with same GRAVY (or pI) value is indicated by different shape symbols. From Figure 3.7A, peptides with different pIs (represented by circle, square, triangular symbols) are more scattered, meaning pI variation influence would well resolve peptides with the same GRAVY value on the second dimension retention. From Figure 3.7B, peptides with different GRAVYs (represented by circle, square, triangular symbols) are not as scattered, means GRAVY variation influence would resolve less for peptides with same pI value for the second dimension RP-HPLC separation. In addition, influence of pI on peptides RT with the same GRAVY seems to resolve well on a large range of peptides with GRAVY from -1.5 to 1.5. Influence of GRAVY on peptides RT with same pI seems to only resolve well in a small range of peptides with pI from 3.5 to 5, and the resolving power is not as significant.

3.4.3 Application of 2D RP-RP-MS/MS for Separation and Detection of Acidic Peptides

One potential application of this 2D RP-RP-MS/MS method is the enrichment of highly acidic peptides in the early fractions during the first dimension high-pH fractionation, thus improving their detections by second dimension MS/MS analysis.

From Figure 3.4, we can observe that, peptides in the first 2-3 early fractions have very low pI value within a narrow range. We had applied this method to enrich for highly acidic C-terminal peptides of tubulin, which is crucial for differentiating tubulin isoforms and responsible for tubulin post-translational modifications (PTMs) [25]. Using this 2D RP-RP-MS/MS method for pepsin digest of tubulin, we successfully isolated and identified C-terminal peptides (SEAREDMAALEKDYEEVGVDSVEGEGEEEGEE, pI 3.38, GRAVY -1.303 and SEAREDMAALEKDYEEVGVDSVEGEGEEEGEEY, pI3.38, GRAVY -1.303) from the first 2 early fractions among a total of 20 first dimension fractions (data not shown). This result suggests the validity of efficient enrichment of acidic peptides by 2D RP-RP-MS/MS method.

3.5 Conclusion

In this study, the 2D-RP-HPLC approach has been investigated with respect to analyzing microtubule interactome by separating peptides under high pH condition for the first dimension and low pH for the second dimension. This method demonstrates high resolving power for the first dimension fractionation and shows orthogonality in terms of peptide separation between dimensions. Such separation efficiency is also consistent across independent biological replicates. With the suitability of this method for large-scale analysis of proteins, we have generated the most comprehensive microtubule interacting protein data set to date, which would help in understanding the biological functions of microtubule network. In addition, we examined peptide retention behavior in RP-HPLC and found peptide separation is influenced not only by its intrinsic

hydrophobicity, but also by its pI value. The experimental pH conditions determine the peptide chargeable extent and this influence could be explained using the solvation theory: water molecule hydration of charged peptide and hydrogen-bonding interaction. Such phenomena lay the foundation for this 2D-RP-HPLC approach with different pH setup to provide high resolution and orthogonal separation for peptides and thus make it a highly competitive choice for multidimensional protein identification technology (MUDPIT). Finally, for proteomic systems with focus on acidic peptides, this method could be useful for potential isolation and improving detection of interested peptides.

References

1. Ham, B.M., et al., *The influence of sample preparation and replicate analyses on HeLa cell phosphoproteome coverage.* Journal of Proteome Research, 2008. **7**(6): p. 2215-2221.
2. Gilar, M., et al., *Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions.* Journal of Separation Science, 2005. **28**(14): p. 1694-1703.
3. Washburn, M.P., D. Wolters, and J.R. Yates, *Large-scale analysis of the yeast proteome by multidimensional protein identification technology.* Nature Biotechnology, 2001. **19**(3): p. 242-247.
4. Hanash, S.M., *Biomedical applications of two-dimensional electrophoresis using immobilized pH gradients: Current status.* Electrophoresis, 2000. **21**(6): p. 1202-1209.
5. Alpert, A.J. and P.C. Andrews, *Cation-Exchange Chromatography of Peptides on Poly(2-Sulfoethyl Aspartamide)-Silica.* Journal of Chromatography, 1988. **443**: p. 85-96.
6. Molnar, I. and C. Horvath, *Separation of amino acids and peptides on non-polar stationary phases by high-performance liquid chromatography.* Journal of Chromatography, 1977. **142**: p. 623-40.
7. Simpson, R.C., H.Y. Mohammed, and H. Veening, *Reversed-Phase Ion-Pairing Liquid-Chromatographic Separation and Fluorimetric Detection of Polyamines.* Journal of Liquid Chromatography, 1982. **5**(2): p. 245-264.
8. Tigaud, S., et al., *Gram-Negative Anaerobes Beta-Lactamases - Difficulty About Their Characterization by Iso-Electric-Focusing - (an 48 Strains Experience).* Medecine Et Maladies Infectieuses, 1990. **20**: p. 141-144.
9. Gilar, M., et al., *Orthogonality of separation in two-dimensional liquid chromatography.* Analytical Chemistry, 2005. **77**(19): p. 6426-6434.
10. Delmotte, N., et al., *Two-dimensional reversed-phase x ion-pair reversed-phase HPLC: An alternative approach to high-resolution peptide separation for shotgun proteome analysis.* Journal of Proteome Research, 2007. **6**(11): p. 4363-4373.
11. Chuong, S.D.X., et al., *Large-scale identification of tubulin-binding proteins provides insight on subcellular trafficking, metabolic channeling, and signaling in plant cells.* Molecular & Cellular Proteomics, 2004. **3**(10): p. 970-983.
12. Patel, P.C., et al., *Proteomic Analysis of Microtubule-associated Proteins during Macrophage Activation.* Molecular & Cellular Proteomics, 2009. **8**(11): p. 2500-2514.

13. Kozielski, F., et al., *Proteome analysis of microtubule-associated proteins and their interacting partners from mammalian brain.* Amino Acids, 2011. **41**(2): p. 363-385.
14. Liska, A.J., et al., *Homology-based functional proteomics by mass spectrometry: Application to the Xenopus microtubule-associated proteome.* Proteomics, 2004. **4**(9): p. 2707-2721.
15. O'Connell, P.A., et al., *Characterization of the microtubule proteome during post-diapause development of Artemia franciscana.* Biochimica Et Biophysica Acta-Proteins and Proteomics, 2006. **1764**(5): p. 920-928.
16. Gache, V., et al., *Xenopus Meiotic Microtubule-Associated Interactome.* PLoS One, 2010. **5**(2): p. -.
17. Kyte, J. and R.F. Doolittle, *A Simple Method for Displaying the Hydropathic Character of a Protein.* Journal of Molecular Biology, 1982. **157**(1): p. 105-132.
18. Gauci, S., et al., *A versatile peptide pI calculator for phosphorylated and N-terminal acetylated peptides experimentally tested using peptide isoelectric focusing.* Proteomics, 2008. **8**(23-24): p. 4898-906.
19. Ure, A.E., et al., *Characterization of the complete genomes of Camelus dromedarius papillomavirus types 1 and 2.* Journal of General Virology, 2011. **92**: p. 1769-1777.
20. Peng, J.M., et al., *Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: The yeast proteome.* Journal of Proteome Research, 2003. **2**(1): p. 43-50.
21. Molnar, I., *Searching for robust HPLC methods - Csaba Horvath and the solvophobic theory.* Chromatographia, 2005. **62**: p. S7-S17.
22. Dorsey, J.G. and K.A. Dill, *The Molecular Mechanism of Retention in Reversed-Phase Liquid-Chromatography.* Chemical Reviews, 1989. **89**(2): p. 331-346.
23. Gritti, F. and G. Guiochon, *Critical contribution of nonlinear chromatography to the understanding of retention mechanism in reversed-phase liquid chromatography.* Journal of Chromatography A, 2005. **1099**(1-2): p. 1-42.
24. Torshin, I.Y., I.T. Weber, and R.W. Harrison, *Geometric criteria of hydrogen bonds in proteins and identification of 'bifurcated' hydrogen bonds.* Protein Engineering, 2002. **15**(5): p. 359-363.
25. Redeker, V., *Mass spectrometry analysis of C-terminal posttranslational modifications of tubulins.* Methods Cell Biol, 2010. **95**: p. 77-103.

Chapter 4

REDUCTIVE DIETHYLATION AND 2D LC-MS/MS TO STUDY THE EFFECT OF 2ME ON THE HUMAN MICROTUBULE INTERACTOME

4.1 Introduction

With the development of proteomic research, there is increasing need for quantification information of proteins other than qualitative information of protein characterizations [1]. Due to the unprecedented sensitivity of mass spectrometry and its continuous improvement of duty cycles, proteins that were not quantifiable on a large scale previously because of low quantity and high complexity are now being investigated using mass spectrometry. Stable isotopic labeling method [2] for quantitative proteomics utilized high mass accuracy to reliably identify peptides through collision induced dissociation (CID) MS/MS fragmentation and then quantify the relative intensity of isotopically-labeled peptide precursor ions. Several isotopic quantification strategies have been developed including *in vivo* isotopic incorporation such as SILAC [3] or chemical labeling method such as isotope-coded affinity tagging [2].

Recently, Hsu et al. developed the dimethylation isotopic labeling strategy [4] by reacting the N-terminus and ε-amino group of lysine residues from peptides through reductive amination to substitute H atoms of free amine with dimethyl groups. This preceding work was followed by a series of studies [5-12] to investigate its multiplex potential because of its simple procedure and low cost. Through the use of isotopic labeling reagents such as formaldehyde (H, D substituted; ^{12}C, ^{13}C substituted) and sodium cyanoborohydride (H, D substituted), mass shift between the incorporated methyl groups would differentiate light and heavy labeled peptides with same sequence but resulting from different samples. Specifically, the quantification ratio between different

forms of labeled peptide would provide protein relative expression information when samples from different origins were differentially labeled and then combined. This method is an inexpensive approach compared to isobaric labeling method such as iTRAQ [13] and is applicable for global quantification. The reported labeling efficiency is close to 100% in a short time (less than 5 minutes) [4] and labeled peptides have comparable ionization to unlabeled peptides. However, due to limited mass shift from ^{12}C and ^{13}C for methyl groups, deuterium was used along with hydrogen to form the isotopic pair. This would result in an HPLC retention time shift [5] for paired peptides, thus making large scale proteomic quantification challenging for complex samples.

In the present work, we sought to use an alternative isotopic labeling strategy by reductive diethylation of the primary amine from N-terminal and ε-amino group of lysine residues from peptides. The advantage of diethylation is the larger mass shift introduced by ethyl groups compared to methyl groups, thus providing the possibility of only using ^{12}C and ^{13}C acetaldehyde for the labeling to eliminate retention shift caused by deuterium and hydrogen pair. Diethylation of peptides had been recently reported [14] but we tested the prospect of this method for quantification of a highly complex sample on the 2D-RP-HPLC scale for shotgun proteomic. In addition, our labeling protocol is different from that reported with the use of sodium cyanoborohydride instead of 2-picoline-borane and without ethanol in reaction buffer.

We started by comparing diethylation to dimethylation with standard BSA peptides to evaluate their labeling performance. After confirming the labeling efficiency of

diethylation to be close to 100% with neglectable non-labeling and partial-labeling, we applied this method to the complex MAPs sample isolated from MCF-7 cells after 2-methoxyestradiol (2ME) treatment for the quantification of protein expression changes. Microtubule-associated proteins (MAPs) had been described previously in chapter 2 with regard to their biological significance as potential biomarkers for anti-cancer drugs [15]. 2ME is an endogenous estradiol metabolite [16], and has been previously studied for tumor treatment. The tumor growth-inhibiting effects of 2ME are results of 2ME binding to microtubule (MT) and affecting MT dynamics, causing cells to arrest in the G2/M phase of the cell cycle [17]. Further studies of 2ME indicate its mechanism for anti-tumor and angiogenic activities not only is associated with microtubules, but also is involved in JNK activation and apoptosis-induced cell signaling pathway [18-22]. A summary of 2ME effects on tumor cells is shown in Figure 4.1 [22].

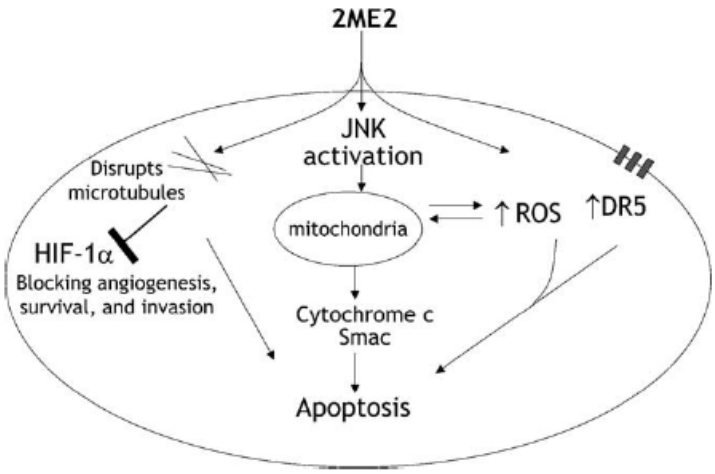


Figure 4.1. Schematic representation of the cellular mechanisms through which 2ME exerts its antitumor and angiogenic activities [22]. The contributions of various apoptotic pathways from 2ME differ substantially depending on cell type.

With the use of ^{12}C and ^{13}C for isotopic peptide pair, even after high resolution separation by 2D-RP-HPLC, our MAPs quantification result showed no retention shift between isotopically diethylated peptides and provide large scale quantification information for MAPs expression change induced by 2ME treatment. Most of the proteins identified do not exhibit significant expression change after 2ME treatment. Nevertheless, we do identify a few candidates that have been documented to be affected by 2ME treatment, and our experimental quantification result is consistent with the expected expression change based on information of these biomarkers.

4.2 Materials and Methods

4.2.1 Materials

Trypsin (sequencing grade) was purchased from Promega (Madison, WI, USA). Bovine serum albumin (BSA), iodoacetamide (IAA), dithiothreitol (DTT), formaldehyde, acetaldehyde, sodium cyanoborohydride ($NaBH_3CN$), ammonium hydroxide solution puriss were purchased from Sigma-Aldrich (St. Louis, MO, USA). Acetaldehyde (^{13}C substituted, 99%) was purchased from Cambridge Isotope Labotories (Andover, MA, USA). Acetonitrile (HPLC grade) and other chemicals were purchased from Fisher Scientific (Pittsburgh, PA, USA).

4.2.2 MCF-7 Cell Culture and 2-Methoxyestradiol (2ME) Treatment

MCF-7 human breast adenocarcinoma cells (passage 6) were cultured in GIBCO minimum essential medium (MEM) with Earle's salts and L-glutamine supplemented

with 1.5mg/mL sodium bicarbonate ($NaHCO_3$), 1mM sodium pyruvate, MEM non-essential amino acids, 10% (vol/vol) fetal bovine serum (FBS), 100 units/mL penicillin, 100μg/mL streptomycin and 10mg/L insulin at 37°C in humidified 95% air and 5% CO_2 to 60-70% confluency. Cells were then washed with sterile PBS and incubated in serum-free medium at 37°C for 24 hours under the same atmospheric conditions.

Following cell synchronization, cells were treated with 2μM 2-methoxyestradiol (2ME) or vehicle (DMSO) and maintained in complete culture medium at 37°C for 24 hours in 5% CO2 humidified air. Cells were then incubated with cell dissociation solution (3mM EDTA, 2.5mM KCl, 150mM NaCl, 1.5mM KH_2PO4, 8mM Na_2HPO4, pH 7.2) at 37°C for 15 minutes under controlled atmospheric conditions. Cells were transferred to 50 mL conical vials and centrifuged at 130 g at 4°C for 5 minutes. The supernatant was discarded and the resulting cellular pellet was washed with sterile PBS and resuspended in MME buffer (0.1 M 2-(N-morpholino) ethanesulfonic acid (MES), 1 mM MgCl2, 1 mM EGTA, pH 6.9) with added phosphatase inhibitor (1mM Na_3VO4, 1mM NaF, 10mM sodium pyrophosphate, 1mM β-glycerophosphate) and EDTA-free complete protease inhibitor cocktail (Roche Applied Science). The cellular suspension was homogenized by pipette, aliquotted (800μL/aliquot) and stored at -80°C until further analysis.

4.2.3 Cell Lysate and Isolation of Microtubule Associated Proteins

Upon use, approximately 2×10^8 (800μL) frozen MCF-7 cells per treatment (2ME treated or DMSO treated as control) in the MME buffer were rapidly thawed and added

DTT to the final concentration of 1 mM. Then cells were lysed by sonicating the cell suspension 7 times for 30 seconds with 30 seconds rest interval on melted ice between sonications. The resulting cell lysates are then centrifuged at 120,000g (Beckman TLA55 rotor) for 1 hour at 4 ℃. Transfer the supernatants (cytosolic extract) to a new tube and discard the pellets (the DNA and cell debris).

For procedure of isolating microtubule-associated proteins (MAPs), refer to previous chapter 2 section **2.3.3**.

The resulting supernatant containing the MAPs released from microtubule pellets is dialyzed against excess H_2O before subsequent use.

4.2.4 Gel-electrophoresis

Gel electrophoresis was conducted on a NuPAGE Novex 4-12% Bis-Tris gel (Invitrogen) in MOPs buffer by following the vendor's procedure. Protein mixtures within each lane would be separated by the molecular weight of each protein in the sample. After gel electrophoresis, gel was stained with Coomassie Blue R250 dye.

4.2.5 Tryptic Digestion of Isolated Microtubule Associated Proteins

Two biological replicates were conducted independently to isolate microtubule associated proteins (MAPs) from individually grown MCF-7 cells with or without 2ME treatment. Each replicate generated approximately 400μg of extracted MAPs after microtubule polymerization, while 120μg were used for tryptic digestion. For each

replicate of MAPs (120μg), ammonium biocarbonate was added to final concentration of 50mM (pH 8). Each sample was then reduced by DTT (DTT : protein = 300 : 1 molar ratio) at 55℃ for 1 hour, followed with reduction by iodoacetamide (IAA : protein = 630 : 1 molar ratio) at room temperature for 45 minutes in the dark. Each sample was then reduced by DTT for the second time under the previous condition and finally digested by sequencing grade trypsin (Promega, Madison, WI, USA) (trypsin : protein = 1 : 50 weight ratio) at 37℃ for 14 hours.

4.2.6 Dimethylation and Diethylation of Peptides

A) Comparison of dimethylation and diethylation labeling performance

Standard BSA peptides were desalted and divided into equal aliquots (7μg / aliquot) for the dimethylation or diethylation labeling. Three replicates were conducted for both dimethylation and diethylation. In each replicate, four aliquots of BSA peptides (7μg) were dissolved in 50μL 0.2M sodium acetate buffer (pH around 5.5). Two of the BSA aliquots were used as unlabeled controls. One aliquot was used for dimethylation while the other was used for diethylation.

For dimethylation, 4μL of 4% (w/w) formaldehyde in water was added to each peptide solution. The solution was mixed briefly and spun down, then mixed with 4μL of 0.6M $NaBH_3CN$ solution in water. The mixture was incubated for 1 hour at room temperature with gentle mixing. Finally, the labeling reaction was quenched by adding

4μL of 4% (w/w) ammonium hydroxide solution in water and further acidified by 8μL of formic acid.

For diethylation, 4μL of 6% (w/w) acetaldehyde in water was added and the solution was incubated for 2 hours at room temperature. After this, 4μL of 0.6M $NaBH_3CN$ solution was added to allow reaction for 1 hour. Then another 4μL of 6% (w/w) acetaldehyde in water was added and the mixture was incubated for an additional hour. Last, at intervals of 1 hour, four more consecutive additions of 4μL of 0.6M $NaBH_3CN$ solution were added and the labeling reaction was incubated overnight. After labeling, the reaction was quenched by addition of 8μL of 4% (w/w) ammonium hydroxide solution in water and further acidified by 16μL of formic acid.

B) Assessment of quantification performance by diethylation labeling

Equal aliquots of desalted peptide mixture (10μg BSA digest and 10μg beta-casein digest) in 60μL 0.2M sodium acetate buffer (pH 5.5) were diethylation labeled using light reagent (acetaldehyde, ^{12}C) for control and heavy reagent (acetaldehyde, ^{13}C substituted) for comparison. Four aliquots were used as the two replicates for diethylation labeling. Each replicate used one aliquot for light diethylation labeling and one aliquot for heavy diethylation labeling. To label the peptide mixtures, 3μL of 40% (w/w) acetaldehyde (^{12}C or ^{13}C) in water was added and the solution was incubated for 2 hours at room temperature. After this, 4μL of 3M $NaBH_3CN$ solution was added to allow reaction for 1 hour. Then another 3μL of 40% (w/w) acetaldehyde (^{12}C or ^{13}C) in water was added and

the mixture was incubated for an additional hour. Last, at intervals of 1 hour, four more consecutive additions of 4μL of 3M $NaBH_3CN$ solution were added and the labeling reaction was incubated overnight. After labeling, the reaction was quenched by addition of 6μL of 25% (w/w) ammonium hydroxide solution in water and further acidified by 8μL of formic acid.

C) Comparative analysis of MAPs by diethylation labeling

For the comparative analysis of MAPs expression change modulated by 2ME treatment, for each replicate, 120μg tryptic digest of isolated MAPs were dissolved in 120μL 0.2M sodium acetate buffer (pH 5.5). Similarly to the above procedure, 20μL of 40% (w/w) acetaldehyde (^{12}C light label) in water was added to MAPs digest (control) while a similar amount of acetaldehyde (^{13}C substituted heavy label) solution was added to MAPs digest (2ME treated). After incubation for 2 hours at room temperature, 16μL of 5M $NaBH_3CN$ was added to allow the proceeded reaction for 1 hour. Then either 20μL of 40% (w/w) ^{12}C acetaldehyde or ^{13}C acetaldehyde was added to the control or 2ME-treated MAPs digest, respectively. The mixture was incubated for another hour, and then four consecutive aliquots of 16μL 5M $NaBH_3CN$ were added with 1 hour intervals. The reaction was incubated overnight and was quenched by 40μL of 25%(w/w) ammonium hydroxide solution. The final sample was acidified by 50μL of formic acid.

4.2.7 Preparation of BSA Digest for Dimethylation/Diethylation Comparison and Diethylation Quantification Validation

For dimethylation/diethylation comparison, equal amounts (7µg) of unlabeled BSA or labeled BSA (dimethylated or diethylated) were prepared. As shown in Figure 4.2, 50% of unlabeled BSA control was mixed 50% of labeled BSA to make a 1-to-1 ratio sample. All remaining samples (control and labeled BSA digest) and the 1-to-1 ratio sample were desalted and reconstituted in 40µL 0.1% acetic acid.

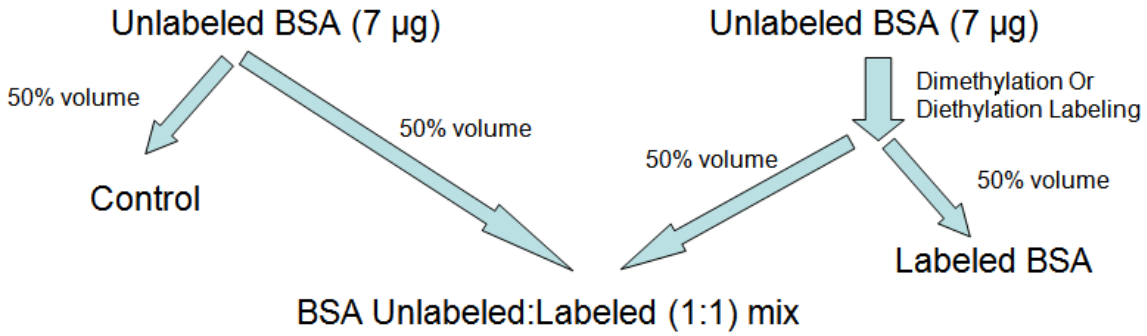


Figure 4.2. Schematic sample preparation of BSA digest for dimethylation/diethylation comparison. For both dimethylation and diethylation, final experimental products have control BSA, labeled BSA and 1-to-1 ratio mix sample.

For diethylation quantification, equal amounts of peptides (10µg BSA digest and 10µg beta-casein digest) were either diethylated with ^{12}C acetaldehyde or ^{13}C acetaldehyde. Labeled peptides were desalted and the light diethylation-labeled peptides were mixed with heavy diethylation-labeled peptides by the following ratios: 1 to 20, 1 to 10, 1 to 5, 1 to 2, 1 to 1, 2 to 1, 5 to 1, 10 to 1, and 20 to 1.

4.2.8 High-pH Fractionation of Tryptic MAPs Digest after Diethylation Labeling

For each replicate of MAPs comparative analysis, about 65μg of tryptic MAPs digest (control) after ^{12}C diethylation was mixed with 65μg of tryptic MAPs digest (2ME treatment) after ^{13}C diethylation and the mixture was then injected onto an Agilent 1200 HPLC system equipped with an Agilent ZORBAX 300Extend-C18 column (150 x 2.1 mm i.d, 3.5 μm) with temperature setting at 35℃. Mobile phase A was composed of 23.5mM ammonium formate, 100% water (pH 7.50) and mobile phase B was 23.5mM ammonium formate, 12% water, 88% acetontrile. Injected peptides were separated by gradient elution 0-70% B in 35 min with a constant flow rate at 0.15mL/min. UV absorbance was monitored at 214 nm and fractions were manually collected every minute. Each collected HPLC fraction (150μL) was centrifuged under vacuum in a CentriVap to reduce sample volume to near dryness and then reconstituted in 40μL 0.1% acetic acid.

4.2.9 LC-MS/MS Analysis

Peptide on-line analysis were performed using Agilent 1100 binary HPLC system (Agilent Technologies, Palo Alto, CA, USA) coupled with Thermo Finnigan LTQ-FT hybrid instrument (ThermoFisher, San Jose, CA, USA). The trapping column was made by packing an 4cm length of reversed-phase C18 material (Monitor C18, 5μm particle, Column Engineering Inc, Downers Grove, IL, USA) into a 75μm ID fused-silica capillary column (Polymecro Technologies, Phoenix, AZ, USA) with self-made frit on one end. Connected to the trapping column, was a self-made analytical column, which was constructed by pulling a 5μm tip from a 75μm ID fused-silica capillary column (Polymecro Technologies, Phoenix, AZ, USA) and packing 10cm reverse-phase C18

material (Monitor C18, 3μm particle, Column Engineering Inc, Downers Grove, IL, USA) from the other end. HPLC mobile phase A was 0.1M HOAc in water while mobile phase B was 0.1M HOAc in acetonitrile. After sample loading, BSA peptides were separated by a gradient of 0 to 70% solvent B in 17 minutes while each high-pH fraction of the labeled MAPs digest was separated by a gradient of 0 to 40% solvent B in 75 minutes followed by 40 to 70% solvent B in 15 minutes. Mass spectrometer settings were configured for the Thermo LTQ-FT-ICR (ThermoFisher, San Jose, CA, USA) as follows: FT for MS mode (m/z 300 to 1700 Da) with resolution of 100,000 and profile data type, MS/MS mode was conducted on IonTrap with normal scan rate and centroid data type. Precursor isolation width was set at 2.0Da (precursor m/z ±1.0Da) and dynamic exclusion of precursor ions was enabled for 30 seconds of ions within 0.1Da range. Precursor ions with +1 charge or unassigned charge were rejected for MS/MS fragmentation. The minimum signal threshold was set as 1,000 counts for CID fragmentation with normalized CE of 35 and activation time of 30.0 ms. BSA peptides were analyzed with 1 FT MS scan followed by 5 MS/MS scans while the fractions of MAPs digest after diethylation were analyzed with 1 FT MS scan followed by 9 MS/MS scan.

The same fractions of MAPs digest after diethylaiton were analyzed again using a Thermo LTQ Orbitrap Velos instrument (ThermoFisher, San Jose, CA, USA) with similar configurations. The major difference is that on the LTQ Orbitrap Velos instrument, FT MS mode resolution was set at 60,000 and data was collected with 1 Orbitrap MS scan followed by 10 LTQ Velos MS/MS scans.

4.2.10 Data Analysis

Raw MS data were converted to mzXML files and then converted to Mascot Generic Format (MGF) using MassMatrix Mass Spec Data File Conversion Tools (version 3.9). The MGF files were then searched by Mascot 2.2.07 (Matrix Science Ltd, UK). BSA MS data were searched against in-house created standard protein database and MAPs MS data were searched against Uniprot human database (Nov. 2011, 73560 protein entries) with the reverse sequence of each entry concatenated to the original database. Mascot search was conducted with the following specifications: trypsin enzyme with 1 miss-cleavage allowed, precursor MS tolerance at 20 ppm and MS/MS tolerance at 0.5 Da. Carbamidomethylation of cysteine was set as a fixed modification while oxidation of methionine, light diethylation with +56 Da (N-term, K, R, NQ) and heavy diethylation with +60 Da (N-term, K, R, NQ) were set as variable modifications.

For BSA search results, five peptides that span the LC elution range were chosen to evaluate dimethylation & diethylation labeling efficiency and ionization efficiency. Peptides' MS/MS spectra were manually verified and their XIC peak areas were quantified by Xcalibur 2.1 Qual Browser (ThermoFisher, San Jose, CA, USA) with 0.1 Da mass tolerance.

ProteoIQ 2.3 (Nusep, Bogart, GA, USA) Isotopic Quantification module was used to quantify MAPs expression change after 2ME treatment using this reductive diethylation strategy. Stringent criteria were used to filter protein results: peptide IDs with Mascot

score of 30 and at least 7 amino acids in the sequence; protein IDs with at least 2 spectrum peptide matches and identification in both biological replicates (for either FT data or Orbitrap data). Finally, peptide and protein FDR were set at the 1% level.

4.3 Results

4.3.1 Comparasion of Dimethylation Labeling Strategy with Diethylation Strategy

Figure 4.3 shows the mechanism of the chemical labeling strategy using formaldehyde or acetaldehyde to react with free amines of the N-terminus or the lysine ε-amino groups of a peptide. After forming the Schiff base, sodium cyanoborohydride is used as a reducing agent and the final labeled peptide has its hydrogen from free amine replaced with a methyl or ethyl group. The labeled peptide will have +28 Da or +56 Da mass shift compared to unlabeled peptide per dimethylation or diethylation label.

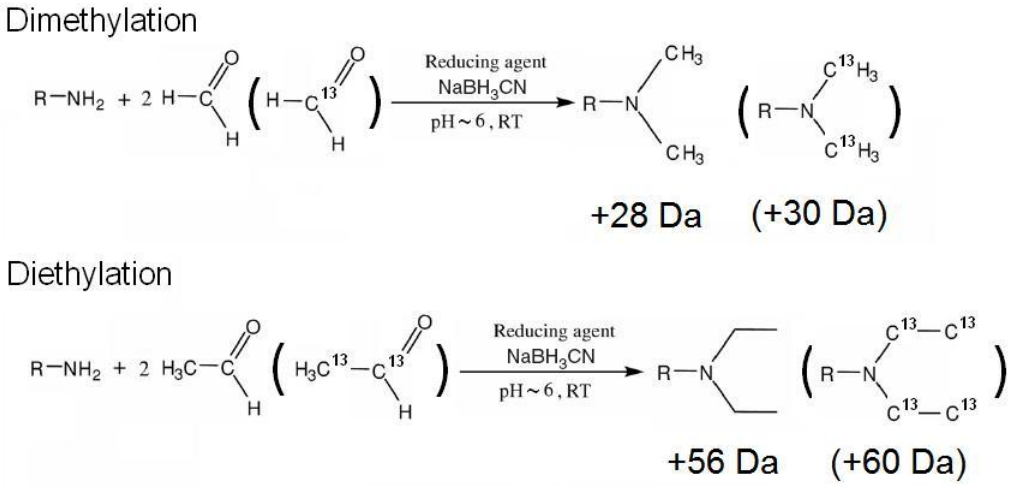


Figure 4.3. Quantification strategy using dimethylation or diethylation labeling.

Acetaldehyde is used in our strategy instead of formaldehyde due to the result of having a greater mass shift between stable isotopically-labeled peptides by carbon (^{12}C and ^{13}C) and hydrogen (^{1}H and ^{2}H) isotopic reagents. With the use of acetaldehyde either in ^{12}C or ^{13}C, for doubly charged tryptic peptides' light and heavy isoforms, mass difference of 2 Da for arginine peptides or 4 Da for lysine peptides would be sufficient for comparative quantitation. In addition, this strategy would not have isotopic effect of reversed-phase retention time shift between paired peptides.

A) Labeling Efficiency

It is crucial to determine the efficiency of the chemical labeling method for quantitative proteomics since the existence of unlabeled peptides after labeling would result in reduced chance for the identification of labeled peptides during competitive ionization of LC-MS analysis. Moreover, comparative quantification result would also be affected because the origin of unlabeled peptides cannot be determined after mixing control sample with drug treated sample. Since we are interested in the utilization of diethylation for isotopic quantification, BSA and beta-casein peptide mixtures were diethylated as described in Materials and Methods for the diethylation quantification validation experiment. Without considering any sample loss during the labeling and desalting steps, 1,600 fmol of BSA peptides and 4,000 fmol of beta-casein peptides after diethylation with ^{12}C acetaldehyde were loaded and analyzed by LC-MS/MS. After searching the data with Mascot with diethylation as a variable modification, 180 peptides (including redundant hits) were identified for BSA giving a sequence coverage of 60.5%

while 88 beta-casein peptides (including redundant hits) were identified with 16.5% sequence coverage. Within these peptide hits, only 6 (3.3%) out of the 180 BSA peptides were unlabeled peptides while all other BSA peptides and all beta-casein peptides were diethylated. This level of sequence coverage performance for each protein is similar to that obtained for unlabeled digests.

In order to more accurately evaluate the labeling efficiency of the diethylation method, the intensity ratio of fully-labeled versus unlabeled peptides were compared. This is accomplished by extracted ion chromatogram (XIC) of peptide precursor ions. BSA peptides were labeled with either dimethylation or diethylation to compare the labeling efficiency of both methods. Without considering any sample loss during the labeling step or desalting step, 600 fmoles of labeled BSA peptides (either dimethylation or diethylation) were loaded and analyzed by LC-MS/MS. Five representative BSA peptides that span the LC elution timeframe were chosen for this purpose. From Table 4.1, for these representative peptides after dimethylation or diethylation labeling, there is almost no existence of unlabeled peptides based on the XIC peak area of possible precursor ions. Labeling efficiency was further evaluated by calculating 1 – XIC ratio (unlabeled form vs fully labeled form) and the result shows that, for both dimethylation and diethylation, the labeling efficiency is at the 99% level. The summary of same analysis for all three experimental replicates is showed in Table 4.2. Across three replicates, our results show that the labeling efficiency of dimethylation and diethylation are highly reproducible at the 99% level with neglectable variation.

Table 4.1. Labeling efficiency comparison of dimethylation and diethylation from representative BSA peptides. (A) Dimethylation result. (B) Diethylation result.

A

	Unlabeled 1+			Unlabeled 2+			Unlabeled 3+		
Peptide	m/z	RT	XIC Area	m/z	RT	XIC Area	m/z	RT	XIC Area
CCTESLVNR	1138.4980	NA		569.7529	17.07	9.32E+05	380.1712	NA	
YICDNQDTISSK	1443.6420	NA		722.3249	17.24	4.14E+06	481.8859	NA	
HLVDEPQNLIK	1305.7161	NA		653.3620	18.46	2.38E+06	435.9106	NA	
LVNELTEFAK	1163.6307	NA		582.3193	19.69	1.45E+06	388.5488	NA	
QTALVELLK	1014.6194	20.09	8.32E+04	507.8136	20.09	1.66E+06	338.8784	NA	

	Dimethylated 1+			Dimethylated 2+			Dimethylated 3+		
Peptide	m/z	RT	XIC Area	m/z	RT	XIC Area	m/z	RT	XIC Area
CCTESLVNR	1166.5293	17.18	9.34E+06	583.7686	17.20	2.21E+08	389.5150	17.22	2.64E+05
YICDNQDTISSK	1499.7046	17.36	2.67E+06	750.3562	17.36	4.23E+08	500.5734	17.36	1.48E+06
HLVDEPQNLIK	1361.7787	18.51	1.31E+06	681.3933	18.51	2.32E+08	454.5981	18.53	3.63E+07
LVNELTEFAK	1219.6933	19.77	6.86E+06	610.3506	19.75	2.27E+08	407.2363	NA	
QTALVELLK	1070.6820	20.17	4.28E+06	535.8449	20.17	1.16E+08	357.5659	20.17	1.57E+05

	Unlabeled Forms	Dimethylated Forms	Dimethylation Efficiency
Peptide	XIC sum	XIC sum	1 - XIC Ratio
CCTESLVNR	9.32E+05	2.31E+08	0.9960
YICDNQDTISSK	4.14E+06	4.27E+08	0.9903
HLVDEPQNLIK	2.38E+06	2.70E+08	0.9912
LVNELTEFAK	1.45E+06	2.34E+08	0.9938
QTALVELLK	1.74E+06	1.20E+08	0.9855

B

	Unlabeled 1+			Unlabeled 2+			Unlabeled 3+		
Peptide	m/z	RT	XIC Area	m/z	RT	XIC Area	m/z	RT	XIC Area
CCTESLVNR	1138.4980	NA		569.7529	16.93	1.90E+05	380.1712	NA	
YICDNQDTISSK	1443.6420	NA		722.3249	17.06	1.84E+06	481.8859	NA	
HLVDEPQNLIK	1305.7161	NA		653.3620	18.23	1.72E+06	435.9106	NA	
LVNELTEFAK	1163.6307	NA		582.3193	19.44	8.75E+05	388.5488	NA	
QTALVELLK	1014.6194	NA		507.8136	19.95	7.31E+05	338.8784	NA	

	Diethylated 1+			Diethylated 2+			Diethylated 3+		
Peptide	m/z	RT	XIC Area	m/z	RT	XIC Area	m/z	RT	XIC Area
CCTESLVNR	1194.5606	17.24	2.71E+06	597.7842	17.25	1.59E+08	398.8588	NA	
YICDNQDTISSK	1555.7672	17.92	9.01E+05	778.3875	17.92	3.23E+08	519.2610	17.92	8.46E+06
HLVDEPQNLIK	1417.8413	18.71	3.33E+05	709.4246	18.73	1.47E+08	473.2857	18.73	3.41E+07
LVNELTEFAK	1275.7559	20.04	4.12E+05	638.3819	20.04	8.55E+07	425.9239	NA	
QTALVELLK	1126.7446	20.44	5.14E+05	563.8762	20.44	5.84E+07	376.2535	NA	

	Unlabeled Forms	Diethylated Forms	Diethylation Efficiency
Peptide	XIC sum	XIC sum	1 - XIC Ratio
CCTESLVNR	1.90E+05	1.62E+08	0.9988
YICDNQDTISSK	1.84E+06	3.32E+08	0.9945
HLVDEPQNLIK	1.72E+06	1.81E+08	0.9905
LVNELTEFAK	8.75E+05	8.59E+07	0.9898
QTALVELLK	7.31E+05	5.89E+07	0.9876

Five representative peptides (1 arginine peptide and 4 lysine peptides) that elute across the BSA digest retention timeframe were selected for illustration. For each peptide after dimethylation or diethylation, we consider three possible charge states (+1, +2, +3) of its unlabeled form or fully-labeled form. XIC peak area and RT information for each possible precursor charge state were manually extracted from raw data using Xcalibur 2.1 based on their theoretical m/z value with mass tolerance of 0.1 Da. Peak detection and peak area calculation was automatically processed by Xcalibur Qual Browser with "Genesis" peak detection algorithm. XIC sum is the summation of XIC peak areas of +1, +2, +3 charge state precursors of the same form of peptide. Labeling efficiency is measured by 1 – XIC ratio which is calculated for each peptide, its XIC sum of unlabeled form versus its XIC sum of fully labeled form (Dimethylated or Diethylated).

Table 4.2. Summary of labeling efficiency analysis for dimethylation and diethylation of representative BSA peptides from three experimental replicates.

Labeling Efficiency			Dimethylation		
Peptides	Rep1	Rep2	Rep3	Average	Stdev
CCTESLVNR	0.9960	0.9984	0.9972	0.9972	0.0012
YICDNQDTISSK	0.9903	0.9932	0.9975	0.9937	0.0036
HLVDEPQNLIK	0.9912	0.9957	0.9968	0.9946	0.0030
LVNELTEFAK	0.9938	0.9947	0.9972	0.9952	0.0018
QTALVELLK	0.9855	0.9890	0.9977	0.9907	0.0063

Labeling Efficiency			Diethylation		
Peptides	Rep1	Rep2	Rep3	Average	Stdev
CCTESLVNR	0.9988	0.9992	0.9948	0.9976	0.0024
YICDNQDTISSK	0.9945	0.9985	0.9916	0.9949	0.0035
HLVDEPQNLIK	0.9905	0.9965	0.9855	0.9908	0.0055
LVNELTEFAK	0.9898	0.9953	0.9853	0.9901	0.0050
QTALVELLK	0.9876	0.9970	0.9820	0.9889	0.0076

Labeling efficiency is measured by 1 – XIC ratio which is calculated for each peptide, its XIC sum of unlabeled form versus its XIC sum of fully labeled form (Dimethylated or Diethylated). For three experimental replicates, the average value and standard deviation of labeling efficiency is calculated for both dimethylation and diethylation.

Because both dimethylation and diethylation would theoretically add four methyl groups or ethyl groups to lysine-containing peptide (two groups for arginine-containing peptide), another critical factor to consider for these chemical labeling methods is the potential partial labeling. For example, partially labeled lysine peptide could potentially have either one, two, or three methyl or ethyl groups while partially labeled arginine peptide could potentially only have one methyl or ethyl group. Depending on the abundance, partially labeled peptides would increase sample complexity and also reduce the chance for identification of fully labeled peptides due to competitive ionization for MS analysis. Table 4.3 compares the partial labeling percentage of dimethylation or diethylation for representative BSA peptides. Because fully labeled peptides are of high abundance, all their charge states (+1, +2, +3) are considered for extracting quantifiable XIC peak areas. Partially labeled peptides are of especially low abundance, only the predominant +2 charge state is considered for extracting XIC peak areas since detection of their +1 and +3 charge precursors are below the limit of detection. From Table 4.3A of the dimethylation result, all forms of the partially labeled peptides are exceptionally low abundance and even their combined XIC intensity is less than 1% of their fully labeled counterparts. For the diethylation result from Table 4.3B, the contribution of partially labeled peptides with either one ethyl group or two ethyl groups is almost negligible compared to the partially labeled peptides with three ethyl groups. By adding the XIC intensity of all partially labeled forms and then comparing to the XIC intensity of fully labeled form, the partial labeling percentage is around 10% to 15% for representative

Table 4.3. Partial labeling comparison of dimethylation and diethylation from representative BSA peptides. (A) Dimethylation result. (B) Diethylation result.

A

	Partial 1 Methyl group 2+			Partial 2 Methyl group 2+			Partial 3 Methyl group 2+		
Peptide	m/z	RT	XIC Area	m/z	RT	XIC Area	m/z	RT	XIC Area
CCTESLVNR	576.7608	17.11	2.39E+05	NA			NA		
YICDNQDTISSK	729.3328	17.26	4.58E+05	736.3406	17.25	4.80E+04	743.3484	17.34	5.72E+05
HLVDEPQNLIK	660.3699	18.46	3.63E+05	667.3777	18.46	7.44E+05	674.3855	18.51	4.65E+05
LVNELTEFAK	589.3272	19.66	3.10E+05	596.3350	19.69	1.04E+05	603.3428	19.75	4.15E+05
QTALVELLK	514.8215	20.12	2.22E+05	521.8293	20.09	1.15E+05	528.8371	20.17	3.20E+05

	Dimethylated 1+			Dimethylated 2+			Dimethylated 3+		
Peptide	m/z	RT	XIC Area	m/z	RT	XIC Area	m/z	RT	XIC Area
CCTESLVNR	1166.5293	17.18	9.34E+06	583.7686	17.20	2.21E+08	389.5150	17.22	2.64E+05
YICDNQDTISSK	1499.7046	17.36	2.67E+06	750.3562	17.36	4.23E+08	500.5734	17.36	1.48E+06
HLVDEPQNLIK	1361.7787	18.51	1.31E+06	681.3933	18.51	2.32E+08	454.5981	18.53	3.63E+07
LVNELTEFAK	1219.6933	19.77	6.86E+06	610.3506	19.75	2.27E+08	407.2363	NA	
QTALVELLK	1070.6820	20.17	4.28E+06	535.8449	20.17	1.16E+08	357.5659	20.17	1.57E+05

	Partial Methylated Forms	Fully Dimethylated Forms	Partial Labeling Percentage
Peptide	XIC sum	XIC sum	% XIC ratio
CCTESLVNR	2.39E+05	2.31E+08	0.10
YICDNQDTISSK	1.08E+06	4.27E+08	0.25
HLVDEPQNLIK	1.57E+06	2.70E+08	0.58
LVNELTEFAK	8.29E+05	2.34E+08	0.35
QTALVELLK	6.57E+05	1.20E+08	0.55

B

	Partial 1 Ethyl group 2+			Partial 2 Ethyl group 2+			Partial 3 Ethyl group 2+		
Peptide	m/z	RT	XIC Area	m/z	RT	XIC Area	m/z	RT	XIC Area
CCTESLVNR	583.7686	17.06	2.69E+06	NA			NA		
YICDNQDTISSK	736.3406	NA		750.3562	17.52	4.84E+04	764.3719	17.75	2.66E+07
HLVDEPQNLIK	667.3777	18.25	1.59E+05	681.3933	18.33	5.41E+05	695.4090	18.62	1.61E+07
LVNELTEFAK	596.3350	19.49	3.36E+05	610.3506	19.71	2.93E+05	624.3663	19.91	1.11E+07
QTALVELLK	521.8293	20.01	1.63E+05	535.8449	20.24	1.89E+05	549.8606	20.33	5.33E+06

	Diethylated 1+			Diethylated 2+			Diethylated 3+		
Peptide	m/z	RT	XIC Area	m/z	RT	XIC Area	m/z	RT	XIC Area
CCTESLVNR	1194.5606	17.24	2.71E+06	597.7842	17.25	1.59E+08	398.8588	NA	
YICDNQDTISSK	1555.7672	17.92	9.01E+05	778.3875	17.92	3.23E+08	519.2610	17.92	8.46E+06
HLVDEPQNLIK	1417.8413	18.71	3.33E+05	709.4246	18.73	1.47E+08	473.2857	18.73	3.41E+07
LVNELTEFAK	1275.7559	20.04	4.12E+05	638.3819	20.04	8.55E+07	425.9239	NA	
QTALVELLK	1126.7446	20.44	5.14E+05	563.8762	20.44	5.84E+07	376.2535	NA	

	Partial Ethylated Forms	Fully Diethylated Forms	Partial Labeling Percentage
Peptide	XIC sum	XIC sum	% XIC ratio
CCTESLVNR	2.69E+06	1.62E+08	1.66
YICDNQDTISSK	2.66E+06	3.32E+08	8.01
HLVDEPQNLIK	1.68E+07	1.81E+08	9.28
LVNELTEFAK	1.17E+07	8.59E+07	13.60
QTALVELLK	5.68E+06	5.89E+07	9.64

Five representative peptides (1 arginine peptide and 4 lysine peptides) that elute across the BSA digest retention timeframe were selected for illustration. For the arginine peptide, we consider one possible partially labeled form (with 1 methyl or ethyl group) and its fully labeled form (with 2 methyl or ethyl groups). For each of the 4 lysine peptides after dimethylation or diethylation, we consider three possible partially labeled forms (with 1, 2, 3 methyl or ethyl groups) and its fully-labeled form (with 4 methyl or ethyl groups). For partially labeled peptides, charge state of +2 is the dominant precursor charge while other charge states (+1 and +3) are neglectable with consideration of their low abundance and instrument detection limit. So only +2 precursor of partially labeled peptides and all possible charge state (+1, +2, +3) precursors of fully labeled peptides were considered for extracting XIC peak area and RT information. XIC peak area and RT information were manually extracted from raw data using Xcalibur 2.1 based on precursor ion theoretical m/z value with mass tolerance of 0.1 Da. Peak detection and peak area calculation was automatically processed by Xcalibur Qual Browser with "Genesis" peak detection algorithm. XIC sum for partially labeled peptide is the summation of XIC peak areas of +2 charged partially labeled precursors. XIC sum for fully labeled peptide is the summation of XIC peak areas of all charge states (+1, +2, +3) of fully labeled precursors. Partial labeling percentage is measured by % XIC ratio which is calculated for each peptide, its XIC sum of partially labeled forms versus its XIC sum of fully labeled form.

BSA peptides with the exception of the arginine peptide at 2% level because it only has one ethyl group for partial labeling. The summary of partial labeling analysis is shown in Table 4.4 for all three experimental replicates. Based on these experiment results, although we used repetitive steps labeling procedure for diethylation method compared to single step labeling procedure for dimethylation method, dimethylation has less than 1% for partially labeled peptides while diethylation has about 10% to 15% of partially labeled peptide with three ethyl groups.

Table 4.4. Summary of partial labeling analysis for dimethylation and diethylation of representative BSA peptides from three experimental replicates.

Partial Labeling Percentage	Dimethylation				Diethylation			
Peptides	Rep1	Rep2	Rep3	Average	Rep1	Rep2	Rep3	Average
CCTESLVNR	0.10%	0.05%	0.04%	0.06%	1.66%	1.36%	2.07%	1.70%
YIC(Carba)DNQDTISSK	0.25%	0.23%	0.03%	0.17%	8.01%	9.45%	9.60%	9.02%
HLVDEPQNLIK	0.58%	0.47%	0.18%	0.41%	9.28%	12.10%	13.20%	11.53%
LVNELTEFAK	0.35%	0.27%	0.62%	0.42%	13.60%	16.50%	15.00%	15.03%
QTALVELLK	0.55%	0.59%	0.79%	0.64%	9.64%	10.30%	12.90%	10.95%
Five Peptide Average				0.34%				9.64%

Partial labeling percentage of each peptide is measured by % XIC ratio of its partially labeled forms versus fully labeled form for both Dimethylation and Diethylation methods. For three experimental replicates, the average value of partial labeling percentage is calculated for each peptide and then averaged across five peptides.

The reason for higher partial labeling percentage of diethylation compared to dimethylation may be attributed to larger steric hindrance when acetaldehyde is being nucleophilically attacked by free amine from peptide than that for formaldehyde. This is the result with single step dimethylation and multiple steps diethylation experiments when the labeling reagent concentrations ([formaldehyde] and [acetaldehyde]) used are the same during each step. In reality, 10% to 15% partial-labeling and less than 1% non-labeling for diethylation of peptides are acceptable for isotopic proteomic quantitation. For the later diethylation quantification validation experiment using ^{12}C acetaldehyde and ^{13}C acetaldehyde, the concentrations of labeling reagents are increased because we used 10μg BSA digest and 10μg beta-casein digest mixture other than the 7μg BSA digest during dimethylation/diethylation comparison experiments. While it's roughly 3 times more peptides, the labeling reagent concentrations used are about 5 times more. Two replicate experiments were conducted and summarized in Table 4.5.

Table 4.5. Partial labeling analysis for diethylation of representative BSA peptides with increased concentration of labeling reagents.

Replicate1

	Partial 1 Ethyl group 2+			Partial 2 Ethyl group 2+			Partial 3 Ethyl group 2+		
Peptide	m/z	RT	XIC Area	m/z	RT	XIC Area	m/z	RT	XIC Area
CCTESLVNR	583.7686	13.39	7.30E+05	NA			NA		
YICDNQDTISSK	736.3406	NA		750.3562	NA		764.3719	14.04	5.91E+05
HLVDEPQNLIK	667.3777	NA		681.3933	NA		695.4090	15.05	5.87E+05
LVNELTEFAK	596.3350	NA		610.3506	NA		624.3663	16.44	1.18E+06
QTALVELLK	521.8293	NA		535.8449	NA		549.8606	17.03	2.01E+05

	Diethylated 1+			Diethylated 2+			Diethylated 3+		
Peptide	m/z	RT	XIC Area	m/z	RT	XIC Area	m/z	RT	XIC Area
CCTESLVNR	1194.5606	13.58	1.03E+07	597.7842	13.58	5.25E+07	398.8588	NA	
YICDNQDTISSK	1555.7672	14.29	6.16E+06	778.3875	14.28	2.69E+08	519.2610	NA	
HLVDEPQNLIK	1417.8413	15.16	4.98E+06	709.4246	15.16	1.46E+08	473.2857	NA	
LVNELTEFAK	1275.7559	16.64	5.93E+06	638.3819	16.62	1.54E+08	425.9239	NA	
QTALVELLK	1126.7446	17.09	3.10E+06	563.8762	17.09	2.62E+07	376.2535	NA	

	Partial Ethylated Forms	Fully Diethylated Forms	Partial Labeling Percentage
Peptide	XIC sum	XIC sum	% XIC ratio
CCTESLVNR	7.30E+05	6.28E+07	1.16
YICDNQDTISSK	5.91E+05	2.75E+08	0.22
HLVDEPQNLIK	5.87E+05	1.51E+08	0.39
LVNELTEFAK	1.18E+06	1.60E+08	0.74
QTALVELLK	2.01E+05	2.93E+07	0.69

Replicate2

	Partial 1 Ethyl group 2+			Partial 2 Ethyl group 2+			Partial 3 Ethyl group 2+		
Peptide	m/z	RT	XIC Area	m/z	RT	XIC Area	m/z	RT	XIC Area
CCTESLVNR	583.7686	14.28	1.42E+06	NA			NA		
YICDNQDTISSK	736.3406	NA		750.3562	NA		764.3719	15.04	9.81E+05
HLVDEPQNLIK	667.3777	NA		681.3933	NA		695.4090	16.09	1.73E+05
LVNELTEFAK	596.3350	NA		610.3506	NA		624.3663	17.74	7.11E+05
QTALVELLK	521.8293	NA		535.8449	NA		549.8606	18.39	2.01E+05

	Diethylated 1+			Diethylated 2+			Diethylated 3+		
Peptide	m/z	RT	XIC Area	m/z	RT	XIC Area	m/z	RT	XIC Area
CCTESLVNR	1194.5606	14.50	1.39E+07	597.7842	14.50	7.93E+07	398.8588	NA	
YICDNQDTISSK	1555.7672	15.27	8.38E+06	778.3875	15.25	4.25E+08	519.2610	NA	
HLVDEPQNLIK	1417.8413	16.23	5.02E+06	709.4246	16.23	2.12E+08	473.2857	NA	
LVNELTEFAK	1275.7559	17.94	4.47E+07	638.3819	17.92	1.26E+08	425.9239	NA	
QTALVELLK	1126.7446	18.45	2.86E+06	563.8762	18.45	5.12E+07	376.2535	NA	

	Partial Ethylated Forms	Fully Diethylated Forms	Partial Labeling Percentage
Peptide	XIC sum	XIC sum	% XIC ratio
CCTESLVNR	1.42E+06	9.32E+07	1.52
YICDNQDTISSK	9.81E+05	4.33E+08	0.23
HLVDEPQNLIK	1.73E+05	2.17E+08	0.08
LVNELTEFAK	7.11E+05	1.71E+08	0.42
QTALVELLK	2.01E+05	5.41E+07	0.37

Quantification analysis about partial labeling percentage of diethylation is same as described in captions of Table 4.3.

Interestingly, the partial labeling percentage for diethylation during these experiments dramatically decreased to 1% level for all four lysine peptides with the only exception of arginine peptide (CCTESLVNR) at slightly above 1%. This decreased partial diethylation is also confirmed in the 1:1 mix sample of ^{12}C and ^{13}C diethylated peptides (data not shown), which is used for light and heavy diethylation quantification. Since the molar ratio of labeling reagents versus peptides only increased about 60%, the reason for this severe decrease in partial labeling (10~15% to 1%) might be contributed to the absolute labeling reagent concentrations because they actually increased by 5 times. For comparative MAPs expression quantitation using diethylation, the concentration of labeling reagents are substantially adjusted with consideration of increasing sample amount (120μg) and this partial labeling experience.

B) Ionization Efficiency

We investigated ionization behavior of labeled peptides after confident determination of unlabeled peptide percentage and partially labeled peptide percentage of the diethylation method to be at the 1% level. For both dimethylation and diethylation, we prepared a 1-to-1 molar ratio sample by combining the same amount of BSA control peptides (unlabeled) and BSA labeled peptides (dimethylated or diethylated). For the BSA labeled peptides, because the contribution of unlabeled peptides after dimethylation or diethylation is less than 1% compared to the fully labeled form, we can consider this 1-to-1 molar ratio sample to be composed of same amount of unlabeled form and fully labeled form of BSA peptides. Thus, we can compare the XIC intensity of unlabeled

peptides to their fully labeled form in order to evaluate their ionization performance. Three experimental replicates were conducted and result for one replicate is shown in Table 4.6. For both dimethylation and diethylation methods, the fully labeled peptides tends to have a slightly higher charge state distribution. For example, while we did not observe +3 ions of unlabeled YICDNQDTISSK peptide in both 1-to-1 mix samples, we observed +3 ions of fully-dimethylated YICDNQDTISSK peptide with intensity of 3.49E+06 and +3 ions of fully-diethylated YICDNQDTISSK peptide with intensity of 9.85E+06, respectively. In addition, for peptide HLVDEPQNLIK, in the dimethylation 1-to-1 mix sample, its unlabeled form has 3.49E+06 intensity for +3 ions but its fully-dimethylated form has 4.87E+07 for +3 ions. In the diethylation 1-to-1 mix sample, the unlabeled form has 1.81E+07 for +3 ions but the fully-diethylated form has 4.53E+07 for +3 ions.

Another apparent phenomenon we observed from Table 4.6 is that there is an overall increased ionization for fully dimethylated peptides comparing to their unlabeled form. However, for fully diethylated peptides, ionization decreased compared to the unlabeled form. The same trend is detected for all three replicates as shown in Table 4.7. From these representative BSA peptides results, dimethylation has an average 78% increased ionization while diethylation has average 27% decreased ionization.

Table 4.6. Ionization performance comparison of dimethylation and diethylation from 1 to 1 molar ratio sample for unlabeled form and fully labeled form of representative BSA peptides. (A) Dimethylation result. (B) Diethylation result.

A

	Unlabeled 1+			Unlabeled 2+			Unlabeled 3+		
Peptide	m/z	RT	XIC Area	m/z	RT	XIC Area	m/z	RT	XIC Area
CCTESLVNR	1138.4980	16.99	4.69E+06	569.7529	16.99	1.58E+08	380.1712	NA	
YICDNQDTISSK	1443.6420	17.11	4.35E+06	722.3249	17.11	2.63E+08	481.8859	NA	
HLVDEPQNLIK	1305.7161	18.27	9.96E+05	653.3620	18.27	1.82E+08	435.9106	18.27	3.49E+06
LVNELTEFAK	1163.6307	19.57	5.14E+06	582.3193	19.57	1.60E+08	388.5488	NA	
QTALVELLK	1014.6194	20.07	4.16E+06	507.8136	20.07	8.74E+07	338.8784	NA	

	Dimethylated 1+			Dimethylated 2+			Dimethylated 3+		
Peptide	m/z	RT	XIC Area	m/z	RT	XIC Area	m/z	RT	XIC Area
CCTESLVNR	1166.5293	17.09	9.90E+06	583.7686	17.09	1.72E+08	389.5150	NA	
YICDNQDTISSK	1499.7046	17.26	1.76E+06	750.3562	17.26	3.68E+08	500.5734	17.26	2.96E+06
HLVDEPQNLIK	1361.7787	18.37	1.32E+06	681.3933	18.40	2.88E+08	454.5981	18.37	4.87E+07
LVNELTEFAK	1219.6933	19.65	7.88E+06	610.3506	19.65	2.47E+08	407.2363	NA	
QTALVELLK	1070.6820	20.19	4.05E+06	535.8449	20.19	1.10E+08	357.5659	NA	

	Unlabeled Forms	Dimethylated Forms	Dimethylated vs Unlabeled
Peptide	XIC sum	XIC sum	XIC ratio
CCTESLVNR	1.63E+08	1.82E+08	1.1166
YICDNQDTISSK	2.67E+08	3.73E+08	1.3970
HLVDEPQNLIK	1.86E+08	3.38E+08	1.8172
LVNELTEFAK	1.65E+08	2.55E+08	1.5455
QTALVELLK	9.16E+07	1.14E+08	1.2445

B

	Unlabeled 1+			Unlabeled 2+			Unlabeled 3+		
Peptide	m/z	RT	XIC Area	m/z	RT	XIC Area	m/z	RT	XIC Area
CCTESLVNR	1138.4980	17.22	3.61E+06	569.7529	17.22	2.19E+08	380.1712	NA	
YICDNQDTISSK	1443.6420	17.33	4.52E+06	722.3249	17.33	3.61E+08	481.8859	NA	
HLVDEPQNLIK	1305.7161	18.53	2.36E+06	653.3620	18.53	3.78E+08	435.9106	18.53	1.81E+07
LVNELTEFAK	1163.6307	19.74	8.68E+06	582.3193	19.74	2.26E+08	388.5488	NA	
QTALVELLK	1014.6194	20.28	4.31E+06	507.8136	20.28	9.41E+07	338.8784	NA	

	Diethylated 1+			Diethylated 2+			Diethylated 3+		
Peptide	m/z	RT	XIC Area	m/z	RT	XIC Area	m/z	RT	XIC Area
CCTESLVNR	1194.5606	17.54	2.85E+06	597.7842	17.54	1.50E+08	398.8588	NA	
YICDNQDTISSK	1555.7672	18.24	1.48E+06	778.3875	18.24	3.79E+08	519.2610	18.26	9.85E+06
HLVDEPQNLIK	1417.8413	19.05	4.46E+05	709.4246	19.05	1.89E+08	473.2857	19.05	4.53E+07
LVNELTEFAK	1275.7559	20.36	8.35E+05	638.3819	20.36	1.25E+08	425.9239	NA	
QTALVELLK	1126.7446	20.80	3.35E+05	563.8762	20.76	4.56E+07	376.2535	NA	

	Unlabeled Forms	Diethylated Forms	Diethylated vs Unlabeled
Peptide	XIC sum	XIC sum	XIC ratio
CCTESLVNR	2.23E+08	1.53E+08	0.6861
YICDNQDTISSK	3.66E+08	3.90E+08	1.0656
HLVDEPQNLIK	3.98E+08	2.35E+08	0.5905
LVNELTEFAK	2.35E+08	1.26E+08	0.5362
QTALVELLK	9.84E+07	4.59E+07	0.4665

Five representative peptides (1 arginine peptide and 4 lysine peptides) that elute across the BSA digest retention timeframe were selected for illustration. For each peptide after dimethylation or diethylation, we consider three possible charge states (+1, +2, +3) of its unlabeled form or fully-labeled form. XIC peak area and RT information for each possible precursor charge state were manually extracted from raw data using Xcalibur 2.1 based on their theoretical m/z value with mass tolerance of 0.1 Da. Peak detection and peak area calculation was automatically processed by Xcalibur Qual Browser with "Genesis" peak detection algorithm. XIC sum is the summation of XIC peak areas of +1, +2, +3 charge state precursors of the same form of peptide. Ionization performance is measured by XIC ratio which is calculated for each peptide, its XIC sum of fully labeled form (dimethylated or diethylated) versus its unlabeled form.

Table 4.7. Summary of ionization performance for unlabeled form and fully labeled form of representative BSA peptides from three experimental replicates of dimethylation and diethylation experiment 1-to-1 molar ratio sample.

XIC ratio (Fully labeled versus Unlabeled) from 1 to 1 molar ratio mix of Unlabeled and Labeled peptides

	Dimethylation					Diethylation				
Peptides	Rep1	Rep2	Rep3	Average	Stdev	Rep1	Rep2	Rep3	Average	Stdev
CCTESLVNR	1.1166	1.3841	1.4686	1.3231	0.1838	0.6861	0.7386	0.6456	0.6901	0.0466
YICDNQDTISSK	1.3970	1.6486	1.5492	1.5316	0.1267	1.0656	1.5132	1.0274	1.2021	0.2701
HLVDEPQNLIK	1.8172	2.4236	3.4726	2.5711	0.8375	0.5905	0.5838	0.6179	0.5974	0.0181
LVNELTEFAK	1.5455	1.9236	1.4822	1.6504	0.2387	0.5362	0.5459	0.6154	0.5658	0.0432
QTALVELLK	1.2445	1.4537	1.3644	1.3542	0.1050	0.4665	0.6098	0.5612	0.5458	0.0729
Five Peptide Average				1.7768	0.3270				0.7278	0.1011

Ionization performance for both dimethylation and diethylation methods is evaluated for each peptide, by calculating the XIC ratio of its fully labeled form versus unlabeled form. For three experimental replicates, the average value and standard deviation of XIC ratio is calculated for each peptide and then averaged across five peptides.

C) Retention behavior and MS/MS fragmentation pattern of labeled peptides

After dimethylation or diethylation, the added methyl groups or ethyl groups would increase the overall hydrophobicity of the fully labeled peptide resulting in an increased retention time during reversed-phase (RP) separation. Illustrated in Figure 4.4A, 4.4B and

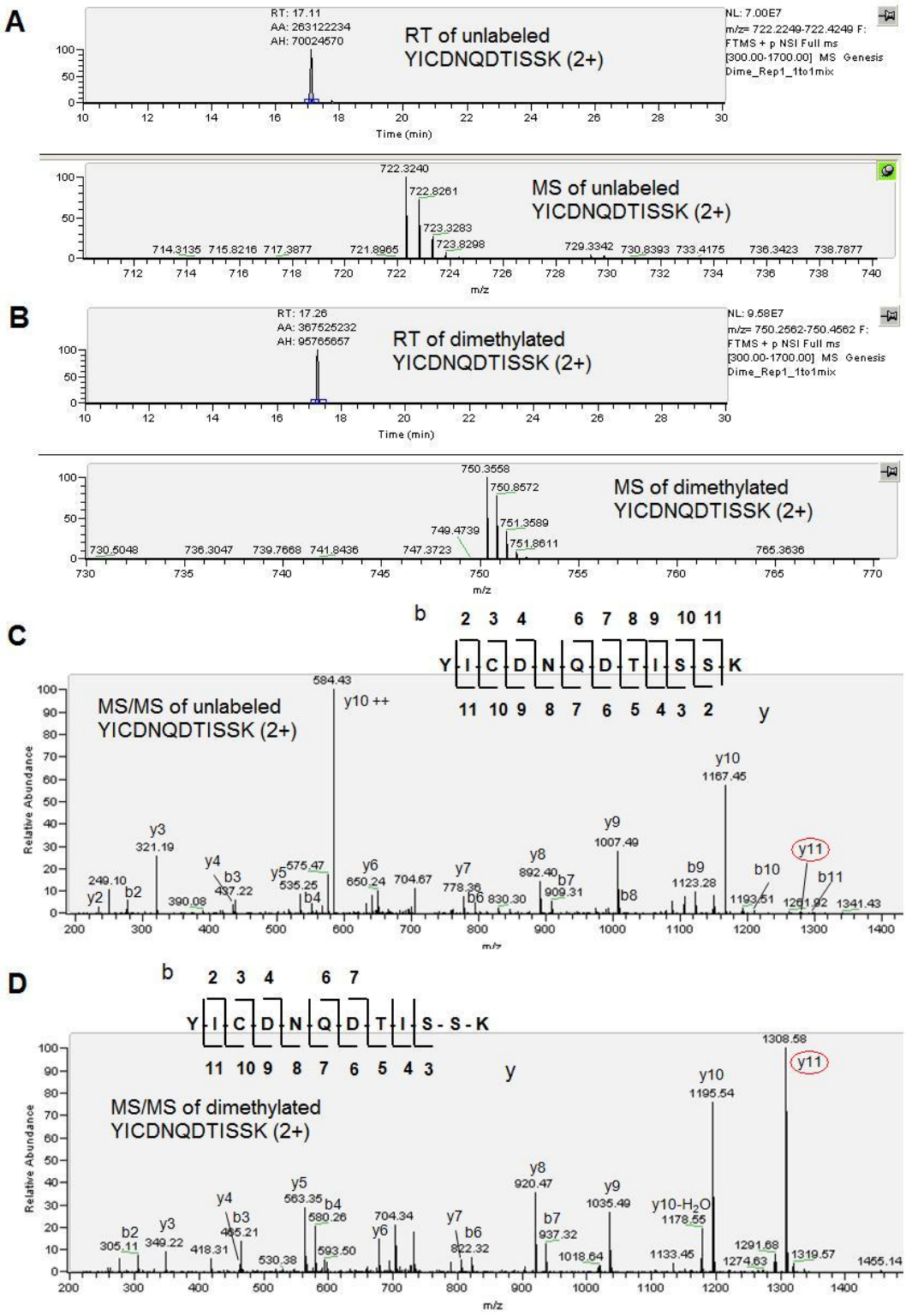


Figure 4.4. Retention behavior and MS/MS pattern of doubly charged BSA peptide YICDNQDTISSK from 1-to-1 molar ratio (unlabeled : dimethylated) sample. (A) RT and MS for the unlabeled form. (B) RT and MS for the fully-dimethylated form. (C) MS/MS for the unlabeled form. (D) MS/MS for the fully-dimethylated form.

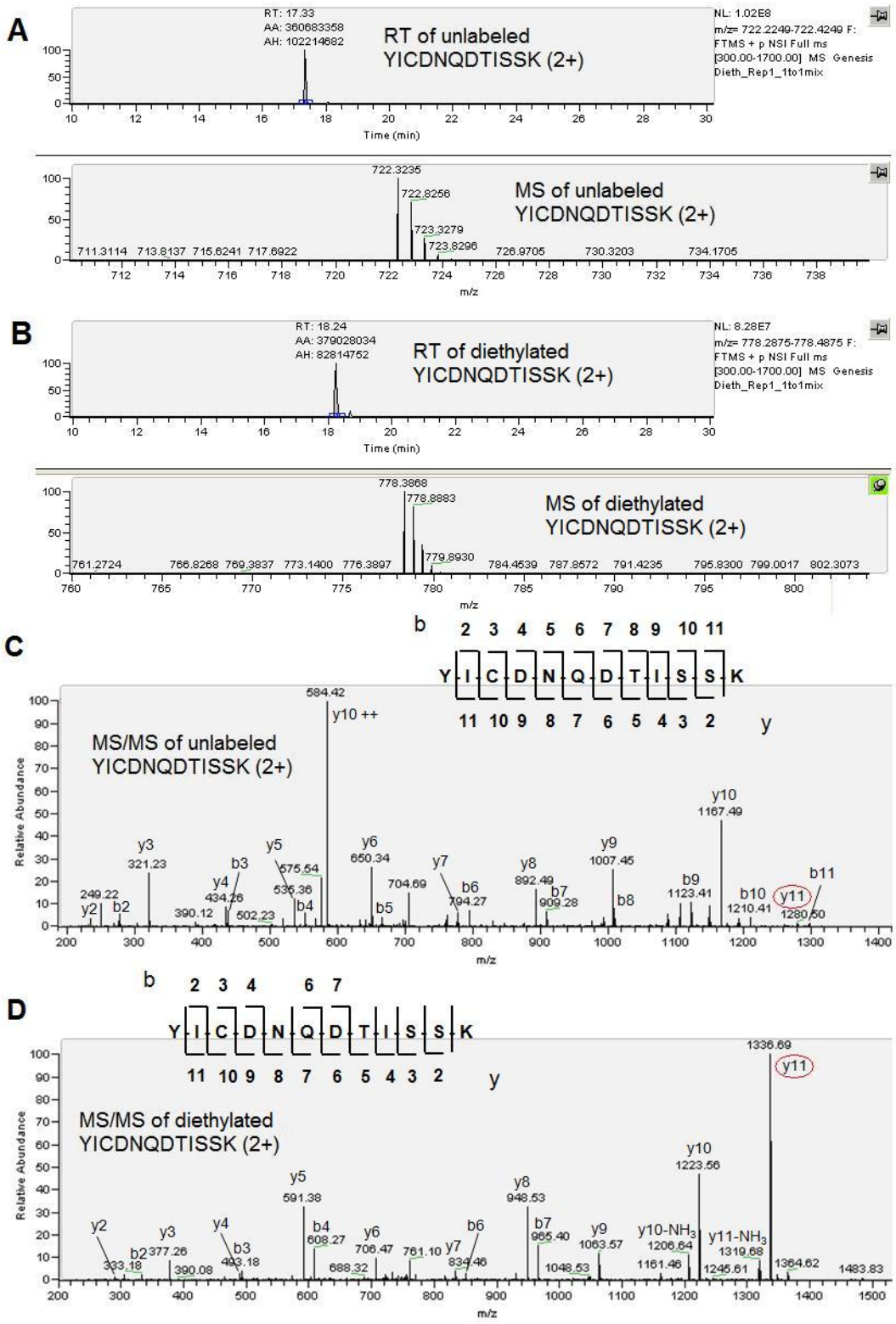


Figure 4.5. Retention behavior and MS/MS pattern of doubly charged BSA peptide YICDNQDTISSK from 1-to-1 molar ratio (unlabeled : diethylated) sample. (A) RT and MS for the unlabeled form. (B) RT and MS for the fully-diethylated form. (C) MS/MS for the unlabeled form. (D) MS/MS for the fully-diethylated form.

Figure 4.5A, 4.5B with doubly charged BSA peptide YICDNQDTISSK, during a 17 minute 0-70% solvent B gradient, we observed retention shifts of 0.15 minute for fully-dimethylated form and 0.52 minute for fully-diethylated form. The longer retention shift for diethylation is due to greater hydrophobicity of ethyl group than the methyl group. There have been reports that dimethylation would enhance a_1 ion during CID fragmentation [12], while we did not observe this phenomena due to low mass cut-off associated with the ion trap instrument used in our studies we observed an tremendous enhancement of y_{n-1} ion (where n is the total number of amino acids in a peptide) for both dimethylation and diethylation shown in Figure 4.4C, 4.4D and Figure 4.5C, 4.5D. The same trend of retention time shifts and enhancement of y_{n-1} ions for MS/MS pattern had also been confirmed for all other representative BSA peptides. In contrast, the y_{n-1} ion has a very low signal or near absent in CID MSMS spectra obtained from unlabeled peptides.

4.3.2 Diethylation Quantification Validation with BSA and Beta-casein Peptides

The advantage of diethylation for isotopic quantification over dimethylation is due to the larger mass shift introduced by ethyl groups. By using ^{12}C acetaldehyde and ^{13}C acetaldehyde for diethylation, the fully labeled peptide would not have retention time shift between the light-labeled form and heavy-labeled form. From Figure 4.6A, 4.6B and Figure 4.7A, 4.7B, both the doubly charged and triply charged BSA peptides HLVDEPQNLIK have identical elution profile between the ^{12}C diethylated form and ^{13}C diethylated form as expected. In addition, because the light labeled peptide is mixed with heavy labeled peptide by 1:1 ratio, they should have same ionization intensity which is

also confirmed by the average relative intensity in Figure 4.6C and 4.7C. The mass difference of 4 Th (mass over charge unit) for +2 charged peptides and 2.67 Th for +3 charged peptides in Figure 4.6C and 4.7C would be enough for isotopic quantification of typical lysine-containing peptides with appropriate precursor isolation window setting.

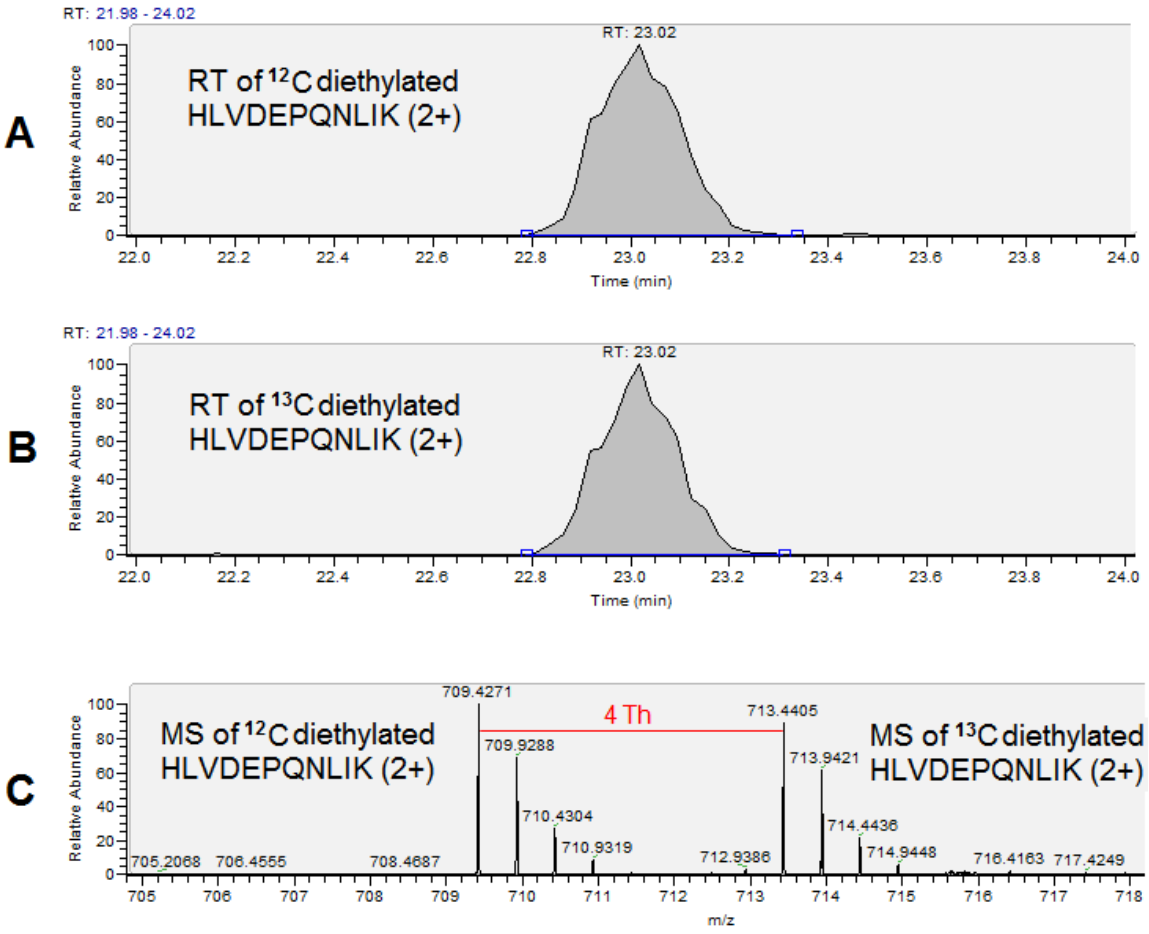


Figure 4.6. Doubly charged BSA peptide HLVDEPQNLIK diethylated with both ^{12}C and ^{13}C isotopes from 1-to-1 molar ratio mixed sample. (A) Extracted ion chromatogram with RT of peptide HLVDEPQNLIK (2+) after ^{12}C diethylation. (B) Extracted ion chromatogram with RT of peptide HLVDEPQNLIK (2+) after ^{13}C diethylation. (C) Mass difference and relative intensity (averaged across retention range 22.80-23.23 min) of 1:1 ratio of ^{12}C and ^{13}C diethylated peptide HLVDEPQNLIK (2+).

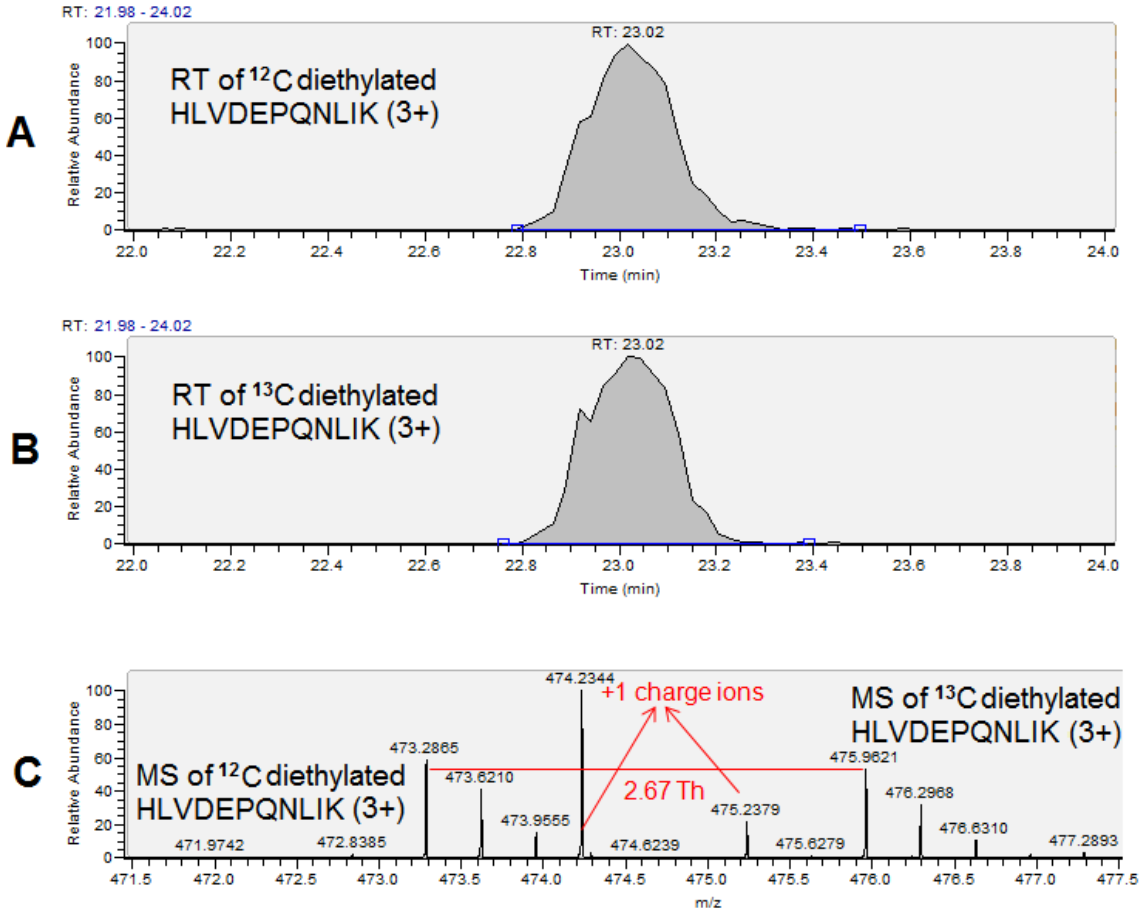


Figure 4.7. Triply charged BSA peptide HLVDEPQNLIK diethylated with both ^{12}C and ^{13}C isotopes from 1-to-1 molar ratio mixed sample. (A) Extracted ion chromatogram with RT of peptide HLVDEPQNLIK (3+) after ^{12}C diethylation. (B) Extracted ion chromatogram with RT of peptide HLVDEPQNLIK (3+) after ^{13}C diethylation. (C) Mass difference and relative intensity (averaged across retention range 22.80-23.23 min) of 1:1 ratio of ^{12}C and ^{13}C diethylated peptide HLVDEPQNLIK (3+).

To validate the diethylation quantitation method, we started with BSA digest and beta-casein digest mixture. After diethylation with ^{12}C or ^{13}C acetaldehyde, we mixed different molar ratios of light diethylated peptides and heavy diethylated peptides. LC-MS/MS analysis data of these samples were searched by Mascot 2.2 and quantified using ProteoIQ 2.3 Isotopic Quantification module. Two experimental replicates were

performed and two technical replicates per sample were analyzed, the results were summarized in Figure 4.8. For both BSA and beta-casein linear regression plot, we see a

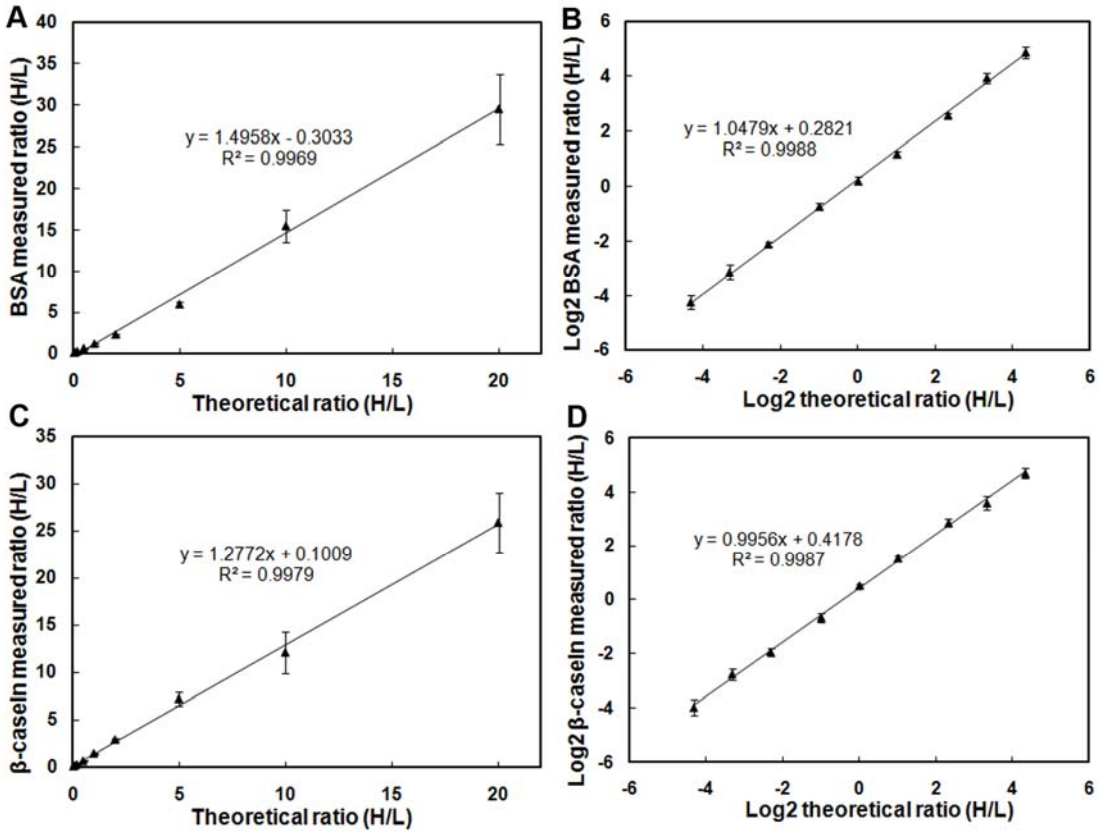


Figure 4.8. Quantification of BSA and beta-casein protein using their labeled peptides after light and heavy diethylation with different molar mix ratio (heavy versus light from 0.05 to 20). (A) Linear regression plot of the measured BSA ratio H/L ($^{13}C/^{12}C$) versus theoretical ratio. (B) Log2 plot of the measured BSA ratio H/L ($^{13}C/^{12}C$) versus theoretical ratio. (C) Linear regression plot of the measured beta-casein ratio H/L ($^{13}C/^{12}C$) versus theoretical ratio. (D) Log2 plot of the measured beta-casein ratio H/L ($^{13}C/^{12}C$) versus theoretical ratio. The measured ratio (H/L) for BSA or beta-casein is the average value of 4 replicates (2 experimental × 2 technical) from ProteoIQ processed result, which consider all quantifiable peptide XIC ratios (H/L) for the same protein. Error bar represents the standard deviation.

good linear fit with R^2 value of about 0.997. The slope is higher than 1 for both BSA and beta-casein, which could be contributed by slight sample carryover between LC-MS runs

and the expected larger error associated with quantification of large H/L ratios. The log2 plot of measured H/L ratio had a better linear fit for theoretical ratio with R^2 value of about 0.999 and the slope is close to 1. The reason is that this calculation method would normalize out initial sample concentration differences and it has been typically used in assessment of relative quantification. From our result, the diethylation method could quantify protein relative abundance change to a magnitude of 20. As expected, like other isotopic quantification methods, this method would be more accurate for quantification of relative abundance change within a magnitude of 5, which can be judged from the error bars in Figure 4.8.

4.3.3 Diethylation of Complex Tryptic MAPs Digest and Quantification of Protein Expression Change after 2D-RP-HPLC-MSMS Analysis

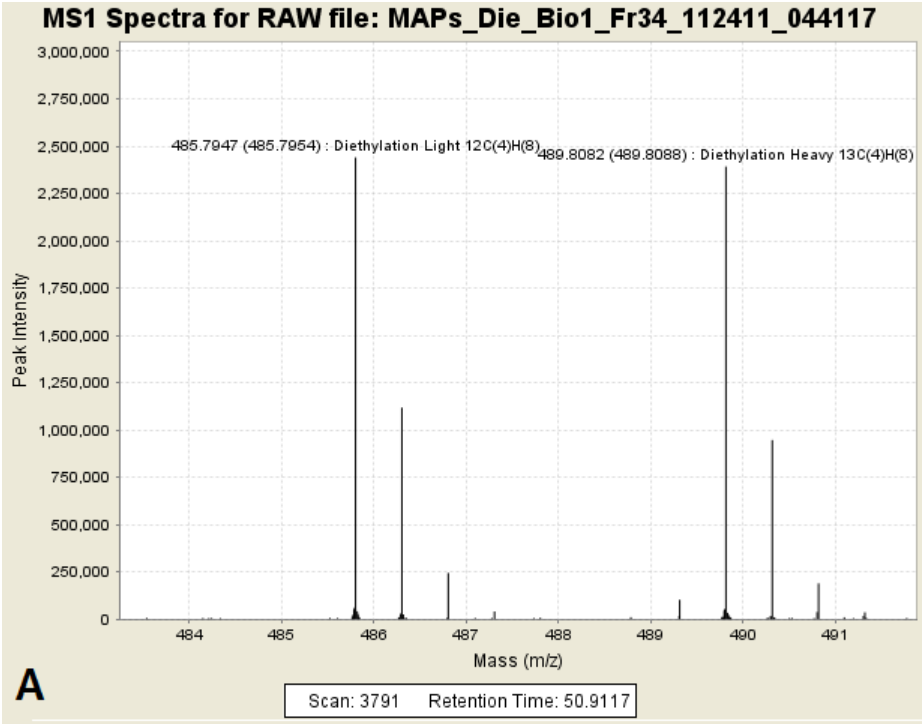
After validating the diethylation quantification method using ^{12}C acetaldehyde and ^{13}C acetaldehyde, we applied this method to detect MAPs expression changes in MCF-7 cells caused by 2ME modulation. MAPs were isolated from MCF-7 breast cancer cells using a modified Taxol-stabilization procedure as described in materials and methods. Tryptic peptides derived from MAPs were then labeled using ^{12}C acetaldehyde (for control cells) or ^{13}C acetaldehyde (for 2ME treated cells). The light/heavy labeled samples were mixed in a 1:1 ratio and fractionated by offline RP-HPLC at pH 7.5. Two biological replicates were conducted and each offline fraction of labeled peptides was analyzed via low-pH LC-MSMS on both the Thermo FT-ICR and the Oribitrap Velos instruments. Data was searched using Mascot (v2.2) and processed by the ProteoIQ (v2.3)

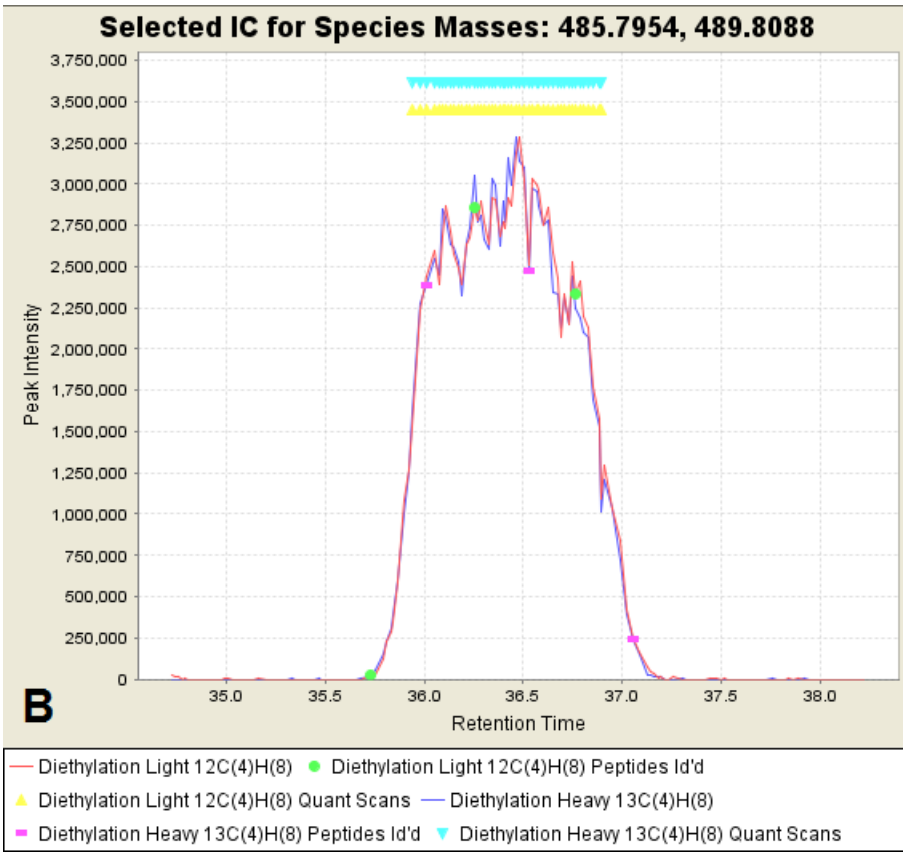
Isotopic Labeling Quantification module. Partial diethylation was not considered during the search since we already verified its percentage to be low based on optimal diethylation result of representative BSA peptides. For quantification of protein expression change induced by 2ME treatment, only the fully diethylated forms of peptides were considered for these quantifiable proteins. For general proteomic level identification, non-labeled peptides were also included for overall protein IDs.

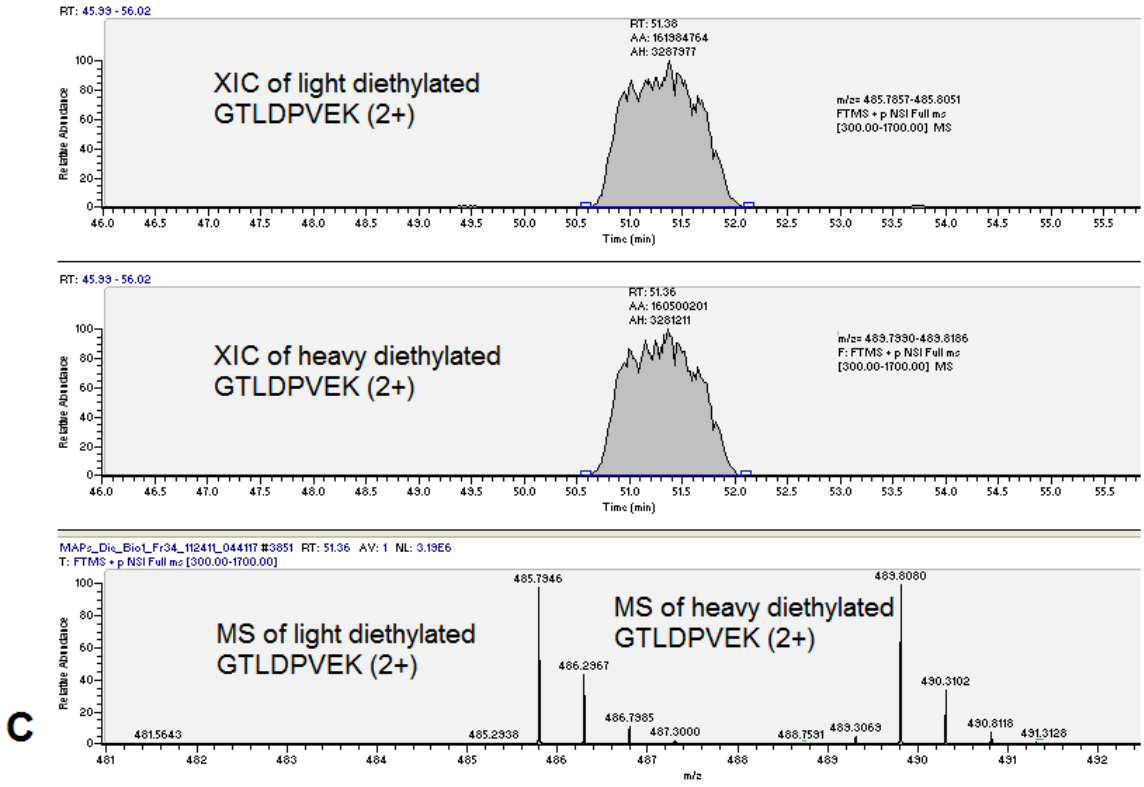
Combining both the FT data and Orbitrap data for all offline fractions from the two biological replicates, with initial filter of protein FDR 5%, minimum protein probability of 0.5, there were 84696 spectrum peptide matches (SPMs), from which, only 275 (0.324%) SPMs have not been diethylation labeled. This result suggests that for the diethylation of complex sample, the labeling efficiency is close to 100%. For protein IDs, we applied the filter of Mascot score of 30, minimum peptide length of 7 amino acids, 2 SPMs per protein, peptide and protein FDR of 1.0%. After the filtering, there are 45282 SPMs (including redundant peptide IDs), 11815 unique peptide IDs and 1466 protein IDs. Of these SPMs, only 136 (0.300%) are the non-labeled form of peptides. So, in terms of peptide identifications, the diethylation labeling efficiency is extremely high with about 99.7% peptide IDs being fully labeled.

All of the observed diethylated $^{13}C/^{12}C$ peptide pairs had similar retention times. ProteoIQ extracted the XIC area for both the light (^{12}C) and heavy (^{13}C) diethylated forms of a peptide and calculate the ratio of H/L (heavy/light). We compared ProteoIQ processed results and manually verified results for some sample peptides.

Figure 4.9 shows the example of a peptide GTLDPVEK that does not have any quantity changes after drug treatment. Both the light form (^{12}C diethylated, control cells) and heavy form (^{13}C diethylated, 2ME treated cells) have the same elution profile as expected in Figure 4.9B (note RT display in Figure 4.9B is wrong due to ProteoIQ software bugs). The intensity information ProteoIQ extracted is 1.197E+08 for heavy form and 1.202E+08 for light form, resulting in a H/L ratio of 0.996. With manual exploration of the raw data, the light and heavy form of peptide GTLDPVEK does have the same XIC elution profile in Figure 4.9C and the XIC area of light form is 1.620E+08 and 1.605E+08 for heavy form. This manual validation of 0.991 H/L ratio is consistent with ProteoIQ result for quantification. Figure 4.10 shows the example of a peptide EAGGGGVGGPGAK that has a significant expression change after drug treatment. While the elution profile for the light and heavy form is still the same as confirmed by ProteoIQ and manual validation, ProteoIQ extracted intensity of 6.130E+05 for heavy form and 7.301E+4 for light form, resulting H/L ratio of 8.396. Manually extracted result is 1.408E+06 for heavy and 2.087E+05 for light with calculated H/L ratio of 6.747. The minor difference between quantification values might be contributed to the fact that for the same elution profile, ProteoIQ integrates MS intensity above certain percentage of peak apex (default at 50% peak apex) while manual extracted XIC is the integration of whole peak area. Overall the ProteoIQ quantification results are consistent with manually verified results for quantification purposes, based on these 2 peptide examples.

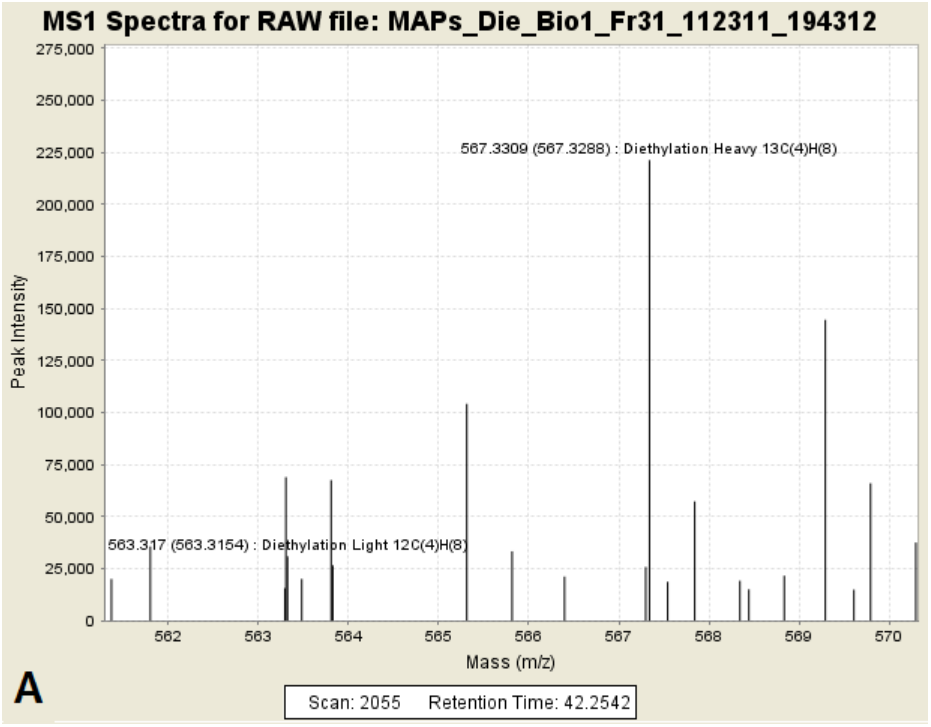

MS1 Spectra for RAW file: MAPs_Die_Bio1_Fr34_112411_044117
485.7947 (485.7954) : Diethylation Light 12C(4)H(8)
489.8082 (489.8088) : Diethylation Heavy 13C(4)H(8)
Peak Intensity
Mass (m/z)
A
Scan: 3791 Retention Time: 50.9117

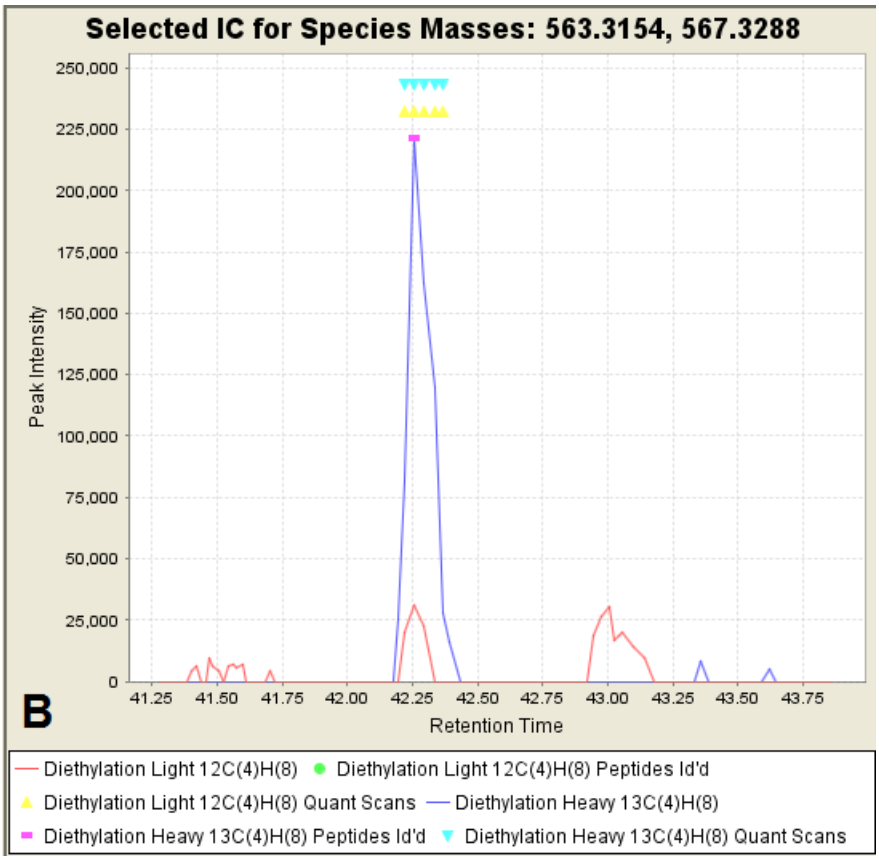

Selected IC for Species Masses: 485.7954, 489.8088
Peak Intensity
Retention Time
B
Diethylation Light 12C(4)H(8)
Diethylation Light 12C(4)H(8) Peptides Id'd
Diethylation Light 12C(4)H(8) Quant Scans
Diethylation Heavy 13C(4)H(8)
Diethylation Heavy 13C(4)H(8) Peptides Id'd
Diethylation Heavy 13C(4)H(8) Quant Scans



C

Figure 4.9. ProteoIQ quantification and manual quantification of doubly charged peptide GTLDPVEK for the light diethylated form (m/z 485.7954) and heavy diethylated form (m/z 489.8088). (A) ProteoIQ regenerated MS spectra at retention time 50.96min when the heavy diethylated form of GTLDPVEK (2+) triggered CID MS/MS fragmentation with Mascot score of 57. (B) ProteoIQ regenerated XIC for the light and heavy diethylated forms of peptide GTLDPVEK. (C) Manual extraction of XIC for light and heavy diethylated forms of GTLDPVEK (2+) ions and FT MS spectra at peptide peak apex of elution profile (RT at 51.36min). Both ProteoIQ XIC extraction and manual XIC extraction are based on theoretical m/z values of peptide precursor ions with mass tolerance of 20 ppm.


MS1 Spectra for RAW file: MAPs_Die_Bio1_Fr31_112311_194312
567.3309 (567.3288) : Diethylation Heavy 13C(4)H(8)
563.3117 (563.3154) : Diethylation Light 12C(4)H(8)
Peak Intensity
Mass (m/z)
A
Scan: 2055 Retention Time: 42.2542


Selected IC for Species Masses: 563.3154, 567.3288
Peak Intensity
Retention Time
B
Diethylation Light 12C(4)H(8)
Diethylation Light 12C(4)H(8) Peptides Id'd
Diethylation Light 12C(4)H(8) Quant Scans
Diethylation Heavy 13C(4)H(8)
Diethylation Heavy 13C(4)H(8) Peptides Id'd
Diethylation Heavy 13C(4)H(8) Quant Scans

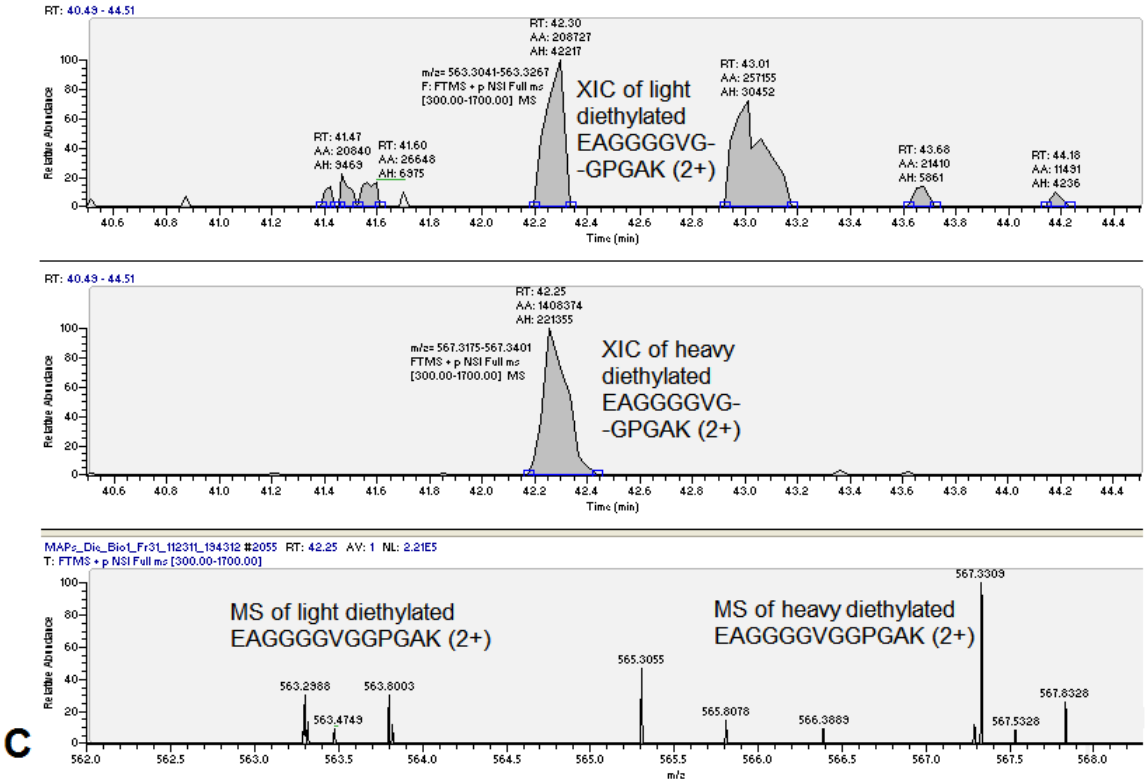


Figure 4.10. ProteoIQ quantification and manual quantification of doubly charged peptide EAGGGGVGGPGAK for the light diethylated form (m/z 563.3154) and heavy diethylated form (m/z 567.3288). (A) ProteoIQ regenerated MS spectrum at retention time 42.25min when the heavy diethylated form of EAGGGGVGGPGAK (2+) triggered CID MS/MS fragmentation with Mascot score of 65. (B) ProteoIQ regenerated XIC for the light and heavy diethylated form of peptide EAGGGGVGGPGAK. (C) Manual extraction of XIC for light and heavy forms of EAGGGGVGGPGAK (2+) ions and FT MS specta at peptide peak apex of elution profile (RT at 42.25min). Both ProteoIQ XIC extraction and manual XIC extraction are based on theoretical m/z values of peptide precursor ions with mass tolerance of 20 ppm.

After determining the ratio of H/L at the peptide level, ProteoIQ consider all assigned peptide IDs for each protein ID and calculated the protein relative expression change. For the FT data of two biological replicates, protein quantification filtering criteria were set as follows: peptide Mascot score at 30 and with at least 7 amino acids in composition; each protein ID with at least 2 SPMs; peptide and protein FDR at 1.0%; and each protein

ID must be identified in both biological replicates. After filtering, there are 13287 SPMs (including redundant peptide IDs), 5089 unique peptides and 791 protein IDs. Of these, 402 protein IDs have quantification information based on XIC intensity. For the Orbitrap data with the same protein quantification filtering criteria, there are 29649 SPMs, 10631 unique peptides and 1337 protein IDs within which 697 IDs are quantified.

In general, the majority of peptides of MAPs isolated from MCF control cells (light label) and 2ME treated cells (heavy label) showed an experimental 1:1 ratio result when they were mixed by the 1 to 1 ratio. Shown in Figure 4.11, we can find that most peptides fall into the region of Log2 relative expression (H/L ratio) close to 0, which means there is no expression change for these peptide IDs after drug treatment, especially for abundant peptide IDs with high intensities. A small percentage of peptides have significant expression changes with Log2 relative expression (H/L ratio) >1 or <-1 from Figure 4.11A and 4.11B, and they are mostly very low intensity peptides. A similar trend is also observed for protein IDs shown in Figure 4.12. Most proteins do not have apparent expression changes after 2ME treatment, as we can see their Log2 relative expression values to be close to 0. However, there are some proteins in Figure 4.12A and 4.12B with a relative expression change greater than 2 (where Log2 value >1 or <-1) after 2ME modulation.

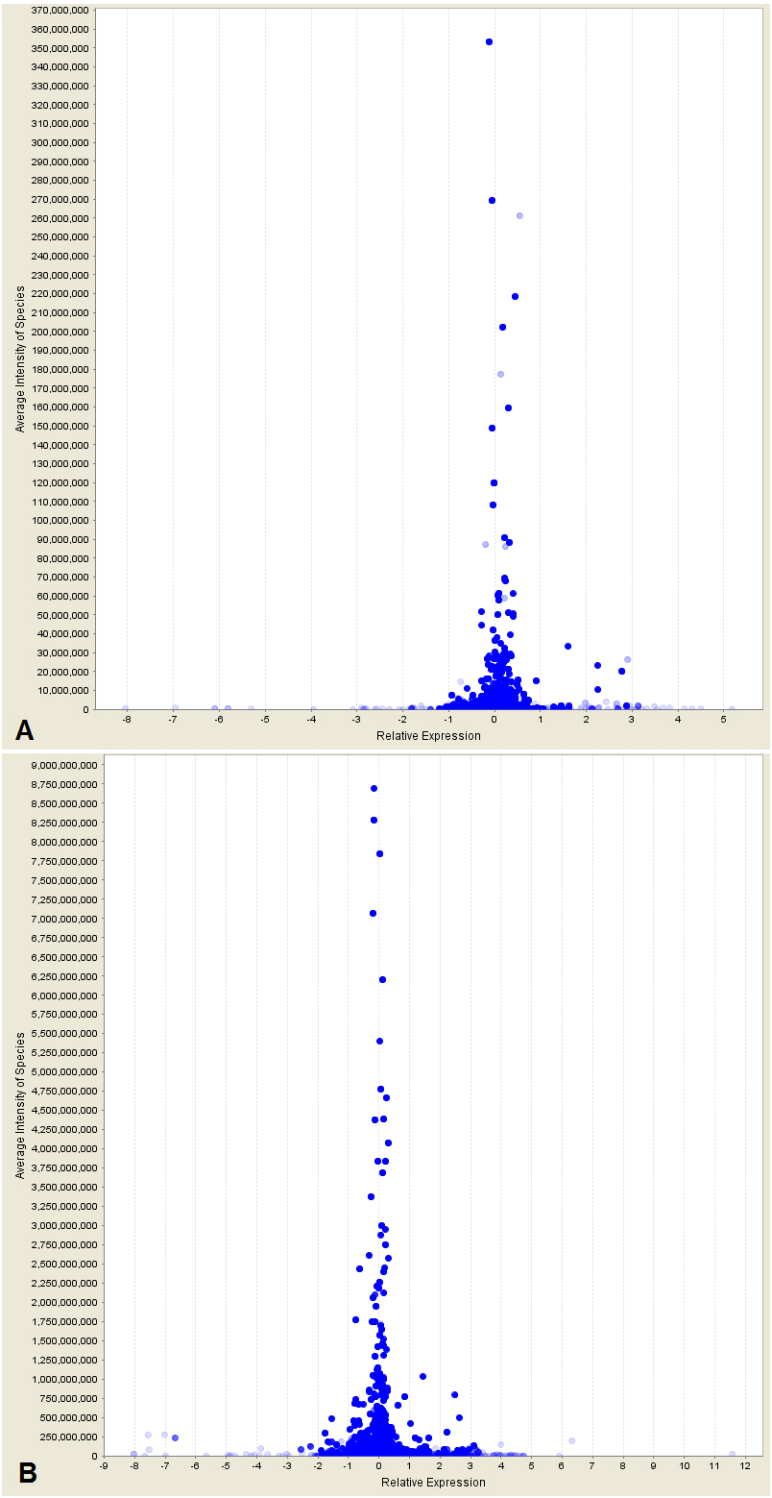


Figure 4.11. ProteoIQ generated peptide average intensity distribution over relative expression (Log2 of H/L ratio) from two biological replicates. (A) FT data analysis result. (B) Orbitrap data analysis result.

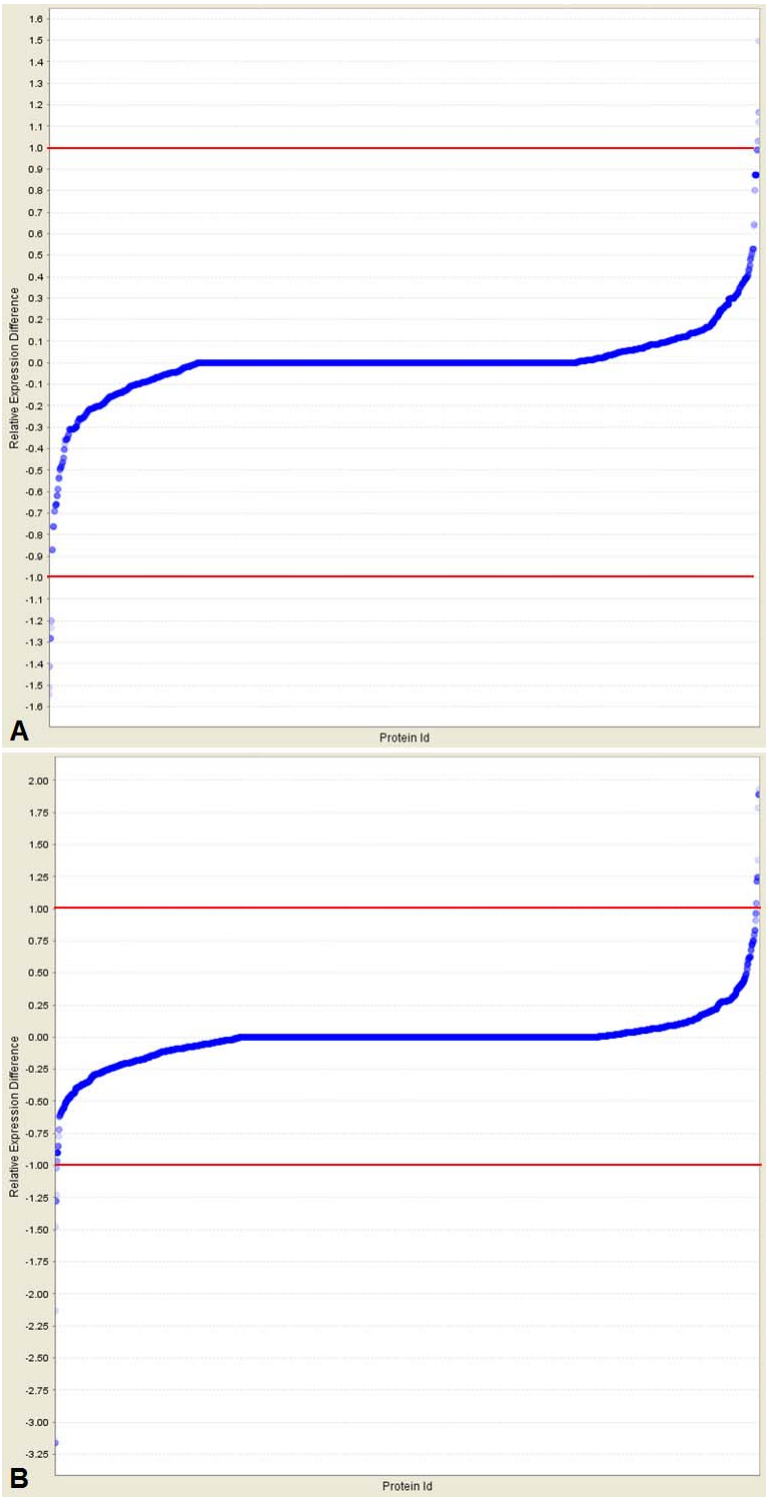


Figure 4.12. ProteoIQ generated plot of protein relative expression difference (2ME treated versus Control) for identified protein IDs. (A) FT data quantification result. (B) Orbitrap data quantification result. Protein abundance is indicated by the degree of blue color.

4.3.4 Proteins Candidates with Up-regulation or Down-regulation Expression Change as Potential Biomarker for 2ME Treatment

Only 5-7% of the quantifiable proteins (402 from FTMS data; 697 from Orbitrap data) exhibited a significant change in abundance with 2ME treatment on MCF-7 cells. We analyzed peptide fractions from each biological replicate with one technical replicate on the FT-ICR instrument and another technical replicate on Orbitrap Velos instrument. As expected, instrument sensitivity and performance differences were observed. Even the sample loading amount on the Orbitrap Velos is about 50% compared to FT-ICR, due to better sensitivity and faster duty-cycle, we had about twice the peptide IDs and 70% increase in terms of protein IDs from the Orbitrap result.

To be considered as protein candidates that have significant up-regulation or down-regulation, we require the Log2 relative expression (Log2 H/L ratio) value to be at least 1 for up-regulation or -1 for down-regulation in either the FT quantification dataset or Orbitrap quantification dataset. Then we consider two levels of candidates, first level candidates require their Log2 relative expression values to be at least 0.50 or -0.50 from the other dataset, second level candidates require their Log2 relative expression values to be positive for up-regulation or negative for down-regulation. Candidates that have one set of qualified Log2 relative expression value but were not quantified in the other dataset due to instrument sensitivity are also included in the list. Those candidates with up- or down-regulation with 2ME modulation are listed in Table 4.8.

Table 4.8. Protein candidates with significant expression changes based on diethylation result of isolated MAPs from MCF-7 cells with 2ME modulation. (A) Up-regulation candidates. (B) Down-regulation candidates.

			Log2 relative expression	
Selction	Up Regulation after 2ME		FT	ORBI
Both	sp|P06454-2|PTMA_HUMAN	Isoform 2 of Prothymosin alpha OS=Homo sapiens GN=PTMA	1.75	0.80
	sp|P55786|PSA_HUMAN	Puromycin-sensitive aminopeptidase OS=Homo sapiens GN=NPEPPS PE=1 SV=	0.63	1.24
FT	sp|O75116|ROCK2_HUMAN	Rho-associated protein kinase 2 OS=Homo sapiens GN=ROCK2 PE=1 SV=4	2.24	0.15
	sp|O15144|ARPC2_HUMAN	Actin-related protein 2/3 complex subunit 2 OS=Homo sapiens GN=ARPC2 PE=1	2.06	0.31
ORBI	tr|A1A441|A1A441_HUMAN	Serine/threonine-protein phosphatase OS=Homo sapiens GN=PPP3CA PE=2 SV=	NA	3.78
	sp|P09493-5|TPM1_HUMAN	Isoform 5 of Tropomyosin alpha-1 chain OS=Homo sapiens GN=TPM1	NA	2.75
	sp|Q99996-2|AKAP9_HUMAN	Isoform 2 of A-kinase anchor protein 9 OS=Homo sapiens GN=AKAP9	NA	2.48
	sp|Q15435-3|PP1R7_HUMAN	Isoform 3 of Protein phosphatase 1 regulatory subunit 7 OS=Homo sapiens GN=PF	NA	2.42
	sp|Q9H0D6-2|XRN2_HUMAN	Isoform 2 of 5'-3' exoribonuclease 2 OS=Homo sapiens GN=XRN2	NA	2.08
	sp|O43719|HTSF1_HUMAN	HIV Tat-specific factor 1 OS=Homo sapiens GN=HTATSF1 PE=1 SV=1	NA	1.92
	sp|P08237|K6PF_HUMAN	6-phosphofructokinase, muscle type OS=Homo sapiens GN=PFKM PE=1 SV=2	NA	1.82
	sp|Q9ULT8|HECD1_HUMAN	E3 ubiquitin-protein ligase HECTD1 OS=Homo sapiens GN=HECTD1 PE=1 SV=3	NA	1.60
	sp|Q7Z6B7-2|SRGP1_HUMAN	Isoform 2 of SLIT-ROBO Rho GTPase-activating protein 1 OS=Homo sapiens GN=S	NA	1.60
	sp|Q9H993|CF211_HUMAN	UPF0364 protein C6orf211 OS=Homo sapiens GN=C6orf211 PE=1 SV=1	NA	1.55
	sp|P62993|GRB2_HUMAN	Growth factor receptor-bound protein 2 OS=Homo sapiens GN=GRB2 PE=1 SV=1	0.07	1.46
	sp|Q5H9R7-2|PP6R3_HUMAN	Isoform 2 of Serine/threonine-protein phosphatase 6 regulatory subunit 3 OS=Homo	NA	1.44
	sp|Q86TU7|SETD3_HUMAN	Histone-lysine N-methyltransferase setd3 OS=Homo sapiens GN=SETD3 PE=1 SV	NA	1.38
	sp|P27694|RFA1_HUMAN	Replication protein A 70 kDa DNA-binding subunit OS=Homo sapiens GN=RPA1 P	NA	1.37
	sp|O43290|SNUT1_HUMAN	U4/U6.U5 tri-snRNP-associated protein 1 OS=Homo sapiens GN=SART1 PE=1 SV	NA	1.36
	sp|P09493-6|TPM1_HUMAN	Isoform 6 of Tropomyosin alpha-1 chain OS=Homo sapiens GN=TPM1	NA	1.35
	sp|Q00765|REEP5_HUMAN	Receptor expression-enhancing protein 5 OS=Homo sapiens GN=REEP5 PE=1 SV	NA	1.26
	sp|Q01831|XPC_HUMAN	DNA repair protein complementing XP-C cells OS=Homo sapiens GN=XPC PE=1 S	NA	1.24
A	sp|P53582|AMPM1_HUMAN	Methionine aminopeptidase 1 OS=Homo sapiens GN=METAP1 PE=1 SV=2	NA	1.22
	sp|P20962|PTMS_HUMAN	Parathymosin OS=Homo sapiens GN=PTMS PE=1 SV=2	NA	1.11
	sp|P50395|GDIB_HUMAN	Rab GDP dissociation inhibitor beta OS=Homo sapiens GN=GDI2 PE=1 SV=2	NA	1.07

			Log2 relative expression	
Selection	Down Regulation after 2ME		FT	ORBI
Both	sp|P60891|PRPS1_HUMAN	Ribose-phosphate pyrophosphokinase 1 OS=Homo sapiens GN=PRPS1 PE=1 SV	-1.08	-1.13
	sp|P11908-2|PRPS2_HUMAN	Isoform 2 of Ribose-phosphate pyrophosphokinase 2 OS=Homo sapiens GN=PRPS	-1.08	-1.13
	sp|Q14008-2|CKAP5_HUMAN	Isoform 2 of Cytoskeleton-associated protein 5 OS=Homo sapiens GN=CKAP5	-1.18	-1.12
	sp|Q9UHX1-2|PUF60_HUMAN	Isoform 2 of Poly(U)-binding-splicing factor PUF60 OS=Homo sapiens GN=PUF60	-1.53	-0.87
	sp|Q9UHD8-2|SEPT9_HUMAN	Isoform 2 of Septin-9 OS=Homo sapiens GN=SEPT9	-1.74	-1.70
FT	sp|P30622-2|CLIP1_HUMAN	Isoform 3 of CAP-Gly domain-containing linker protein 1 OS=Homo sapiens GN=CL	-1.00	-0.29
	sp|P22234-2|PUR6_HUMAN	Isoform 2 of Multifunctional protein ADE2 OS=Homo sapiens GN=PAICS	-1.24	-0.14
	sp|Q15019-2|SEPT2_HUMAN	Isoform 2 of Septin-2 OS=Homo sapiens GN=SEPT2	-1.33	NA
	sp|Q9UNM6-2|PSD13_HUMAN	Isoform 2 of 26S proteasome non-ATPase regulatory subunit 13 OS=Homo sapiens	-1.38	-0.40
	sp|O75044|FNBP2_HUMAN	SLIT-ROBO Rho GTPase-activating protein 2 OS=Homo sapiens GN=SRGAP2 PE	-3.09	NA
ORBI	sp|Q15833-2|STXB2_HUMAN	Isoform 2 of Syntaxin-binding protein 2 OS=Homo sapiens GN=STXBP2	NA	-1.01
	sp|Q01813|K6PP_HUMAN	6-phosphofructokinase type C OS=Homo sapiens GN=PFKP PE=1 SV=2	-0.44	-1.02
	sp|Q6P1J9|CDC73_HUMAN	Parafibromin OS=Homo sapiens GN=CDC73 PE=1 SV=1	-0.20	-1.04
	sp|P32322|P5CR1_HUMAN	Pyrroline-5-carboxylate reductase 1, mitochondrial OS=Homo sapiens GN=PYCR1	NA	-1.05
	sp|O95817|BAG3_HUMAN	BAG family molecular chaperone regulator 3 OS=Homo sapiens GN=BAG3 PE=1 S	NA	-1.08
	sp|O75122|CLAP2_HUMAN	CLIP-associating protein 2 OS=Homo sapiens GN=CLASP2 PE=1 SV=2	NA	-1.11
	sp|P26583|HMGB2_HUMAN	High mobility group protein B2 OS=Homo sapiens GN=HMGB2 PE=1 SV=2	NA	-1.12
	sp|Q01082|SPTB2_HUMAN	Spectrin beta chain, brain 1 OS=Homo sapiens GN=SPTBN1 PE=1 SV=2	NA	-1.14
	sp|Q13409-2|DC1I2_HUMAN	Isoform 2B of Cytoplasmic dynein 1 intermediate chain 2 OS=Homo sapiens GN=D	NA	-1.18
	sp|O95782-2|AP2A1_HUMAN	Isoform B of AP-2 complex subunit alpha-1 OS=Homo sapiens GN=AP2A1	NA	-1.23
	sp|Q96HY7|DHTK1_HUMAN	Probable 2-oxoglutarate dehydrogenase E1 component DHKTD1, mitochondrial OS	NA	-1.25
	sp|Q9NYF8-2|BCLF1_HUMAN	Isoform 2 of Bcl-2-associated transcription factor 1 OS=Homo sapiens GN=BCLAF	NA	-1.44
	sp|P49748|ACADV_HUMAN	Very long-chain specific acyl-CoA dehydrogenase, mitochondrial OS=Homo sapien	-0.41	-1.80
	sp|P41091|IF2G_HUMAN	Eukaryotic translation initiation factor 2 subunit 3 OS=Homo sapiens GN=EIF2S3 F	NA	-1.93
	sp|Q9BUQ8|DDX23_HUMAN	Probable ATP-dependent RNA helicase DDX23 OS=Homo sapiens GN=DDX23 PE=	-0.23	-2.46
B	sp|Q92541|RTF1_HUMAN	RNA polymerase-associated protein RTF1 homolog OS=Homo sapiens GN=RTF1 I	NA	-2.96
	sp|Q9UBD5-2|ORC3_HUMAN	Isoform 2 of Origin recognition complex subunit 3 OS=Homo sapiens GN=ORC3	NA	-6.33

4.4 Discussion

In this work, we investigated the possibility of utilizing diethylation labeling as isotopic quantification method. With optimal labeling condition, both the unlabeled peptides and partially labeled peptides were less than 1% level from representative BSA peptides compared to their fully diethylated forms. Such high labeling efficiency is similar to what we found for the dimethylation method, but diethylation has the advantage of higher mass shift and the same elution profile during RP-HPLC separation when ^{12}C and ^{13}C isotopic labeling reagents are used.

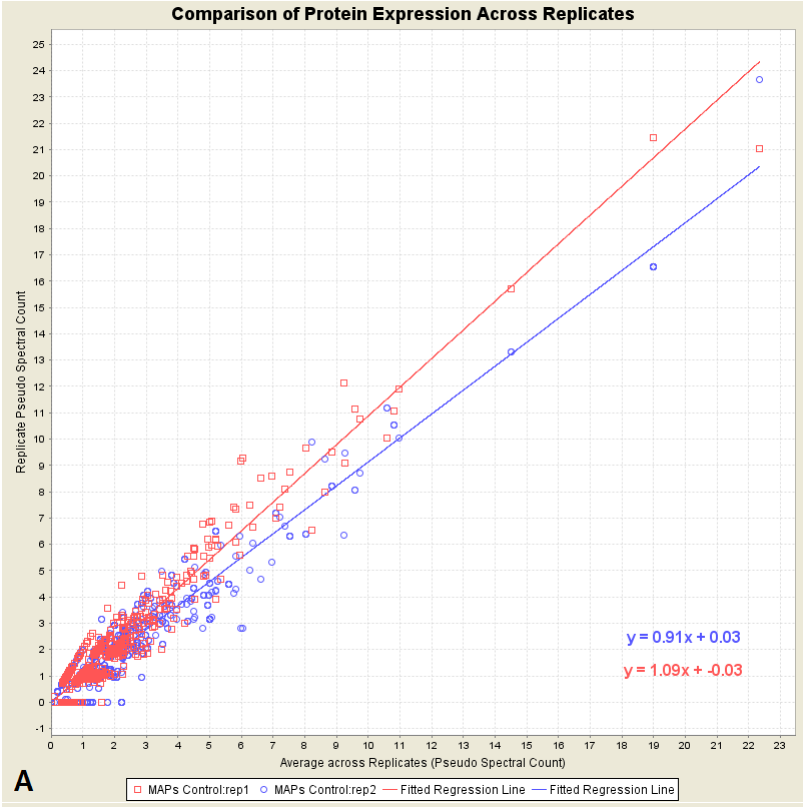
For labeled peptides after dimethylation or diethylation, we observed significant ionization increase for dimethylation and slight ionization decrease for diethylation. The difference might be explained in electron density change around the nitrogen atom of the free amine where the labeling occurs. For unlabeled peptides, this nitrogen atom would share an electron pair with hydrogen atoms. After being dimethylated, the nitrogen would share electron pair with sp^3 hybridized orbital from the carbon atom of the methyl group and possibly has a hyperconjugation effect with the C-H bond. This profound mechanism could potentially increase the electron density around nitrogen atom and make it easier to attract protons for ionization. However, for the diethylation case, the attached carbon atom shares electron pairs with the other carbon atom from the ethyl group and the hyperconjugation effect would be different. This complicated mechanism might not increase electron density of nitrogen atom. And thus would not increase ionization of diethylated peptides. Another phenomenon we noticed is the enhancement of y_{n-1} ions

during CID MSMS fragmentation after dimethylation or diethylation. This could be explained by the mobile photon theory that CID fragmentation happens at the location of moving photon along peptide backbone. However, after dimethylation or diethylation, the mobile photon might have a biased preference for the labeled N-terminal, resulting in increased fragmentation at this location.

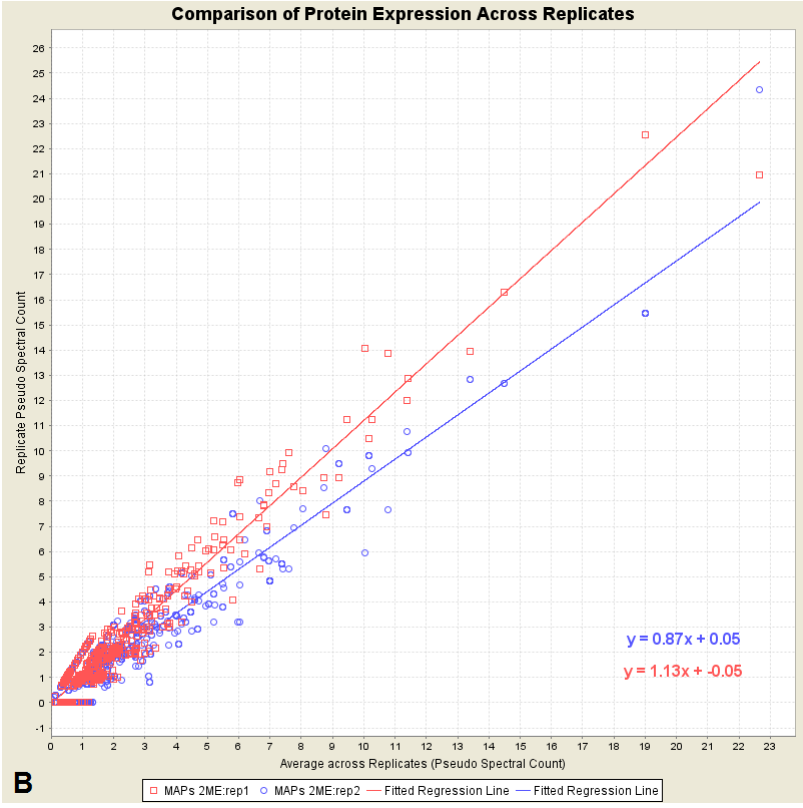
We applied this diethylation strategy for quantifying highly complex MAPs sample from MCF-7 cells with 2ME modulation. One thing to consider is that after diethylation with light and heavy reagents, the complexity of peptide mixture would double when mixing 1:1 ratio of isotopically diethylated peptides. Thus, high resolution separation of peptides is necessary prior to MS analysis. We utilized a previously developed 2D-RP-HPLC separation strategy to separate peptides at high-pH and low-pH conditions. From our result, after 2D-LC separation, the paired peptides (light and heavy) exhibits identical elution profile when using ^{12}C and ^{13}C acetaldehyde as labeling reagents. Another thing associated with increased sample complexity after diethylation is the potential labeling sites other than N-terminal and lysine of peptides. We also considered NH_2 groups coming from arginine (R), asparagine (N) and glutamine (Q), although they are usually neglected for both dimethylation and diethylation reactions. To illustrate, we have 45282 SPMs (including redundant peptide IDs) after combining all analysis data and applying the protein identification criteria described before. From these SPMs, 146 were diethylated at arginine (R), 502 at asparagine (N), and 381 at glutamine (Q). Total of these minor diethylation sites constitute a number of 1029 SPMs, which is 2.27% of total

redundant peptides IDs. Although these peptides would have tiny chance to be quantified due to low abundance, their existence is still worthy of being kept in mind.

For quantification experiments of biological samples, an important measure of the experiment efficacy is the reproducibility. We conducted two biological replicates and analyzed each replicate on both FT-ICR and Orbitrap Velos instruments. Figure 4.13 compares the experimental reproducibility between quantification results of these two biological replicates. For both the FT and Orbitrap results, quantified protein expression was converted to pseudo spectral count by ProteoIQ. Then the protein expression value from both biological replicates is compared to the average value of these two replicates. Theoretically, for identical replicates, the slope of the plot would be 1. From Figure 4.13A and 4.13B, the quantified MAPs protein expression from both MCF-7 control cells and 2ME-treated cells are highly reproducible between the two biological replicates conducted with slopes ranging from 0.87 to 1.13. Interestingly, from Figure 4.13C and 4.13D, the Orbitrap data for same samples showed slightly higher reproducibility with slopes ranging from 0.93 to 1.07. This might be associated with the better sensitivity of Orbitrap instrument, which could result in higher ionization intensity and improved quantitation.


Comparison of Protein Expression Across Replicates
Replicate Pseudo Spectral Count
Average across Replicates (Pseudo Spectral Count)
y = 0.91x + 0.03
y = 1.09x + -0.03
MAPs Control:rep1
MAPs Control:rep2
Fitted Regression Line
Fitted Regression Line

A


Comparison of Protein Expression Across Replicates
Replicate Pseudo Spectral Count
Average across Replicates (Pseudo Spectral Count)
y = 0.87x + 0.05
y = 1.13x + -0.05
MAPs 2ME:rep1
MAPs 2ME:rep2
Fitted Regression Line
Fitted Regression Line

B

Figure 4.13. ProteoIQ generated experimental reproducibility across two biological replicates. (A) Protein expression comparison of MAPs control from FT data of two replicates. (B) Protein expression comparison of MAPs 2ME treated from FT data of two replicates. (C)

Protein expression comparison of MAPs control from Orbitrap data of two replicates. (D) Protein expression comparison of MAPs 2ME treated from Orbitrap data of two replicates.

For protein candidates that were up-regulated or down-regulated as potential biomarkers after 2ME treatment, about 60% of them were only quantified by the Orbitrap data. This may be due to increased sensitivity and faster duty-cycle of Orbitrap instrument compared to FT-ICR instrument. Of these candidates, Pyrothymosin-alpha (PMTA) had been reproducibly recognized for up-regulation from both FT and Orbitrap data. After literature search, it had been reported to be associated with death regulatory pathway of cancer cells [23-25]. In addition, 2ME treatment of cancer cells had been shown inducement of death receptor pathway [26, 27]. So the observed up-regulation of PMTA might be associated with the inducement of apoptosis by 2ME on MCF-7 cells. Another candidate cytoskeleton-associated protein 5 (CKAP5) had been reproducibly identified as down-regulation from both FT and Orbitrap data. CKAP5 is a microtubule-associated protein typically over-expressed in tumor cells and it is involved in microtubule dynamics required for mitotic spindle formation [28-30]. In addition, 2ME tumor-inhibiting effects had been reported as results of binding to microtubule (MT) and affecting MT dynamics [17, 22]. So the down-regulation of CKAP5 might be contributed to 2ME effect on MCF-7 cells with disruption of microtubule dynamics. From these examples, the quantified protein expression changes from our diethylation strategy are consistent with literature evidence. Other protein candidates in the differentially expressed protein list from our quantitation result may also be potential biomarkers, but

would require further literature search or other experimental evidence to support their validity.

In this work, while we focused on the capability of diethylation strategy, the biological significance of MAPs as potential biomarkers for 2ME treatment on tumor cells is also important. For selection of protein candidates, we chose 2-fold expression change as the threshold. However, this limits the sensitivity of our result since biologically relevant changes induced by 2ME modulation cannot be detected if it's below the threshold. An alternative approach would be analyzing all quantified protein expression changes using statistical test tools and report these protein candidates with significant expression changes in a statistical perspective. Common statistical tests are listed in Table 4.9. Because technical replicates would inherit differences from parent biological replicates,

Table 4.9. Common statistical tests and their applications in quantitative proteomics [31].

Test	Requirements	Statistical power	Application
Tests for experiments with replicates			
t-test	Replications, $n>3$ Data normally distributed	+++	All quantitation methods
LPE-test	Replications, $n>1$	++	All quantitation methods 2–3 replicates Strong changes
Tests for experiments without replicates			
G-test	(Very) large number of peptide spectra	+	Spectrum counting
Fischer's exact test	(Very) large number of peptide spectra	+	Spectrum counting
AC-test	(Very) large number of peptide spectra	+	Spectrum counting

it's not credible for performing an appropriate Student's t-Test [32]. Considering the fact that we conducted two biological replicates, the local-pooled-error test (LPE) [33-35]

could be useful for determining protein candidates with statistically significant expression changes after the 2ME treatment on MCF-7 cells.

4.5 Conclusion

In this work, we investigated the potential of an alternative isotopic labeling strategy by reductive diethylation of the primary amines from peptides for quantitative proteomics. The advantage of diethylation due to larger mass shift from ethyl groups allows the possibility of only using ^{12}C and ^{13}C reagents for labeling, thus eliminating the challenge of retention shift for high resolution quantification profiling. From our quantification result for the microtubule-associated proteome, the light and heavy labeled peptide pair showed no retention shift even after two-dimensional reversed-phase HPLC separation. In addition, this method has high labeling efficiency and adequate ionization for labeled peptides, thus making it applicable for global proteomic quantitation. At last, we identified a few protein candidates that are potential biomarkers for 2ME drug treatment by this method. Some of these protein candidates have literature implications for the up-regulation or down-regulation that we observed, thereby supporting the validity of this quantification strategy.

References

1. Ong, S.E. and M. Mann, *Mass spectrometry-based proteomics turns quantitative.* Nature Chemical Biology, 2005. **1**(5): p. 252-262.
2. Julka, S. and F. Regnier, *Quantification in proteomics through stable isotope coding: A review.* Journal of Proteome Research, 2004. **3**(3): p. 350-363.
3. Mann, M., *Functional and quantitative proteomics using SILAC.* Nature Reviews Molecular Cell Biology, 2006. **7**(12): p. 952-958.
4. Hsu, J.L., et al., *Stable-isotope dimethyl labeling for quantitative proteomics.* Analytical Chemistry, 2003. **75**(24): p. 6843-6852.
5. Boersema, P.J., et al., *Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics.* Nature Protocols, 2009. **4**(4): p. 484-494.
6. Boersema, P.J., et al., *Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates.* Proteomics, 2008. **8**(22): p. 4624-4632.
7. Guo, K., C.J. Ji, and L. Li, *Stable-isotope dimethylation labeling combined with LC-ESI MS for quantification of amine-containing metabolites in biological samples.* Analytical Chemistry, 2007. **79**(22): p. 8631-8638.
8. Hsu, J.L., S.Y. Huang, and S.H. Chen, *Dimethyl multiplexed labeling combined with microcolumn separation and MS analysis for time course study in proteomics.* Electrophoresis, 2006. **27**(18): p. 3652-3660.
9. Ji, C.J., et al., *Identification and quantification of differentially expressed proteins in E-cadherin deficient SCC9 cells and SCC9 transfectants expressing E-cadherin by dimethyl isotope labeling, LC-MALDI MS and MS/MS.* Journal of Proteome Research, 2005. **4**(4): p. 1419-1426.
10. Ji, C.J. and L. Li, *Quantitative proteome analysis using differential stable isotopic labeling and microbore LC-MALDI MS and MS/MS.* Journal of Proteome Research, 2005. **4**(3): p. 734-742.
11. Ji, C.J., N. Guo, and L. Li, *Differential dimethyl labeling of N-termini of peptides after guanidination for proteome analysis.* Journal of Proteome Research, 2005. **4**(6): p. 2099-2108.
12. Hsu, J.L., et al., *Beyond quantitative proteomics: Signal enhancement of the a(1) ion as a mass tag for peptide sequencing using dimethyl labeling.* Journal of Proteome Research, 2005. **4**(1): p. 101-108.
13. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.* Molecular & Cellular Proteomics, 2004. **3**(12): p. 1154-1169.
14. Barrios-Llerena, M.E., et al., *The use of a novel quantitation strategy based on Reductive Isotopic Di-Ethylation (RIDE) to evaluate the effect of glufosinate on*

the unicellular algae Ostreococcus tauri. Journal of Proteomics, 2011. **74**(12): p. 2798-2809.
15. Bhat, K.M.R. and V. Setaluri, *Microtubule-associated proteins as targets in cancer chemotherapy.* Clinical Cancer Research, 2007. **13**(10): p. 2849-2854.
16. Shang, W., I. Konidari, and D.W. Schomberg, *2-Methoxyestradiol, an endogenous estradiol metabolite, differentially inhibits granulosa and endothelial cell mitosis: a potential follicular antiangiogenic regulator.* Biol Reprod, 2001. **65**(2): p. 622-7.
17. Schumacher, G. and P. Neuhaus, *The physiological estrogen metabolite 2-methoxyestradiol reduces tumor growth and induces apoptosis in human solid tumors.* J Cancer Res Clin Oncol, 2001. **127**(7): p. 405-10.
18. Mueck, A.O. and H. Seeger, *2-Methoxyestradiol-Biology and mechanism of action.* Steroids, 2010.
19. Kamath, K., et al., *2-methoxyestradiol suppresses microtubule dynamics and arrests mitosis without depolymerizing microtubules.* Molecular Cancer Therapeutics, 2006. **5**(9): p. 2225-2233.
20. Montgomery, R.B., et al., *Estrogen effects on tubulin expression and taxane mediated cytotoxicity in prostate cancer cells.* Prostate, 2005. **65**(2): p. 141-150.
21. Gokmen-Polar, Y., et al., *beta-Tubulin mutations are associated with resistance to 2-methoxyestradiol in MDA-MB-435 cancer cells.* Cancer Research, 2005. **65**(20): p. 9406-9414.
22. Mooberry, S.L., *Mechanism of action of 2-methoxyestradiol: new developments.* Drug Resist Updat, 2003. **6**(6): p. 355-61.
23. Malicet, C., et al., *Regulation of apoptosis by the p8/prothymosin alpha complex.* Proc Natl Acad Sci U S A, 2006. **103**(8): p. 2671-6.
24. Jiang, X.J., et al., *Distinctive roles of PHAP proteins and prothymosin-alpha in a death regulatory pathway.* Science, 2003. **299**(5604): p. 223-226.
25. Martini, P.G.V., et al., *Prothymosin alpha selectively enhances estrogen receptor transcriptional activity by interacting with a repressor of estrogen receptor activity.* Molecular and Cellular Biology, 2000. **20**(17): p. 6224-6232.
26. Van Zijl, C., et al., *In vitro effects of 2-methoxyestradiol on MCF-12A and MCF-7 cell growth, morphology and mitotic spindle formation.* Cell Biochemistry and Function, 2008. **26**(5): p. 632-642.
27. LaVallee, T.M., et al., *2-methoxyestradiol up-regulates death receptor 5 and induces apoptosis through activation of the extrinsic pathway.* Cancer Research, 2003. **63**(2): p. 468-475.
28. Royle, S.J., *The role of clathrin in mitotic spindle organisation.* Journal of Cell Science, 2012. **125**(1): p. 19-28.
29. Cappell, K.M., et al., *Symplekin Specifies Mitotic Fidelity by Supporting Microtubule Dynamics.* Molecular and Cellular Biology, 2010. **30**(21): p. 5135-5144.
30. Francone, V.P., et al., *The microtubule-associated protein tumor overexpressed gene/cytoskeleton-associated protein 5 is necessary for myelin basic protein*

expression in oligodendrocytes. Journal of Neuroscience, 2007. **27**(29): p. 7654-7662.

31. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: a critical review.* Analytical and Bioanalytical Chemistry, 2007. **389**(4): p. 1017-1031.
32. Karp, N.A. and K.S. Lilley, *Design and analysis issues in quantitative proteomics studies.* Proteomics, 2007. **7 Suppl 1**: p. 42-50.
33. Jain, N., et al., *Local-pooled-error test for identifying differentially expressed genes with a small number of replicated microarrays.* Bioinformatics, 2003. **19**(15): p. 1945-1951.
34. Jain, N., et al., *Rank-invariant resampling based estimation of false discovery rate for analysis of small sample microarray data.* BMC Bioinformatics, 2005. **6**.
35. Zhang, B., et al., *Detecting differential and correlated protein expression in label-free shotgun proteomics.* Journal of Proteome Research, 2006. **5**(11): p. 2909-2918.