2016

Proteomics method development and application for interaction of influenza virus and cells

Hanzhi Wu
Hong Kong Baptist University

Follow this and additional works at: http://repository.hkbu.edu.hk/etd_oa

Recommended Citation
http://repository.hkbu.edu.hk/etd_oa/162

This Thesis is brought to you for free and open access by the Electronic Theses and Dissertations at HKBU Institutional Repository. It has been accepted for inclusion in Open Access Theses and Dissertations by an authorized administrator of HKBU Institutional Repository. For more information, please contact repository@hkbu.edu.hk.
Proteomics Method Development and Application for Interaction of
Influenza Virus and Cells

WU Hanzhi

A thesis submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Principal Supervisor: Prof. CAI Zongwei
Hong Kong Baptist University
January 2015
Declaration

I hereby declare that this thesis represents my own work which has been done after registration for the degree of PhD at Hong Kong Baptist University, and has not been previously included in a thesis or dissertation submitted to this or any other institution for a degree, diploma or other qualifications.

Signature:

Date: January 2015
Abstract

Influenza virus H1N1 is a huge threat on human health. Influenza occurs with seasonal variations and reaches peak prevalence in winter, with many people killed worldwide every year. In the research of interaction between influenza virus and cells, four major parts were in the range of our consideration, namely the proteins of virus, the proteome of host cell, the method of proteomic and the potencial medicine related with those significant proteins.

Hemagglutinin (HA), as an envelope protein, plays an important role in influenza A virus. It was found that HA has a series of isoforms in two dimensional gels in this study. For the investigation of HA, firstly, virus was purified by sucrose density-gradient centrifugation, followed by the separation of virus proteins through electrophoresis method, and then these proteins were digested by different enzymes and analyzed through MALDI-TOF MS and ESI-Q-TOF MS. Database searching was used for identification of sequences. The results of the virus samples digested by different enzymes were compared, and the isoforms of HA were proved to be related with the glycan and their glycosylation sites.

A novel strategy of stable-isotope N-phosphorylation labeling was developed for peptide de novo sequencing and protein quantification based on organic phosphorus chemistry. Different from other stable-isotope labeling reagents that needed to be activated in advance for peptide coupling, N-phosphorylation labeling reagents were activated in situ to form labeling intermediates with high activity and selectivity targeting on N-terminus and -amino group of lysine under various reaction conditions. The obtained results showed excellent correlation of the measured ratios to theoretical ratios with errors that ranging from 0.5 to 6.7 %
and relative standard deviation of less than 10.6 %, indicating the reproducibility and preciseness of the developed method. The method development based on organic phosphorus chemistry offered a new approach for quantitative proteomics by using novel stable-isotope labeling reagents.

A method combining hydrazide chemistry, stable isotope labeling and mass spectrometry analysis was developed and applied to study glycoproteins of H1N1 (A/Purto Rico/8/1934) infected cell line (A549). The result showed that some glycoproteins were significant in influenza virus infected cells. In these glycoproteins, RPC1_HUMAN, RHG25_HUMAN, RPTOR_HUMAN, ARHGC_HUMAN, ROCK1_HUMAN, DOCK3_HUMAN were down-regulated. Protein named TITIN_HUMAN, DESP_HUMAN, PTN13_HUMAN were up-regulated.

High dose of N-acetylcysteine (NAC) was recently reported for a therapy of H1N1 influenza pneumonia. NAC was used as a small-molecule organic probe to investigate the protein expression of human lung carcinoma cell line (A549) infected by influenza virus H1N1. The obtained results showed that NAC kept cells away from apoptosis. Virus-infected cells were arrested in G0/G1 phase. The lowest cell population of G0/G1 phase was detected when the cells were treated by 10 mM NAC for one day. Software analysis showed that 4 proteins had close relationship. The results indicated that NAC as a small-molecule probe might effect the proteins expression of A549 cells infected by the H1N1 virus.
Acknowledgement

First and foremost, I would like to express my sincere gratitude to my supervisor, Prof. Zongwei CAI for his invaluable advice and support both in lab and life. Prof. Cai provides me a lot of opportunity to explore and develop my potentials in scientific inquiry. Without his inspiration and supervision, I could not have completed my PhD studies. His guidance, inspiration, confidence and continued encouragement are greatly appreciated.

I would also like to thank my co-supervisor Dr. KWONG, Daniel W. J. for his helpful discussion. Many thanks must go to Mr. David T.W. CHIK and Ms Silvia T. MO for providing MS technical support. My sincere thanks must go to all of my research group members, Dr. Shuhai LIN, Dr. Yongquan LAI, Mr. Yang SHEN, Mr. Zhi TANG, Mr. Shangfu LI, Ms Manwen ZHANG, Ms Li XIAN, Mr. Hemi LUAN, Ms Xiaona LI and so on, for their support and helpful discussion.

I would also like to give my wordless and deepest gratitude and love to my wife, daughter and parents for their love, support and encouragement.
Table of Contents

Declaration........................................................................................................... I
Abstract.............................................................................................................II
Acknowledgements...............................................................................................IV
Table of Contents.....................................................................................................V
List of Tables.........................................................................................................X
List of Figures.......................................................................................................XII
List of Abbreviations.............................................................................................XVII

Chapter 1 Introduction.............................................................................................1
   1.1 Overview of influenza A virus.........................................................................1
   1.2 Proteomics in influenza virus research.........................................................4
   1.3 Potential medicine study in influenza virus infection.................................8

Chapter 2 Identification of hemagglutinin isoforms difference of influenza A virus H1N1...........................................................................................................10
   2.1 Introduction......................................................................................................10
   2.2 Materials and methods.....................................................................................11
      2.2.1 Virus cultivation.......................................................................................11
      2.2.2 Isolation of virus by ultracentrifugation.................................................12
      2.2.3 SEM detection, PNGase F digest and PAGE electrophoresis ...............12
      2.2.4 In-gel digestion.......................................................................................13
      2.2.5 MALDI-TOF MS analysis......................................................................14
      2.2.6 Nanoelectrospray-ESI-MS/MS analysis..............................................15
   2.3 Result.............................................................................................................15
      2.3.1 Purification of influenza virus H1N1 ....................................................16
Chapter 3 Stable-isotope N-phosphorylation labeling for peptide \textit{de novo} sequencing and protein quantification based on organic phosphorus chemistry

3.1 Introduction

3.2 Materials and methods

3.2.1 Materials and reagents

3.2.2 Synthesis of deuterium-labeled dimethyl phosphite

3.2.3 Trypsin digestion

3.2.4 N-phosphorylation of standard peptide and tryptic peptides

3.2.5 Mass spectrometry analysis

3.2.6 Nano LC-chip/Q-TOF MS analysis

3.2.7 Protein identification and quantification

3.3 Results and discussion

3.3.1 Selection of phosphorus chemistry

3.3.2 Stable isotopic N-phosphorylation labeling of standard peptide and proteins

3.3.3 Peptide \textit{de novo} sequencing

3.3.4 Quantification of standard peptide and protein

3.3.5 Quantification of protein mixtures by nano LC-chip/TOF MS

3.3.6 Validation of deuterium atom effect of labeled peptides
Chapter 4 Identification and quantification of N-linked glycoproteins of influenza virus infected cells using hydrazide chemistry, stable isotope labeling and mass spectrometry

4.1 Introduction

4.2 Materials and methods

4.2.1 Materials

4.2.2 Standard proteins preparation

4.2.3 Virus and cell line culture

4.2.4 Separation of N-glycosylation protein

4.2.5 N-phosphorylation of tryptic peptides

4.2.6 Mass spectrometry analysis

4.2.7 Nano LC-Chip Q-TOF MS

4.2.8 Protein identification, quantification and bio-information analysis

4.3 Result and discussion

4.3.1 Standard protein testing

4.3.2 Glycoprotein separation and quantification

4.3.3 Relationship with those different proteins

4.4 Chapter summary

Chapter 5 Proteomics study of N-acetylcysteine response in H1N1-infected cells by using mass spectrometry

5.1 Introduction

5.2 Materials and methods

5.2.1 Virus and cell line culture
List of Tables

Table 2.1 Matched sequence of spot 1………………………………… ..…… 24
Table 2.2 Matched sequence of spot 2………………………………… ..…… 24
Table 2.3 Matched sequence of spot 3………………………………… ..…… 24
Table 2.4 Matched sequence of spot 4……………………………… .… .…… 25
Table 2.5 Matched sequence of spot 5………………………………… ..…… 25
Table 2.6 Q-TOF MS database searching result of spot 1………………25
Table 2.7 Q-TOF MS database searching result of spot 2…………… ....…… 25
Table 3.1 MALDI MS data for tryptic peptides of lysozyme. ………….44
Table 3.2 Quantification of the protein mixture under various ratios by nano
     LC-chip/TOF MS. .........................................................61
Table 4.1 Identification of the significantly regulated glycoproteins by database
     searching.......................................................................79
Table 5.1 Confirmation of proteins differentially expressed in A549 cells from
     MALDI-TOF MS and Q-TOF MS/MS analyses. ..................104-105
Table A1.1 Total phenolic (TP) and total monomeric anthocyanin (TMA) content
     of elderberry juice samples of different genotypes harvested from the
     same growing location and year (Jefferson City, MO, 2011)………118
Table A1.2 Total phenolic (TP) and total monomeric anthocyanin (TMA) content
     of elderberry juice of the ‘Bob Gordon’ genotype at different
     locations and during different seasons.................................119
Table A1.3 Parent-daughter ion masses, retention times and concentrations of
     individual anthocyanins and anthocyanidins present in a sample of
     ‘Wyldewood’ elderberry harvested in 2011. .........................126
Table A2.1  Applied Nano-LC gradient parameters for the elution of the peptides. .................................................................133
Table A2.2  MS/MS results of elderberry juice analyzed by MALDI-TOF/TOF MS ..................................................................................135
Table A2.3  Elderberry juice peptide sequences with ALC scores greater than 60 for Group A.........................................................142
Table A2.4  Elderberry juice peptide sequences with ALC scores greater than 60 for Group B.........................................................143-146
List of Figures

Figure 1.1 A schematic diagram of the structure of the influenza A virus. ................................................................. 1

Figure 1.2 Analysis of virion composition by mass spectrometry. Typical approaches for analyzing virion protein composition by MALDI-TOF MS and LC-MS/MS are shown................................. 7

Figure 2.1 3D structure of HA from PDB.................................................. 11

Figure 2.2 Influenza virus purified by gradient sucrose centrifugation.............. 16

Figure 2.3 Virus particles detected by SEM.............................................. 17

Figure 2.4 SDS-PAGE separation of influenza virus H1N1 (PR8) proteins..... 17

Figure 2.5 Two dimensional electrophoresis separation of H1N1 (PR8) proteins............................................................. 18

Figure 2.6 Two dimensional electrophoresis separation of H1N1 (PR8) proteins which digested by PNGase F .............................................. 19

Figure 2.7 Peptide mass spectrogram (MALDI-TOF MS) of HA digested by Trypsin................................................................. 20

Figure 2.8 Peptide mass spectrogram (MALDI-TOF MS) of HA digested by Asp-N................................................................. 21

Figure 2.9 Peptide mass spectrogram (MALDI-TOF MS) of HA digested by Glu-C................................................................. 22

Figure 2.10 HA sequence of H1N1 (PR8). The detected sequences were marked in red color. The signal sequence was marked in blue........... 23

Figure 3.1 Chemical modification reactions involving the incorporation of stable-isotope labeling for quantitative proteomics. ............... 30

Figure 3.2 Chemical structures of organophosphorus reagents for stable-isotope
labeling of tryptic peptides. ..................................................32

Figure 3.3 Brief workflow of experiment ........................................33

Figure 3.4 Phosphorus chemistry for stable-isotope N-phosphorylation labeling..........................................................35

Figure 3.5 Stable-isotope N-phosphorylation labeling (SIPL) strategy for protein quantification based on organic phosphorus chemistry.........................................................40

Figure 3.6 Nano ESI-MS spectra of the standard Fibrinopeptide ADSGEGDFLAEGGGVR. (A) native peptide; (B) N-terminal H-DMP labeling; (C) N-terminal stable-isotope D-DMP labeling. .........................................................42

Figure 3.7 MALDI-TOF MS spectra of the tryptic peptides of lysozyme. (A) native; (B) H-DMP labeling; (C) stable-isotope D-DMP labeling. (*, the number of labeled phosphoryl group).................................43

Figure 3.8 Identification of digest peptides KVFGGR, HGLDNSR, and CELAAAMKR derived from lysozyme by nano ESI-MS/MS. (A) H-DMP-labeled KVFGGR (1+, \( m/z \) 822.3691); (B) H-DMP-labeled HGLDNSR (2+, \( m/z \) 545.7138); (C) H-DMP-labeled CELAAAMKR (2+, \( m/z \) 633.2594). ........................................45-46

Figure 3.9 MALDI-TOF MS analysis of the tryptic peptides derived from BSA. (A) MALDI-TOF MS spectrum of native BSA digest; (B) sequence coverage searched using Mascot; (C) MALDI-TOF MS spectrum of H-DMP-labeled BSA digest; (D) sequence coverage searched using Mascot after H-DMP-labeling. ........................................47

Figure 3.10 Identification of the deaminated peptide CKGTDVQAWIR (2+, \( m/z \)
712.3379) derived from lysozyme by nano ESI-MS/MS.........48

Figure 3.11 Nano ESI-MS/MS of tryptic peptide CKGTDVQAWIR derived from lysozyme. (A) MS/MS spectrum of native peptide (2+, m/z 667.3344); (B) MS/MS spectrum of H-DMP-labeled peptide (2+, m/z 775.3282). ..........................................................................................51

Figure 3.12 (A) Extract ion chromatogram for a linearity assay of H-DMP- and stable-isotope D-DMP-labeled Fibrinopeptide A standard peptide ADSGEGFLAEGGGVR (2+, m/z 822.82 and m/z 825.83) under various ratios (1:0.5, 1:1, 1:2, 1:3, 1:4, 1:5, 1:10, and 1:20). (B) The linearity plot of the H-DMP- and D-DMP-labeled peptide ADSGEGFLAEGGGVR........................................................................54

Figure 3.13 Relative quantification of standard protein lysozyme under various ratios using LC-MS..................................................565

Figure 3.14 Identification and quantification of protein mixes by nano LC-chip/TOF MS and MS/MS. The ratios of all isotopic pairs were calculated from the extracted peak area in the LC-MS chromatograms and averaged. ........................................58-60

Figure 3.15 Comparison of extracted ion chromatograms of peptide pairs labeled with H-DMP or D-DMP...........................................64

Figure 4.1 Work flow of glycoprotein separation and quantification.........73

Figure 4.2 MALDI-TOF MS results from analysis of individual and mixture standard proteins. A. digested peptides of OVA. B. peptides of mixture standard proteins (OVA and BSA) after separation and digestion. C. digested peptides of BSA............................................78
Figure 4.3  Relationship of the 9 differential proteins by STRING software....83

Figure 5.1  Histograms of the A/Puerto Rico/8/1934 (H1N1)-infected A549 cells with the treatments using different NAC concentrations.............92

Figure 5.2  Histogram of difference A549 cell life cycle phase in NAC-treated cells with different concentrations. ...........................................93

Figure 5.3  Two dimensional electrophoresis gel of NAC-treated A549 cell proteins.................................................................95

Figure 5.4  Protein spots of different A549 cell cycle phase in NAC-treated cells with different concentrations. Control, A, B and C refers the differential protein spots in 0, 1 mM, 10 mM and 100 mM NAC-treated A549 cells, respectively. ............................................96

Figure 5.5  Optical density of the differential protein spots. ....................97

Figure 5.6  Relationship of the 12 differential proteins (A) and the 8 proteins with optical density alteration rule same as that for G0/G1 phase of cells by STRING software (B). The names of proteins are listed in Table 5.1. .................................................................103

Figure A1.1 Experimental scheme for the purification and isolation of elderberry juice peptides.........................................................113

Figure A1.2 Structures of the 6 common anthocyanins at pH of 1.0 (I), pH of 4.5 (II), and anthocyanindins (III)..............................................117

Figure A1.3 UPLC-MS/MS calibration curve using Cyanidin-3-O-glucoside standard.................................................................121

Figure A1.4 Parent ion scans of 6 anthocyanidins by ESI-MS. A: Pelargonidin (m/z 271), B: Cyanidin (m/z 287), C: Malvidin (m/z 331), D: Peonidin (m/z 301), E: Delphinidin (m/z 303), F: Petunidin (m/z 323)
Figure A1.5 Individual anthocyanin chromatograms observed in elderberry juice utilizing multiple reaction monitoring.

Figure A2.1 Experimental scheme for the purification and isolation of elderberry juice peptides.

Figure A2.2 Sample clarified by centrifugation, MALDI-TOF MS (Voyager).

Figure A2.3 MALDI-TOF MS result for mass range of m/z 809-3187.

Figure A2.4 MALDI-TOF MS result for mass range of m/z 2960-5100.

Figure A2.5 MALDI-TOF MS result from the analysis using matrix CHCA.

Figure A2.6 (A) De novo ALC scores of Group A, MS/MS scans: 2530. Peptides after filter: 1789. (B) De novo ALC scores of Group B, MS/MS Scans: 1833. Peptides after filter: 1233.
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CID</td>
<td>Collision induced dissociation</td>
</tr>
<tr>
<td>EIC</td>
<td>Extracted ion chromatogram</td>
</tr>
<tr>
<td>ESI</td>
<td>Electrospray ionization</td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>LC–MS</td>
<td>Liquid chromatography-mass spectrometry</td>
</tr>
<tr>
<td>Q-TOF MS</td>
<td>Quadrupole time-of-flight mass spectrometry</td>
</tr>
<tr>
<td>MALDI-TOF MS</td>
<td>Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry</td>
</tr>
<tr>
<td>RSD</td>
<td>Relative standard deviation</td>
</tr>
<tr>
<td>SPE</td>
<td>Solid-phase extraction</td>
</tr>
<tr>
<td>SRM</td>
<td>Selected reaction monitoring</td>
</tr>
<tr>
<td>TIC</td>
<td>Total ion chromatogram</td>
</tr>
<tr>
<td>UPLC</td>
<td>Ultra performance liquid chromatography</td>
</tr>
<tr>
<td>NAC</td>
<td>N-acetylcysteine</td>
</tr>
<tr>
<td>GAE</td>
<td>Gallic acid equivalents</td>
</tr>
<tr>
<td>C3GE</td>
<td>Cyanidin 3-O-glucoside equivalents</td>
</tr>
<tr>
<td>H-DMP</td>
<td>Dimethyl phosphite</td>
</tr>
<tr>
<td>D-DMP</td>
<td>Deuterium-labeled dimethyl phosphite</td>
</tr>
<tr>
<td>DEP</td>
<td>Diethyl phosphite</td>
</tr>
<tr>
<td>DPP</td>
<td>Dipropyl phosphite</td>
</tr>
<tr>
<td>DIPP</td>
<td>Diisopropyl phosphite</td>
</tr>
<tr>
<td>EPP</td>
<td>Ethoxylphenyl phosphinate</td>
</tr>
<tr>
<td>PR8</td>
<td>A/Puerto Rico/8/1934</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Overview of influenza A virus

Influenza A virus, a member of the family Ortho-myxoviridae, is a small enveloped virus with a genome consisting of 8 segments of negative-sense single-stranded RNA. It encodes for 10 to 11 proteins depending on the strain. The segmented genome and highly error-prone viral replication lead to enormous genetic plasticity, mediated by nucleotide or genome segment exchange, termed genetic drift or genetic shift, respectively. Genomic changes control the differences in virulence and host range observed among influenza virus isolated. Influenza viruses are serologically categorized by 2 surface proteins: hemagglutinin (HA), of which there are currently 16 types (H1 to H16), and neuraminidase (NA), of which there are currently 9 types (N1 to N9). Virtually every possible H-N combination has been found in water fowl, the generally accepted reservoir, but only a few H-N types have circulated in humans, namely H1N1 (Figure 1.1), H2N2 and H3N2 \[1\].

![Figure 1.1 A schematic diagram of the influenza Avirus \[2\].](image-url)
Among the combinatorial diversity of 144 possible A/HN subtypes, relatively few subtypes have been identified as causes of human disease. Four pandemic outbreaks in the last century, one catastrophic, appear to be introduced the subsequently prevalent seasonal human influenza virus subtypes A/H1N1 (Spanish flu, 1918), A/H2N2 (Asian flu, 1957), A/H3N2 (Hong Kong flu, 1968), and A/H1N1 again (Swine flu, 1976; Russian flu, 1977) [3].

As with the few common subtypes of human type A influenza viruses, few subtypes of type A influenza viruses are found to associate with most influenza infections of swine, horses or dogs. In contrast, wildfowl species are natural hosts and a global reservoir for the majority of possible influenza A/HN subtypes. Many of these variant strains appear to be associated with endemic infections, often asymptomatic in avian hosts. Incidental infections of humans by avian influenza viruses have been documented for avian influenza subtypes A/H5N1, A/H7N2, A/H7N3, A/H7N7, A/H9N2, A/H10N7 and A/H11N9. Recent outbreaks of “bird flu” may foreshadow an eventual pandemic outbreak, in the emergence of strains and variants with enhanced pathogenicity, virulence and transmissibility in human hosts. Examples of such outbreaks include A/H5N1 Hong Kong, 1997; H9N2 Hong Kong, 1999; A/H7N7 Netherlands, 2003; A/H5N1 Southeast Asia, 2004. Some avian A/H5 and A/H7 strains of influenza virus are recognized as highly pathogenic (HP) in domestic poultry and concerns arise that this phenotype may carry over to infections of humans. Since 1997, human infections associated with the Eurasian-African lineage of A/H5N1 HP avian influenza virus have been associated with 467 documented cases in 15 countries with high mortality (282 deaths). Fortunately, infectious transmission of such avian influenza virus strains between humans continues to be limited. However, history suggests that further
evolution of these or other types of A influenza strains could emerge as a next pandemic strain. Similarly, variant type A influenza virus strains have emerged from time to time, imposing serious costs and burdens upon poultry and livestock production. As for the natural history and the molecular biology of influenza viruses reflect such viral genome diversity, there is a critical need for rapid, sensitive, specific, and informative assays to detect and characterize the subtype of influenza virus. Currently, many benchmark standard methods are available, such as employing propagation of virus in cell culture or in embryonating chicken eggs, with assays using panels of specific serological reagents, or reverse transcriptase polymerase chain reaction (RT-PCR)-based assays, using panels of short oligonucleotide primers and probes. However, these methods are either slow or time consuming, or expensive. As prevailing strains of avian influenza continue to evolve and diverge, diagnostic assays which are based on specific recognition of short signature sequences or peptide biomarker loci will have a probability to fail to determine the virus, through false-positive and/or false-negative results, which affect critical decision making adversely.

In early April 2009, an influenza A (H1N1) virus emerged among humans in California and Mexico, quickly spreaded worldwide through human-to-human transmission, and generated the first influenza pandemic of the 21st century. It has been marked by a late season pandemic-scale emergence of a novel A/H1N1 outbreak strain, raising immediate concerns for public health. This virus was found to be antigenically unrelated to human seasonal influenza viruses but genetically related to viruses known to circulate in pigs. In view of its likely swine origin, it is often referred to be as ‘swine-origin influenza virus’ A/H1N1, or pandemic influenza A (H1N1) 2009 virus \[^{[4]}\].
A number of antiviral strategies, including vaccines and small molecule inhibitors, have been developed to combat this virus, but its genetic plasticity often leads to resistance to virus-targeted antiviral strategies. Because of its small genome, the influenza virus, like other viruses, is an obligate parasite and must make extensive use of host cell machinery. Thus, an alternate antiviral strategy could be to better understand the critical host factors that are influenced and required by the virus for its efficient propagation. While a cell’s genome generally remains relatively constant (except for certain epigenetic events), the cell’s proteome (the total protein repertoire, including how any given protein may be cotranslationally or posttranslationally modified) varies greatly due to its biochemical interactions with the genome, as well as the cell’s interactions with the environment. A cell’s protein expression is dependent on the location of the cell, different stages of its life cycle, and different environmental conditions. In the case of viruses, which require the host cell’s machinery and metabolism to replicate, the cell’s proteome also reflects the specific alterations of the pathways induced by virus infection. Previous analyses of how cells respond to influenza virus infection have used microarray technologies which measure the cellular “transcriptome”. However, little concordance is found between microarray and protein data. It is partly because mRNA levels cannot provide complete information about levels of protein synthesis or extents of posttranslational modifications. Thus, proteomic analyses have also been employed to better understand host alterations to virus infection.

1.2 Proteomics in influenza virus research

In the past decade, there have been remarkable advances in proteomic technologies. Mass spectrometry (MS) has emerged as the preferred method for
in-depth characterization of the protein components of biological systems. Using mass spectrometry, key insights into the composition, regulation and function of molecular complexes and pathways have been gained. It is clear that mass spectrometry based proteomics is now a powerful ‘hypothesis-generating engine’ that, when combined with complementary molecular, cellular and pharmacological techniques, provides a framework for translating large data sets into an understanding of complex biological processes.

Proteomics combined with bioinformatics has emerged as an important tool to extract detailed information of cellular signaling mechanisms. Modern mass spectrometry based approaches are possible to identify and quantify thousands of proteins from cellular samples \cite{5, 6}. Most of the large scale quantitative proteomics experiments, however, have focused on studying changes in protein abundances in whole cell lysates or individual organelles. The use of subcellular proteomics provides a deeper insight into cellular events as protein abundancies can be studied on the level of different subcellular compartments and protein translocations between different cell parts can also be detected. Moreover, pathway and network analyses can provide mechanistic insights by subsequently linking the proteins found to be differentially regulated to the underlying cellular functions and other key players known to be involved in these events. Combining the quantitative subcellular proteomics with bioinformatics provides a global view of host-pathogen-interactions during influenza A virus infection of human primary macrophages \cite{7}. It shows that viral infection regulates the expression and/or subcellular localization of more than one thousand host proteins at early phases of infection \cite{8}.

The ability to fractionate protein samples to enhance the dynamic range of
detectable proteins is a key issue when identifying the components of a protein complex by mass spectrometry. Two complementary techniques often used in virus analysis, one of which is based on separation of proteins and the other is based on separation of peptides. For the first method, virus preparations were separated by SDS-PAGE on an 8–16% gradient gel. Following Coomassie Blue staining, each lane was cut into successive slices from top to bottom and the individual slices were subjected to in-gel trypsin digestion. This procedure was repeated on a gel and gel slices were excised, so as to maximize the chances of detecting small molecular weight proteins. The peptides in each gel slice were then analyzed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometric (MALDI-TOF MS), quadrupole-time of flight tandem mass spectrometric (Q-TOF MS/MS) or liquid chromatography tandem mass spectrometry (LC-MS/MS) and the resulting fragment ion spectra were searched against protein databases for identification. The second method employed in the study was multidimensional protein identification technology (MudPIT). The deglycosylated and purified virus was digested with trypsin and the peptides in the mixture were separated by two dimensional chromatography, first on the basis of charge and then on hydrophobicity. The chromatography separation step was directly coupled to the mass spectrometer detector and the resulting spectra were searched against the database for protein identification. The disadvantage of MudPIT is that there is no information on the size of the proteins which is useful for confirmation of protein identity. However, the method allows for the detection of low abundance proteins and extremely small molecular weight proteins that are often lost during gel-separation or gel-extraction steps.
Several different mass spectrometric approaches have been used successfully to analyze the composition of a variety of virions. These methods have proven to be complementary and together have led to a more complete picture of the viral particle. While the data sets obtained by different methods largely agree, each method has its own strengths and weaknesses, and this can result in missed or additional protein identifications. The enveloped DNA viruses in general are composed of a large number of viral proteins, including many that are not strictly structural but are known or predicted to function once released into the host cell \[11\]. It is not yet known whether these cellular proteins play a conserved functional role in these virions, are passenger proteins that are accidentally packed...
due to their location and high abundance, or are contaminants that copurify with the virions. The latter possibility seems unlikely, since the different types of virions analyzed were prepared using different methods and, in each case, checked for purity prior to analysis. Whether or not the above-described host proteins are purposefully packaged in the virions will require more extensive investigation. However, it is not hard to imagine that host chaperone proteins may be required to chaperone various viral proteins during packaging and/or release from the virions and hence are packaged along with their viral partners.

1.3 Potencial medicine study in influenza virus infection

Currently, two classes of anti-influenza virus drugs, M2 ion-channel inhibitors (i.e., amantadine and rimantadine) and neuraminidase inhibitors (i.e., oseltamivir and zanamivir) are widely used clinically. The former is effective only against type A influenza virus, whereas the latter is effective against both type A and B viruses. Although these drugs reduce the time to alleviation of illness by 1–2 days, resistance to these drugs and adverse effects has been reported. For those reasons, novel anti-influenza virus drugs are needed. Many antioxidant agents are potential medicines in virus infection study because they have been used in treatment of virus infections. The antioxidants create oxidative stress in cells, which have emerged as a suspected component in the pathogenesis of some diseases in recent years [12]. In the earliest stage of virus infection, reductive-oxidative may be important to cell recovery and duplication [13]. The increased damage causing reactive oxygen intermediates are generated when stores of naturally occurring antioxidant reducing agents are depleted. The presence of uncontrolled oxygen-containing molecules may cause damage to cell membranes, proteins and nucleic acids, as well as the alterations in the intra- and inter-cellular
environments. The net effect of this damage has been termed as oxidative stress. Plants have a long evolutionary history with respect to developing resistance against viruses and are increasingly drawing attention as potential sources for development of antiviral drugs. In our research, small molecule anti-oxidant was incubated into the cell’s culture medium to test their effort on influenza virus infected cells. In addition, the investigation of elderberry juice that was known to contain a variety of polyphenols and peptides was introduced in Appendix. The total phenolic and total monomeric anthocyanin concentrations of elderberry were found to vary highly among different cultivars. More than 1,000 peptides were identified successfully in elderberry juice.
Chapter 2

Identification of Hemagglutinin isoforms difference of influenza A virus

H1N1

2.1 Introduction

Influenza is a highly contagious, acute, viral respiratory disease that causes significant morbidity and mortality worldwide each year. Global outbreaks of human influenza (pandemics) arise from influenza A viruses with novel hemagglutinin (HA) and/or neuraminidase (NA) molecules to which humans have no immunity. HA, the envelope glycoprotein of influenza virus, is one of the best-studied viral fusion proteins. Proteolysis of the precursor HA generates two polypeptides, HA1 and HA2, linked by a disulfide bond [14-17]. Following binding of HA1 to sialic acid-containing receptors on the host membrane, the virus is internalized by endocytosis [18, 19]. Specifically, the structural changes and antigenic variations in the HA molecule are the main obstacle to the control of viral transmission. The HA1 domain of HA almost contains all of the antigenic sites, and is more prone to mutations and antigenic drifts than the rest of protein [20]. The N-terminal domain of the influenza hemagglutinin (HA) is the only portion of the molecule that inserts deeply into membranes of infected cells to mediate the viral and the host cell membrane fusion. This domain constitutes an autonomous folding unit in the membrane, causes hemolysis of red blood cells and catalyzes lipid exchange between juxtaposed membranes in a pH-dependent manner. HA molecules, which form trimers, attach the virus to sialic acid receptors on the cell surface and mediate the release of viral ribonucleoprotein particles into the cytoplasm. HA1 contains the sialic acid binding site. After binding, the virus is internalized in endosomes.
Three dimensional structure of HA has been reported \cite{21, 22}. But in this experiment, isoforms of HA were detected \cite{23-27}. These isoform sequences were important for vaccine development. These sequences also provided the different domains for some medicine design to cure the influenza.

![3D structure of HA from PDB \cite{21}.](image)

**Figure 2.1** 3D structure of HA from PDB \cite{21}.

### 2.2 Materials and methods

HPLC grade ACN and methanol were purchased from Fisher (Fairlawn, NY). Sequencing grade trypsin was obtained from Promega (Madison, WI). All other chemicals were purchased from Sigma-Aldrich (St. Louis, MO). Avian influenza virus A/Puerto Rico/8/1934(H1N1) was propagated in a biosafety level 3 (BL-3) containment facility. The genome of this strain of duck-origin influenza virus was partially sequenced, that has been preserved in State Key Laboratory for Emerging Infectious Diseases of the University of Hong Kong.

#### 2.2.1 Virus cultivation
Avian influenza virus strain A/Puerto Rico/8/1934(H1N1) was harvested from allantoic fluid of chicken embryos inoculated as 10-day old embryos and collected 48 h later. Several serial passages in chicken embryos were carried out to promote potential mutations in virus. Due to the highly pathogenicity of the virus possess, the virus-containing allantoic fluid was inactivated with 0.03 % formalin at 4 °C for 72 h. All the experiments using the active virus were carried out in a biosafety level 3 laboratory. The inactivated virus-containing allantoic fluid was then subjected to further experiments.

2.2.2 Isolation of virus by ultracentrifugation

The virus-containing allantoic fluid was cleared from large debris by centrifugation at 5,000 g for 30 min. The virus was collected from the supernatant by pelleting it through a 5-step discontinuous gradient cushion consisting of 20, 30, 40, 50, and 60 % (wt/vol) sucrose, in a Beckman Rotor SW40 Ti at 155,000 g (35,000 rpm) at 4 °C for 60 min. The resulted virus band between 40% and 50% sucrose was carefully aspirated out from the SW40 tubes, followed by dilution with four volumes of buffer (100 mM Tris, 10 mM MgCl₂, pH 7.4) to form a suspension. The virus was then applied to a two steps discontinuous gradient of 30% and 60% (wt/vol) sucrose in Ultra clear SW60 tubes. Centrifugation was run in a Beckman Rotor SW60 Ti at 100,000 g (38,800 rpm) at 4 °C for 60 min. The virus band was carefully collected, and then freezed at -80 °C until use. The purity of the virus was checked by SDS-PAGE.

2.2.3 SEM detection, enzyme digestion and PAGE electrophoresis

After purification, some viruses were detected by SEM. The virus sample was then separated into two groups. One group was separated by SDS PAGE electrophoresis directly. The N-glycan on the proteins was able to be released by
enzyme named peptide-N-glycosidase F (PNGase F). For this reason, the other group was digested by PNGase F and then separated by SDS PAGE electrophoresis. Electrophoretic analyses were performed by using a Mini-Cell system (Bio-Rad, Hercules, CA), and run in 12 % tris-glycine-SDS polyacrylamide gels with a 5 % stacking gel at a constant voltage setting of 100 V. Samples were prepared by dilution with loading buffer and boiled at 95 °C for 5 min. After electrophoretic separation, the gels were stained with colloidal Coomassie G250 and scanned with a calibrated densitometer (GS800, Bio-Rad).

IEF was carried out on Protean IEF Cell (BioRad). Samples containing 0.4 mg proteins were diluted to 300 mL with rehydration solution (6 M urea, 2% w/v CHAPS, 65 mM DTT, 0.5% v/v pH 3–10 Bio-Lyte, trace bromophenol blue), and applied to IPG gel strips (pH 3–10, 17 cm, BioRad) for 14 h in a passive mode. Proteins absorbed into IPG gel strips were focused for 80 kVh. After equilibrated in the equilibration solution, gel strips were applied on second-dimensional PAGE with 12 % polyacrylamide. Separation was then carried out on a Protean II xi electrophoresis system (BioRad) at a current setting of 10 mA/gel for the initial 1h and 15 mA/gel until the bromophenol blue reached the bottom of the gel.

2.2.4 In-gel digestion

The protein spots were sliced. Three different endoproteinase, namely Trypsin, Asp-N, Glu-C were used to digest these protein spots. The in-gel digestion methods of three kinds of endoproteinase are different.

The protein sample (approximate 1.0 μg proteins) was prepared by dilution with equal volume of 25 mM ammonium bicarbonate. After reduction and alkylation, the sample was separated into three parts. One part was digested by addition of sequeling grade modified porcine trypsin (Promega, Madison, WI) in
a trypsin to protein weight ratio of 1:20. The mixture was incubated for 16 h or overnight at 37 °C. The second part was digested by endoproteinase Asp-N (sequencing grade from \textit{Pseudomonas fragi} mutant, Roche). Protein is subjected to digestion by Asp-N at a ratio of 20:1 respectively for 16 hours at 37°C in Asp-N Reaction Buffer (50 mM Tris-HCl, 2.5 mM ZnSO₄, pH 8.0). The third part was digested by endoproteinase Glu-C (sequencing grade from \textit{Staphylococcus aureus} V8, Roche), Protein was subjected to digestion by Glu-C at a ratio of 20:1 for 16 hours at 25°C in Glu-C Reaction Buffer (25mM Na₃PO₄, pH 7.8). The resulted tryptic peptide mixture was stored at -20 °C until analysis.

2.2.5 MALDI-TOF MS analysis

The matrix solution was freshly prepared by dissolving 10 mg CHCA in 1 mL of 0.5 % TFA and 50 % ACN. Peptides from in-gel digested proteins were mixed with matrix solution (1:1 v/v) and loaded on the anchorchip sample plate. Air-dried PMFs of protein spots were obtained on an Autoflex MALDI-TOF mass spectrometer (Bruker Daltonics, Bremen, Germany) operating in the reflection mode with an accelerating voltage of 19 kV and a delayed extraction time of 150 ns. All MALDI-TOF MS spectra were calibrated internally using trypsin autolysis products and/or calibrated externally by a set of peptide standards. Peak lists were created from raw data using flexAnalysis software (version 2.0) with the SNAP peak detection algorithm. Data searching against the Swiss-Prot database in the entries of Homo sapiens and/or other viruses were performed using an in-house MASCOT server (Matrix Science, London, UK). Peptide tolerance was set at 100 ppm with one missing cleavage allowed. N-terminal acetylation, asparagine/glutamine deamidation, and methionine oxidation were selected as variable modifications, while cysteine carbamidomethylation was as fixed
modification. For positive identification, the score of result had to be over the significance threshold level (p<0.05)

2.2.6 Nanoelectrospray-ESI-MS/MS analysis

The tryptic peptides were desalted using ZipTip C\textsubscript{18} pipette tips (Millipore, Bedford, MA). Full scan and MS/MS experiments were carried out with a quadrupole orthogonal acceleration TOF mass spectrometer with an external nanoelectrospray ion source (Protana A/S, Odense, Denmark). Renin substrate was used to calibrate the instrument in both MS and MS/MS mode. Approximately 2 \textmu L of sample was loaded into a nanoelectrospray tip. After the full mass spectra (parent ions) of trypsin-digested peptides were obtained in TOF MS mode, the charge states of the peptides were assigned by their stable isotope spacing. Then parent ions of interest were subject to sequence analysis in product ion mode. After the MS/MS analysis, the MASCOT generic files were generated by using a script embedded in the Analyst QS 1.1 software (MDS Sciex). The followed database searching was applied for the generated files.

Against the Swiss-Prot database, in the entries of H. sapiens and/or other viruses, the parameter settings on a local MASCOT server were showed as the following: one missed cleavage; peptide tolerance, 0.2 Da; MS/MS tolerance, 0.1 Da; fixed carbamidomethylation of cysteine, and variable oxidation of methionine. Uninterpreted peptide tandem mass spectra were subjected to manual \textit{de novo} sequencing with the aid of sequence tag searching tool in MASCOT server to expand the confirmation of identified proteins. A minimum of two significant MS/MS peptide fragments was considered sufficient to assign positive identification for a protein spot.

2.3 Result
2.3.1 Purification of influenza virus H1N1

The purified influenza virus H1N1 by gradient sucrose is as the Figure 2.2. The white layer between 40 % and 50 % is the virus layer.

![Image](image_url)

**Figure 2.2** Influenza virus purified by gradient sucrose centrifugation.

2.3.2 SEM detection and electrophoresis separation

After centrifugation, the virus layer was collected carefully. The virus samples were stored in 4°C. One part of virus sample were detected by SEM (Figure 2.3), the other part of virus samples was separated by SDS-PAGE. The stained gels were showed in Figure 2.4. Two dimensional electrophoresis separation of H1N1 (PR8) proteins was shown in Figure 2.5
**Figure 2.3** Virus particles detected by SEM.

**Figure 2.4** SDS-PAGE separation of influenza virus H1N1 (PR8) proteins.
One dimensional electrophoresis is not enough to separate the protein with different isoelectro-point. Two dimensional electrophoresis is often needed for separation of protein isoforms especially those with similar molecular weigh and different isoelectro-point. As the Figure 2.5 showed, the HA was composed in a serious of protein isoforms. PNGase F can eliminate the glycan from glycoproteins. After the elimination, HA was separated also by 2DE (Figure 2.6). Only one protein spot was observed in the 2DE gel.

Figure 2.5 Two dimensional electrophoresis separation of H1N1 (PR8) proteins.
Figure 2.6 Two dimensional electrophoresis separation of H1N1 (PR8) proteins who were digested by PNGase F. The spot marked with arrow was the HA protein.

2.3.3 Mass spectrometry detection

Five HA spots of influenza A virus H1N1 (PR8) protein that separated by 2DE method were digested by three kinds of proteases respectively. These proteases are Trypsin, Asp-N and Glu-C. Their digested peptides mass spectra detected by MALDI-TOF MS were show in Figures 2.7-2.9.
**Figure 2.7** Peptide mass spectrogram (MALDI-TOF MS) of HA digested by Trypsin.
Figure 2.8 Peptide mass spectrogram (MALDI-TOF MS) of HA digested by Asp-N.
Different enzymes break down the proteins into peptide fragments on the different amino acid residues. If amino acids or their orders of HA isoforms are different, specific peaks would be detected in the spectra. After three enzymes’ digestion (Figure 2.7-2.9), peaks of peptides were almost the same in the mass spectra of each spot, indicating that components of amino acids and their orders of the five protein spots were identical.
2.3.4 Database analysis

Database searching confirmed these protein isoforms separated by 2DE. The result showed that they were with the same Swiss-Prot accession ID (HEMA_I34A1) and name (Hemagglutinin). The Origin Source (OS) that was influenza A virus (strain A/Puerto Rico/8/1934 H1N1). The nominal molecular mass (Mr) was 64225 Dalton and the calculated pI value was 6.52.

Table 2.1-2.5 presented the matched sequences of spot 1-5 from the database searching of the MS result, while Table 2.6 and 2.7 were the matched sequences of spot 1 and 2 from the database searching of MS/MS result. Figure 2.10 showed the amino acid sequences of intact hemagglutinin of H1N1 (PR8). The detected sequences were highlighted with red color. The signal sequence was marked in blue.

| HA1 | MKANLLVLLC ALAAADDTI CIGYHANNST DTVDTLEKN VTVTHSVNLL
| 51 | EDSHINGKLR IK1LTKIPLQG KINAGWLLG NPECPLLVPH RSWSYIVETP
| 101 | NSENGICYPG DFIDYELRE QLSYSSFER FEIFPK
| 151 | ACSHEGKSSF YRNLWLTLTI EGSPKLKNS YVKKLAEVL VLGHIHPFN
| 201 | KSEQQNLQYN ENAYSVVTS NVNRRFTEPI AERPKVQDQA GRNNYWTLL
| 251 | KPGDITIEFA NGNLIAPMYA FALSRGFSGG ITTSNASMHE CNTKQTPLG
| 301 | AINSSLPYQN IHPVTIGECP KVVRSAKLRI MVTGLF

| HA2 | RGLFGAIA
| 351 | GF1EGGWTGM IDGWGYHHQ NEQGSGYAAD QKSTQNAING ITNKYNTVIE
| 401 | KNIQFTAVG KENKLEKRM ENLNNKVDGG FLDIWYNAE LLVLLENERT
| 451 | LDFHDSNKN LYEYVSQLK NNAEIGNGC FEFYHKCDNE CMSVRNGTY
| 501 | DYPKYSEESK LRKKEVGVK LESMGYQIL AIYSTVASSL VLLVSLGAI
| 551 | FWMSNGLSQ CRIC

Figure 2.10 HA sequence of H1N1 (PR8). The detected sequences were marked in red color and framed. The signal sequence was marked in blue and underlined.
Table 2.1 Matched sequence of spot 1.

<table>
<thead>
<tr>
<th>Observed</th>
<th>Mr(expt)</th>
<th>Mr(calc)</th>
<th>ppm</th>
<th>Start</th>
<th>End</th>
<th>Miss</th>
<th>Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>676.3522</td>
<td>675.3449</td>
<td>675.3738</td>
<td>-42.74</td>
<td>330</td>
<td>335</td>
<td>0</td>
<td>R.MVTGLR.N</td>
</tr>
<tr>
<td>780.3751</td>
<td>779.3678</td>
<td>779.4218</td>
<td>-69.23</td>
<td>131</td>
<td>136</td>
<td>0</td>
<td>R.FEIIPK.E</td>
</tr>
<tr>
<td>801.4098</td>
<td>800.4026</td>
<td>800.4253</td>
<td>-28.39</td>
<td>236</td>
<td>242</td>
<td>1</td>
<td>K.VRDQAGRM</td>
</tr>
<tr>
<td>896.5173</td>
<td>895.5100</td>
<td>895.5491</td>
<td>-43.60</td>
<td>63</td>
<td>71</td>
<td>0</td>
<td>K.GIAPLQLGK.C</td>
</tr>
<tr>
<td>1016.5580</td>
<td>1015.5507</td>
<td>1015.5702</td>
<td>-19.21</td>
<td>163</td>
<td>170</td>
<td>0</td>
<td>R.NLLWLTEKE</td>
</tr>
<tr>
<td>1187.6774</td>
<td>1186.6702</td>
<td>1186.6346</td>
<td>30.0</td>
<td>226</td>
<td>235</td>
<td>1</td>
<td>R.FTPAERP.KV</td>
</tr>
<tr>
<td>1268.6648</td>
<td>1267.6575</td>
<td>1267.6044</td>
<td>41.9</td>
<td>120</td>
<td>130</td>
<td>0</td>
<td>R.EQLSSVSSF.V</td>
</tr>
<tr>
<td>1343.7875</td>
<td>1342.7802</td>
<td>1342.7357</td>
<td>33.1</td>
<td>225</td>
<td>235</td>
<td>1</td>
<td>R.FTPAERP.KV</td>
</tr>
<tr>
<td>2633.3654</td>
<td>2632.3581</td>
<td>2632.2153</td>
<td>54.2</td>
<td>203</td>
<td>224</td>
<td>0</td>
<td>K.EQONLYQENAYVSVVTSNYNR.R</td>
</tr>
</tbody>
</table>

Table 2.2 Matched sequence of spot 2.

<table>
<thead>
<tr>
<th>Observed</th>
<th>Mr(expt)</th>
<th>Mr(calc)</th>
<th>ppm</th>
<th>Start</th>
<th>End</th>
<th>Miss</th>
<th>Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>780.3324</td>
<td>779.3251</td>
<td>779.4218</td>
<td>-124.06</td>
<td>131</td>
<td>136</td>
<td>0</td>
<td>R.FEIIPK.E</td>
</tr>
<tr>
<td>801.3679</td>
<td>800.3607</td>
<td>800.4253</td>
<td>-80.74</td>
<td>236</td>
<td>242</td>
<td>1</td>
<td>K.VRDQAGRM</td>
</tr>
<tr>
<td>896.4439</td>
<td>895.4367</td>
<td>895.5491</td>
<td>-125.54</td>
<td>63</td>
<td>71</td>
<td>0</td>
<td>K.GIAPLQLGK.C</td>
</tr>
<tr>
<td>1016.4978</td>
<td>1015.4905</td>
<td>1015.5702</td>
<td>-78.50</td>
<td>163</td>
<td>170</td>
<td>0</td>
<td>R.NLLWLTEKE</td>
</tr>
<tr>
<td>1187.6236</td>
<td>1186.6164</td>
<td>1186.6346</td>
<td>-15.37</td>
<td>226</td>
<td>235</td>
<td>0</td>
<td>R.FTPAERP.KV</td>
</tr>
<tr>
<td>1268.6084</td>
<td>1267.6011</td>
<td>1267.6044</td>
<td>-2.62</td>
<td>120</td>
<td>130</td>
<td>0</td>
<td>R.EQLSSVSSF.V</td>
</tr>
<tr>
<td>1343.7287</td>
<td>1342.7215</td>
<td>1342.7357</td>
<td>-10.60</td>
<td>225</td>
<td>235</td>
<td>1</td>
<td>R.FTPAERP.KV</td>
</tr>
<tr>
<td>2633.1526</td>
<td>2632.1453</td>
<td>2632.2153</td>
<td>-26.61</td>
<td>203</td>
<td>224</td>
<td>0</td>
<td>K.EQONLYQENAYVSVVTSNYNR.R</td>
</tr>
</tbody>
</table>

Table 2.3 Matched sequence of spot 3.

<table>
<thead>
<tr>
<th>Observed</th>
<th>Mr(expt)</th>
<th>Mr(calc)</th>
<th>ppm</th>
<th>Start</th>
<th>End</th>
<th>Miss</th>
<th>Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>801.4574</td>
<td>800.4501</td>
<td>800.4253</td>
<td>31.0</td>
<td>236</td>
<td>242</td>
<td>1</td>
<td>K.VRDQAGRM</td>
</tr>
<tr>
<td>1137.7667</td>
<td>1136.7595</td>
<td>1136.7281</td>
<td>27.6</td>
<td>61</td>
<td>71</td>
<td>1</td>
<td>R.LKGIAPLQLGK.C</td>
</tr>
<tr>
<td>1187.7076</td>
<td>1186.7004</td>
<td>1186.6346</td>
<td>55.4</td>
<td>226</td>
<td>235</td>
<td>0</td>
<td>R.FTPAERP.KV</td>
</tr>
<tr>
<td>1343.8001</td>
<td>1342.7928</td>
<td>1342.7357</td>
<td>42.5</td>
<td>225</td>
<td>235</td>
<td>1</td>
<td>R.FTPAERP.KV</td>
</tr>
<tr>
<td>1677.8991</td>
<td>1676.8918</td>
<td>1676.8773</td>
<td>8.61</td>
<td>163</td>
<td>176</td>
<td>1</td>
<td>R.NLLWLTEKEGSYPK.L</td>
</tr>
<tr>
<td>2633.1706</td>
<td>2632.1633</td>
<td>2632.2153</td>
<td>-19.77</td>
<td>203</td>
<td>224</td>
<td>0</td>
<td>K.EQONLYQENAYVSVVTSNYNR.R</td>
</tr>
<tr>
<td>2789.2411</td>
<td>2788.2338</td>
<td>2788.3164</td>
<td>-29.62</td>
<td>203</td>
<td>225</td>
<td>1</td>
<td>K.EQONLYQENAYVSVVTSNYRR.F</td>
</tr>
</tbody>
</table>
Table 2.4 Matched sequence of spot 4.

<table>
<thead>
<tr>
<th>Observed</th>
<th>Mr(expt)</th>
<th>Mr(calc)</th>
<th>ppm</th>
<th>Start</th>
<th>End</th>
<th>Miss</th>
<th>Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>780.3681</td>
<td>779.3608</td>
<td>779.4218</td>
<td>-78.24</td>
<td>131</td>
<td>136</td>
<td>0</td>
<td>R.FEIFPK.E</td>
</tr>
<tr>
<td>896.5032</td>
<td>895.4959</td>
<td>895.5491</td>
<td>-59.36</td>
<td>63</td>
<td>71</td>
<td>0</td>
<td>K.GIAPLQLGK.C</td>
</tr>
<tr>
<td>1016.5649</td>
<td>1015.5576</td>
<td>1015.5702</td>
<td>-12.44</td>
<td>163</td>
<td>170</td>
<td>0</td>
<td>R.NLLWLTEK.E</td>
</tr>
<tr>
<td>1187.6998</td>
<td>1186.6925</td>
<td>1186.6346</td>
<td>48.8</td>
<td>226</td>
<td>235</td>
<td>0</td>
<td>R.FTPEIAERP.K.V</td>
</tr>
<tr>
<td>1268.6764</td>
<td>1267.6691</td>
<td>1267.6044</td>
<td>51.0</td>
<td>120</td>
<td>130</td>
<td>0</td>
<td>R.EQLSSVSFER.F</td>
</tr>
<tr>
<td>1343.8159</td>
<td>1342.8086</td>
<td>1342.7357</td>
<td>54.3</td>
<td>225</td>
<td>235</td>
<td>1</td>
<td>R.RFTPEIAERP.K.V</td>
</tr>
<tr>
<td>1726.0033</td>
<td>1724.9960</td>
<td>1724.9362</td>
<td>34.7</td>
<td>188</td>
<td>202</td>
<td>0</td>
<td>K.EVLWLGIHHPPNSK.E</td>
</tr>
<tr>
<td>1911.1189</td>
<td>1910.1116</td>
<td>1910.0526</td>
<td>30.9</td>
<td>186</td>
<td>202</td>
<td>1</td>
<td>K.GKEVLWLGIHHPPNSK.E</td>
</tr>
<tr>
<td>2633.3590</td>
<td>2632.3517</td>
<td>2632.2153</td>
<td>51.8</td>
<td>203</td>
<td>224</td>
<td>0</td>
<td>K.EQQNLQONENAYVSVVTSNYNR.R</td>
</tr>
</tbody>
</table>

Table 2.5 Matched sequence of spot 5.

<table>
<thead>
<tr>
<th>Observed</th>
<th>Mr(expt)</th>
<th>Mr(calc)</th>
<th>ppm</th>
<th>Start</th>
<th>End</th>
<th>Miss</th>
<th>Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>780.3059</td>
<td>779.2986</td>
<td>779.4218</td>
<td>-158.05</td>
<td>131</td>
<td>136</td>
<td>0</td>
<td>R.FEIFPK.E</td>
</tr>
<tr>
<td>896.4641</td>
<td>895.4568</td>
<td>895.5491</td>
<td>-103.08</td>
<td>63</td>
<td>71</td>
<td>0</td>
<td>K.GIAPLQLGK.C</td>
</tr>
<tr>
<td>1016.5620</td>
<td>1015.5547</td>
<td>1015.5702</td>
<td>-15.28</td>
<td>163</td>
<td>170</td>
<td>0</td>
<td>R.NLLWLTEK.E</td>
</tr>
<tr>
<td>1187.7551</td>
<td>1186.7478</td>
<td>1186.6346</td>
<td>95.4</td>
<td>226</td>
<td>235</td>
<td>0</td>
<td>R.FTPEIAERP.K.V</td>
</tr>
<tr>
<td>1268.7522</td>
<td>1267.7450</td>
<td>1267.6044</td>
<td>111</td>
<td>120</td>
<td>130</td>
<td>0</td>
<td>R.EQLSSVSFER.F</td>
</tr>
<tr>
<td>1725.9696</td>
<td>1724.9623</td>
<td>1724.9362</td>
<td>15.1</td>
<td>188</td>
<td>202</td>
<td>0</td>
<td>K.EVLWLGIHHPPNSK.E</td>
</tr>
<tr>
<td>2293.7495</td>
<td>2292.7422</td>
<td>2293.1347</td>
<td>-171.16</td>
<td>72</td>
<td>91</td>
<td>0</td>
<td>K.CNIAGWLLGNPECPLL PVR.S</td>
</tr>
</tbody>
</table>

Table 2.6 Q-TOF MS database searching result of spot 1.

<table>
<thead>
<tr>
<th>Observed</th>
<th>Mr(expt)</th>
<th>Mr(calc)</th>
<th>Delta</th>
<th>Miss</th>
<th>Score</th>
<th>Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>508.8000</td>
<td>1015.5854</td>
<td>1015.5702</td>
<td>0.0152</td>
<td>0</td>
<td>36</td>
<td>R.NLLWLTEK.E</td>
</tr>
<tr>
<td>878.4000</td>
<td>2632.1782</td>
<td>2632.2153</td>
<td>-0.0372</td>
<td>0</td>
<td>62</td>
<td>K.EQQNLQONENAYV SVVTSNYNR.R</td>
</tr>
</tbody>
</table>

Table 2.7 Q-TOF MS database searching result of spot 2.

<table>
<thead>
<tr>
<th>Observed</th>
<th>Mr(expt)</th>
<th>Mr(calc)</th>
<th>Delta</th>
<th>Miss</th>
<th>Score</th>
<th>Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>509.3000</td>
<td>1016.5854</td>
<td>1015.5702</td>
<td>1.0152</td>
<td>0</td>
<td>25</td>
<td>R.NLLWLTEK.E</td>
</tr>
<tr>
<td>878.4000</td>
<td>2632.1782</td>
<td>2632.2153</td>
<td>-0.0372</td>
<td>0</td>
<td>57</td>
<td>K.EQQNLQONENAYVSVVTSNYNR.R</td>
</tr>
</tbody>
</table>

2.4 Discussion

Influenza virus was pandemic in the world due to the structure change of virus protein. It has been difficult to find some effective and long-term medicine or
vaccine because virus changed their gene sequence frequently. Thus the peptides which are translated by altered gene are different original peptide sequence. In addition, the difference of gene sequence between viruses was also compared to predict their origin source.

Genetic sequencing is quick method to detect the difference of virus in gene level, but it is difficult to find the change at protein level. Although people can predict the protein sequence from their gene sequence, the result may not absolutely correct. If using predict protein to design the medicine or vaccine, it is may not provide the vaccine with sufficient high potency to defend the virus. HA, an important protein of virus, is mostly located at outer wall of H1N1 viral particles. Some medicines were designed to interfere the fusion of virus through the change of HA structure.

In this study, gradient sucrose centrifugation, electrophoresis separation, MS detection and database searching methods were used to analyze the HA of H1N1 (PR8). In one-dimensional electrophoresis method, the HA only presented one band of protein in the gel. However, when using two-dimensional electrophoresis method to separate HA, the obtained result showed that the protein had more than five protein isoforms. After digestion with different enzymes and MALDI-TOF MS, Q-TOF MS analysis, the comparison on the sequences showed that there are some identical sequences of H1N1 (PR8). In Figure 2.10, the whole length of HA sequence were presented. The detected sequences were showed in frame. The underlined one is the signal sequence.

Different methods have been used to identify the mutant protein sites of influenza virus\textsuperscript{[28, 29]}\textsuperscript{28, 29}. But the domain of conserved sequences of protein could provide better information for medicine or vaccine design. These obtained
sequences agreed well with those from database searching. In the five protein isoforms, their determined sequences were also compared to each other. The results indicated that those sequences were same in HA isoforms. Therefore, the method could be applied for discovery of medicines or vaccines that might interact with the conserved sequence in influenza virus.

From the 2DE gel of virus protein sample whose N-glycan were eliminated by PNGase F, only one protein spot was shown. After the MS detection, this protein spot was also identified as HA. The result indicated that isoforms of HA in 2DE gels were related with the glycan or their glycosylation sites. Different glycan or glycosylation sites resulted in different isoelectro-point of HA isoforms. When virus infected the cells, the HA attached to the surface receptor of cells first. The glycan played a major role in this attachment. Therefore, glycan structure and glycosylation sites are significant in virus infection process.

2.5 Chapter summary

In this chapter, influenza virus was purified by ultra-centrifuge methods. The SEM result showed that the virus was well purified. After one dimensional and two dimensional electrophoresis, hemagglutinins of virus were separated into a series of isoforms. These isoforms of proteins were digested by different enzymes. The MS detection result of these protein isoforms showed that they were almost the same in peptides. Glycan elimination enzyme, such as PNGase F, was used to cut off the glycan of HA and then separated by two dimensional electrophoresis. The result indicated that the isoforms of HA in 2DE gels existed because of different glycan structure or glycosylation site of the HA. Those differences in isoforms may play significant role in virus infection process.
Chapter 3

Stable-isotope N-phosphorylation labeling for peptide de novo sequencing
and protein quantification based on organic phosphorus chemistry

3.1 Introduction

Analysis of proteins on a genome-wide scale is increasingly important for providing valuable information on biomarker discovery and metabolic pathway analysis, accelerating the development of powerful new diagnostic tools and new drugs, and leading to a better understanding of the molecular mechanisms that control cell behaviors. The systematic analysis of all proteins in biological samples such as a tissue or cell was popularized under the name of proteomics.[30] Mass spectrometry plays an important role not only in rapidly identifying proteins and determining details of their covalent structures and posttranslational modifications in a biological mixture, but also in high-throughput proteome-wide quantification of proteins, their variants and protein-protein interactions in cell or tissue in response to a variety of conditions of interest.[31] For quantitative analysis, stable-isotope labeling coupled with mass spectrometry have been developed and used extensively in the comparison of relative abundances of expressed proteins.[32] Numerous methods including mass-difference labeling and isobaric labeling methodologies have been developed for quantification analysis in proteomics, peptidomics, and glycomics over the past decade.[33-36] Relative quantification is based on the introduction of a chemically equivalent but mass-differential stable-isotope tags that can be used to determine accurately the abundance of proteins in one sample to another. Isotope incorporation can be performed at the peptide or protein level using $^2$H, $^{13}$C, $^{15}$N, or $^{18}$O as heavy
isotopes. Based on the characteristic isotope patterns, the “light” and “heavy” isotopic peptide pairs from both samples shared the same physicochemical properties may be compared and distinguished in the mass spectra. Mass spectrometry-based stable-isotope labeling methods are mainly divided by the way of stable-isotope labels introduced into the peptide or protein. Generally, stable-isotope tags can be introduced metabolically, enzymatically, chemically, or labeled synthetic peptides as internal standard spiked into the sample at defined concentrations. Metabolic *in vivo* labeling of proteins is achieved by supplying isotope-enriched compounds, such as $^{13}$C-, $^{15}$N-, or $^{18}$O-enriched nutrients\cite{37,38} or stable isotope labeled amino acids (SILAC),\cite{39-44} to an organism or cell culture in a way that can be metabolized and exclusively incorporated into proteins. Another global quantitative method is based on labeling the terminal carboxylic groups of the tryptic peptides with $^{18}$O atoms during or after proteolysis in $^{18}$O-enriched water.\cite{45-48}

Alternative applications of various types of stable isotope tagging by *in vitro* chemical derivatization of functional groups in proteins or peptides have been reported, mainly on N-terminal amino group, but also to a less extent C-terminal carboxylic group or specific amino acid residues with reactive side chains, such as cysteine, lysine, and tryptophan, in order to achieve the quantification using MS or tandem mass spectrometry (MS/MS).\cite{49,50} The chemical labeling strategies are particularly suited for any source of biological protein sample, such as tissue samples, derived from animals or humans where metabolic incorporation is difficult. Since the introduction of the isotope-coded affinity tag (ICAT) described in 1999 by Aebersold and co-workers,\cite{51} stable-isotope labeling strategies have been widely used in quantitative proteomics with many approaches and reagents.
over the past decade. These methods can be divided into two main categories. The first one is mass difference labeling, such as reductive formaldehyde dimethylation,[52-55] acetylation (H3/D3) and propionylation (H5/D5),[56] ICPL,[57-59] ICAT, and DSIC[60]. The second group gained popularity recently is isobaric tags that allow the determination of multiple pools of proteins in one single analysis, such as TMT,[61-63] iTRAQ,[64-67] DiLeu,[68] DiART,[69, 70] IPTL[71] and so on. Most of the labeling methods are based on traditional protein chemistry established during 60s-70s for the coupling of stable isotopic tags with functional group of proteins or peptides (Fig. 3.1).

For example, ICAT has three elements: biotin for affinity purification, stable isotopic linker, and thio-specific reactive group contained an iodoacetyl reactive
group for alkylating efficiently on thiol group of cysteine residues. This labeling chemistry is derived from the protein reduction and alkylation widely used in protein preparations. During the past decade, the chemical structures of the functional group of the tags have been changed and designed rationally to achieve different properties with novel functions, such as the acid cleavable ICAT (cICAT)\textsuperscript{[72, 73]} and visible-coded affinity tags (VICAT)\textsuperscript{[74]}. However, the reactive groups of ICAT-based approaches have not been changed. Similarly, iTRAQ is composed of a primary amine reactive group succinimidyl ester (NHS) for peptide coupling, an isotopic reporter group, and an isotopic balancer group for the normalization of the total mass of the tags. For most of the quantitative tags targeting on N-terminal amino group, the carboxylic groups of stable-isotope tags are traditionally activated to form an anhydrides or NHS esters, which can subsequently be attacked by primary amines through $S_N_2$ reaction pathways and led to the formation of C-N amide bond. Fundamentally, it should be kept in mind that virtually all successful quantitative labeling approaches are dependent on the step of coupling reactions. However, it is not surprising that the coupling chemistry for stable-isotope labeling still relies on classic reactions because of the chemical reactions involved not only need to be extremely specific, but also need to proceed to completion and involve minimal sample handling under mild reaction conditions.

Recently, new reagents and principles have been reported for chemical labeling for quantitative proteomics, including the soluble polymer-based isotopic labeling\textsuperscript{[75]} and the Caltech isobaric tags (CITs) based on click chemistry.\textsuperscript{[76]} In our previous work, a novel chemical labeling concept based on organic phosphorus chemistry was developed for peptide sequencing by using tandem ESI mass
spectrometry. Chemical structures of organophosphorus reagents were shown in Figure 3.2.

**Figure 3.2** Chemical structures of organophosphorus reagents for stable-isotope labeling of tryptic peptides. H-DMP: dimethyl phosphite; D-DMP: deuterium-labeled dimethyl phosphite; DEP: diethyl phosphite; DPP: dipropyl phosphite; DIPP: diisopropyl phosphite; EPP: ethoxylphenyl phosphinate.

A serial of stable organophosphorus reagents, such as dimethyl phosphite (DMP-H),[77] diethyl phosphite (DEP-H),[78] diisopropyl phosphite (DIPP-H),[79, 80] and ethyloxy (phenyl)phosphinate (EPP-H),[81, 82] have been synthesized and applied to peptide *de novo* sequencing. A phosphoramidate P-N bond was formed and a neutral phosphoryl group can be easily incorporated into the N-terminus of peptides through a one-pot specific reaction with quantitative yield under mild reaction conditions. It was found that the introduction of phosphoryl group onto the N-terminus of amino acids and peptides could not only improve the sensitivity in ESI-MS and FAB-MS analyses, but also simplify the fragmentation pathways and interpretation of peptide sequences from the mass spectra of the derivatized peptides, providing a convenient and rapid method for peptide sequencing. In this
paper, we present a novel stable-isotope N-phosphorylation labeling (SIPL) method that may be applied for both peptide sequencing and quantitative proteomics based on organic phosphorus chemistry. By proceeding through a modified Atherton-Todd reaction, a neutral phosphoryl group was specifically incorporated into tryptic peptides with high yields under mild conditions. A deuterium-labeled dimethyl phosphite (D-DMP-D) was synthesized and employed on a standard peptide, a model protein, and protein mixtures for demonstrating the capability of the proposed method for relative quantitative proteomics. It was found that this one-pot phosphorylation reaction was simple, high efficient, and selective, which might provide a very rapid, straightforward, and cost-effective mean both for peptide de novo sequencing and for the comparison of the relative abundance of proteins. The brief workflow of experiment is shown in Figure 3.3.

**Figure 3.3** Brief workflow of experiment of N-phosphorylation labeling for peptide de novo sequencing and protein quantification.
3.2 Materials and methods

3.2.1 Materials and reagents

Lysozyme, myoglobin, hemoglobin, β-lactoglobulin B, bovine serum albumin (BSA), DL-dithiothreitol (DTT), iodoacetamide (IAA), ammonium bicarbonate, formic acid, trifluoroacetic acid (TFA), Triethylamine (TEA), tetrachloromethane, and α-cyano-4-hydroxycinnamic acid (CHCA) were purchased from Sigma (St. Louis, MO, USA) and used without further purification. Sequencing grade trypsin was obtained from Promega (Madison, WI). The Fibrinopeptide A (human, ADSEGEGLAEGGGVR) was purchased from GL Biochem. Ltd. (Shanghai, China). Dimethyl phosphate (H-DMP-H) and CD3OD (99.9 at% D) were obtained from Alfa Aesar Chemical Ltd. (Tianjin, China). HPLC-grade acetonitrile (ACN) and methanol were purchased from Tedia (Fairfield, OH, USA). Deionized water (18 MΩ) was produced by Milli-Q system (Millipore, Bedford, MA, USA). Unless specified otherwise, all chemicals and solvents were analytical reagents.

3.2.2 Synthesis of deuterium-labeled dimethyl phosphite

Deuterium-labeled dimethyl phosphite (D-DMP-D) was synthesized using modified Michaelis-Arbuzov reaction according to a literature method (Figure 3.4 A). Briefly, CD3OD (30 mmol) in 5 mL anhydrous CH2Cl2 was added dropwise to PCl3 (10 mmol) in 5 mL anhydrous CH2Cl2 at 0 °C under argon. The solution was warmed to room temperature and stirred for 30 min. The progress of the reaction was traced by 31P NMR spectroscopy. Once the starting material had disappeared, the reaction solvent was removed under reduced pressure and the residue was distilled to give the target compound as colourless oil (yield, 75 %). The NMR data for D-DMP: 31P NMR (162 MHz, CDCl3): δ = 10.1 (t, 1J_P,D = 107.0 Hz) ppm. HRMS (ESI-TOF, positive ion mode): calcd. for [M+H]+ m/z
118.0650; found m/z 118.0658.

**Figure 3.4.** Phosphorus chemistry for stable-isotope N-phosphorylation labeling. (A) Synthetic pathway of deuterium-labeled dimethyl phosphite (D-DMP) by Michaelis-Arbuzov rearrangement; (B) The reaction mechanism of peptide labeling for protein quantification using modified Atherton-Todd reaction.

### 3.2.3 Trypsin digestion

Standard proteins (1 mg/mL) or protein mixtures were prepared in 50 mM ammonium bicarbonate solution (pH 7.5). The digest of proteins was carried out as follows: the protein solutions (100 µL) were reduced by adding DTT (5 µL, 200 mM in 100 mM ammonium bicarbonate solution) and boiled for 10 min. The resulting cysteine residues were alkylated with IAA (4 µL, 1 M in 100 mM ammonium bicarbonate solution) for 1 h in the dark at room temperature. The excess of IAA was neutralized by adding 20 µL DTT stock. Freshly prepared trypsin (0.2 µg/µL) was added (the ratio of trypsin to the treated protein was 1:50), and the digestion was performed overnight at 37 °C. Prior to MS analysis, the digested samples were desalted using reverse-phase C₁₈ ZipTip pipette tips.
(Millipore, Billerica, MA, USA) according to the manufacture’s instructions.

3.2.4 N-Phosphorylation of standard peptide and tryptic peptides

The individual standard peptide (2 µL, 1 mg/mL) or protein digests (2 µL) were dissolved in a mixture of H₂O (10 µL), ethanol (5 µL), and triethylamine (5 µL) in a 0.2 mL PCR tube. Gently vortex and spin the sample to the bottom of the tube. The reaction mixture was cooled to 0 °C in an ice-water bath. Then, H-DMP-H or D-DMP-D (1 µL) in tetrachloromethane (5 µL) was added into the above reaction mixture, which was mixed periodically. The reaction was allowed to proceed 40 min at room temperature before evaporating the mixtures to dryness under N₂ gas. The residues were dissolved in 50 µL H₂O (0.1 % formic acid) and desalted using ZipTip tips mentioned above.

3.2.5 Mass spectrometry analysis

MALDI-TOF MS analyses were performed on a Bruker Autoflex II mass spectrometer (Bruker Daltonics, Bremen, Germany) in positive reflection mode. The mass spectrometer was equipped with a pulsed nitrogen laser operated at 337 nm with 3 ns duration pulses and employed stainless steel targets (MTP 384 target ground steel, Bruker Daltonics). Voltage impressed on the ion one and two was 20.0 and 19.0 kV, respectively. The laser power energy was adjusted as needed. The acceleration voltage, grid voltage, and delayed extraction time were set as 19 kV, 90 %, and 150 ns, respectively. Each mass spectrum was acquired as an average of 200 laser shots at 10.0 Hz frequency. The protein digests after desalting were mixed with matrix solution (20 mg CHCA in 1 mL 50 % ACN/water with 0.5 % TFA) and loaded on the AnchorChip sample plate. After drying, the sample mixture was analyzed in the reflection ion mode and peptide mass fingerprints of the protein were obtained for protein identification.
ESI-MS analyses were carried out with a quadrupole orthogonal acceleration TOF mass spectrometer (API QStar Pulsar, MDS Sciex, Toronto, Canada) with an external nanoelectrospray ion source (Protana, Odense, Denmark). The ion source voltage, mirror voltage, plate voltage and liner voltage were 1, 0.980, 0.346, 0.428, and 4.000 kV, respectively. Rennin peptide substrate was used to calibrate the instrument. Approximately 2 µL of sample was loaded into the Proxeon nanoelectrospray needles. The needles were originally closed at the tip, and would be opened by gently touching the tips on the orifice plate. The scan range was generally from m/z 500 to m/z 2000. The precursor ions of interest were then subjected to MS/MS analysis in the product ion mode. The collision energy applied was adjusted during experiments.

3.2.6 Nano LC-Chip/Q-TOF MS analysis

Quantification of the standard peptide, protein, and protein mixtures were carried out with a Agilent 6200 HPLC-chip/TOF MS system (Agilent Technologies, Santa Clara, CA), which equipped with a Agilent 1100 nanopump as analytical pump for sample separation, a capillary pump as the loading pump for sample enrichment and desalting, a microwell-plate autosampler maintained at 6 °C by the thermostat, HPLC-Chip cube as interface, and Agilent 6210 TOF MS as detection. Chromatographic separations were performed on an enrichment column with a volume of 40 nL and an analytical reverse-phase column (Zorbax C18, 5 µm, 43 mm x 75 µm). Mobile phase buffer A was 5.0 % ACN/water (v/v) with 0.1 % formic acid. Mobile phase buffer B was 95 % ACN/water in 0.1 % formic acid. A flow rate of 4 µL/min of solvent A was used for sample loading and desalting with 6 µL injection volume for 3 min. Following this, the stream-select module was switched to a positive where the enrichment column became inline.
with the analytical nano column. The tryptic peptides were separated using a linear gradient of 5-85 % solvent B over 60 min at a flow rate of 200 nL/min. The mass spectrometer was performed in positive ion mode. The instrument settings were adjusted during autotune and set to: fragmentor voltage 175 V, skimmer 65 V, and the octopole 1 RF voltage 750 V. The temperature and flow rate of drying gas were at 325 °C and 3.5 L/min respectively. The data acquisition was set to auto MS/MS mode. The precursor ion was isolated based on relative abundances. The doubly charged ions were given the first priority followed by singly charged ions, triply charged, and then other multiply charged ions. The collision energy applied was based on the mass-to-charge (m/z) ratio of the ion. The scan range was generally from m/z 500 to m/z 2000. The data analysis was performed with the Agilent MassHunter qualitative analysis software (Agilent Technologies, Santa Clara, CA).

3.2.7 Protein identification and quantification

The peptide sequences were identified using Mascot search. The peak lists that generated using the respective software of each mass spectrometer were searched using an inhouse-licensed Mascot search engine (version 2.2.0, Matrix Sciences, London, UK) against UniProt-Swiss-Prot database for the standard protein analyses. Mascot generic files were generated from raw data using default values in Distiller (version 2.1.1.0, Matrix Science, UK). Carbamidomethyl cysteine and N-terminal dimethyl phosphorylation were set as fixed modifications, while oxidized methionines, and dimethyl phosphorylation of lysine and tyrosine were set as variable modifications (N-dimethyl phosphorylations were specifically added to the list of modifications on an in-house Mascot server). Trypsin was specified as the proteolytic enzyme with one missing cleavage allowed. The mass
tolerance of the precursor ion and fragment ions were set to 200 ppm and 0.5 Da respectively. The search results that were within the list of significant hit were regarded as identified proteins, and all results were further verified by manual interpretation. For positive identification, the score of the result had to be over the significance threshold (p < 0.05). A minimum of three tryptic peptides confirmed by MS/MS was selected manually in order to quantify specific protein in the sample. The relative quantification of a protein from two different samples was determined by averaging the ratios of the H-DMP- and D-DMP-labeled peptides derived from the same protein.

3.3 Results and discussion

3.3.1 Selection of phosphorus chemistry

Phosphorus as a key biological element plays important roles in biological processes, such as replication and information (DNA and RNA), metabolism (ATP, NADPH, and other coenzymes), and membrane structures (phospholipids). For example, reversible phosphorylation is one of the most common and important regulatory modifications of proteins and serves as a primary mechanism of cellular signal transduction.\[^{84}\] Phosphoryl group has been selected by evolution for biochemical transformations because of several key properties of phosphorus as phosphate, such as its thermodynamic instability coupled to kinetic stability, charge and coordination state, and a constant oxidation state at living system conditions.\[^{85, 86}\] Besides utilizing by lives from the very beginning of the life origins, phosphorus chemistry has also been used by scientists to investigate the life itself. For example, Staudinger ligation of azide and arylphosphine has been applied for the selective formation of an amide bond with complex biological environs for glycan labeling.\[^{87}\] Furthermore, some phosphonium bromides (TMP)
have been widely used for charge derivatization of peptides for sequencing\cite{88} and MALDI-TOF MS tissue imaging.\cite{89} Indeed, molecules containing the ionized phosphoric acid group are more difficult to be detected and analyzed by mass spectrometry because of the low ionization efficiency. However, it was found that the neutral phosphoryl group intrinsically has a very high proton affinity leading to exhibit good ionization efficiency for amino acids analysis during ESI or FAB processes.\cite{90} Under mild reaction conditions, primary amine groups of amino acids or peptides could be modified by an active phosphate chloride intermediate formed \textit{in situ} in basic aqueous solutions through a modified Atherton-Todd reaction (Figure 3.4 B),\cite{91,92} which has been widely used for synthesis of small pharmaceutical compounds.\cite{93-96} This modified Atherton-Todd reaction for peptide coupling has several notable features,\cite{97} such as proceeding rapidly to completion, tagging selectively, and involving simple handling steps, which has the potential to be an ideal global strategy based on phosphorus chemistry for introducing stable-isotope labeled tags for quantitative proteomics (Figure 3.5).

\textbf{Figure 3.5} Stable-isotope N-phosphorylation labeling (SIPL) strategy for protein
quantification based on organic phosphorus chemistry. After digestions of the samples, peptides of different states were separately labeled with either of the dimethyl phosphite isotopomers. Then, the labeled samples were combined, desalted and subsequently analyzed by mass spectrometry. In MS, pairs of stable-isotope labeled peptides could be recognized by the known mass difference between them. The relative quantifications could be performed by comparison of the signal intensity of the differentially peptides.

3.3.2 Stable isotopic N-Phosphorylation labeling of standard peptide and proteins

To examine the labeling efficiency and selectivity of N-phosphorylation labeling, a synthetic peptide Fibrinopeptide A with the sequence ADSGEGDFLAEGGGVR as model was investigated first by nano ESI MS. As shown in Figure 3.6 (A), ions for the native peptide were observed at \( m/z \) 1536.3761 and 768.7190 corresponded to the protonated ion \([M+H]^+\) and doubly charged ion \([M+2H]^{2+}\), respectively. After labeling, an \( m/z \) shift compared with native peptide \((m/z \ 1536.3761)\) of 108.2746 and 114.4987 is found for the H-DMP- (light) and D-DMP- (heavy) labeled peptide ions, respectively. This is consistent with the fact that the model peptide has one site available to be labeled at N-terminal amine of alanine residue.
Figure 3.6 Nano ESI-MS spectra of the standard Fibrinopeptide ADSGEGDFLAEGGGVR. (A) native peptide; (B) N-terminal H-DMP labeling; (C) N-terminal stable-isotope D-DMP labeling.
Moreover, Figure 3.6 (C) shows that [M+H]^+ and [M+H+Na]^{2+} ions of the D-DMP-labeled peptide at m/z 1650.8748 and 836.9596 are 6 and 3 mass units higher relative to the H-DMP-labeled ions at m/z 1644.6570 and 833.8426, respectively, indicating that the mass difference imparted by isotopic labeling (at least 3 or 4 Da mass shift is needed) is enough to minimize quantitative errors resulting from isotopic overlapping. Furthermore, complete labeling was obtained as no peaks at the m/z corresponding to native peptide could be observed in mass spectra of H-DMP- and D-DMP-labeled peptides.

Figure 3.7 MALDI-TOF MS spectra of the tryptic peptides of lysozyme. (A) native; (B) H-DMP labeling; (C) stable-isotope D-DMP labeling. (*, the number of labeled phosphoryl group).
Table 3.1 MALDI-TOF MS data for tryptic peptides of lysozyme.

<table>
<thead>
<tr>
<th>No.</th>
<th>Peptide Position</th>
<th>Peptide sequence</th>
<th>Calc</th>
<th>Expt</th>
<th>N-Phosphoryl labeled peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>H-DMP-labeled</td>
</tr>
<tr>
<td>1</td>
<td>97-112</td>
<td>(K)IYSDGNGMNAWAVWR(N)</td>
<td>1803.89</td>
<td>1803.87</td>
<td>2019.92 (2)</td>
</tr>
<tr>
<td>2</td>
<td>46-61</td>
<td>(R)NTDGSTDYGLIQINSR(W)</td>
<td>1753.83</td>
<td>1753.87</td>
<td>1861.93 (1)</td>
</tr>
<tr>
<td>3</td>
<td>98-112</td>
<td>(K)IYSDGNGMNAWAVWR(N)</td>
<td>1675.80</td>
<td>1675.79</td>
<td>1793.87 (1)</td>
</tr>
<tr>
<td>4</td>
<td>62-73</td>
<td>(R)WNCNDGRTPGSR(N)</td>
<td>1491.65</td>
<td>1491.66</td>
<td>1599.73 (1)</td>
</tr>
<tr>
<td>5</td>
<td>34-45</td>
<td>(K)FESNFNTQATNR(N)</td>
<td>1428.65</td>
<td>1428.70</td>
<td>1536.75 (1)</td>
</tr>
<tr>
<td>6</td>
<td>115-125</td>
<td>(R)CKGTDVQAWIR(G)</td>
<td>1333.72</td>
<td>1333.73</td>
<td>1549.76 (2)</td>
</tr>
<tr>
<td>7</td>
<td>22-33</td>
<td>(R)GYSGLNWVCAK(F)</td>
<td>1325.68</td>
<td>1325.69</td>
<td>1541.74 (2)</td>
</tr>
<tr>
<td>8</td>
<td>115-125</td>
<td>(R)CKGTDVQAWIR(G)</td>
<td>1316.70</td>
<td>1316.68</td>
<td>1424.73 (1)</td>
</tr>
<tr>
<td>9</td>
<td>117-125</td>
<td>(K)GTDVQAWIR(G)</td>
<td>1045.54</td>
<td>1045.59</td>
<td>1153.63 (1)</td>
</tr>
<tr>
<td>10</td>
<td>62-68</td>
<td>(R)WNCNDGR(T)</td>
<td>993.41</td>
<td>993.42</td>
<td>1101.48 (1)</td>
</tr>
<tr>
<td>11</td>
<td>6-14</td>
<td>(R)CELAAAMR(H)</td>
<td>1049.49</td>
<td>-</td>
<td>1265.61 (2)</td>
</tr>
<tr>
<td>12</td>
<td>15-21</td>
<td>(R)HGLDNY(R)</td>
<td>874.41</td>
<td>-</td>
<td>1090.51 (2)</td>
</tr>
<tr>
<td>13</td>
<td>1-5</td>
<td>KVFGR(C)</td>
<td>606.36</td>
<td>-</td>
<td>822.50 (2)</td>
</tr>
</tbody>
</table>

a Start and end position of the peptides in the protein sequence;

b Peptide sequence identified by peptide mass mapping using Mascot;

c Value in parentheses is the number of phosphoryl groups labeled for each tryptic peptide observed;

d The N-terminal amino group of cysteine was lose during digestion process.

Figure 3.7 shows the MALDI-TOF mass spectra of the native tryptic peptides as well as H-DMP- and D-DMP-labeled digest of standard lysozyme. In addition, the assigned peptide sequences were summarized in Table 3.1 based on database search. As shown in Figure 3.7 and Table 3.1, thirteen peptides were observed both for H-DMP- and D-DMP-labeling as compared to ten native peptides detected without N-phosphorylation in the same mass range from m/z 700 to 2500. It is worth noting that all of the signals are protonated species and no metal ion adducts are found. Three additional tryptic peptides are observed at m/z 822.5003, 1090.5111, and 1265.6172 for H-DMP-labeling and at m/z 834.5506, 1102.5773, and 1277.6785 for D-DMP-labeling, respectively, indicating that there are two
dimethyl phosphoryl group attached to each peptide because of the 12 Da mass shifts between H-DMP- and D-DMP-labeling. Indeed, the newly observed peaks are determined by nano ESI-MS/MS, and the sequences of the peptides are KVFGR (13), HGLDNYR (12), and CELAAAMKR (11), respectively (Figure 3.8).
Figure 3.8 Identification of peptides KVFG, HGLDN, and CELAAAMKR derived from lysozyme by nano ESI-MS/MS. (A) H-DMP-labeled KVFG (1+, m/z 822.3691); (B) H-DMP-labeled HGLDN (2+, m/z 545.7138); (C) H-DMP-labeled CELAAAMKR (2+, m/z 633.2594).
Compared to the peptide sequence coverage found from the unlabeled peptide digest, the N-phosphorylation labeling can increase the coverage by 15% from 61% to 76% searched by Mascot. The N-phosphorylation labeling of several other test proteins with dimethyl phosphite shows that the labeling reaction is highly efficient. For example, the sequence coverage of BSA with the higher molecular weight than lysozyme can also be increased by 13% from 41% to 54% (Figure 3.9 and 3.10).

**Figure 3.9** MALDI-TOF MS analysis of the tryptic peptides derived from BSA. (A) MALDI-TOF MS spectrum of native BSA digest; (B) sequence coverage searched using Mascot; (C) MALDI-TOF MS spectrum of H-DMP-labeled BSA digest; (D) sequence coverage searched using Mascot after H-DMP-labeling.
Figure 3.10 Identification of the deaminated peptide CKGTDVQAWIR (2+, \( m/z \) 712.3379) derived from lysozyme by nano ESI-MS/MS. The side chain amino group of lysine was labeled with a dimethyl phosphate group. The inset was the possible fragmentation pathways of b2 ion for the formation of fragment ions at \( m/z \) 255.1049 and 192.0745.

It is worth noting that all N-terminal primary amine can be labeled quantitatively with H-DMP-H or D-DMP-D reagents in 40 min by a one-pot reaction under mild reaction conditions. In most cases, some interfering components, such as the primary amine-containing molecules, needed to be removed by desalting the samples before labeling or performing the digestion in buffers without primary amines. However, under our conditions, the ammonium bicarbonate remaining in the reaction mixture after in-solution protein digestion did not affect the labeling efficiency. The peak of the unmodified peptide...
disappeared completely in both spectra of labeled lysozyme and BSA digests. Furthermore, the introduction of neutral phosphoryl group with high proton affinity could not only provide improved ionization efficiency and concomitant signal enhancement but also increase its retention capability on the reverse-phase C_{18} column. For example, the relative intensity of [M+H]^+ ion of H-DMP-labeled Fibrinopeptide A peptide is increased about 50 times than that of unlabeled peptide in nano ESI-MS under same conditions. For MALDI-TOF MS analysis, the signal intensities of most of the labeled small peptides are comparable or slightly enhanced as compared to those of the native peptides. Tryptic peptide KVFGR derived from lysozyme can not be detected without phosphorylation labeling. However, after incorporations of two phosphoryl groups at N-terminus and side chain amino group of lysine, very intensive peak corresponded to KVFGR was observed. This is mainly due to two possible reasons. First, the labeled peptides are easier to be trapped on desalting C_{18} column (SPE using ZipTips) than native peptides because two primary amine groups that could form the salt under acidic desalting conditions have been protected by phosphoryl groups, which can increase the retention capacity of the peptides. Second, two phosphoryl groups can significantly improve the ionization efficiency during the ESI process because of its intrinsic high surface affinity and proton affinity. In addition, the hydroxyl group of tyrosine could also be derivatized dependently on the peptide structure as a side reaction, which might increase the complexity of the peptide mixtures but provide some useful information for peptide identification and improve intensity. Overall, N-phosphorylation labeling based on phosphorus chemistry with many notable features, such as high efficiency, quantitative yield, one-pot and one-step reaction, minimal handling steps, and
enhanced intensity, makes an excellent performance in derivatization of complex peptide samples.

3.3.3 Peptide de novo sequencing

In our preliminary studies, several unique advantages of N-phosphorylation labeling for *de novo* peptide sequencing are noted and demonstrated using some small synthetic peptides with different structures. Specifically, for singly charged peptides, N-phosphorylation labeling approach can significantly enhance a and b ions series that are usually missing in MS/MS spectra and simplify the MS/MS fragmentation patterns, thus providing more confident sequence assignment and facilitating peptide identifications. Furthermore, it can be used to effectively distinguish the isobaric amino acid residues glutamine (Q, 128.0585 Da) versus lysine (K, 128.0949 Da) and enable differentiation of overlapping masses with one amino acid from the combination of two amino acids, such as glycine and alanine (G+A, 128.0585 Da). Although N-phosphorylation labeling has a great promise for *de novo* sequencing, it is needed to be further validated using complex tryptic peptide mixtures of protein samples.

Unlike the synthetic model peptides being singly charged in ESI-MS, the peptides derived from protein digests usually have some common features, such as containing basic residue either arginine or lysine residues at their C-terminus for tryptic peptides and being multiply charged easily, which might have a great effect on the fragmentation pathways of N-phosphoryl labeled peptides.
Figure 3.11 Nano ESI-MS/MS of tryptic peptide CKGTDVQAWIR derived from lysozyme. (A) MS/MS spectrum of native peptide (2+, m/z 667.3344); (B) MS/MS spectrum of H-DMP-labeled peptide (2+, m/z 775.3282).
As an example, Figure 3.11 shows MS/MS sequencing of tryptic peptide CKGTDVQAIR derived from lysozyme by Nano ESI-MS. As shown in Figure 3.11 (A), the native peptide displays intensive y ions and complex fragment ions with low abundance in the mass range of m/z 100 to 700. In contrast, the incorporation of phosphoryl group results in much cleaner fragmentation spectrum as shown in Figure 3.11 (B). The b ion series can be greatly enhanced and provide complete sequence coverage despite the presence of a C-terminal arginine that strongly favors the formation of y ions. It is also worth noting that most of the small fragments in the low mass range are disappeared as well as two intense ions, y₁-H₂O-CO ion at m/z 129.0993 and y₂-NH₃ ion at m/z 271.1187 in Figure 3.11 (A), were not observed in Figure 3.11 (B) after N-phosphorylation labeling. Furthermore, a mass shift of 216 Da was observed for b ions but not for y ions, which is consistent the fact that two dimethyl phosphoryl groups are attached to the N-terminus and ε-NH₂ of lysine residue respectively. Interestingly, for MS/MS spectrum of the deaminated peptide CKGTDVQAIR (2⁺, m/z 712.3379), the b ions were not enhanced because of the loss of N-terminal amine group of cysteine residue resulting in only one phosphoryl group attached at ε-NH₂ of lysine residue, which suggests that the N-terminal phosphoryl group plays an essential role on the fragmentation pathways of peptides (Figure 3.12).
Indeed, the b ions can also be enhanced even through the sites of phosphoryl group labeling are at the C-terminus of the peptides, such as H-DMP-labeled HGLDNYR (2+, m/z 547.7138) and CELAAAMKR (2+, m/z 633.2594) (Figure 3.8 (B)-(C)). In contrast to the doubly charged ions, the y ions were missing or weak in the MS/MS of singly charged KVFGGR peptide while the yields of a and b ions are greatly increased (Figure 3.8 (A)).

Major advantages of N-phosphorylation labeling are the great enhancement of b ions and reduces the complexity of the MS/MS fragmentation patterns of peptides, particularly for singly charged peptides, which is very helpful for interpreting data from spectrum with completely unknown amino acid sequence. The incorporation of neutral phosphoryl group can increase the relative proton affinities (PA) of the modified N-terminus, which determine the relative abundance of b and y ions formed through a charge direct fragmentation mechanism. According to the ‘pathways in competition model’, the N-terminal phosphoryl group can compete with other basic sites, such as arginine, histidine or lysine, to localize the charges favorably. For singly charged peptides in which the charge is mainly retained at the N-terminal phosphoryl group, the C-terminal fragments will be neutral. Thus, a high relative abundance of b ions are formed since the y-type cleavage is suppressed. On the other hand, for multiply charged peptides, y ions will be formed because of the C-terminal fragments will be charged due to the remaining charge at the C-terminal basic residues in the sequence. Overall, by using N-phosphorylation labeling method discussed above, the entire sequence can be easily and rapidly extrapolated from the spectra, suggesting that this method as a useful tool is likely to have a greater degree of success for the de novo peptide sequencing.
3.3.4 Quantification of standard peptide and protein

In order to investigate the feasibility of N-phosphorylation labeling for quantitative analysis, peptide Fibrinopeptide A mentioned above was investigated first as model using LC-ESI-TOF MS (Figure 3.12).

![Figure 3.12](image)

**Figure 3.12** (A) Extract ion chromatogram for a linearity assay of H-DMP- and stable-isotope D-DMP-labeled Fibrinopeptide A standard peptide ADSGEGDFLAEGGGVR \( (2^+, \ m/z \ 822.82 \text{ and } m/z \ 825.83) \) under various ratios (1:0.5, 1:1, 1:2, 1:3, 1:4, 1:5, 1:10, and 1:20). (B) The linearity plot of the H-DMP- and D-DMP-labeled peptide ADSGEGDFLAEGGGVR. The ratios are plotted by their expected values on the x-axis and their measured ratios on the y-axis. A red diagonal line indicates the ideal distribution. Every value was analyzed in triplicates. The error bars denote the minimum and maximum range of the data acquired from three individual experiments.
Two solutions of the peptide standard with different concentration ratios within a linear range from 1:0.5 to 1:20 were labeled using H-DMP-H and D-DMP-D, respectively, and then they were mixed and analyzed by LC-MS. Every value was analyzed in triplicates. One challenge associated with stable-isotope labeling is the presence of peptides overlapping that may contribute to quantitative errors. Although peptide ADSGEGDFLAEGGGVR is doubly charged with one labeled site at N-terminus (6 mass units for each labeling), the isotopic pair can be completely resolved by MS since the mass difference imparted by the stable isotope atoms is 3 Da between the light- (2+, m/z 822.82) and heavy- (2+, m/z 825.83) labeled isotopic pair. As shown in Figure 3.12, the average values of the three experimentally acquired ratios were plotted against their expected ratios and shows that the N-phosphorylation labeling method was linear across a 20-fold range of concentration ratios from 0.5 to 20 with an R² value of 0.9984, indicating that this quantification approach is accurate. Furthermore, the relative standard derivations obtained from three repeated measurements are in the range of 2.5 % to 9.4 %, implying that this method is reproducible and precise for peptide quantification.

Figure 3.13 shows the relative quantification of lysozyme as model protein under various ratios. As described previously, two versions of sample containing different ratios (1:0.5, 1:1, 1:2, 1:3, 1:4, 1:5, and 1:10) of the tryptic peptides are labeled with light- and heavy-labeled phosphorus reagents, respectively, and then mixed together and subsequently analyzed by LC-ESI-MS. As shown in Figure 3.13, two tryptic peptides, namely FESNFNTQATNR (2+, one labeling, 3 Da of mass shift) and WWCNDGR (1+, one labeling, 6 Da of mass shift), are selected manually to estimate the linearity of this method. The measured ratios of isotopic
pairs are consistent with the known ratios of the two labeling versions, resulting in a linear dynamic range with $R^2$ values of 0.9951 and 0.9919, respectively. Rationally, it is believed that the higher confidence level and accuracy may be obtained if more tryptic peptide pairs are selected for quantification.

Figure 3.13 Relative quantification of standard protein lysozyme under various ratios using LC-MS. (A) tryptic peptide FESNFNTQATNR labeled with H-DMP ($2^+$, m/z 768.79) and D-DMP ($2^+$, m/z 771.81); (B) tryptic peptide WWCNDGR labeled with H-DMP ($1^+$, m/z 1101.34) and D-DMP ($1^+$, m/z 1107.37); (C) The linearity plot of H-DMP- and D-DMP-labeled peptide FESNFNTQATNR; (D) The linearity plot of H-DMP- and D-DMP-labeled peptide WWCNDGR. The ratios are plotted by their expected values on the x-axis and their measured ratios on the y-axis. The red and blue diagonal lines indicate the ideal distributions of the corresponded peptides respectively.
3.3.5 Quantification of protein mixtures by nano LC-chip/TOF MS

To test the usefulness of stable-isotope N-phosphorylation labeling for relative quantification in a relatively complex sample, the relative abundances of six purified proteins in two samples were analyzed using full-scan MS and the sequencing for protein identification was based on data-dependent LC-MS/MS on a nano LC-chip/TOF MS system and a database search. Two tryptic digests of six standard proteins were combined in different amounts and different ratios to give two samples, in which the protein abundance ratios (H-DMP- and D-DMP-labeling) between the two mixtures were 1, 2, 10, 4, 0.5, and 1 for BSA, β-casein, myoglobin, β-lactoglobulin, lysozyme, and hemoglobin, respectively. Three peptide pairs, including double and triply charged ions, were selected manually to calculate peptide ratios for each protein (Figure 3.14).
(C) Hemoglobin (ratio of H+/D+, 1:10)

(D) β-Lactoglobulin B (ratio of H+/D+, 1:4)
Lysozyme (ratio of H-/D-, 1:0.5)

Myoglobin (ratio of H-/D-, 1:1)

Figure 3.14 Identification and quantification of protein mixes by nano LC-chip/TOF MS and MS/MS. The ratios of all isotopic pairs were calculated from the extracted peak area in the LC-MS chromatograms and averaged.
Table 3.2 Quantification of the protein mixture under various ratios by nano LC-chip/TOF MS.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Peptide position</th>
<th>Peptide sequence</th>
<th>Charge state</th>
<th>No. of labeling</th>
<th>Ratio</th>
<th>Mean ±SD</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSA</td>
<td>205-209</td>
<td>IETMR</td>
<td>1</td>
<td>1</td>
<td>1.00</td>
<td>1.10</td>
<td>1.01 ±0.10</td>
</tr>
<tr>
<td></td>
<td>360-371</td>
<td>RHPEYA VSVLLR</td>
<td>3</td>
<td>1</td>
<td>1.00</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>469-482</td>
<td>MPCTE DYLSLI LNR</td>
<td>2</td>
<td>2</td>
<td>1.00</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>β-Casein</td>
<td>48-63</td>
<td>FQpSEE QQQTE DELQD K</td>
<td>2</td>
<td>1</td>
<td>2.00</td>
<td>2.04</td>
<td>1.99 ±0.15</td>
</tr>
<tr>
<td></td>
<td>192-198</td>
<td>AVPYPQ R</td>
<td>2</td>
<td>1</td>
<td>2.00</td>
<td>2.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>199-217</td>
<td>DMPiQA FLLYQE PVLGPR</td>
<td>2</td>
<td>1</td>
<td>2.00</td>
<td>1.82</td>
<td></td>
</tr>
<tr>
<td>Hemoglobin</td>
<td>32-41</td>
<td>LLVVYP WTQR</td>
<td>2</td>
<td>2</td>
<td>10.00</td>
<td>9.65</td>
<td>9.57 ±0.78</td>
</tr>
<tr>
<td></td>
<td>97-105</td>
<td>LHVDPE NFR</td>
<td>2</td>
<td>1</td>
<td>10.00</td>
<td>8.74</td>
<td></td>
</tr>
<tr>
<td></td>
<td>134-145</td>
<td>VVAGV ANALA HK</td>
<td>2</td>
<td>1</td>
<td>10.00</td>
<td>10.31</td>
<td></td>
</tr>
<tr>
<td>β-Lactoglobulin</td>
<td>108-117</td>
<td>VLVLDT DYKK ALPMHI R</td>
<td>2</td>
<td>2</td>
<td>4.00</td>
<td>4.12</td>
<td>3.73 ±0.33</td>
</tr>
<tr>
<td></td>
<td>158-164</td>
<td>LSFNPT QLEEQC HI</td>
<td>2</td>
<td>1</td>
<td>4.00</td>
<td>3.53</td>
<td></td>
</tr>
<tr>
<td></td>
<td>165-178</td>
<td></td>
<td>2</td>
<td>1</td>
<td>4.00</td>
<td>3.54</td>
<td></td>
</tr>
<tr>
<td>Lysozyme</td>
<td>52-63</td>
<td>FESNFN TQATNR</td>
<td>2</td>
<td>1</td>
<td>0.50</td>
<td>0.44</td>
<td>0.47 ±0.05</td>
</tr>
<tr>
<td></td>
<td>80-86</td>
<td>WWCND GR</td>
<td>2</td>
<td>1</td>
<td>0.50</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td></td>
<td>135-143</td>
<td>GTDVQ AWIR</td>
<td>2</td>
<td>1</td>
<td>0.50</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>Myoglobin</td>
<td>2-17</td>
<td>GLSDGE WQQVL NVWGK</td>
<td>2</td>
<td>1</td>
<td>1.00</td>
<td>1.02</td>
<td>0.96 ±0.05</td>
</tr>
<tr>
<td></td>
<td>18-32</td>
<td>VEA DIA GHQQE VLIR ALELFR</td>
<td>2</td>
<td>1</td>
<td>1.00</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td></td>
<td>135-140</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1.00</td>
<td>0.91</td>
<td></td>
</tr>
</tbody>
</table>
Meanwhile, the means and standard deviations of peptide ratios were used to quantify proteins in the mixture. As shown in Table 3.2, the N-phosphorylation labeling strategy is able to determine the relative abundance ratios of the six target proteins in the sample and shows excellent correlation of the measured ratios with theoretical ratios with errors that range from 0.5 to 6.7 % and relative standard deviation of less than 10.6 %, which are sufficient for many proteomic applications. Additionally, the isotopic mass pair of phosphopeptide FQpSEEQQTEDELQDK (2+, m/z 1085.4145 and 1088.4350) with one phosphoric acid group on the serine residue was selected for the quantification of β-casein as a phosphorylated protein (Figure 3.14 (B)). It was found that the measured ratio (H-DMP-/D-DMP-, 1:2.04) precisely reflected the true ratio (1:2.00) and no side reactions occurred at the phosphoryl group on serine residue during labeling, indicating that this labeling strategy can also be used for analysis the posttranslational protein phosphorylation. In addition, it is worth noting that the average cost of N-phosphorylation labeling for 100 µg of protein typically digested peptide is estimated to be about 8 dollars, whereas the same amount of iTRAQ reagents costs more than two hundred dollars. The overall data demonstrate that differential N-phosphorylation labeling can be applied to quantify tryptic peptides present at different concentration ratios within the biological sample for quantitative proteomics.

3.3.6 Validation of deuterium atom effect of labeled peptides

Generally, deuterium labeling has many attractive features, such as being less expensive, simpler to use in the preparation of coding agents, and allowing smaller coding agents. However, it is known that deuterium atom is slightly more hydrophilic than hydrogen atom, which results a slightly or even completely
separation of the light- and heavy-labeled peptides in reversed phase chromatography. Moreover, the deuterium isotope effect is really an issue with small- to intermediate-sized peptides, which may affect the accuracy of quantification. Fortunately, it has been reported that the structural features plays a key role for controlling the deuterium isotope effects in reversed phase separations, and deuterium effect can be minimized to acceptable values by grouping deuterium atoms around polar functional groups in the labeling reagents. To monitor the elution profiles of both N-phosphorylation labeled peptides, two small peptides KVFGR and HGLDNYR derived from lysozyme were investigated as extreme case, in which twelve deuterium atoms for the heavy labeled peptides were incorporated and located around the phosphoryl groups (Figure 3.15).
Figure 3.15 Comparison of extracted ion chromatograms of peptide pairs labeled with H-DMP or D-DMP. (A) KVFGGR (ratio of H-/D- is 1:1, two phosphoryl groups labeling, singly charged); (B) HGLDNYR (ratio of H-/D- is 1:2, two phosphoryl groups labeling, doubly charged). The up-right insets were the summed spectra across the eluting peaks.
As an example, the extracted chromatograms of the $\text{D}_{12}\text{-DMP}$- and
$\text{H}_{12}\text{-DMP}$-labeled KVFGR ions at $m/z$ 834.4260 and 822.3497 almost coelute
with the comparative signal intensity at retention times of 16.28 and 16.32 min,
respectively, as shown in Figure 3.15 (A). The slightly isotope effect observed
may also affect the measurement accuracy. For this reason, the quantitative ratios
can be determined based on the integration of the entire extracted ion peaks of the
heavy and the light peptides. Indeed, the isotope effect is negligible for other
labeled peptides with higher molecular weight and small number of deuterium
atoms. The detailed mechanism is not clear, but we suspect that two factors may
contribute to the insignificant deuterium effect. The first is that the coding agent
size of dimethyl phosphate group is moderately small, which is considered to be
correlated inversely with the magnitude of the isotope effect.[101] The second is
that the deuterium atoms are placed adjacent to the neutral phosphoryl group with
higher electrophilicity, which may minimize the hydrophilicity difference between
deuterium atom and hydrogen atom under acidic separation conditions, and
therefore reduce the differential interactions with the reversed-phase
chromatography that leads to chromatographic resolution.[102]

3.4 Chapter summary

We have successfully developed a novel stable isotopic N-phosphorylation
labeling method for peptide de novo sequencing and protein quantification based
on organic phosphorus chemistry. The N-phosphorylation labeling reaction was
performed easily within 40 min in a one-pot reaction without additional cleanup
procedures reducing sample handling and thus the potential of experimental errors.
The method was amenable to various types of sample, including cell culture,
biological fluids and tissues. The new approach has several advantages. First, the
labeling reaction is a global labeling with quantitative yield and high selectivity. The primary amine group of N-terminus of peptides may be labeled efficiently. Furthermore, the labeling reagent is more stable than most of the other stable isotopic labeling reagents, in which the tags need to be activated in advance. By changing the reaction conditions, the phosphite may be activated in situ to form a labeling intermediate with high activity for peptide coupling. Second, the N-terminal labeled phosphoryl group with highly relative proton affinity not only improves the ionization efficiency for peptide mass fingerprints, but also greatly enhances the intensities of b ions and reduces the complexity of the MS/MS fragmentation patterns of peptides, which allows more confident peptide de novo sequencing and protein identification. Moreover, N-terminal phosphorylation adds a moderately hydrophobic phosphoryl group that increases retention on stationary phase of reversed-phase chromatography, which can increase the detection capability of short or polar tryptic peptides that might otherwise elute near the void volume and escape detection. Third, the labeling reagent is inexpensive compared to other current labeling reagents and can be easily synthesized by a one-step reaction at high yield with a commercially available isotope reagent. Finally, the N-phosphorylation labeling tags have small molecular weight of 108 Da with simple chemical structures, which may not generate fragments interfering the MS/MS sequencing. Additionally, the isotope effect is insignificant in most cases even though six deuterium atoms will be incorporated for each labeling. The isotope effect can be further minimized by introducing other isotope atoms, such as $^{13}$C, $^{15}$N, and $^{18}$O, into the coding reagents. Further investigations may be conducted to design series of phosphorus reagents with different chemical structures as well as novel functions and to optimize the coupling conditions of
the reactions in order to improve the applicability of the N-phosphorylation labeling for quantitative proteomics.
Chapter 4
Identification and quantification of N-linked glycoproteins of influenza virus infected cells using hydrazide chemistry, stable isotope labeling and mass spectrometry

4.1 Introduction

Influenza viruses cause yearly epidemics with significant morbidity and mortality globally. Influenza A virus is a rapidly evolving single-stranded negative-sense RNA virus \cite{1}. More than one hundred antigenic subtypes, each containing many variants, are maintained in wild animal reservoir populations \cite{2–4}. Certain subtypes have jumped species and established endemic infections among humans, pigs, horses, and other land and sea mammals. The 2009 pandemic H1N1 virus arose from a reassortment event between two swine origin viruses. Additionally, the 2013 H7N9 virus responsible for the outbreak in Mainland China and Taiwan is a reassortant virus deriving gene segments from avian H9N2 and H7N9 viruses.

Host–virus interactions involve protein expression changes within both the host and the virus. An understanding of the nature of these interactions provides insight into metabolic processes and critical regulatory events of the host cell as well as into the mechanisms of pathogenesis by infectious viruses. Virus exposure induces changes in host proteins at many functional levels including cell signaling pathways, protein degradation, cytokines and growth factor production, phagocytosis, apoptosis, and cytoskeletal rearrangement. Since proteins are responsible for the cell biological functions, viruses have evolved to manipulate the host cell proteome to achieve optimal replication. Intracellular viruses can also
change their proteome to adapt to the host cell and escape from immune surveillance, or can incorporate cellular proteins to invade other cells \[103\]. Given that the interactions of intracellular infectious agents with host cells are mainly at the protein level, proteomics is the most suitable tool for investigating these changes of proteins. Proteomics is the systematic analysis of proteins, particularly their interactions, modifications, localization and functions, that permits the study of the association between pathogens with their host cells as well as complex interactions such as the host-vector–pathogen interplay \[11\].

Protein glycosylation has long been recognized as a common posttranslational modification. Glycosylation is the covalent linkage of an oligosaccharide side chain to a protein. Typically, carbohydrates are linked to serine or threonine residues (O-linked glycosylation) or to asparagine residues (N-linked glycosylation). N-linked glycosylation sites generally fall into the N-X-S/T sequence motif in which X denotes any amino acid except proline. Protein glycosylation, in particular N-linked glycosylation, is prevalent in proteins destined for extracellular environments. These glycoproteins include the some proteins on the extracellular side of the plasma membrane, secreted proteins and proteins contained in body fluids (such as blood serum, cerebrospinal fluid, urine, breast milk, saliva, lung lavage fluid or pancreatic juice). These glycoproteins also happen to be the proteins in the human body that are most easily accessible for diagnostic and therapeutic purposes. It is therefore no surprise that many clinical biomarkers and therapeutic targets are glycoproteins. Eukaryote plasma membrane proteins are often heavily posttranslationally modified in the extracellular domains (e.g. N-glycosylation). Because of their extensive hydrophobic regions, integral membrane proteins present a special challenge to
proteome analysis because they are insoluble under standard aqueous conditions. Several promising MS-based proteomic strategies for the identification and characterization of membrane proteins have been reported. These approaches do not need to separate the membrane protein first [104-109].

There are many glycoprotein enrichment methods. Lectins are carbohydrate binding proteins that recognize specific carbohydrate structures, and thus are used to enrich for glycoproteins and glycopeptides. Lectin-based affinity enrichment of glycopeptides combine with glycosidase-catalyzed 18-O stable isotope labeling and MS/MS permits isolation, detection and sequencing of N-glycosylated peptides. Hydrolytic incorporation of 18-O at the glycosylation site facilitates MS- and MS/MS-based assignment of glycopeptides and exact determination of N-glycosylation sites. This method was applied to investigate 400 N-glycosylation sites in 250 glycoproteins in a caenorhabditis elegans protein extract. Recently, a new method was used in relative quantification of proteins. In this method, N-phosphorylation labeling reaction was performed within 40 min in a one-pot reaction without additional cleanup procedures, which reduced sample handling and thus the potential experimental errors. This new approach has some advantages. The reaction is a global labeling with quantitative yield and high selectivity. The primary amine group of the N-terminus of peptides may be labeled efficiently. The labeling reagent is more stable than most of the other stable isotopic labeling reagents, in which the tags need to be activated in advance [110-114].

In this study, methods with hydrazide chemistry, stable isotope N-phosphorylation labeling and mass spectrometry analysis [115] were applied to study the glycoproteins of H1N1 (A/Puerto rico/8/1934) infected cell line (A549).
After mass spectrometry determination, glycoproteins were identified with database searching. The results indicated that quantity of some glycoproteins was significantly regulated in influenza virus infected cells.

4.2 Materials and methods

4.2.1 Materials

Bovine serum albumin (BSA), Ovalbumine (OVA), ammonium bicarbonate, DL-dithiothreitol (DTT), iodoacetamide (IAA), trifluoroacetic acid (TFA), triethylamine (TEA), formic acid, α-cyano-4-hydroxycinnamic acid (CHCA), and tetrachloromethane were purchased from Sigma (St. Louis, MO, U.S.A.). CD3OD (99.9 atom % D) and Dimethyl phosphite (H-DMP) were obtained from Alfa Aesar Chemical Ltd. (Tianjin, China). Deuterium-labeled dimethyl phosphite (D-DMP) was synthesized using modified Michaelis–Arbuzov reaction according to a reported method. HPLC grade acetonitrile (ACN) and methanol were purchased from Tedia (Fairfield, OH, U.S.A.). Deionized water (18 MΩ) was produced by a Milli-Q system (Millipore, Bedford, MA, U.S.A.). Unless specified otherwise, all chemicals and solvents were analytical reagents.

4.2.2 Preparation of standard proteins

Standard proteins, BSA and OVA, were used in method development. OVA is a protein with only one glycosylated site and BSA has non glycosylated site. Same quantities of these two proteins were solved in water to make the testing sample.

4.2.3 Virus and cell line culture

Human lung adenocarcinoma epithelial cells (A549) were cultured in Ham’s F12K medium with 2 mM L-glutamine adjusted to contain 1.5 g/L sodium bicarbonate containing 10% FBS, 200 U/mL penicillin G and 200 μg/mL streptomycin. The cells were cultured in Petri dishes (145 × 20 mm). The
influenza virus strain was inoculated onto the confluent monolayer of the A549 cell line at a multiplicity of infection of approximately 0.1. After the 24 h cell culture, the infected cells were collected for use. Three parallel experiments were performed simultaneously.

4.2.4 Separation of N-glycosylation proteins

Cells’ proteins were dissolved by lysis buffer (8 M urea, 4 % w/v CHAPS, 50 mM DTT, and 0.2 % w/v Bio-Lyte at pH 3–10). CarboLink™ Immobilization Kit (Thermo Scientific) was used to separate the glycoproteins with other proteins. The glycoprotein coupled with resin was followed by the protocol of the Kit. After that, coupled glycoproteins were digested with trypsin. The peptides were stored in -4 °C. The workflow was showed in Figure 4.1.
Figure 4.1 Work flow of glycoprotein separation and quantification.
4.2.5 N-phosphorylation of tryptic peptides

The digested peptides (2 µL) were dissolved in a mixture of H₂O (10 µL), ethanol (5 µL), and triethylamine (5 µL) in a 0.2 mL PCR tube. The sample was gently vortexed and spun to the bottom of the tube. The reaction mixture was cooled to 0°C in an ice-water bath. Then, H-DMP-H or D-DMP-D (1 µL) in tetrachloromethane (5 µL) was added into the above reaction mixture, which was mixed periodically. The reaction was allowed to proceed 40 min at room temperature before evaporating the mixtures to dryness under N₂ gas. The residues were dissolved in 50 L H₂O (0.1 % formic acid) and desalted using ZipTip tips (Millipore, Billerica, MA, USA) according to the manufacture’s instructions.

4.2.6 Mass spectrometry analysis

MALDI-TOF MS analysis was performed on a Bruker Autoflex II mass spectrometer (Bruker Daltonics, Bremen, Germany) in positive reflection mode. The mass spectrometer was equipped with a pulsed nitrogen laser operated at 337 nm with 3 ns duration pulses and employed stainless steel targets (MTP 384 target ground steel, Bruker Daltonics). Voltage impressed on the ion one and two was 20.0 and 19.0 kV, respectively. The laser power energy was adjusted as needed. The acceleration voltage, grid voltage, and delayed extraction time were set as 19 kV, 90 %, and 150 ns, respectively. Each mass spectrum was acquired as an average of 500 laser shots at 10.0 Hz frequency. The protein digested after desalting were mixed with matrix solution (20 mg CHCA in 1 mL 30 % ACN/water with 0.5 % TFA) and loaded on the AnchorChip sample plate. After auto-drying, the sample mixture was analyzed in the reflection ion mode and peptide mass fingerprints of the protein were obtained for protein identification.
4.2.7 Nano LC-chip Q-TOF MS analysis

Quantification of the peptide mixtures were carried out with an Agilent 6200 HPLC-chip/TOF MS system (Agilent Technologies, Santa Clara, CA), which equipped with an Agilent 6210 TOF MS as detection, Agilent 1100 nanopump as analytical pump for sample separation, a capillary pump as the loading pump for sample enrichment and desalting, a microwell-plate autosampler maintained at 6 °C by the thermostat, and HPLC-Chip cube as interface. Chromatographic separations were performed on an enrichment column with a volume of 40 nL and an analytical reverse-phase column (Zorbax C18, 5 m, 43 mm x 75 m). Mobile phase buffer A was 5.0 % ACN/water (v/v) with 0.1 % formic acid. Mobile phase buffer B was 95 % ACN/water in 0.1 % formic acid. A flow rate of 4 μL/min of solvent A was used for sample loading and desalting with 6 μL injection volume for 3 min. Following this, the stream-select module was switched to a positive where the enrichment column became inline with the analytical nano column. The tryptic peptides were separated using a linear gradient of 5-85 % solvent B over 60 min at a flow rate of 200 nL/min. The mass spectrometer was performed in positive ion mode. The instrument settings were adjusted during autotune and set to: fragmentor voltage 175 V, skimmer 65 V, and the octopole 1 RF voltage 750 V. The temperature and flow rate of drying gas were at 325 °C and 3.5 L/min respectively. The data acquisition was set to auto MS/MS mode. The precursor ion was isolated based on relative abundances. The doubly charged ions were given the first priority followed by singly charged ions, triply charged, and then other multiply charged ions. The collision energy applied was based on the mass-to-charge (m/z) ratio of the ion. The scan range was generally from m/z 500 to m/z 2000. The data analysis was performed with the Agilent MassHunter
qualitative analysis software (Agilent Technologies, Santa Clara, CA).

4.2.8 Protein identification, quantification and bio-information analysis

The peptide sequences were identified using Mascot search. The peak lists that generated using the respective software of each mass spectrometer were searched using an inhouse-licensed Mascot search engine (version 2.2.0, Matrix Sciences, London, UK) against UniProt-Swiss-Prot database for the standard protein analyses. Mascot generic files were generated from raw data using default values in Distiller (version 2.1.1.0, Matrix Science, UK). Carbamidomethyl cysteine and N-terminal dimethyl phosphorylation were set as fixed modifications, while oxidized methionines, and dimethyl phosphorylation of lysine and tyrosine were set as variable modifications (N-dimethyl phosphorylations were specifically added to the list of modifications on an in-house Mascot server). Trypsin was specified as the proteolytic enzyme with one missing cleavage allowed. The mass tolerance of the precursor ion and fragment ions were set to 200 ppm and 0.5 Da respectively. The search results that were within the list of significant hit regarded as identified proteins, and all results were further verified by manual interpretation. For positive identification, the score of the result had to be over the significance threshold (p < 0.05). A minimum of two tryptic peptides confirmed by MS/MS was selected manually in order to quantify specific protein in the sample. The relative quantification of a protein from two different samples was determined by averaging the ratios of the H-DMP- and D-DMP-labeled peptides derived from the same protein. Mascot distiller was also used to analyze the data. Each scan result of peptide included data of charge, mass, area, and S/N.

Protein relationship was analyzed by using software STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) version 9.0 (http://string.embl.de/).
4.3 Results and discussion

4.3.1 Analysis of standard proteins

The data in Figure 4.2 showed that OVA was separated from mixture standard proteins successfully by CarboLink™ Immobilization Kit. Figure 4.2A was MALDI-TOF MS result of the digested peptides of the OVA; B was the MS result of mixture standard proteins (OVA and BSA) after separation and digestion; C showed MS result of the digested peptides of the BSA. BSA does not contain glycosylation site and OVA only has one glycosylation site. The comparison of peptide peaks of A, B and C in the Figure 4.2 showed the detection of the peptides coupling on the beads that were peptides derived from OVA, indicating that OVA was coupled on the beads successfully. Immobilizing glycoproteins through oxidized sugar groups was an efficient way to separate glycoprotein from non-glycosylated proteins.
Figure 4.2 MALDI-TOF MS result of individual and mixture standard proteins. A. digested peptides of OVA. B. peptides of mixture standard proteins (OVA and BSA) after separation and digestion. C. digested peptides of BSA

Glycoprotein of cells was coupled to the CarboLink™ Immobilization Kit’s beads. Quantification of these glycoproteins by identifying their proteolyses peptides was performed. The glycopeptides on the beads were also able to be released by other enzyme such as peptide-N-glycosidase F (PNGase F).

4.3.2 Cellular glycoprotein separation and quantification

Glycoproteins of H1N1 infected cell line (A549) were separated by CarboLink™ Immobilization Kit. After successful separation, stable isotope N-phosphorylation labeling (SIPL) approach was used for peptide labeling and protein quantification. N-phosphorylation labeling reagents were activated in situ to form labeling intermediates with high reactivity targeting on N-terminus and ε-amino groups of lysine under mild reaction conditions. Nano liquid chromatography chip/time-of-flight mass spectrometry (nano LC-chip/TOF MS)
was performed to detect these labeled peptides. The results showed average number of peptide was 12727. Nine significant differential proteins were identified after quantification (Table 4.1).

Table 4.1 Identification of the significantly regulated glycoprotein by database searching.

<table>
<thead>
<tr>
<th>NO.</th>
<th>Swissprot ID</th>
<th>Mass Weight</th>
<th>Regulated</th>
<th>Matched sequences</th>
<th>STRING ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RPC1_HUMAN</td>
<td>157644</td>
<td>down</td>
<td>R.TVISPDPNLRI.K</td>
<td>POLR3A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>K.SPEEMR.Q</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>RHG25_HUMAN</td>
<td>72955</td>
<td>down</td>
<td>R.RTMSQDLR.Q.K</td>
<td>ARHGAP25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MGWLKKMGWLKK.Q</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>RPTOR_HUMAN</td>
<td>150653</td>
<td>down</td>
<td>R.RMALSECR.V.R</td>
<td>RPTOR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VYDRRVYDRR.M</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>TITIN_HUMAN</td>
<td>3843119</td>
<td>up</td>
<td>K.AEGILTD.R.K</td>
<td>TTN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>K.AVVQIEISQK.A.K</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>K.IEGDLRAMLK.R.K</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R.KPVPR.K</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>ARHGCF_HUMAN</td>
<td>174094</td>
<td>down</td>
<td>R.AGVQTGDRIIKR.V</td>
<td>ARHGEF12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R.MAASVMAASV.K.E</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>DESP_HUMAN</td>
<td>334023</td>
<td>up</td>
<td>R.QVQLNLVNSK.K.R</td>
<td>DSP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SEDDLRSEDLLR.Q</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ROCK1_HUMAN</td>
<td>159102</td>
<td>down</td>
<td>R.MKAEDYEVVK.R</td>
<td>ROCK1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R.ÉKAENREKAENR.V</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>PTN13_HUMAN</td>
<td>278643</td>
<td>up</td>
<td>K.QESSSVNTSNMNT.F</td>
<td>PTN13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>K.GSLGFTVTKSLGFTVTK_G</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>DOCK3_HUMAN</td>
<td>235006</td>
<td>down</td>
<td>K.AFPDRMDVLGR.R.K</td>
<td>DOCK3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GPKDGPKDK.E</td>
<td></td>
</tr>
</tbody>
</table>
These differential glycoproteins were \textit{RPC1\textsubscript{HUMAN}}, \textit{RHG25\textsubscript{HUMAN}}, \textit{RPTOR\textsubscript{HUMAN}}, \textit{TITIN\textsubscript{HUMAN}}, \textit{ARHGC\textsubscript{HUMAN}}, \textit{DESP\textsubscript{HUMAN}}, \textit{ROCK1\textsubscript{HUMAN}}, \textit{PTN13\textsubscript{HUMAN}}, \textit{DOCK3\textsubscript{HUMAN}}. In these glycoproteins, \textit{RPC1\textsubscript{HUMAN}}, \textit{RHG25\textsubscript{HUMAN}}, \textit{RPTOR\textsubscript{HUMAN}}, \textit{ARHGC\textsubscript{HUMAN}}, \textit{ROCK1\textsubscript{HUMAN}}, \textit{DOCK3\textsubscript{HUMAN}} were down-regulated. Other proteins such as \textit{TITIN\textsubscript{HUMAN}}, \textit{DESP\textsubscript{HUMAN}}, \textit{PTN13\textsubscript{HUMAN}} were up-regulated. \textit{RPC1\textsubscript{HUMAN}} is DNA-directed RNA polymerase III subunit RPC1. \textit{RHG25\textsubscript{HUMAN}} is Rho GTPase-activating protein 25. \textit{RPTOR\textsubscript{HUMAN}} is regulatory-associated protein of mTOR. \textit{ARHGC\textsubscript{HUMAN}} is Rho guanine nucleotide exchange factor 12. \textit{ROCK1\textsubscript{HUMAN}} is Rho-associated protein kinase 1. \textit{DOCK3\textsubscript{HUMAN}} is dedicator of cytokinesis protein 3. \textit{RPC1\textsubscript{HUMAN}} is \textit{POLR3A} - polymerase (RNA) III (DNA directed) polypeptide A, 155kDa; DNA-dependent RNA polymerase catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates. Largest and catalytic core component of RNA polymerase III that synthesizes small RNAs, such as 5S rRNA and tRNAs. It forms the polymerase active center together with the second largest subunit. A single stranded DNA template strand of the promoter is positioned within the central active site cleft of Pol III. \textit{RHG25\textsubscript{HUMAN}} is Rho GTPase activating protein 25; GTPase activator for the Rho-type GTPases by converting them to an inactive GDP-bound state (by similarity). \textit{RPTOR\textsubscript{HUMAN}} is regulatory associated protein of MTOR, complex 1; involved in the control of the mammalian target of rapamycin complex 1 (mTORC1) activity that regulates cell growth and survival, and autophagy in response to nutrient and hormonal signals; functions as a scaffold for recruiting mTORC1 substrates. mTORC1 is activated in response to growth factors or
amino-acids. Amino-acid-signaling to mTORC1 is mediated by Rag GTPases, which cause amino-acid-induced relocalization of mTOR within the endomembrane system. ARHGC_HUMAN is Rho guanine nucleotide exchange factor (GEF) 12. It plays a role in the regulation of RhoA GTPase by guanine nucleotide-binding alpha-12 (GNA12) and alpha-13 (GNA13). Acts as guanine nucleotide exchange factor (GEF) for RhoA GTPase and may act as GTPase-activating protein (GAP) for GNA12 and GNA13. DESP_HUMAN is desmoplakin, major high molecular weight protein of desmosomes. It is involved in the organization of the desmosomal cadherin-plakoglobin complexes into discrete plasma membrane domains and in the anchoring of intermediate filaments to the desmosomes. ROCK1_HUMAN is ROCK1-Rho-associated, coiled-coil containing protein kinase 1, phosphorylates and activates DAPK3, which then regulates myosin light chain phosphatase through phosphorylation of MYPT1 thereby regulating the assembly of the actin cytoskeleton, cell migration, invasiveness of tumor cells, smooth muscle contraction and neurite outgrowth. It is required for centromere positioning and centromere-dependent exit from mitosis and necessary for apoptotic membrane blebbing. PTN13_HUMAN is protein tyrosine phosphatase, non-receptor type 13 (APO-1/CD95 (Fas)-associated phosphatase). It regulates negatively FAS-induced apoptosis and NGFR-mediated pro-apoptotic signaling. DOCK3_HUMAN is dedicator of cytokinesis 3, potential guanine nucleotide exchange factor (GEF). GEF proteins activate some small GTPases by exchanging bound GDP for free GTP. Its interaction with presenilin proteins as well as its ability to stimulate Tau/MAPT phosphorylation suggests that it may be involved in Alzheimer disease. Ectopic expression in nerve cells decreases the secretion of beta-amyloid APBA1 protein and lowers the rate of
cell-substratum adhesion, suggesting that it may affect the function of some small GTPase involved in the regulation of actin cytoskeleton or cell adhesion receptors.

Some reports showed that there are large differences in the sensitivities of influenza A virus strains for inhibition of the IFN-induced Mx-GTPase proteins. The avian influenza virus strains are very sensitive, whereas strains of human origin are more resistant \[116\]. And this protein accumulates in distinct nuclear domains and inhibits influenza A virus in cells that lack promyelocytic leukaemia protein nuclear bodies \[117\]. Rho-associated, coiled-coil-containing protein kinase 1 (ROCK) is recruited as downstream mediators of the HSV-1-induced cofilin inactivation pathway. Inhibitors specific for these kinases significantly reduce the virus infectivity without affecting virus binding to the target cells \[118\]. Inhibition of Rho-ROCK signaling also resulted in a perturbation of the cell arrangement and a rounding of plaques. These opposing effects of Rho and Rac pathways in MDV cell-to-cell spread were validated for two parental MDV recombinant viruses with different ex vivo spread efficiencies \[119\]. Respiratory syncytial virus reduced protein tyrosine phosphates expression and activity \[120\].

### 4.3.3 Relationship of the different proteins

Protein-protein relationship of the detected proteins was generated by using software STRING. The relationship map of all differential proteins was shown in Figure 4.3. In Figure 4.3, RPTOR_HUMAN, ROCK1_HUMAN (Number 3 and 7 in Table 4.1) were found to have close relationship in these proteins.
Figure 4.3 Relationship of the 9 differential proteins by STRING software. POLR3A is RPC1_HUMAN, ARHGAP25 is RHG25_HUMAN, RPTOR is RPTOR_HUMAN, TTN is TITIN_HUMAN, ARHGEF12 is ARHGC_HUMAN, DSP is DESP_HUMAN, ROCK1 is ROCK1_HUMAN, PTPN13 is PTN13_HUMAN, DOCK3 is DOCK3_HUMAN. MTOR, TCAP, EIF4EBP1, RPS6KB1 and JUP are predicted functional partners.
4.4 Chapter summary

Hydrazide chemistry, stable isotope labeling and mass spectrometry methods were applied to investigate the glycoproteins in H1N1 (A/Purto Rico/8/1934) infected cell line (A549). The experiment was designed and performed with the conjugation of glycoproteins to a solid support using hydrazide chemistry, stable isotope labeling of glycopeptides and the specific releasing of non-linked glycosylated peptides via enzyme of trypsin continuously. The analyses combined with database searching were applied for the protein identification. The result showed that some glycoproteins were significant in influenza virus infected cells. In these glycoproteins, RPC1_HUMAN, RHG25_HUMAN, RPTOR_HUMAN, ARHG1_HUMAN, ROCK1_HUMAN, DOCK3_HUMAN were down-regulated. The proteins TITIN_HUMAN, DESP_HUMAN, PTN13_HUMAN were up-regulated.
Chapter 5
Proteomics study of N-acetylcysteine response in H1N1-infected cells by using mass spectrometry

5.1 Introduction

Influenza occurs with seasonal alterations and reaches peak prevalence in winter. A typical example of 2009 pandemic influenza A virus (H1N1) broke out from Mexico and the United States \[121-123\]. In recent years, significant efforts have been made to investigate interaction mechanisms of influenza virus and host cells. Numerous studies have focused either on biomarker research, pathogenesis in humans, antiviral therapies and discovery of novel drug targets, or on analysis of virulence and adaptation strategies of influenza virus \[124-125\]. Mass spectrometric analysis of cellular proteome alterations in human cell lines infected by A/Puerto Rico/8/1934 (H1N1) was reported \[126\]. Coombs \textit{et al.} analyzed influenza virus-infected cultured human lung cells by using quantitative proteomic methods and discovered some significant differential proteins. Emmott \textit{et al.} \[127\] used stable isotope labeling amino acids in cell culture combined with LC-MS/MS analysis for investigating the changes of proteome in A/Puerto Rico/8/1934 (H1N1)-infected cells. Vester \textit{et al.} \[128\] used two-dimensional difference in gel electrophoresis and identified 8 significantly altered host proteins in influenza virus A/Puerto Rico/8/1934 (H1N1)-infected MDCK and human A549 cells. Liu \textit{et al.} applied a similar mass spectrometry-based approach to identify about 25 significantly altered host proteins in avian influenza A/Hong Kong/108/2003 (H9N2)-infected human gastric carcinoma cells \[129\].

Probe has been found useful in target research of drug discovery and active
sites study of protein analysis. However, application of probes used in proteomic study of influenza virus A/Puerto Rico/8/1934 (H1N1) infect A549 cell line has not been reported. Many antioxidant agents are potential probes in virus infection study because they have been used in treatment of virus infections. The antioxidants create oxidative stress in cells, which has emerged as a suspected component in the pathogenesis of some diseases in recent years [130]. In the earliest stage of virus infection, reductive-oxidative may be important to cell recovery and duplication [131]. The increased damage causing reactive oxygen intermediates are generated when stores of naturally occurring antioxidant reducing agents are depleted. The presence of uncontrolled oxygen-containing molecules may cause damage to cell membranes, proteins and nucleic acids, as well as the alterations in the intra- and inter-cellular environments. The net effect of this damage has been termed as oxidative stress [132-133].

Recently, N-acetylcysteine (NAC), one of the famous antioxidant agents, was successfully used for a therapy of H1N1 (2009) influenza pneumonia [134]. NAC is the pre-acetylated form of the simple amino acid cysteine, which has been served as a premier antitoxin and immune support substance. This small-molecule compound is essential for the synthesis of glutathione (GSH) in the body. NAC has been used as a possible treatment for HIV infection for several years [135]. The rational for the use of NAC in HIV treatment is based on evidence from in vitro studies that cells deficient in GSH are particularly sensitive to inflammatory cytokines (tumor necrosis factor alpha). NAC also inhibits virus replication and expression of pro-inflammatory molecules in A549 cells infected with highly pathogenic H5N1 (2004) influenza A virus [136]. The role of NAC in the treatment of H1N1 influenza pneumonia may involve overdose treatment of acetaminophen.
Thus, NAC may be a useful reagent in H1N1 treatment and syndrome therapy. The concentration of NAC in biological systems may play a key role in the process of medical treatment. However, little is known about the mechanism of NAC activity on A549 cells infected by A/Puerto Rico/8/1934 (H1N1).

Small-molecule probes provide ability of activity-based protein profiling to monitor active species of a panel of enzymes in cells. They are also good at profiling signal transduction pathways in development and differentiation [137-142]. By using proteomics, profiling or quantification of the proteins from virus-infected cells have been reported [143-146]. In this study, mass spectrometry-based proteomics method was applied to study the proteins of NAC-treated A549 cells infected by A/Puerto Rico/8/1934 (H1N1). Cell and protein level studies were both performed. The obtained data suggested that the impact study of NAC as a small-molecule probe on the pathogenesis of influenza virus-infected cells might be useful for analyzing the protein presentation of virus-cell interaction.

5.2 Materials and methods

Propidium iodide (PI) was bought from Becton, Dickinson and Company (BD) (Franklin Lakes, NJ, USA). HPLC grade ACN, methanol and formic acid were purchased from Fisher (Fairlawn, NY, USA). Sequencing grade trypsin was obtained from Promega (Madison, WI, USA). Precast IPG dry strips, pH 3–10 non-linear were purchased from BioRad (Hercules, CA, USA). All other chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA). Influenza virus A/Puerto Rico/8/1934 (H1N1) was propagated in a biosafety level-1 containment facilities. The influenza virus was preserved in State Key Laboratory for Emerging Infectious Diseases, The University of Hong Kong.
5.2.1 Virus and cell line culture

Human lung adenocarcinoma epithelial cells (A549) were cultured in Ham’s F12K medium with 2 mM L-glutamine adjusted to contain 1.5 g/L sodium bicarbonate containing 10% FBS, 200 U/mL penicillin G and 200 µg/ml streptomycin. The cells were cultured in the petri dishes (145×20 mm). Influenza virus strain was inoculated onto confluent monolayer of the A549 cell line at a multiplicity of infection of approximately 0.1. After the 24 h cell culture, the infected cells were added differential doses NAC (0, 1 mM, 10 mM and 100 mM). Three parallel experiments were performed simultaneously. After another incubation of 24 h, the same dose NAC treated A549 cells were collected and mixed for following analysis.

5.2.2 Flow cytometric analysis of apoptosis/necrosis

Apoptosis and nuclear DNA content were measured by propidium iodide (PI) staining and fluorescence-activated cell sorting. Adherent A549 cells were collected by treatment with trypsin and then washed with phosphate-buffered saline. The cells were fixed in 1 mL of cold 70% ethanol overnight at 4°C and resuspended in staining buffer (50 µg/mL PI with RNase) for 20 min at room temperature. PI-stained cells were analyzed on a flow cytometer (FACS Calibur, BD), and 20,000 cells were counted for each sample. Each NAC concentration treatment group was triple tested. Data analysis was performed by using ModFit LT, version 2.0 (Verity Software House).

5.2.3 Two dimensional electrophoresis separation

Cell samples from the control and experimental groups were frozen in liquid nitrogen and then crushed by using a mortar and pestle. All samples were lysed with a lysis buffer containing 8 M urea, 4 % w/v CHAPS, 50 mM DTT, and 0.2 %
w/ v Bio-Lyte at pH 3-10. Each sample was centrifuged at 40,000g for 60 min. The supernatant was collected and the protein concentration was determined by using the DC RC protein assay kit following the manufacturer’s instructions. The samples were then aliquoted and stored at -80 °C for further analysis.

Isoelectric focusing (IEF) was carried out on protein IEF cell. Samples containing 0.5 mg protein were diluted to 300 uL with rehydration solution (6 M urea, 2 % w/v CHAPS, 65 mM DTT, 0.5 % v/v Bio-Lyte at pH 3-10, and trace bromophenol blue), and applied to IPG gel strips for 14 h in a passive mode. The non-linear IPG was used for the first dimensional IEF. Proteins absorbed into IPG gel strips were focused for 80 kV·h at 17 °C. After equilibrated in the equilibration solution, gel strips were applied on second-dimensional PAGE with 12 % polyacrylamide.

The gels were stained by blue silver staining method [147]. The stained gels were scanned with a calibrated densitometer (GS800, BioRad) and the gel images were processed by PDQuest software (Version 7.1.1; BioRad). At least three replicates from independent samples were analyzed. A spot was registered only if it was detected in two or more replicated experiments. The gel spot patterns of replicates from each dose of NAC were summarized in a master gel after the spot matching. To accurately compare spot quantity between gels, a normalization based on the total density on each gel was applied for each gel image and normalized spot intensities were expressed in ppm. Data were analyzed using a student’s t-test. The significantly different expressed protein spots were selected and subject to the identification by MS.

5.2.4 Mass spectrometry analyses and database search

Protein spots of interest were excised from stained gels and digested by the
enzyme solution containing sequencing grade modified porcine trypsin for 16 h at 37 °C. The resulted peptides were extracted three times by 20 µL aliquots of aqueous solution containing 5% TFA and 50% ACN. The extracts were pooled and then dried in a vacuum centrifuge and redissolved in an aqueous solution with 0.5% TFA and 5% ACN for MS analysis.

The matrix solution was freshly prepared by dissolving 10 mg HCCA in 1 mL of aqueous solution with 0.5% TFA and 50% ACN. Peptides from in-gel digested proteins were mixed with matrix solution (1:1 v/v) and loaded on the anchorchip sample plate and air-dried. Peptide Mass Fingerprints (PMFs) of protein spots were obtained on an Autoflex MALDI-TOF mass spectrometer (Bruker Daltonics, Bremen, Germany) operating in the reflection mode with an accelerating voltage of 19 kV and a delayed extraction time of 150 ns. The MALDI-TOF MS spectra were calibrated internally using trypsin autolysis products and calibrated externally by a set of peptide standards. Peak lists were created from raw data using FlexAnalysis software (version 2.0) with the SNAP peak detection algorithm. Data search against the Swiss-Prot database (version 2011_07) in the entries of Homo sapiens (human) was performed by using an in-house MASCOT server (Matrix Science, London, UK). Peptide tolerance was set at 100 ppm with one missing cleavage allowed. N-terminal acetylation, asparagine/glutamine deamidation, and methionine oxidation were selected as variable modifications, while cysteine carbamidomethylation as fixed modification. For positive identification, the score of the result had to be over the significance threshold level (p<0.05).

The tryptic peptides were desalted by using ZipTip C_{18} pipette tips (Millipore, Bedford, MA, USA). Full-scan MS and MS/MS experiments were carried out
with a quadrupole orthogonal acceleration TOF mass spectrometer (QStar, Applied Biosystems, CA, USA) with an external nanoelectrospray ion source (Protana A/S, Odense, Denmark). Renin substrate was used to calibrate the instrument in both MS and MS/MS modes. The ion source voltage, mirror voltage, plate voltage, grid voltage and liner voltage are 1 kV, 0.980 kV, 0.346 kV, 0.428 kV and 4.000 kV respectively. Approximately 4 µL of sample was loaded into a nanoelectrospray tip. After the full-scan mass spectra (parent ions) of trypsin-digested peptides were obtained, the charge states of the peptides were assigned from the observation of their stable isotope spacing. The parent ions of interest were then subject to sequence analysis in product ion mode. Following the MS/MS analysis, the MASCOT generic files were generated by using a script embedded in the Analyst QS 1.1 software (MDS Sciex). The files were used to search against the Swiss-Prot database in the entries of Homo sapiens (human) on an in-house MASCOT server with the following parameters: one missed cleavage; peptide tolerance, 0.2 Da; MS/MS tolerance, 0.1 Da; fixed carbamidomethylation of cysteine, and variable oxidation of methionine. A minimum of two significant MS/MS peptide fragments was considered sufficient to assign positive identification for a spot.

Protein relationship was analyzed by using software STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) version 9.0 (http://string.embl.de/).

5.3 Results and discussion

5.3.1 Flow cytometric analysis of the cells

The cell life cycle histograms of the A/Puerto Rico/8/1934 (H1N1)-infected A549 cells with the treatments using different NAC concentrations were shown in Figure5.1.
Figure 5.1 Histograms of the A/Puerto Rico/8/1934 (H1N1)-infected A549 cells with the treatments using different NAC concentrations

Populations of different cell life cycle phase were determined and demonstrated in Figure 5.2. The results of cell life cycle histogram and different phases’ population percentage showed that apoptotic population was low. In different NAC-treated concentrations, apoptosis numbers of A/Puerto Rico/8/1934 (H1N1)-infected A549 cells increased slightly, but not significantly. The ratios of apoptosis cells were less than 3% of the total cells number. Even at the high NAC concentration the ratio was only 2.92%. The results showed that NAC did not promote apoptosis. The population of G0/G1 phase decreased when the NAC treatment concentration was increased from 0 mM to 10 mM, but increased from 10 mM to 100 mM. The G0/G1 phase population was the lowest at the 10 mM
dose of NAC, indicating that the effect of virus to the cells was the least at the treatment point. Synthesis phase (S-phase) is part of the cell cycle in which DNA is replicated, occurring between G1 phase and G2 phase. The characteristic of S-phase cell population was found to have opposite effect on the G0/G1 phase. For the NAC treatments from 10 mM to 100 mM, the population of G0/G1 phase increased, while the S-phase population decreased.

**Figure 5.2** Histogram of difference A549 cell life cycle phase in NAC-treated cells with different concentrations.

It was found that population of G0/G1 phase was higher than other phases (S-phase and G2/M) in the NAC-treated cells with different concentrations. When A549 cells were infected by influenza virus A/Puerto Rico/8/1934 (H1N1) the cell life cycle population was mainly located in G0/G1 phase, indicating that cells were almost arrested in this phase. The obtained results agreed with the literature report that arresting of G0/G1 phase cell cycle was observed not only in this kind of
A/Puerto Rico/8/1934 (H1N1) virus but also in other strains and subtypes such as A/Donghu/312/2006 (H3N2), A/Guangzhou/333/99 (H9N2) and pandemic swine A/WSN/1933 (H1N1) viruses \cite{[148]}. In the control group, the number of G2/M phase cells was very small because most of the virus-infected cells were arrested in G0/G1 phase. After adding NAC, G2/M phase of the cell number increased significantly, indicating that NAC played an important role in A549 cell reproduction. At the concentration of 10 mM, the NAC-treated cells for about one day had the lowest G0/G1 phase population. Clearly, the NAC treatment at 10 mM allowed the increase of virus-infected cells in S-phase and G2/M phase.

Viral protein expression and progeny virus production increased in cells synchronized in the G0/G1 phase have been reported, compared to those in either unsynchronized cells or cells synchronized in other phase. For virus production and viral protein expression, influenza A virus replication might induce cell cycle arrest in G0/G1 phase. The cell cycle arrest in G0/G1 phase is considered as the common way for indicating the successful virus infection in cells \cite{[148]}. Thus, the G0/G1 phase might be an important phase during the A/Puerto Rico/8/1934 (H1N1) infection. The alteration rules of cell population in this phase were compared to the alteration rule of protein concentration of cells.

5.3.2 Mass spectrometry-based proteomics analysis

The protein analysis was conducted with 2DE-PAGE separation, MS detection and data mining. The 2DE-PAGE separation from the blue silver staining method showed approximately 1,000 protein spots (Figure 5.3-).
Figure 5.3 Two dimensional electrophoresis gel of NAC-treated A549 cell proteins.

Protein spots of different A549 cell cycle phase in different NAC-treated cells were showed in Figure 5.4. Groups of control, A, B and C were shown with optical densities for differential proteins spots of 0 mM, 1 mM, 10 mM and 100 mM NAC-treated A549 cells, respectively (Figure 5.5).
Figure 5.4 Protein spots of different A549 cell cycle phase in NAC-treated cells with different concentrations. Control, A, B and C refers the differential protein spots in 0, 1 mM, 10 mM and 100 mM NAC-treated A549 cells, respectively.
Figure 5.5 Optical density of the differential protein spots.
Twelve differential protein spots were detected by comparing 2DE-PAGE gels of the control group with those from various NAC treatments. The differential proteins were successfully identified using MALDI-TOF MS and Q-TOF MS/MS as well as database searching. The results were listed in Table 1 with the protein spot ID, Swiss-Prot accession No., protein information, Mr/pI values, MASCOT scores and sequence coverage. The significant value scores were higher than 56 (p<0.05), based on the Swiss-Prot database using the MASCOT searching program. For each protein, at least two sequences were indentified in the protein database searching. The significant value scores of the identified proteins were higher than 34 (p<0.05), including 14-3-3 protein epsilon, eukaryotic translation initiation factor 6, tropomyosin alpha-3 chain, 14-3-3 protein gamma, prohibitin, L-lactate dehydrogenase B chain, annexin A4, ubiquitin carboxyl-terminal hydrolase isozyme L1, 3,2-trans-enoyl-CoA isomerase, flavin reductase, coflin-1 and peptidyl-prolyl cis-trans isomerase A. The protein identification results from Q-TOF MS/MS analysis agreed well with those from MALDI-TOF MS analysis. The obtained result showed that these proteins were significant regulated in the cell.

Characteristics of some identified differential proteins and their relationship with virus infection have been reported in the literatures. Specific interaction of eukaryotic translation initiation factor (spot 2) with the 5’ nontranslated regions of hepatitis C virus and classical swine fever virus RNAs were detected [149]. Double-stranded RNA-dependent protein kinase was activated but phosphorylation of the subunit of eukaryotic translation initiation factor 2 was observed in human cells infected with herpes simplex virus [150]. Tropomyosin (spot 3) is a protein within the troponin complex and is associated with actin and
responsible for mediating calcium in the contraction-generating interaction of actin and myosin. Down-regulation of tropomyosin may destabilize microfilament architecture and contribute to changes in cell shape and mobility that usually accompany virus infection [151]. Prohibitin (spot 5) has previously been found to be differentially expressed in AGS cells infected with influenza A/Hong Kong/108/2003 (H9N2) virus. Another protein of interest induced by the A/Puerto Rico/8/1934 (H1N1) virus was annexin (spot 7). Annexins are calcium-dependent phospholipid-binding proteins that have been proposed to act as scaffolding proteins at certain membrane domains. Annexin A1 has been shown to prevent fusion of raft-associated vesicles. Interestingly, annexin A2 that interacts with A1 has the opposite effect and is required for the apical transport of raft-associated vesicles. Since influenza virus also buds from raft domains, this indicates a potential regulatory role of annexin A1 [152]. 3,2-trans-enoyl-CoA isomerase (spot 9) is involved in mitochondrial β-oxidation of unsaturated fatty acids. The defects in distribution of polyunsaturated fatty acids in healthy and cancerous breast tissues have been documented. Decreased levels of this enzyme might have impact on the aberrant behavior of cancer cells [153]. According to the classical pathway of mitochondrial β-oxidation, all unsaturated fatty acids are channeled via their 3-cis- and/or 3-trans-enoyl-CoA isomers to their respective 2-trans-enoyl-CoA intermediates, which are regular intermediates of the β-oxidation spiral. Mitochondrial 3,2-trans-enoyl-CoA isomerase is the essential link between saturated and unsaturated β-oxidation. The peroxisomal trifunctional enzyme contains an isomerase subunit, which carries out the equivalent reaction when peroxisomal β-oxidation is challenged, e.g. by lipid-lowering drugs [154].

In the identified differential proteins, the intensity of spot optical density
representing relative protein concentration was different. The obtained protein concentration data showed that they were not simply up-regulated or down-regulated. In different kinds of protein and different range of concentration, their regulation was different. For eukaryotic translation initiation factor 6, tropomyosin, tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein gamma polypeptide, prohibitin, lactate dehydrogenase, annexin, 3,2-trans-enoyl-CoA isomerase and cofilin (spots 2, 3, 4, 5, 6, 7, 9 and 11), the intensity alteration rule in different NAC treatment was found to be the same as the alteration rule of G0/G1 phase population. The intensity of spot optical density decreased from 0 mM to 10 mM NAC, while increased from 10 mM to 100 mM NAC, indicating that the concentrations of the eight proteins were lowest at 10 mM NAC treatment. The obtained results suggested that these proteins might be related with the cell G0/G1 phase and virus replication because the cells needed lower protein concentrations to provide their activities in the 10 mM treatment than other doses of NAC.

NAC is a precursor of glutathione. It is a functional agent in cell replication for getting rid of reactive oxygen species (ROS). Through redox-dependent signal transduction pathways, ROS regulates the strength and duration of signaling via the cyclic oxidation and reduction of cysteine residues in enzymes and other regulatory factors, and promotes the apoptosis of cell \cite{133}. It has been reported that eukaryotic elongation factor 2 was down-regulated in INS-1E cells after the sustained exposure to high glucose for 24 h \cite{155}. Atrial fibrillation (AF) is thought to be caused by oxidative stress. Gene expression profiles in AF patients showed that tyrosine 3-monooxygenase gene was up-regulated \cite{156}. Additionally, the expression of prohibitin is also down-regulated after the induction of oxidative
stress in epithelial cells in vitro as well as in diseases linked to enhanced levels of ROS such as ulcerative colitis and Crohn's disease. In cells, the increase of ROS production upon loss of prohibitin was eliminated, clearly indicating mitochondrial respiration as the source of elevated ROS levels upon loss of prohibitin [157]. The ROS-dependent activation of the Slingshot-1L-cofilin pathway stimulates the Slingshot-1L-dependent formation of cofilin-actin rods in cofilin-GFP–expressing HeLa cells [158]. The identified eight proteins in this study were related with the function of cells to remove ROS, at the same time they might be closely related to each other in influenza virus-infected cells. NAC appeared to have significant impact on the process of infecting cells by influenza virus.
5.3.3 Protein-protein relationship

Protein-protein relationship of the detected differential proteins was generated by using software STRING (Figure 5.6 and Table 5.1). While the relationship map of all differential protein spots was shown in Figure 5.6 A, Figure 5.6 B provided the relationship map of the eight proteins whose intensity alteration rule of optical density behaved the same as the alteration rule of G0/G1 phase of A549 cell population with different NAC treatments.

Proteins in Figure 5.6 A such as tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein epsilon polypeptide, eukaryotic translation initiation factor 6, tropomyosin, tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein gamma polypeptide, prohibitin, cofilin, and peptidyl-prolyl cis-trans isomerase A (spots 1, 2, 3, 4, 5, 11 and 12) were found to have significant score of relationship. In Figure 5.6 B, four proteins, namely eukaryotic translation initiation factor 6, tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein gamma polypeptide, prohibitin and cofilin (spots 2, 4, 5 and 11), had higher scores than others.
Figure 5.6 Relationship of the 12 differential proteins (A) and the 8 proteins with optical density alteration rule same as that for G0/G1 phase of cells by STRING software (B). The names of proteins are listed in Table 1.
Table 5.1 Confirmation of proteins differentially expressed in A549 cells from MALDI-TOF MS and Q-TOF MS/MS analyses.

<table>
<thead>
<tr>
<th>Spot no.</th>
<th>Swiss-Prot accession ID</th>
<th>Protein information</th>
<th>Mr/ pI</th>
<th>MALDI-TOF MS Analysis</th>
<th>Q-TOF MS/MS Analysis</th>
<th>STRING ID in Figure 5.6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MASCOT scores</td>
<td>MASCOT scores</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sequence coverage</td>
<td>Matched sequences</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1433E_HUMAN</td>
<td>14-3-3 protein epsilon</td>
<td>29326/4.63</td>
<td>83</td>
<td>29%</td>
<td>95</td>
</tr>
<tr>
<td>2</td>
<td>IF6_HUMAN</td>
<td>Eukaryotic translation initiation factor 6</td>
<td>27095/4.56</td>
<td>78</td>
<td>34%</td>
<td>145</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3. LNEAQPSTIATSMR[224 – 237] + Oxidation (M)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4. ASFNENCEIGCFAK[5 – 18]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5. LSALGNVTTCDYVALVHPDLDR[101 – 123]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>TPM3_HUMAN</td>
<td>Tropomyosin alpha-3 chain</td>
<td>32856/4.68</td>
<td>58</td>
<td>20%</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1. IQVLQQQADDAEER[14 – 27]</td>
<td>TPM3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1433G_HUMAN</td>
<td>14-3-3 protein gamma</td>
<td>28456/4.80</td>
<td>98</td>
<td>43%</td>
<td>220</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1. YLAEVATGEK[133 – 142]</td>
<td>2. DSTLIMQLLR[218 – 227] + Oxidation (M)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3. NVTENPLESNEER[29 – 42]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4. TAFDDAIAELDTLNEDSYK[199 – 217]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>PHB_HUMAN</td>
<td>Prohibitin</td>
<td>29843/5.57</td>
<td>136</td>
<td>34%</td>
<td>213</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3. NITYLPAGQSVLQLPQ[256 – 272]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4. AAEELIANSLATAGDLIELR[220 – 239]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>LDHB_HUMAN</td>
<td>L-lactate dehydrogenase B chain</td>
<td>36900/5.71</td>
<td>59</td>
<td>19%</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1. MVVESAYEVK[234 – 244]</td>
<td>K.MVVESAYEVK[234 – 244] + Oxidation (M)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LDHB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ANXA4_HUMAN</td>
<td>Annexin A4</td>
<td>36088/5.84</td>
<td>174</td>
<td>44%</td>
<td>266</td>
</tr>
<tr>
<td>---</td>
<td>-------------</td>
<td>------------</td>
<td>------------</td>
<td>------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>8</td>
<td>UCHL1_HUMAN</td>
<td>Ubiquitin carboxyl-terminal hydrolase isozyme L1</td>
<td>25151/5.33</td>
<td>94</td>
<td>34%</td>
<td>55</td>
</tr>
</tbody>
</table>
5.4 Chapter Summary

The pathology of A/Puerto Rico/8/1934 (H1N1) infection associated with the interaction of virus and its host cells is not clear. N-acetylcysteine (NAC) is an antioxidant as well as a premier antitoxin and immune support substance. High dose of NAC was recently reported for a therapy of H1N1 (2009) influenza pneumonia. NAC was used as a small-molecule organic probe to investigate the protein expression of human lung carcinoma cell line (A549) infected by influenza virus A/Puerto Rico/8/1934 (H1N1). Differential proteins were identified from MALDI-TOF MS and Q-TOF MS/MS analyses. The obtained results showed that NAC kept cells away from apoptosis. Virus-infected cells were arrested in G0/G1 phase. The lowest cell population of G0/G1 phase was detected when the cells were treated by 10 mM NAC for one day. Application of mass spectrometry-based proteomics allowed the identification of the differential proteins. Software analysis showed that 4 proteins had close relationship. The results indicated that NAC as a small-molecule probe might effect the proteins expression of A549 cells infected by the A/Puerto Rico/8/1934 (H1N1) virus.
Conclusion

Influenza virus H1N1 is a huge threat on human health. The above five chapters reported investigations on relationship between influenza A virus H1N1 and host cells. Mass Spectrometry has been proved to be a key analytical technology for identifying the components and studying modification of protein complex. Different mass spectrometric approaches have been used to analyze the important protein of H1N1 virions. HA is one of the important proteins of influenza A virus. MALDI-TOF MS and ESI-Q-TOF MS analyses and database searching were used to confirm the protein sequence. After different enzyme digestion, the difference of HA isoforms were found to be closely related with their glycan structure and glycosylation sites.

Mass spectrometry plays an important role not only in rapid identification of proteins and determination of their covalent structures and posttranslational modifications in a biological mixture, but also in high-throughput proteome-wide quantification of proteins. Mass spectrometry-based stable-isotope labeling methods are an important strategy of quantitative proteomics. In our research, the method development based on organic phosphorus chemistry offered a new approach for quantitative proteomics by using novel stable-isotope labeling reagents. It was found that this method was simple, high efficient, and selective, which might provide a rapid, straightforward, and cost-effective mean for both peptide de novo sequencing and the comparison of the relative abundance of proteins The new method combining hydrazide chemistry was developed and applied to study glycoproteins of H1N1 infected cell line successfully.

Cellular proteome alterations in human cell lines infected by influenza virus were analyzed by mass spectrometry based proteomics methods. N-acetylcysteine
(NAC) was a useful reagent in H1N1 treatment and syndrome therapy. Differential proteins were identified from MALDI-TOF MS and Q-TOF MS/MS analyses. These proteins were potential biomarkers in influenza virus infection. Small-molecule probes provide ability of activity-based protein profiling to monitor active species of a panel of enzymes in cells. They are also good at profiling signal transduction pathways in development and differentiation.
Appendix chapters

Abstract

This part of research work is not closely related to the major objective of the PhD thesis. However, the research was completed during the exchange study in the U.S. in the PhD study period. Therefore, this part was suggested by the thesis evaluation panel to be presented as the appendix.

Elderberry juices have produced anti-influenza activity in human clinical. The juice from elderberry fruit is known to contain a variety of polyphenols such as anthocyanins. Concentration of anthocyanins can change with elderberry cultivar. In this study, juice samples are examined in a profile mode. Multiple reaction monitoring (MRM) mode in LC-MS was selected for reducing false positives, maximizing selectivity, and achieving reliable quantification. The results suggest that this method may help characterize and profile the anthocyanins composition of elderberry juices for quality control, assessment of dietary intake, and provision of support on biochemical studies. The total phenolic and total monomeric anthocyanin concentrations of elderberry were found to vary highly among different cultivars. The method was used by solid phase extraction to separate anthocyanins and peptides, followed by HPLC using a C\textsubscript{18} column to purify the peptides. More than 1,000 peptides were identified successfully in elderberry juice.
Chapter A1

Determination of anthocyanins and total polyphenols in a variety of elderberry juices by UPLC-MS/MS and other methods

A1.1. Introduction

Elderberry (Sambucus) is a widespread species that grows on sunlight-exposed locations in most parts of Europe, Asia, North Africa, and USA. Elderberries are consumed for their nutritional profile, as dietary supplements as well as medicinal purposes [159]. The juice pressed from elderberry fruit contains many primary metabolites including various sugars and organic acids. Elderberry juices have produced anti-influenza activity in human clinical [160]. Its use as an anti-influenza drug can be traced back to the Roman Times. Elderberry has garnered recent research interest because of its high polyphenolic content. Polyphenols, including anthocyanins, have been associated with several potential health benefits. These include anti-oxidant, anti-carcinogenic, anti-viral, and antibacterial benefits [161-163]. They also manifest an array of health promoting benefits such as protecting cells against oxidative stress, anti-inflammatory effects and inhibition of some human tumor cells [164-168]. Therefore, their consumption may contribute to protect against cardiovascular, neurodegenerative, and gastrointestinal diseases, cancer, obesity, and diabetes. These compounds are well known free radical scavengers, reported as potential chemo-preventive agents. Therefore, elderberry cultivars containing higher concentrations of polyphenols are considered more important.

Anthocyanins are proposed to be responsible for the aforementioned health benefits; however a mechanism has yet to be proven. A major concern with
The instability of anthocyanins, especially at alkaline pH levels, exposure to heat, light, and un-ideal storage conditions have been reported to cause anthocyanin degradation. The anthocyanins are also an important fruit quality indicator, because their pigment is responsible for the rich fruit color, and varies as a function of pH. They have gained interest as functional compounds in food colorants and as potent agents against oxidative stress, reducing oxidative damage to the human body.

Anthocyanins are polyphenolic phytochemicals with structures comprising an anthocyanidin aglycone and different sugar moieties. Usually, anthocyanidins are linked to sugars that confer stability and water solubility to the molecule. Anthocyanins are also the most common pigmented flavonoids. They are responsible for most of the red, pink, purple and blue colors observed in plants [169] and play a role in two important plant processes. First, they attract animals for pollination and seed dispersion. Their second role is protection of plant tissues against harmful ultraviolet radiation. Anthocyanin profiles for fruits and vegetables can be used for genuses and/or cultivars for control of the quality and authenticity of related food. Anthocyanin as well as quercetin profiles of the selected plant material were established. Since elderberry is an important fruit species for processing, internal quality parameters were compared to other fruit cultivated for processing.

Quantification methods of total phenolic and anthocyanin concentrations from elderberry have been proposed as a way to analyze different types of fruit. Ozgen et al. concluded that different elderberry accessions can vary in total phenolic concentration twofold, and total monomeric anthocyanin concentration threefold. Certain variables such as cultivar and growing location are not included in this
study. Therefore it is difficult to speculate what the large deviation in phenolic and anthocyanin concentration is from. This study aims to use high-throughput UV-Vis spectrophotometry to analyze the effects cultivar and year-to-year harvesting has on the total phenolic and total monomeric anthocyanin concentration in elderberries.

Separation methods for anthocyanins range from simple solvent extractions to various forms of chromatography. Solid phase extraction (SPE) methods have gained popularity due to a balance of efficiency and cost. SPE sorbents adsorb analytes via non-selective interactions normally. Using different purification solvents can help elucidate the best conditions for SPE separation and sample loading volume. In this paper, a mixed-mode cation exchange SPE column was chosen and the most available SPE sorbent and sample loading volume were studied. Ultra performance liquid chromatography electrospray ionization mass spectroscopy (UPLC–ESI MS) method was developed for the quantitative and qualitative analysis of phenolic compounds in plant material. UPLC is a valuable tool for detecting analytes present at trace levels and screening a large quantity of samples. This technique, which has a higher sensitivity, higher resolution and a shorter run-time than HPLC, has recently been used to study flavonoids in plant materials. Unfortunately, there are few UPLC MS/MS methods used in elderberry juice anthocyanin profiling and quantification. In this study, a UPLC MS/MS method was developed and verified for anthocyanin profiling and quantification.

A1.2 Materials and methods

Elderberry samples were harvested from two Missouri (USA) growing locations (Mt. Vernon and Jefferson City) during 2009 - 2011. Nine different
elderberry genotypes were included: ‘Bob Gordon’, ‘Dallas’, ‘Marge’, ‘Ocoee’, ‘Ozark’, ‘Ozone’, ‘Sperandio’, ‘Wyldewood’, and ‘York’. All genotypes are American elderberry [Sambucus nigra L. subsp. canadensis (L.) Bolli] except for ‘Marge’, which is of European origin (Sambucus nigra L. subsp. nigra). Each site included four randomized field replications (plots) of each genotype. Details on genotypes used and field production methods (sites, soils, climate, planting, management, harvesting, etc.) can be found in Thomas et al. (2014a; 2014b). Immediately upon harvest, fruit was placed into zippered plastic freezer bags and frozen (-20°C). Samples were later de-stemmed, thawed, juiced, filtered through a 0.45 µm nylon filter, and re-frozen (-20°C) before analysis. The total phenolics and monomeric anthocyanin content of the juice were measured in March 2013 and May 2013, respectively. The work flow was showed in Figure A1.1.

![Figure A1.1](image)

**Figure A1.1** Experimental scheme for the purification and isolation of elderberry juice peptides.

**A1.2.1 Chemicals**
Methanol, formic acid, ammonium hydroxide, sodium carbonate, potassium chloride, acetonitrile, and water were purchased from Fisher Scientific (HPLC grade, Fair Lawn, NJ, USA). The cyanidin 3-O-glucoside, gallic acid, Folin and Ciocalteu phenol reagent (FCR), and sodium acetate trihydrate were purchased from Sigma (Saint Louis, MO, USA). Oasis (MCX) mixed-mode reversed-phase/strong cation exchange SPE cartridges (3 cc/60 mg) were purchased from Waters (Milford, MA, USA).

A1.2.2 Total phenolic testing
The Folin-Ciocalteu total phenolic assay was followed with slight modifications (Singleton and Rossi, 1965). Samples were equilibrated for two hours at room temperature as opposed to heating. Gallic acid standards were prepared and treated in the same manner as the samples. Absorbance values were measured in triplicate at $\lambda$ of 760 nm using a PerkinElmer Enspire 2300 multimode plate reader. Total phenolic concentration is expressed as gallic acid equivalents (GAE mg/mL).

A1.2.3 Total monomeric anthocyanin
The pH differential method [175] was used to estimate the total monomeric anthocyanin content of elderberry juice. Absorbance values were measured in triplicate at $\lambda$ of 520 nm and 700 nm. Cyanidin 3-O-glucoside standards were prepared and treated in the same manner as the samples. Total monomeric anthocyanin values are represented as cyanidin 3-O-glucoside (C3G) equivalents (C3GE mg/mL).

A1.2.4 Solid phase extraction
A MCX cation-exchange solid phase extraction (SPE) method [176] was used for anthocyanin separation from other matrix compounds. The results showed that
MCX separation method worked well for separating anthocyanins from other compounds present in elderberry juice. The cartridge was first washed with 6 mL of water (0.1% formic acid), after 100 µL of elderberry juice in aqueous solution was added to the SPE cartridge. The fraction of other phenols was collected by elution with 6 mL of methanol (0.1% formic acid). Subsequently, the anthocyanins were eluted with 2 mL of methanol and 2 mL of water/methanol (40:60, v/v), both containing 1% NH₄OH. Immediately, 250 µL of formic acid was added to the combined alkaline eluent. Samples were dried in a Buchi rotary evaporator (35°C), and the fractions were re-dissolved in acidified water.

**A1.2.5 UPLC-MS/MS analysis**

Separation of anthocyanins was achieved on a C₁₈ column (Acquity BEH, 1.7µm, 50 mm × 2.1 mm, Waters, Milford, MA, USA) at room temperature (~20°C) using a Waters Xevo TQ-S UPLC-MS/MS system. The mobile phase included 4.5% formic acid in LCMS grade water (mobile phase A) and 0.1% formic acid in HPLC grade acetonitrile (mobile phase B). The flow rate was 400 µL/min. The gradient started at 95% mobile phase A and was lowered to 5% mobile phase A over 4 minutes, followed by a 30 sec isocratic step and a 30 sec re-equilibration in 95% mobile phase A.

The following conditions were used for the electrospray ionization (ESI) source: source temperature 150°C, desolvation temperature 350°C, capillary voltage 2.0 kV, cone voltage 12 V, and nebulizer N₂ gas 500 L h⁻¹. Argon was used as the collision gas. The collision energies were optimized and ranged from 10 to 40 eV for individual analytes. The ESI source was operated in the positive ion mode. Instrument control and data processing were performed by using MassLynx software (version 4.1, Waters, Milford, MA, USA). Cyanidin 3-O-glucoside
standard solutions were prepared with concentrations ranging from 1 ng/mL to 1 μg/mL in methanol. Analyte identity was determined based on retention time and mass spectra, and quantitation was based on analyte to C3G area ratios.

**A1.3 Results and discussion**

**A1.3.1 Total Phenolic and Total Monomeric Anthocyanin**

Phosphomolybdenum complexes in a 5+ oxidation state are present in the FCR stock solution. These complexes are reduced to a 4+ oxidation state in the presence of polyphenols. This oxidation state has a maximum absorption wavelength of 760 nm; at this wavelength little optical interference exists. Therefore, by adding the FCR in excess to elderberry juice, one can estimate the total polyphenol content.

Elderberry juice has been shown to contain a variety of anthocyanins (Figure A1.2). Although the molecules vary based on the number and type of sugar moieties attached, their absorption spectra are very similar. Most anthocyanins exhibit a maximum wavelength of absorption ranging from 500-530 nm\(^{[175]}\); and deviations among samples in this range are negligible due to their broad maximum peak. The total monomeric anthocyanin content of juice samples can be estimated by taking absorbance measurements of juice samples at 520 nm, which is the maximum wavelength of absorption of the cyanidin 3-\(O\)-glucoside standard in the solvent used.
Figure A1.2 The structures of the 6 common anthocyanins at pH of 1.0 (I), pH of 4.5 (II), and anthocyanindins (III).

Table A.1.1 shows that the total phenolic (TP) content of elderberry juice varies highly among genotypes grown at the same location and growing season (Jefferson City, MO, 2011, in this case). For this table, juice was analyzed from either 3 or 4 replicated field plots per genotype, with each sample quantified in triplicate (n = 9 or 12). The juice of the ‘Wyldewood’ cultivar contained the highest TP content (8.59 mg/mL GAE), and the juice of ‘Ocoee’ the lowest (2.16 mg/mL GAE). A two-variable t-test, at the 95% confidence level, was performed on the data in Table A1.1. Five of the nine genotypes evaluated (‘Wyldewood’, ‘Bob Gordon’, ‘Ozark’, ‘Ozone’, and ‘Ocoee’) had TP levels (in descending order) that were statistically different from each other, underscoring the influence of genetics on these polyphenol levels.

Total monomeric anthocyanin content of elderberry also varied highly among
samples. The ‘Ozark’ and ‘Sperandio’ genotypes displayed the highest and lowest total monomeric anthocyanin concentrations with 5.25 and 0.09 mg/mL C3GE, respectively (Table A1.1). For the total monomeric anthocyanin content, the genotypes can be divided into four statistically equivalent groups, in descending order: (1) ‘Ozark’, ‘Wyldewood’, and ‘Bob Gordon’; (2) ‘Marge’ and ‘Ozone’; (3) ‘Dallas’, ‘Ocoee’, and ‘York’; and (4) ‘Sperandio’. Each group is statistically different from the other groups. It is interesting that the polyphenol content does not necessarily correlate with anthocyanin content. Certain genotypes possess copious amounts of anthocyanins, while others have other polyphenols largely contributing to their TP profile.

Table A1.1 Total phenolic (TP) and total monomeric anthocyanin (TMA) content of elderberry juice samples of different genotypes harvested from the same growing location and year (Jefferson City, MO, 2011).

<table>
<thead>
<tr>
<th>Genotype</th>
<th>n</th>
<th>TP (mg/mL GAE ± SE)</th>
<th>TMA (mg/mL C3GE ± SE)</th>
<th>Percent anthocyanin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob Gordon</td>
<td>12</td>
<td>7.21 ± 0.11</td>
<td>4.37 ± 0.56</td>
<td>60.6</td>
</tr>
<tr>
<td>Dallas</td>
<td>12</td>
<td>4.89 ± 0.28</td>
<td>0.75 ± 0.07</td>
<td>15.3</td>
</tr>
<tr>
<td>Marge</td>
<td>12</td>
<td>3.14 ± 0.55</td>
<td>1.31 ± 0.14</td>
<td>41.7</td>
</tr>
<tr>
<td>Ocoee</td>
<td>12</td>
<td>2.16 ± 0.23</td>
<td>0.68 ± 0.11</td>
<td>31.5</td>
</tr>
<tr>
<td>Ozark</td>
<td>12</td>
<td>6.10 ± 0.05</td>
<td>5.25 ± 0.70</td>
<td>86.1</td>
</tr>
<tr>
<td>Ozone</td>
<td>9</td>
<td>5.62 ± 0.06</td>
<td>1.15 ± 0.04</td>
<td>20.5</td>
</tr>
<tr>
<td>Sperandio</td>
<td>9</td>
<td>4.30 ± 0.79</td>
<td>0.09 ± 0.01</td>
<td>2.1</td>
</tr>
<tr>
<td>Wyldewood</td>
<td>12</td>
<td>8.59 ± 1.02</td>
<td>4.67 ± 0.44</td>
<td>54.4</td>
</tr>
<tr>
<td>York</td>
<td>12</td>
<td>3.98 ± 1.19</td>
<td>0.64 ± 0.06</td>
<td>16.1</td>
</tr>
</tbody>
</table>

Abbreviations: GAE = gallic acid equivalents; C3GE = cyanidin 3-O-glucoside equivalents; SE = standard error
Growing locations and growing seasons were examined for their impact on TP and TMA content of elderberry juice (Table A1.2). Our preliminary data suggest that the TP and TMA content of juice of the same genotype can vary significantly from year-to-year and site-to-site, likely due to variations in environmental, climatological, and ecological factors. A larger study involving different genotypes, growing location and season is currently underway utilizing the methods described here.

**Table A1.2** Total phenolic (TP) and total monomeric anthocyanin (TMA) content of elderberry juice of the ‘Bob Gordon’ genotype at different locations and during different seasons.

<table>
<thead>
<tr>
<th>Fruit harvest year and location</th>
<th>n</th>
<th>TP (mg/mL GAE ± SE)</th>
<th>TMA (mg/mL C3GE ± SE)</th>
<th>Percent anthocyanin</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009 Mt. Vernon</td>
<td>12</td>
<td>2.57 ± 0.41</td>
<td>0.19 ± 0.03</td>
<td>7.4</td>
</tr>
<tr>
<td>2010 Mt. Vernon</td>
<td>12</td>
<td>7.13 ± 0.14</td>
<td>3.40 ± 0.49</td>
<td>47.6</td>
</tr>
<tr>
<td>2010 Jeff. City</td>
<td>9</td>
<td>4.31 ± 0.52</td>
<td>2.04 ± 0.31</td>
<td>47.3</td>
</tr>
</tbody>
</table>

Abbreviations: GAE = gallic acid equivalents; C3GE = cyanidin 3-O-glucoside equivalents; SE = standard error
A1.3.2 UPLC optimization

The UPLC conditions were optimized to separate compounds. Optimal separation was achieved with methanol as an organic phase and 10 mM formic acid as an aqueous phase. Formic acid, which is the first choice MS-compatible additive, was used to minimize peak tailing. The ionization efficiency and limits of detection were slightly reduced in its presence however. The combination of methanol and water balanced with formic acid has previously been used to study anthocyanin. Because anthocyanin needed separate condition is low pH 2 and UPLC C\textsubscript{18} column, Acquity BEH, 1.7\textmu m, 50 mm $\times$ 2.1 mm, the optimum pH value is between 1-12. The mobile phase was 4.5\% formic acid in LCMS grade water which pH is about 1.5. UPLC-MS/MS detection is a time saving method\textsuperscript{[177-179]}. In this experiment, the running time was shortened to 5 min.

A1.3.3 Method development

The Cyanidin 3-O-glucoside calibration curve is shown in Figure A1.3. Parent scans of 6 anthocyanidin are shown in Figure A1.4. A: Pelargonidin (\textit{m/z} 271), B: Cyanidin (\textit{m/z} 287), C: Malvidin (\textit{m/z} 331), D: Peonidin (\textit{m/z} 301), E: Delphinidin (\textit{m/z} 303), F: Petunidin (\textit{m/z} 317).
Figure A1.3 UPLC-MS/MS calibration curve using Cyanidin-3-O-glucoside standard.
A: m/z 271 (Pelargonidin)

B: m/z 287 (Cyanidin)

C: m/z 331 (Malvidin)
Figure A1.4 Parent ion scans of 6 anthocyanidins by ESI-MS. A: Pelargonidin (m/z 271), B: Cyanidin (m/z 287), C: Malvidin (m/z 331), D: Peonidin (m/z 301), E: Delphinidin (m/z 303), F: Petunidin (m/z 317).
Cyanidin’s parent’s spectra, which include seven peaks, possess the highest intensity. It has a high signal to noise ratio in those six parent spectra. This means cyanidin and anthocyanin bases on cyanidin are the most abundant compounds in the elderberry test juice. These results agree with literature reports. From the intensity of the spectra, the six anthocyanidin based anthocyanin from max to min are: Cyanidin, Peonidin, Malvidin, Pelargonidin, Delphinidin, Petunidin.

A1.3.4 Profiling and quantification of anthocyanin

Careful adjustment of the MS/MS parameters enabled the positive ion mode to be used for the whole set of analytes with reasonable sensitivity. The optimal MS/MS parameters, such as capillary and cone voltages, and collision energy, were estimated from the maximum peak areas for the analytes. To further increase the sensitivity of the method, the chromatographic were divided into different retention windows, which covered from each individual MRM transitions with prolonged dwell times. By considering the peak widths, latency and number of MRM transitions, the collision energy were adjusted to ensure fit for per peak. From the result of MS detection, the extracted ion chromatograms (Figure A1.5) of each analytes are good enough (Table A1.3) showed anthocyanins and anthocyanidin quantification in sample of elderberry juice. From these 16 anthocyanins, the concentrations showed the abundant ones are cyanidin based anthocyanins. Delphinidin based anthocyanin, peonidin based anthocyanin and pelargonidin based anthocyanin are detected but malvidin and petunidin are not found in the elderberry juice sample. In some literature reports, the major content of anthocyanin in American and European elderberry are cyanidin based anthocyanin, trace Delphinidin based anthocyanin and little peonidin based anthocyanin and pelargonidin-based anthocyanin. Our results closely agree with
previous researchers.

Figure A1.5 Several individual anthocyanin chromatograms observed in elderberry juice utilizing multiple reaction monitoring. Table A1.3 contains the names of the anthocyanins for a given \([\text{M-H}]^+ \rightarrow \text{MS}^n\) transition. The names/transitions are listed in Table A1.3 in the order of top to bottom within the Figure.
Table A1.3 Parent-daughter ion masses, retention times and concentrations of individual anthocyanins and anthocyanidins present in a sample of ‘Wyldewood’ elderberry harvested in 2011.

<table>
<thead>
<tr>
<th>Anthocyanin</th>
<th>[M-H]+ (Da)</th>
<th>MSn (Da)</th>
<th>Retention time (min)</th>
<th>Collision energy (eV)</th>
<th>Concentration (ng/mL C3GE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delphinidin</td>
<td>271.1</td>
<td>121.0</td>
<td>4.65</td>
<td>30</td>
<td>ND</td>
</tr>
<tr>
<td>Cyanidin</td>
<td>287.1</td>
<td>137.1</td>
<td>3.88</td>
<td>20</td>
<td>28.8</td>
</tr>
<tr>
<td>Peonidin 3- O-arabinoside</td>
<td>433.1</td>
<td>301.1</td>
<td>0.50</td>
<td>20</td>
<td>16.2</td>
</tr>
<tr>
<td>Cyanidin based anthocyanin-1</td>
<td>595.1</td>
<td>287.1</td>
<td>3.91</td>
<td>30</td>
<td>ND</td>
</tr>
<tr>
<td>Peonidin based anthocyanin-1</td>
<td>579.0</td>
<td>301.1</td>
<td>4.39</td>
<td>40</td>
<td>ND</td>
</tr>
<tr>
<td>Cyanidin 3- O-coumaroyl-sambubioside</td>
<td>727.4</td>
<td>287.1</td>
<td>3.92</td>
<td>40</td>
<td>1190.2</td>
</tr>
<tr>
<td>Pelargonidin based anthocyanin-1</td>
<td>873.4</td>
<td>271.0</td>
<td>3.91</td>
<td>35</td>
<td>ND</td>
</tr>
<tr>
<td>Delphinidin 3- O-coumaroyl-sambubioside-5-gluco side</td>
<td>889.4</td>
<td>287.1</td>
<td>3.92</td>
<td>35</td>
<td>3649.2</td>
</tr>
<tr>
<td>Peonidin based anthocyanin-2</td>
<td>975.6</td>
<td>301.1</td>
<td>4.55</td>
<td>35</td>
<td>ND</td>
</tr>
<tr>
<td>Delphinidin based anthocyanin-1</td>
<td>975.6</td>
<td>303.1</td>
<td>4.53</td>
<td>40</td>
<td>ND</td>
</tr>
<tr>
<td>Cyanidin 3- O-glucoside</td>
<td>449.1</td>
<td>287.1</td>
<td>2.15</td>
<td>20</td>
<td>107.0</td>
</tr>
<tr>
<td>Cyanidin 3- O-sambubioside</td>
<td>581.0</td>
<td>287.1</td>
<td>2.12</td>
<td>30</td>
<td>2509.2</td>
</tr>
<tr>
<td>Peonidin based anthocyanin-3</td>
<td>607.2</td>
<td>301.1</td>
<td>4.55</td>
<td>40</td>
<td>ND</td>
</tr>
<tr>
<td>Delphinidin 3- O-rutinoside</td>
<td>610.9</td>
<td>303.1</td>
<td>3.60</td>
<td>40</td>
<td>651.6</td>
</tr>
<tr>
<td>Pelargonidin based anthocyanin-2</td>
<td>727.4</td>
<td>271.0</td>
<td>2.06</td>
<td>35</td>
<td>13.9</td>
</tr>
<tr>
<td>Cyanidin 3-sambubioside-5-glucoside</td>
<td>743.2</td>
<td>287.1</td>
<td>1.57</td>
<td>20</td>
<td>8192.7</td>
</tr>
<tr>
<td>Cyanidin based anthocyanin-2</td>
<td>785.3</td>
<td>287.1</td>
<td>2.93</td>
<td>35</td>
<td>651.6</td>
</tr>
<tr>
<td>Peonidin based anthocyanin-2</td>
<td>857.3</td>
<td>301.1</td>
<td>4.22</td>
<td>35</td>
<td>411.2</td>
</tr>
</tbody>
</table>

Abbreviations: [M-H]+ = parent ion mass; MSn (Da) = daughter ion mass; ND = not detected
A1.4 Chapter summary

The juice from elderberry fruit is known to contain a variety of polyphenols, especially anthocyanins, the concentration of which can change with elderberry cultivar. In this study, reversed-phase chromatography was optimized to achieve separation of analytes. Positive-ion electrospray mass spectrometry was used for compound detection. Juice samples are examined in a profile mode to obtain an overall chemical “fingerprint.” Multiple reaction monitoring (MRM) is then used. MRM was selected for the reduction of false positives, maximizing selectivity, and reliable quantification. The quantitative performance of the method was validated. The results suggest that this method will help to characterize and profile the anthocyanins composition of elderberry juices for quality control, assessment of dietary intake, and anthocyanins biochemical studies. The total phenolic and total monomeric anthocyanin concentrations of elderberry were found to vary highly among different cultivars. From these results, we hope to be able to correlate the anthocyanins content of elderberry juices with different regions and varieties.
Chapter A2
Peptidomics study in elderberry juice by LTQ-Orbitrap MS/MS

A2.1 Introduction

Thousands of peptides have been isolated and purified from microorganisms, animals, and plants. Plant peptides are often key players in cell-to-cell communication governing plant development. They also play a role in communication between plants and other organisms. Plants are an enormously rich source of peptides, with the potential to be developed as nutrients, medicines and biomarkers [180, 181]. Some secreted peptides are recognized as important hormones that coordinate and specify cellular functions in plants. Peptides are of interest as they possess several advantages including low molecular weight, relatively simple structure, ease of absorption, and lower antigenicity [182-184]. The first functional plant peptide to be identified was tomato systemin, an 18 amino acid polypeptide, which acts in the rapid expression of defense-responsive genes via cellular communication [185]. In the 1990s, a disulfated pentapeptide was isolated by Matsubayashi and Sakagami, and named as phytosulfokine. It is a potent mitogenic factor from conditioned medium derived from cultures of asparagus mesophyll cells [186]. Studies over the past decade have highlighted the diverse nature of peptides and explored their possible uses.

Peptidomics is a fast, efficient method that can detect low and transient concentrations of peptides and identify their posttranslational modifications. It is a branch of proteomics that has been developing over several years. There are several approaches to peptide separation and identification [187]. Two-dimensional
gel electrophoresis can be extended to detect smaller peptides (1 kDa and larger) by the use of tris-tricine in the second dimension. After extraction from the gel, the isolated peptides can be measured by MALDI-TOF MS. The combination of liquid chromatography coupled with MS/MS can help to overcome complex peptide mixtures and aid in the identification of less abundant peptides. Nano-liquid chromatography can increase the signal-to-noise ratio and sensitivity of MS for the detection of peptides that are complex or present in low concentrations [188-191].

Peptidomics studies the role of peptides and bridges the space between proteomics and metabolomics [192]. The full spectrum of peptide functions in plants is unknown. Peptide hormone signaling in plants is an emerging area of research and peptides have been shown to affect cell division, development, reproduction, nodulation and defense. Peptidomics relies heavily on a combination of liquid chromatography for separation of the highly complex peptide pools to ease the process of peptide fragmentation and identification with tandem mass spectrometry (MS/MS). The major difference between proteomics and peptidomics, regardless of peptide size, is that peptidomics identifies native peptides. Thus, it can be referred to as a version of top-down proteomics which considers only the peptidome (peptides).

Elderberry has been used as a traditional medicine whose use can be traced back to the Roman Times. The juice is known to be rich in anthocyanin compounds. There are also many peptides in it. Elderberry juices are expected to vary with cultivar and growing region. Because of the different atmospheric and environmental situations, the peptides of different elderberry cultivars are expected to vary. To date, no peptide profiling studies of elderberry have been
reported \cite{193-197}. In this chapter we described a method to remove the anthocyanin compounds and purify the peptides from elderberry juice by solid phase extraction. We then use LC-MS/MS to profile and de novo sequence the peptides. Analysis of the peptide sequences is then used to identify proteins present in the juice samples.

A2.2 Materials and methods

A2.2.1 Materials and instruments

Methanol, formic acid, and ammonium hydroxide were purchased from Fisher Scientific (Fair Lawn, NJ, USA). Elderberry (Sambucus canadensis) juice was obtained from the MU Center for Botanical Interaction Studies. It was thawed and aliquoted (pH 4.51), and was filtered through a 0.45µm nylon filter to remove any suspended solids. The color of the juice was dark-violet. Solid phase extraction (SPE) columns: Oasis MCX extraction cartridge 3cc, 60mg and C$_{18}$ Certified Sep-Pak Vac, 3cc, 500mg, both were purchased from Waters Corp., Milford, MA. MALDI-TOF MS spectra were acquired on an Applied Biosystems 4700. LC-MS/MS spectra were acquired on a LTQ Orbitrap XL mass spectrometer (Thermo Scientific) and an interfaced nano-LC (Proxean Easy nLC-1000).

A2.2.2 Peptide separation and purification

The Oasis MCX sorbent contains two types of support material; strong cation-exchange with a C$_{18}$-reversed phase. Two cartridge volumes (vol) of elderberry juice in aqueous solution (12 mL) were applied to an MCX SPE cartridge (6 mL, 1 g sorbent; Waters Corp., Milford, MA). After washing with 2 vol of 0.1% FA, the peptide fraction was collected by elution with 2 vol of methanol (0.1% FA). Anthocyanins remained on the column. The anthocyanin fraction can then be eluted with 1 vol of methanol and 1 vol of water/methanol
(40:60 v/v), both containing 1% NH$_4$OH. The combined alkaline eluate was immediately mixed with an aliquot (250 µL) of formic acid (99%) to lower the pH to less than 2 to degradation of the anthocyanin fraction.

The peptide fraction was briefly evaporated in a rotovaporator at 90°C for 10 minutes to remove the organic solvent, redissolved in 0.1% FA acidified water, filtered through 0.45 µm polypropylene filter, and then separated into two groups for analysis; one for analysis by MS directly (Group A) and the other group was further purified using a C$_{18}$ SPE cartridge (Group B) following the protocol provided by Waters. Figure A2.1 is an overall flow chart of the procedure used for the separation of the peptides from the elderberry juice.
Figure A2.1 Experimental scheme for the purification and isolation of elderberry juice peptides.
**A2.2.3 MALDI-TOF MS/MS detection**

Elderberry juice samples were tested by MALDI-TOF MS/MS. Each of the samples was combined 1/1 (V/V) with alpha-cyano-4-hydroxycinnamic acid (CHCA), 2,5-dihydroxybenzoic acid (DHB), or sinapinic acid (SA) matrix for application to the target for MALDI-TOF/TOF MS analysis. MS spectra were acquired in the positive ion reflector delayed-extraction mode with external calibration.

**A2.2.4 Peptide identification**

All data were acquired on an LTQ Orbitrap XL. A nano-HPLC system was connected to the mass spectrometer for LC-MS/MS. Nano-liquid chromatography and MS settings are as follows: Mobile phase A: flow rate: 20uL/min. Sample loading volume: 30 µL, flow 25 µL/min, max pressure 200.00 Bar. The gradient used is shown as Table A2.1. Acquire time: 120 min. Scan Event 1, Analyzer: FTMS, Mass Range: normal, Resolution: 30000, Scan type: Full, Polarity: positive, Data type: profile. Scan event 2, Analyzer: ion trap, Mass range: normal, Scan rate: normal, Data type: centroid.

**Table A2.1** Applied Nano-LC gradient parameters for the elution of the peptides.

<table>
<thead>
<tr>
<th>Time (min)</th>
<th>Duration(min)</th>
<th>Solvent B %</th>
<th>Flow Rate (nL/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>N/A</td>
<td>5</td>
<td>400</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>5</td>
<td>400</td>
</tr>
<tr>
<td>107</td>
<td>105</td>
<td>40</td>
<td>400</td>
</tr>
<tr>
<td>108</td>
<td>1</td>
<td>90</td>
<td>400</td>
</tr>
<tr>
<td>116</td>
<td>8</td>
<td>90</td>
<td>400</td>
</tr>
<tr>
<td>117</td>
<td>1</td>
<td>5</td>
<td>400</td>
</tr>
<tr>
<td>120</td>
<td>30</td>
<td>5</td>
<td>400</td>
</tr>
</tbody>
</table>
The peptide mixture was loaded onto a trap column (C$_{18}$, 75 µm × 100 mm) and separated on a MonoSpray C$_{18}$ tip using a 120 min gradient from 10 to 50% acetonitrile in 0.5% formic acid at a flow rate of 200 nL/min. A protonated ion of polycyclodimethylsiloxane with 445.120025 Da was used for internal calibration throughout. The mass spectrometer alternated between a full FT-MS scan (300-1500 Da) and subsequent MS/MS scans. Cations were isolated with an isolation window of 5 Da and provided a dynamic exclusion list for 2400 seconds after selected for at least two MS/MS scans. Singly charged precursors were excluded. Monoisotopic selection was disabled with an exclusion window setting of 1 Da. The four most intense ions were chosen for CID fragmentation. Automatic gain control was used to accumulate sufficient fragment ions (MS/MS target value: 2E5; maximum injection time: 1000 ms)\[188\]. A sample blank was run before the real sample. The blank solution included 5% ACN and 1% FA. Peptide de novo sequencing was performed using the proteomics software PEAKS. A trial version of PEAKS 6 was downloaded from the bioinformatics solutions website (http://www.bioinfor.com).

A2.2.5 Protein sequence similarity searches

BLAST protein sequence searches were performed using the measured peptide sequences (http://www.uniprot.org/blast). The parameter settings were as follows: Program blastp (BLASTP 2.2.28+), Database: uniprotkb_viridiplantae (Protein) generated for BLAST, Sequences: 1,711,995 sequences consisting of 584,837,709 letters, Matrix: blosum62, Threshold: 0.1, Filtering: None, Gapped: Yes.

A2.3 Result and discussion

A2.3.1 MALDI-TOF MS/MS detection

The original elderberry juice and its diluted solution were used for detection by
MALDI-TOF MS in different matrix and mass range. As the Table A2.2 showed, the presence of the MS/MS fragment ion at 287.03 Da corresponded to cyanidin (expected 287.055 Da, monoisotopic), and this result indicated that many of the precursor ions were various modified forms of cyanidin. Fragmentation patterns were not similar to those expected for peptides. Instead fragmentation patterns were observed corresponding to the loss of hexose ($\Delta = 162.053$ Da, monoisotopic), deoxyhexose (146.058 Da), pentose (132.042 Da), deoxypentose (116.047 Da) and the disaccharide rutinose (308.111 Da).

**Table A2.2** Results of elderberry juice analyzed by MALDI-TOF/TOF MS

<table>
<thead>
<tr>
<th>Precursor Ion (m/z)</th>
<th>Fragment Ions (m/z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>449.13</td>
<td>287.02 (and weak ions at 389.07, 383.05, 353.02, 329.02)</td>
</tr>
<tr>
<td>611.19</td>
<td>449.07, 287.02</td>
</tr>
<tr>
<td>727.23</td>
<td>287.03 (most intense), 595.1, 581.11, 433.06, 419.06, 147.02</td>
</tr>
<tr>
<td>743.24</td>
<td>581.1, 449.07, 287.03</td>
</tr>
<tr>
<td>889.29</td>
<td>727.1, 449.08, 287.03</td>
</tr>
<tr>
<td>927.25</td>
<td>765.11, 727.21 (weak), 487.04, 449.11 (weak), 287.04 (weak)</td>
</tr>
<tr>
<td>1051.34</td>
<td>889.21, 727.14, 611.1, 449.07, 287.05</td>
</tr>
<tr>
<td>1078.35</td>
<td>916.19, 449.08, 287.03</td>
</tr>
<tr>
<td>1329.45</td>
<td>1167.31, 889.22, 727.14, 449.08, 287.06</td>
</tr>
<tr>
<td>1337.43</td>
<td>1175.27, 1167.31, 1013.21, 889.24, 7274, 449.08, 287.02</td>
</tr>
<tr>
<td>1469.47</td>
<td>1307.37, 1175.3, 1167.32, 1013.22</td>
</tr>
<tr>
<td>1615.52</td>
<td>1453.4, 1175.28, 1167.35, 1013.22, weak 287.01</td>
</tr>
</tbody>
</table>
Based on these results, it appears that simple dilution of the elderberry juice is not sufficient to permit the analysis of peptides and proteins directly. The interfering problem is maybe caused by high concentrations of anthocyanins compared relative to the concentrations of any peptides in the juice (Figure A2.2-A2.5). Another possibility for not-detected peptides might be that the peptide signal was suppressed by the presence of anthocyanins. Therefore, a solid phase extraction (SPE) cleanup procedure step was developed to remove interference and try to enhance the peptide analysis.
Figure A2.2  Sample Clarified by Centrifugation, MALDI-TOF MS

(Voyager)

Top Panel: SA matrix, crystals washed, laser 2151
Center Panel: CHCA matrix, crystals washed, laser 2151
Bottom Panel: DHB matrix, crystals not washed, laser 2401
Figure A2.3 MALDI-TOF MS result, Mass (m/z) range: 809-3187.
Figure A2.4 MALDI-TOF MS result, Mass (m/z) range: 2960-5100.
Figure A2.5 MALDI-TOF MS result analyzed by using matrix CHCA.
A2.3.2 Mixed SPE method developing

Cation-exchange MCX SPE cartridges were used to remove the anthocyanins. This step was used to remove a wide range of anthocyanins contained in the elderberry juice. Anthocyanins are adsorbed to the MCX sorbent allowing for easy separation. The anthocyanin molecules acquire a positive charge to become flavylium cations at acidic pH while at basic pH, they convert to negatively charged quinonoidal bases. Such a transformation does not occur with the peptides in the juice. Peptides are then easily eluted by using the SPE cartridge and lowering the pH of the juice. After removal of the anthocyanins from the juice, the peptides were extracted with organic solvent and dried at 90 °C. Under these conditions, the activity of endogenous proteases is decreased to its lowest levels. Sample collected after the cation-exchange SPE (Group A) was either analyzed by LC-MS or subject to further cleanup using C18 SPE to produce Group B sample for the LC-MS analysis. The use of C18 SPE helped remove highly polar interference and ionic compounds.

A2.3.3 Peptide identification

It is relatively easy to get a peptide profile by MALDI-TOF MS, but this method will also fail to detect many trace peptides. The use of nano-LC MS using the Orbitrap will acquire peptide information that is of a larger quantity and higher quality. After analyzing the sample from Group A (Figure A2.6) on the Orbitrap instrument, a total of 2530 MS/MS spectra were obtained. These spectra were then analyzed using the PEAKS program. A total of 1789 peptides were found by using de novo sequencing testing in the program. The de novo sequencing rate was 69.84 %. In Group B (Figure A2.6), 1833 MS/MS spectra were acquired and a total of 1233 peptides were found. The de novo sequencing rate for Group B was
78.94%. The detailed peptide sequences information is showed in Tables A2.3 and A2.4.

**Table A2.3** Elderberry juice peptide sequences with ALC scores greater than 60 for Group A.

<table>
<thead>
<tr>
<th>No.</th>
<th>M/Z</th>
<th>Charge</th>
<th>DENOVO</th>
<th>ALC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>372.120</td>
<td>2</td>
<td>EEMTCM</td>
<td>60.03</td>
</tr>
<tr>
<td>2</td>
<td>394.1062</td>
<td>2</td>
<td>CCCCAGPM</td>
<td>65.34</td>
</tr>
<tr>
<td>3</td>
<td>400.0883</td>
<td>2</td>
<td>CCCHGACC</td>
<td>62.05</td>
</tr>
<tr>
<td>4</td>
<td>490.1198</td>
<td>2</td>
<td>CMCGWACCC</td>
<td>61.73</td>
</tr>
<tr>
<td>5</td>
<td>386.1172</td>
<td>2</td>
<td>DSSDAACC</td>
<td>60.11</td>
</tr>
<tr>
<td>6</td>
<td>352.6239</td>
<td>2</td>
<td>YMEMM</td>
<td>61.47</td>
</tr>
<tr>
<td>7</td>
<td>408.1661</td>
<td>2</td>
<td>YCCGAGNK</td>
<td>60.79</td>
</tr>
<tr>
<td>8</td>
<td>320.1586</td>
<td>2</td>
<td>YPTMK</td>
<td>67.92</td>
</tr>
</tbody>
</table>
**Table A2.4** Elderberry juice peptide sequences with ALC scores greater than 60 for Group B.

<table>
<thead>
<tr>
<th>No.</th>
<th>M/Z</th>
<th>Charge</th>
<th>DENovo</th>
<th>ALC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>336.1234</td>
<td>2</td>
<td>EDHACP</td>
<td>70.2</td>
</tr>
<tr>
<td>2</td>
<td>357.6659</td>
<td>2</td>
<td>MAHAGLD</td>
<td>71.81</td>
</tr>
<tr>
<td>3</td>
<td>428.1912</td>
<td>2</td>
<td>SSGPAGGHW</td>
<td>71.25</td>
</tr>
<tr>
<td>4</td>
<td>472.2460</td>
<td>2</td>
<td>ATGSGNVQLP</td>
<td>73.37</td>
</tr>
<tr>
<td>5</td>
<td>472.2461</td>
<td>2</td>
<td>DGSGGNVKPL</td>
<td>73.76</td>
</tr>
<tr>
<td>6</td>
<td>678.8597</td>
<td>2</td>
<td>PSTSDGLPVTLRG</td>
<td>71.03</td>
</tr>
<tr>
<td>7</td>
<td>332.5206</td>
<td>3</td>
<td>RGTSPKVGH</td>
<td>62.73</td>
</tr>
<tr>
<td>8</td>
<td>497.7386</td>
<td>2</td>
<td>DGGLDHPAR</td>
<td>61.3</td>
</tr>
<tr>
<td>9</td>
<td>368.6776</td>
<td>2</td>
<td>ESGTNKT</td>
<td>63.56</td>
</tr>
<tr>
<td>10</td>
<td>323.5389</td>
<td>3</td>
<td>TTPKKVVPV</td>
<td>66.92</td>
</tr>
<tr>
<td>11</td>
<td>497.7383</td>
<td>2</td>
<td>KVHNMPDGPN</td>
<td>60.67</td>
</tr>
<tr>
<td>12</td>
<td>328.1086</td>
<td>2</td>
<td>ATGCCCGS</td>
<td>60.95</td>
</tr>
<tr>
<td>13</td>
<td>405.0637</td>
<td>2</td>
<td>CCCDGCCC</td>
<td>73.84</td>
</tr>
<tr>
<td>14</td>
<td>478.1087</td>
<td>2</td>
<td>YDCDMCCC</td>
<td>60.01</td>
</tr>
<tr>
<td>15</td>
<td>357.6664</td>
<td>2</td>
<td>MAHAGLD</td>
<td>65.71</td>
</tr>
<tr>
<td>16</td>
<td>457.2598</td>
<td>2</td>
<td>LVEAAPGLSG</td>
<td>67.26</td>
</tr>
<tr>
<td>17</td>
<td>629.3033</td>
<td>2</td>
<td>HSNKTVPAMWS</td>
<td>62.44</td>
</tr>
<tr>
<td>18</td>
<td>357.6658</td>
<td>2</td>
<td>MAGQNP</td>
<td>68.99</td>
</tr>
<tr>
<td>19</td>
<td>357.6656</td>
<td>2</td>
<td>MAYSSR</td>
<td>68.90</td>
</tr>
<tr>
<td>20</td>
<td>471.2622</td>
<td>2</td>
<td>LNGNGTLPR</td>
<td>64.63</td>
</tr>
<tr>
<td>21</td>
<td>472.2463</td>
<td>2</td>
<td>NMAGGRQLP</td>
<td>62.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-------</td>
<td>---</td>
<td>----------------</td>
<td>------</td>
</tr>
<tr>
<td>22</td>
<td>472.2459</td>
<td>2</td>
<td>TASGGNVQLP</td>
<td>64.53</td>
</tr>
<tr>
<td>23</td>
<td>522.7886</td>
<td>2</td>
<td>RTPADGSVNK</td>
<td>66.67</td>
</tr>
<tr>
<td>24</td>
<td>522.7887</td>
<td>2</td>
<td>RTGNLATALQ</td>
<td>64.05</td>
</tr>
<tr>
<td>25</td>
<td>472.2461</td>
<td>2</td>
<td>DGGSGNVKPL</td>
<td>69.18</td>
</tr>
<tr>
<td>26</td>
<td>472.2466</td>
<td>2</td>
<td>DGGSGNVKPL</td>
<td>66.17</td>
</tr>
<tr>
<td>27</td>
<td>428.1910</td>
<td>2</td>
<td>SSGPANHW</td>
<td>69.41</td>
</tr>
<tr>
<td>28</td>
<td>546.1499</td>
<td>2</td>
<td>PTCGCCAGACCC</td>
<td>63.40</td>
</tr>
<tr>
<td>29</td>
<td>310.0999</td>
<td>2</td>
<td>ECHGCA</td>
<td>62.19</td>
</tr>
<tr>
<td>30</td>
<td>381.1158</td>
<td>2</td>
<td>YGDCCGSG</td>
<td>65.37</td>
</tr>
<tr>
<td>31</td>
<td>529.1376</td>
<td>2</td>
<td>YDCCCSASCC</td>
<td>64.10</td>
</tr>
<tr>
<td>32</td>
<td>689.3488</td>
<td>2</td>
<td>YYRPAAADHW</td>
<td>62.21</td>
</tr>
<tr>
<td>33</td>
<td>381.0850</td>
<td>2</td>
<td>CDACCCSG</td>
<td>61.81</td>
</tr>
<tr>
<td>34</td>
<td>543.1358</td>
<td>2</td>
<td>DEECCHCCSG</td>
<td>60.66</td>
</tr>
<tr>
<td>35</td>
<td>398.0957</td>
<td>2</td>
<td>YSSCCCN</td>
<td>62.48</td>
</tr>
<tr>
<td>36</td>
<td>513.7769</td>
<td>2</td>
<td>GRLTHEDKA</td>
<td>67.09</td>
</tr>
<tr>
<td>37</td>
<td>303.1001</td>
<td>2</td>
<td>SHSGDGC</td>
<td>61.06</td>
</tr>
<tr>
<td>38</td>
<td>455.1508</td>
<td>2</td>
<td>MDCSVDHCC</td>
<td>61.64</td>
</tr>
<tr>
<td>39</td>
<td>454.1415</td>
<td>2</td>
<td>MMMCNYD</td>
<td>68.50</td>
</tr>
<tr>
<td>40</td>
<td>414.1455</td>
<td>2</td>
<td>MGVDSSCE</td>
<td>60.51</td>
</tr>
<tr>
<td>41</td>
<td>399.0947</td>
<td>2</td>
<td>CHDGGCCCG</td>
<td>61.65</td>
</tr>
<tr>
<td>42</td>
<td>504.1649</td>
<td>2</td>
<td>YMYCCHMG</td>
<td>60.06</td>
</tr>
<tr>
<td>43</td>
<td>365.1260</td>
<td>2</td>
<td>AMCCASGS</td>
<td>68.16</td>
</tr>
<tr>
<td>44</td>
<td>536.2009</td>
<td>2</td>
<td>PTMTEMDCGS</td>
<td>61.54</td>
</tr>
<tr>
<td>45</td>
<td>618.2335</td>
<td>2</td>
<td>MGGPNTSSHCGGM</td>
<td>60.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>--------</td>
<td>---</td>
<td>---------</td>
<td>----</td>
</tr>
<tr>
<td>46</td>
<td>455.1838</td>
<td>2</td>
<td>MMSSCLTH</td>
<td>60.05</td>
</tr>
<tr>
<td>47</td>
<td>361.2049</td>
<td>2</td>
<td>LKKVMC</td>
<td>67.29</td>
</tr>
<tr>
<td>48</td>
<td>589.7829</td>
<td>2</td>
<td>CELPVLEQTF</td>
<td>61.51</td>
</tr>
<tr>
<td>49</td>
<td>304.0888</td>
<td>2</td>
<td>HAACCC</td>
<td>64.55</td>
</tr>
<tr>
<td>50</td>
<td>500.1712</td>
<td>2</td>
<td>YMCGEFF</td>
<td>63.46</td>
</tr>
<tr>
<td>51</td>
<td>557.1770</td>
<td>2</td>
<td>APCGCCVWDG</td>
<td>61.12</td>
</tr>
<tr>
<td>52</td>
<td>365.6508</td>
<td>2</td>
<td>YNDVYG</td>
<td>68.38</td>
</tr>
<tr>
<td>53</td>
<td>379.1231</td>
<td>2</td>
<td>GMCMGDGS</td>
<td>63.09</td>
</tr>
<tr>
<td>54</td>
<td>570.2802</td>
<td>2</td>
<td>WGVFAANEFG</td>
<td>63.80</td>
</tr>
<tr>
<td>55</td>
<td>540.1861</td>
<td>2</td>
<td>SHWGCMEED</td>
<td>64.47</td>
</tr>
<tr>
<td>56</td>
<td>527.2130</td>
<td>2</td>
<td>EEPCEAGRY</td>
<td>63.29</td>
</tr>
<tr>
<td>57</td>
<td>315.1779</td>
<td>2</td>
<td>MLKTH</td>
<td>65.44</td>
</tr>
<tr>
<td>58</td>
<td>438.2373</td>
<td>2</td>
<td>RATKAAGSN</td>
<td>61.16</td>
</tr>
<tr>
<td>59</td>
<td>409.1599</td>
<td>2</td>
<td>MTTPGDPC</td>
<td>62.05</td>
</tr>
<tr>
<td>60</td>
<td>319.1409</td>
<td>2</td>
<td>DSKSGGS</td>
<td>60.89</td>
</tr>
<tr>
<td>61</td>
<td>362.1406</td>
<td>2</td>
<td>NQAGSCSG</td>
<td>61.07</td>
</tr>
<tr>
<td>62</td>
<td>517.2629</td>
<td>2</td>
<td>YGRRATYF</td>
<td>61.48</td>
</tr>
<tr>
<td>63</td>
<td>362.1413</td>
<td>2</td>
<td>TSTGGGED</td>
<td>62.44</td>
</tr>
<tr>
<td>64</td>
<td>363.2208</td>
<td>2</td>
<td>KKDVHV</td>
<td>66.63</td>
</tr>
<tr>
<td>65</td>
<td>416.1786</td>
<td>2</td>
<td>QCHGGSVGS</td>
<td>64.63</td>
</tr>
<tr>
<td>66</td>
<td>456.2362</td>
<td>2</td>
<td>KKHSGGGPSG</td>
<td>65.52</td>
</tr>
<tr>
<td>67</td>
<td>333.1389</td>
<td>2</td>
<td>YMGASH</td>
<td>62.52</td>
</tr>
<tr>
<td>68</td>
<td>328.1347</td>
<td>2</td>
<td>NGCAGPH</td>
<td>60.44</td>
</tr>
<tr>
<td>69</td>
<td>309.0813</td>
<td>2</td>
<td>MTGCCC</td>
<td>61.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>--------</td>
<td>----------</td>
<td>-----------------</td>
<td>------</td>
</tr>
<tr>
<td>70</td>
<td>301.0926</td>
<td>FDGGCC</td>
<td>66.66</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>430.2498</td>
<td>FKATAPPK</td>
<td>68.72</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>315.2166</td>
<td>LLAVTL</td>
<td>60.02</td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>376.1566</td>
<td>TCCSAPGL</td>
<td>62.17</td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>302.1472</td>
<td>YGVGHA</td>
<td>68.29</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>376.1566</td>
<td>SCCTAAVP</td>
<td>65.83</td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>345.1477</td>
<td>FHPSDS</td>
<td>64.60</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>345.1473</td>
<td>FHPSDS</td>
<td>67.98</td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>541.1912</td>
<td>MMDGHAGCPY</td>
<td>68.68</td>
<td></td>
</tr>
</tbody>
</table>

The average local confidence (ALC) score for the de novo sequencing of the peptides in Group A and B are shown as Figure A2.6 A and B, respectively. ALC considers each individual amino acid score and is divided by the number of amino acids. The ALC is simply reporting the number of correctly identified sequenced peptides, whether there are 3 or 30 amino acids. There is good confidence in peptides with an ALC score > 30 %. And the manufacturer of the PEAKS program states that ALC scores >55 % have the highest confidence. This does not mean that the whole sequence is the correct sequence, as the beginning and end of the MS/MS spectrum are more difficult to interpret when it comes to de novo sequencing due to decreased signal-to-noise.

In this experiment, the number of peptides detected in Group B was less than Group A. In Group B, the de novo sequencing rate and number of high confidence-level peptides (ALC > 30 %) is greater than Group A. Group B was subjected to C18 SPE cleanup prior to analysis. This C18 cleanup resulted in fewer measured peptides, but raised the corresponding MS signal, and led to higher
confidence-level peptide sequences. The total number of peptides detected was
more than one thousand and is a larger number than animal or human tissue
peptidome results reported in previous studies. The combination of
cation-exchange MCX and C_{18} SPE cleanup proved to be an effective way to
separate peptides from anthocyanins abundant in the elderberry juice.

We used a relatively high temperature to denature the enzymes from sample
itself and some loss of peptides due to protein self-degradation or enzyme
digestion is unavoidable [198].

Figure A2.6 A. De novo ALC scores of Group A, MS/MS Scans: 2530.
Peptides after filter: 1789. B. De novo ALC scores of Group B, MS/MS Scans:
A2.3.4 Protein sequence similarity searches

A BLAST search using the peptide information (ALC >60) found 56 proteins from the Group A and 114 proteins were found for Group B. Proteins with high confidence levels are matched to proteins from several kinds of plants. The number of founded proteins is considerable and larger than some other methods of peptidomics [199]. The function of the proteins and peptides needs further characterization and is beyond the scope of the present study.

A2.4 Chapter summary

Biologically active peptides play a role in plant signaling and defense. Elderberry juice is known to contain a variety of anthocyanin compounds, a sub-set of polyphenols, which are responsible for the deep purple color of the juice. In this paper, we describe a method utilizing solid phase extraction (SPE) to separate anthocyanins and peptides, followed by HPLC using a C_{18} column to purify the peptides. Liquid chromatography in combination with tandem mass spectrometry is used to identify the peptides and their sequences. We successfully identified more than 1,000 peptides present in elderberry juice. These results show that the use of SPE is an effective method to separate peptides and anthocyanins from samples with high polyphenol content.
Conclusion of appendix chapters

Elderberry juices have produced anti-influenza activity in human clinical. The juice from elderberry fruit is known to contain a variety of polyphenols and peptides. Waters Xevo TQ-S UPLC-MS/MS system and LTQ Orbitrap XL mass spectrometer were used to analyze polyphenols and peptides respectively. In this study, juice samples are examined in a profile mode. Multiple reaction monitoring was selected for the reduction of false positives, maximizing selectivity, and reliable quantification. The results showed that this method was helpful to characterize and profile the anthocyanins composition of elderberry juices for quality control, assessment of dietary intake, and anthocyanins biochemical studies. The total phenolic and total monomeric anthocyanin concentrations of elderberry were found to vary highly among different cultivars. There was also described a method utilizing solid phase extraction to separate anthocyanins and peptides, followed by HPLC using a C18 column to purify the peptides. It was successfully identified more than 1,000 peptides present in elderberry juice.
Reference


24. Clantin, B.; Hodak, H.; Willery, E.; Locht, C.; Jacob-Dubuisson, F.; Villeret, V., The crystal structure of filamentous hemagglutinin secretion domain and


33. Iliuk, A.; Galan, J.; Tao, W. A. Playing tag with quantitative proteomics.
Analytical and Bioanalytical Chemistry, 2009, 393, 503-513.


42. Soufi, B.; Kumar, C.; Gnad, F.; Mann, M.; Mijakovic, I.; Macek, B. Stable


novo peptide sequencing based on phosphorylation strategy and mass

79. Chen, J.; Chen, Y.; Gong, P.; Jiang, Y.; Li, Y. M.; Zhao, Y. F. Novel phosphoryl
derivatization method for peptide sequencing by electrospray ionization mass
531-536.

Xu, P. X.; Zhao, Y. F. An isotope ($^{18}$O, $^{15}$N, and $^2$H) technique to investigate
the metal ion interactions between the phosphoryl group and amino acid side
chains by electrospray ionization mass spectrometry. *Journal of

81. Ai, H. W.; Fu, H.; Zhao, Y. F. Novel synthetic method of phosphonamidate
peptides and its application in peptide sequencing via multistage mass

82. Bao, J. Y.; Ai, H. W.; Fu, H.; Jiang, Y. Y.; Zhao, Y. F.; Huang, C. Peptide
sequencing through N-terminal phosphorylation and electrospray ionization

83. Sun, X. B.; Kang, J. X.; Zhao, Y. F. One-pot synthesis of hydrogen
phosphonate derivatives of d4T and AZT. *Chemical Communications*, **2002**, 
2414-2415.

84. Hunter, T. Protein kinases and phosphatases: the yin and yang of protein

1173-1178.


100. Paizs, B.; Suhai, S. Fragmentation pathways of protonated peptides. *Mass
Spectrometry Reviews, 2005, 24, 508-548.


108. Zhang, J.; Wang, Y., Deuterium isobaric amine-reactive tags for quantitative


114. Sun, S.; Wang, Q.; Zhao, F.; Chen, W.; Li, Z., Prediction of Biological Functions on Glycosylation Site Migrations in Human Influenza H1N1


1851-1858.


150. He, B.; Gross, M.; Roizman, B., The gamma(1)34.5 protein of herpes simplex virus 1 complexes with protein phosphatase 1alpha to dephosphorylate the alpha subunit of the eukaryotic translation initiation factor 2 and preclude the shutoff of protein synthesis by double-stranded RNA-activated protein kinase. *Proceedings of the National Academy of Sciences USA*, 1997, 94, 843-848.


152. Vester, D.; Rapp, E.; Klugeb, S.; Genzel, Y.; Reichl, U., Virus–host cell interactions in vaccine production cell lines infected with different human


174. Nicoletti, I.; Bello, C.; Rossi, A. D.; Corradini, D., Identification and Quantification of Phenolic Compounds in Grapes by HPLC-PDA-ESI-MS on a Semimicro Separation Scale. *Journal of Agricultural and Food Chemistry*,


Curriculum vitae

Academic qualifications of the thesis author, Mr. WU Hanzhi

• Received the degree of Bachelor of Science from Fuzhou University, Fuzhou, China, July 1999;

• Received the degree of Master of Science from Xiamen University, Xiamen, China, November 2004.

January 2015
Outcome of the thesis work

A. Publications


B. Conference and symposium presentations

1. HKSMS workshop 2014 March “Bioinformatics and Data Analysis”.

2. HKSMS Programme of the HKSMS Symposium and the 14th AGM on 16 June 2012.

3. HKPS 2011 proteomics workshop on 7 and 8 Jan 2011.

4. HKSMS Workshop 2010 - Proteomics by Mass Spectrometry.