New Approaches to Trace LC/MS Analysis of Biological Samples

by Siyang Li
B.S. in Chemistry, China Agricultural University

A dissertation submitted to

The Faculty of
the College Science of
Northeastern University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

December 12, 2013

Dissertation directed by

Barry L. Karger
Professor of Chemistry and Chemical Biology
Director of Barnett Institute
Acknowledgements

I want to express my sincere gratitude to the following professors, lab mates, friends and many people who supported me to complete this dissertation. I know it is impossible for me to credit all of them.

First of all, I would like to acknowledge my advisor, Dr. Barry Karger, who accepted me as a Ph.D student and led me into his leading research area. Under his careful and patient guidance, I matured as a scientist in this area. I will never forget the importance of being “enthusiastic and responsible” in research, and this principle will benefit me throughout my career no matter whether I am in academics or industry.

Many thanks to Dr. Alexander Ivanov, Dr. Billy Wu, Dr. Marina Hincapie and Dr. Tomas Rejtar who advised me on my projects. I gratefully thank Dr. Shujia (Daniel) Dai, Dr. Dongdong Wang, Mr. Somak Ray for their advice in the discussion of my experiments.

I would like to give my thanks to my dissertation committee members, Profs. William Hancock, Roger Giese and Paul Vouros for their valuable time and suggestions on my dissertation and defense.

Many thanks to the former and current lab mates working with me in the past years, from both Dr. Karger’s and Dr. Hancock’s groups: Dr. Dipak Thakur, Dr. Zhenke (Jack) Liu, Dr. Chen Li, Dr. Wenqin Ni, Dr. Yue Zhang, Dr. Fangfei Yan, Dr. Suli Liu, Dr. Zhi Zeng, Xianzhe Wang, Siyuan Liu, for their help, encouragement and friendship.

I would also want to express my gratitude to the staffs of the Department of Chemistry and Chemical Biology and the Barnett Institute: Dr. Graham Jones, Dr. David A. Forsyth, Nancy
Carbone, Jeffrey Kesilamn, Felicia Hopkins, Jean Harris, Andrew Bean, Cara Shockley, and Richard Pumphrey for their help during my graduate study.

Last but not least, I want to dedicate this dissertation to my parents. I am grateful for their selfless love and nurture in my life. I could not make this achievement without all their heart and love of me.
Abstract of Dissertation

Many clinical samples from human or animals are available in limited amounts. For example, the number of cell type subpopulations (e.g. stem cells or circulating tumor cells), which are related to cancer may be very low in whole blood. Additionally, rare blood proteins of diagnostic interest may need to be isolated from numerous blood proteins with a very wide dynamic range. The analysis of such limited sample availabilities requires an integration of highly specific sample preparation with high performance and sensitive liquid chromatography/mass spectrometry (LC/MS). The purpose of this dissertation is to develop new LC/MS approaches for trace analysis of a limited number of cells as low as few hundred and targeted blood proteins in low levels.

In Chapter 1, current analytical technologies and approaches in the trace analysis of limited samples reviewed. The significances and benefits of gradient elution and nano flow LC are discussed in detail. In addition, proteomic and bioinformatic strategies for identification and quantitation of large set of proteins are discussed. Finally, technologies in preparing limited clinical samples are presented.

In Chapter 2, the performance of 10 µm i.d. porous layer open tubular (PLOT) columns, based on the polymer, poly(styrene-divinylbenzene) (PS-DVB), was studied at the ultralow flow rate of less than 20 nL/min. Increased sensitivity over higher flow rates was observed for electrospray in nanoLC/MS for peptide determination. A model of the PLOT column in gradient LC separation was established based on the band broadening and solvent-strength theories, which was used in the optimization of PLOT in proteomic separation of complex samples. Further, the
nanoLC/MS PLOT platform was demonstrated to be a more efficient and sensitive approach in one dimensional or two dimensional analysis of limited amounts of cell lysate in comparison to packed columns at a conventional nanoflow rates. In addition, attempts were made to improve the LC performance and hydrophobicity of PLOT columns, such as using even narrower bore columns, incorporating carbon nanotubes and co-polymerizing alkyl monomer with the PS-DVB.. Improved loading capacity and efficiency were demonstrated in the modified monolithic SPE columns.

In Chapter 3, the ultra-performance and sensitive nanoLC/MS platform was integrated with Adaptive Focused Acoustics™ (AFA) sample preparation technology and advanced microfluidic cell capture methods. The PLOT column coupled with a high resolution/accurate mass (HR/AM) Q Exactive mass spectrometer showed a impressive depth in the microproteomic analysis of limited cell lysates. The sample preparation approach, including cell lysis and digestion, was optimized based on AFA ultrasonication in one tube without sample dilution and transfer. High recovery in the preparation of limited starting materials as low as 2,000 cells was shown. A microfluidic device using magnetophoretic cell isolation is demonstrated as a specific and quantitative approach in capturing 1,000 to 10,000 MCF-7 cells. To illustrate the potential clinical use of the approach, endothelial progenitor cells (EPCs) and hematopoietic stem cells (HSCs) were isolated from human whole blood drawn from 5 non-smokers and 2 smokers. Systematic biological differences are shown.

In Chapter 4, in collaboration with Alnylam Pharmaceutical, a fast and quantitative LC/MS method was established to study the pharmacodynamic (PD) behavior of siRNA drugs in reducing the level of the peptide hormone, hepcidin, in blood. A fast and high throughput and
precise hepcidin extraction method was achieved using solid phase extraction on an automated liquid handling platform. For quantitation, an LC-MS method, using SRM approach on a QQQ mass spectrometer, was developed to measure hepcidin from monkey and mouse sera. In the PD study of monkey sera, it was shown that high doses of siRNA drugs could efficiently suppress the hepcidin level. In addition to the SRM approach, an alternative quantitative approach using targeted-MS/MS mode on Q Exactive resulted in a more than 10-fold increase in sensitivity and specificity, which is helpful for future preclinical studies where less sample may be available, particularly for smaller size of animals such as mice.
# Table of Contents

Acknowledgements ........................................................................................................................... ii

Abstract of Dissertation .................................................................................................................... iv

Table of Contents ............................................................................................................................... vii

List of Figures ..................................................................................................................................... xii

List of Tables ...................................................................................................................................... xvii

List of Abbreviations ......................................................................................................................... xvii

Chapter 1 Overview of Analytical Methods and Technologies on Trace Analysis of Limited Biological Samples ............................................................................................................................. 1

Abstract ............................................................................................................................................... 2

1.1 High-performance liquid chromatography and Mass Spectrometry ........................................... 3

1.1.1 High-Performance Liquid Chromatography (HPLC) ................................................................. 3

1.1.2 Mass Spectrometry ...................................................................................................................... 18

1.2 Proteomics ..................................................................................................................................... 30

1.2.1 Overview of Proteomics .............................................................................................................. 30

1.2.2 Protein Quantitation .................................................................................................................... 30

1.2.3 Bioinformatics ............................................................................................................................ 36

1.3 Sample Preparation of Limited Biological Samples ......................................................................... 40

1.3.1 Sample Collection ..................................................................................................................... 40

1.3.2 Sample Purification/Isolation ..................................................................................................... 43

1.3.3 Sample Digestion ........................................................................................................................ 45

1.4 Pharmacokinetics (PK) and Pharmacodynamics (PD) of Biotech Drug ......................................... 46
Chapter 2 The Fundamental Theory of Porous Layer Open Tubular Column and Optimization of LC Separation ...................................................................................................................... 65

Abstract .................................................................................................................................................................................. 66

2.1 Introduction ......................................................................................................................................................................... 67

2.2 Experimental ......................................................................................................................................................................... 69

2.2.1 Materials and Reagents .................................................................................................................................................. 69

2.2.2 PLOT Column Preparation ........................................................................................................................................... 69

2.2.3 Monolithic Column (SPE) Preparation ............................................................................................................................ 71

2.2.4 PLOT Columns Platforms ............................................................................................................................................... 72

2.2.4 LC and Mass Spectrometry .......................................................................................................................................... 77

2.3 Results and Discussion .......................................................................................................................................................... 79

2.3.1 ESI-MS Sensitivity at Ultralow nL/min Flow Rates ...................................................................................................... 79

2.3.2 PLOT Column Efficiency at Different Flow Rates .......................................................................................................... 80

2.3.3 Theoretical Study of Peak Capacity Optimization of PLOT Column ............................................................................. 84

2.3.4 On-line 1D SPE-PLOT Analysis ..................................................................................................................................... 92

2.3.5 On-Line 2D SCX-PLOT/LC-ESI-MS Analysis .................................................................................................................... 96

2.3.6 Hydrophobicity Modified PLOT Columns ....................................................................................................................... 97

2.3.7 Narrower Bore PLOT Columns ....................................................................................................................................... 103

2.3.7 Alkyl Monomers Modified Monolithic Columns ........................................................................................................... 107
Chapter 3 Ultrasensitive Microproteomic Profiling of Rare-cell Populations Using Microfluidic Magnethophoretic Isolation from Whole Blood followed by Porous Layer Open Tubular Liquid Chromatography - Tandem Mass Spectrometer

Abstract

3.1 Introduction

3.2 Methods

3.2.1 Cell Handling

3.2.2 Microfluidic Device Design and Fabrication

3.2.3 Microparticle Modification

3.2.4 Microfluidic Capture of Spiked MCF-7 cells

3.2.5 Whole Blood Cancer Cell Isolation

3.2.6 Isolation of Hematopoietic Stem Cells and Endothelial Progenitor Cells from Whole Blood

3.2.7 Sample Processing Prior to Lysis and LC/MS

3.2.8 Cell Lysis and Protein Digestion

3.2.9 Preparation of PLOT and Monolithic Columns

3.2.10 PLOT Nanoflow Liquid Chromatography-Tandem Mass Spectrometry (PLOT nanoLC-MS/MS)

3.2.11 Database Search

3.2.12 Gene Ontology (GO) Annotation Analysis
4.2.6 LC-MS for QQQ study ................................................................. 180
4.2.7 Data process for Q-Exactive method ........................................ 181
4.2.8 Data process for QQQ method .................................................... 181
4.3 Results and Discussion ................................................................. 182
  4.3.1 Development of hepcidin extraction from serum ....................... 182
  4.3.2 Development of LC-MS method ................................................ 183
  4.3.3 Establishment of linear curves, LOD and LOQ ............................ 184
  4.3.4 Detection of endogenous hepcidin in monkey and mouse species .... 191
  4.3.5 Determination of the siRNA drug dosage ................................... 191
  4.3.6 Consistency of siRNA drug effect .......................................... 194
  4.3.7 Development of LC-MS method using Q-Exactive ....................... 196
4.4 Conclusions ................................................................................. 204
4.5 References ................................................................................. 205
Copyrights ....................................................................................... 207
List of Figures

Figure 1.1. General elution problem in isocratic condition (A) and the improved separation by gradient elution (B) ............................................................................................................................................... 7

Figure 1.2. Flow rate vs. ESI signal intensity .......................................................................................................................................................................................... 10

Figure 1.3. Eddy diffusion in the van Deemter equation ......................................................................................................................................................... 12

Figure 1.4. van Deemter plot of HETP as function of linear velocity .................................................................................................................................. 15

Figure 1.5. Scanning electron micrographs (SEM) of the 10-µm-i.d. PS- DVB PLOT column .......................................................................................... 17

Figure 1.6. Ionization of Electrospray in conventional flow and nanoflow ......................................................................................................................... 20

Figure 1.7. Schematic diagram of a LIT consisting of three quadrupole ion traps ........................................................................................................ 23

Figure 1.8. Three-dimensional schematic of an Orbitrap cell ........................................................................................................................................... 26

Figure 1.9. Chemical bond cleavages in MS/MS Fragmentations ................................................................................................................................. 28

Figure 1.10. Schematic of the hybrid LTQ (a) and LTQ Orbitrap XL with octopole collision cell (b) .................................................................................................................................................................................. 29

Figure 1.11. Design of a tandem mass tag (TMT) ................................................................................................................................................................. 32

Figure 1.12. Ranked protein abundances calculated by iBAQ intensities ....................................................................................................................... 34

Figure 1.13. Bioinformatics analysis paths for MS-based proteomics ............................................................................................................................. 36

Figure 1.14. Examples of CTC capture technologies that combine multiple enrichment principles .................................................................................................................. 42

Figure 1.15. Schematic of a laser capture microdissection process .............................................................................................................................. 42

Figure 1.16. Illustration of two targeted analysis approaches ........................................................................................................................................ 48
Figure 2.1. Schematic diagram of split-injection PLOT design. ......................................................... 73
Figure 2.2. Schematic diagram of SPE-PLOT 1D system. ................................................................. 75
Figure 2.3. Schematic diagram of (PR-SCX)-SPE-PLOT 2D system. ............................................. 76
Figure 2.4. Plots of absolute signal intensities of angiotensin I (green, round) and bradykinin 2-9 fragment (red, square) gain as a function of flow rate on a 3m PLOT column. ......................... 80
Figure 2.5. van Deemter plot of 10 μm i.d. PLOT and 20 μm i.d. packed C_{18} columns with 1.7 μm beads at ultra-nano flow ........................................................................................................ 84
Figure 2.6. Prediction of effects of different parameters on peak capacity. ................................. 90
Figure 2.7. Separation of standard peptide mixture by SEP-PLOT 1D system ................................ 94
Figure 2.8. Analysis of 500 MCF-7 cells using 4.2m PLOT column and 4 hr long gradient ......... 95
Figure 2.9. Advantage of SPE-PLOT system over 75 μm i.d. C_{18} column ................................. 95
Figure 2.10. 2D SCX-SPE-PLOT analysis of 500 MCF-7 cells and the comparison to 1D analysis ........................................................................................................................................ 96
Figure 2.11. Incorporation of carbon nanotubes into PLOT column. 500 amol 6 bovine protein digest was split-injected on each 3.7 m PLOT column. ............................................................... 100
Figure 2.12. PS-DVB surface modified PLOT columns with alkyl-alkene. 500 amol 6 bovine protein digest was split-injected on each 3.7 m PLOT column. .................................................. 102
Figure 2.13. 2nd layer PS-DVB PLOT column. 500 amol 6 bovine protein digest was split-injected on each 3.7 m PLOT column. .......................................................... 104
Figure 2.14. Changes in the thickness of the porous PS-DVB layer in a 5 μm i.d. capillary by varying the ratio of monomers to porogen ................................................................. 106
Figure 2.15. PS-DVB monolithic column modified with alkyl-alkene ........................................ 109
Figure 2.16. Effect of pore size of monolithic polymer on the column efficiency ......................... 110
Figure 3.1. Performance characterization of a porous layer open tubular column coupled with a Q Exactive mass spectrometer .......................................................... 136

Figure 3.2. Parallel product ion monitoring of 8 targeted precursor ions from 6 bovine mix digest. .................................................................................................................. 137

Figure 3.3. Proteomic analysis of limited amount of MCF-7 cell lysate prepared by a conventional urea-based lysis and in-solution digestion. ...................................................... 140

Figure 3.4. Improved protein IDs in different optimized sample preparation approaches for limited starting material analysis in comparison to conventional sample preparation from millions cells. ......................................................................................... 144

Figure 3.5. Optimization of sample preparation for analysis of limited samples. ............... 145

Figure 3.6. Work flow of proteomic analysis of limited amount cells captured by microfluidic device. .................................................................................................................. 146

Figure 3.7. Evaluation of the microfluidic cell capture approach using proteomic profiling of MCF-7 cells isolates from blood (“spiked-in MCF-7 cells in blood”) or positive control (“MCF-7 cells in DPBS buffer”) and proteomes recovered in negative control experiments. ............... 150

Figure 3.8. Titration analysis of spiked-in MCF-7 cells captured by the magnetophoretic microfluidic device. ........................................................................................................ 154

Figure 3.9. iBAQ Label-free Quantification........................................................................ 155

Figure 3.10. Proteomic analysis of HSCs and ECPs isolated from healthy non-smoking and smoking donors. ........................................................................................................ 158

Figure 3.11. IPA analysis of associated network functions in dermatological diseases and Conditions, organismal injury and abnormalities and inflammatory disease. ................. 159
Figure 4.1. Differences in the hepcidin sequences between mouse, human and monkey species.
.................................................................................................................................................. 185
Figure 4.2. SRM method development using QQQ for detection of monkey hepcidin. ........... 185
Figure 4.3. SRM method development using QQQ for detection of mouse hepcidin............. 186
Figure 4.4. Using optimized SRM method developed in QQQ as m/z 564.3 (precursor ion) and
m/z 120.0 (product ion) at 24 eV collision energy for detection of monkey hepcidin at low levels..
.................................................................................................................................................. 187
Figure 4.5. Linear regression curve of monkey hepcidin standards measured from 0.5 ng/mL to
250 ng/mL by the optimized SRM method described in Figure 4.2........................................... 188
Figure 4.6. Linear calibration curve for mouse standards in rabbit blood at 0.2 ng/mL to 250
ng/mL concentration measured using QQQ. ................................................................. 188
Figure 4.7. Endogenous hepcidin levels in 40 cynomolgus monkeys and 40 mice (measured by
the optimized SRM method described in Figure 4.2 and Figure 4.3)...................................... 190
Figure 4.8. Effect of hepcidin reduction by three different doses of siRNA drug measured at 0,
24hr, and 48 hr after injection of the drugs............................................................................ 192
Figure 4.9. Effect of hepcidin reduction by three different batches of siRNA drugs, and a non-
siRNA drug measured at 1, 2, 4, 7, 10, and 14 days after injection of the drugs. .................. 195
Figure 4.10. SRM method development using Q-Exactive for monkey hepcidin detection. ..... 197
Figure 4.11. SRM method development using Q-Exactive for mouse hepcidin detection....... 198
Figure 4.12. Linear calibration curve for monkey and mouse standards in rabbit blood measured
using target-MS/MS mode on Q Exactive. .............................................................................. 200
Figure 4.13. Using optimized targeted-MS² method described in Figure 4.10 for detection of
monkey hepcidin at low levels. .......................................................................................... 200
Figure 4.14. Effect of hepcidin reduction by one siRNA drugs, and a non-siRNA drug measured at 1, 2, 4, 7, 10, and 14 days after injection of the drugs ........................................................... 201

Figure 4.15. SRM-like validation of precursor ion by multiple product ions in the method development using Q-Exactive for mouse hepcidin detection. Product ions, y_{21}^{2+}, b_{3}^{+} and b_{4}^{+} of mouse hepcidin were extracted in the level of (A) 0.05 ng/mL and (B) 200 ng/mL................. 203
List of Tables

Table 1.1. Classification of different diameters of HPLC columns ........................................ 12
Table 2.1. Comparisons of theoretical plates and peak capacities at different flow rate 91
Table 2.2. Summary of the peptide separation by modified PLOT columns ......................... 106
Table 3.1: Cells Losses During the Isolation and Purification Process ................................. 127
Table 3.2: MCF-7 Cells captured for the quantitative analysis. The numbers of captured MCF-7 cells were counted by flow cytometry. ................................................................. 153
Table 3.3: Numbers of hematopoietic stem cells (HSCs) and endothelial progenitor cells (EPCs) captured from the microfluidic device counted by flow cytometry. ................................. 153
Table 4.1. Precision and accuracy of SRM using QQQ for quantitation of monkey hepcidin standards. ......................................................................................................................... 189
Table 4.2. Precision and accuracy of SRM using QQQ for quantitation of mouse hepcidin standards. ......................................................................................................................... 190
Table 4.3. Dose Effect of siRNA drug in reduction of hepcidin in monkey model ............... 193
Table 4.4. Precision and accuracy of SRM using Q-Exactive for quantitation of monkey hepcidin standards. ......................................................................................................................... 199
Table 4.5. Quantitation of spiked-in monkey hepcidin level in rabbit serum by Q-Exactive .... 201
**List of Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D</td>
<td>three-dimensional</td>
</tr>
<tr>
<td>ACD</td>
<td>anemia of chronic disease</td>
</tr>
<tr>
<td>ACN</td>
<td>acetonitrile</td>
</tr>
<tr>
<td>AFA</td>
<td>Adaptive Focused Acoustics</td>
</tr>
<tr>
<td>AGC</td>
<td>automatic gain control</td>
</tr>
<tr>
<td>CID</td>
<td>collision induced dissociation</td>
</tr>
<tr>
<td>CM</td>
<td>mass transfer from mobile phase to stationary phase</td>
</tr>
<tr>
<td>CNT</td>
<td>carbon nanotube</td>
</tr>
<tr>
<td>CS</td>
<td>mass transfer stationary phase to mobile phase</td>
</tr>
<tr>
<td>CTC</td>
<td>circulating tumor cell</td>
</tr>
<tr>
<td>CV</td>
<td>coefficients of variation</td>
</tr>
<tr>
<td>ELISA</td>
<td>enzyme-linked immunosorbent assays</td>
</tr>
<tr>
<td>EPC</td>
<td>endothelial progenitor cell</td>
</tr>
<tr>
<td>ESI</td>
<td>electrospray ionization</td>
</tr>
<tr>
<td>ETD</td>
<td>electron transfer dissociation</td>
</tr>
<tr>
<td>FA</td>
<td>formic acid</td>
</tr>
<tr>
<td>FASP</td>
<td>filter-aided sample preparation</td>
</tr>
<tr>
<td>FDR</td>
<td>false discovery rates</td>
</tr>
<tr>
<td>FTICR</td>
<td>fourier transform ion cyclotron resonance</td>
</tr>
<tr>
<td>FWHM</td>
<td>mass difference being the full width at half maximum height</td>
</tr>
<tr>
<td>GC</td>
<td>gas chromatography</td>
</tr>
</tbody>
</table>
GO   Gene Ontology
GSEA   Gene Set Enrichment Analysis
HCD   higher energy collision dissociation
HETP   height equivalent to a theoretical plate
HILIC   hydrophilic interaction chromatography
HPLC   high-performance liquid chromatography
HR/AM   high resolution/accurate mass
HSC   hematopoietic stem cell
IAA   iodoacetic acid
iBAQ   intensity-based absolute quantification
ICR   ion cyclotron resonance
IEC   ion exchange chromatography
IEF   isoelectric focusing
IP   immunoprecipitation
IPA   Ingenuity Pathways Analysis
iTRAQ   isobaric tags for relative and absolute quantification
iTRAQ   isobaric tags for relative and absolute quantification
$k$   retention factor
$k^*$   retention factor in gradient elution
LCM   laser capture microdissection
LIT   linear ion trap
LOD   limit of detection
LOQ  limit of quantitation
mAb  monoclonal antibody
MALDI matrix-assisted laser desorption ionization
MRM  multiple reaction monitoring
MS  mass spectrometry
MS/MS tandem mass spectrometry
N  theoretical plate
N*  theoretical plate in gradient elution
NCE  normalized collision energy
NPC  normal phase chromatography
OD  octadecene
PAGE polyacrylamide gel electrophoresis
PC  peak capacity
PD  pharmacodynamics
PK  pharmacokinetics
PLOT porous layer open tubular
PRM  parallel reaction monitoring
PS-DVB poly(styrene-divinylbezene)
PTM posttranslational modifications
QIT  quadrupole ion trap
QQQ  triple quadrupole
Q-TOF quadrupole time-of-flight
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RE</td>
<td>accuracy</td>
</tr>
<tr>
<td>RF</td>
<td>radio frequency</td>
</tr>
<tr>
<td>RPC</td>
<td>reversed phase chromatography</td>
</tr>
<tr>
<td>S/N</td>
<td>signal to noise</td>
</tr>
<tr>
<td>SC</td>
<td>stem cell</td>
</tr>
<tr>
<td>SCX</td>
<td>strong cation exchange</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SEM</td>
<td>scanning electron micrographs</td>
</tr>
<tr>
<td>SILAC</td>
<td>stable isotope labeling by amino acids in cell culture</td>
</tr>
<tr>
<td>siRNA</td>
<td>small interference RNA</td>
</tr>
<tr>
<td>SISCAPA</td>
<td>stable isotope capture by anti-peptide antibodies</td>
</tr>
<tr>
<td>SPE</td>
<td>solid phase extraction</td>
</tr>
<tr>
<td>SRM</td>
<td>selected reaction monitoring</td>
</tr>
<tr>
<td>ST</td>
<td>styrene</td>
</tr>
<tr>
<td>( t_0 )</td>
<td>dead time</td>
</tr>
<tr>
<td>( t_R )</td>
<td>retention time</td>
</tr>
<tr>
<td>W</td>
<td>peak width</td>
</tr>
<tr>
<td>( W_{0.5} )</td>
<td>peak width at half height</td>
</tr>
<tr>
<td>XCorr</td>
<td>cross-correlation</td>
</tr>
<tr>
<td>( \phi )</td>
<td>volume fraction of organic or stronger solvent</td>
</tr>
</tbody>
</table>
Chapter 1

Overview of Analytical Methods and Technologies on Trace Analysis of Limited Biological Samples
Abstract

The trace analysis of proteins in biological or clinical samples is a challenging task for analytical chemists. A current interest in the clinical biological samples is the analysis of subpopulation material, in many cases that the starting amount of target analyte is limited; hence, the development of sensitive and reproducible analytical approaches is essential for the characterization, identification and quantitation of the limited sample availabilities.

The separation performance of chromatography can be improved by using a long analytical column with reduced particle size in the stationary phase or an alternative high performance column, e.g. the porous layer open tubular (PLOT) column our group developed\(^1\). Increased sensitivity in MS detection can be achieved by reducing the mobile phase flow rate to nanoliter/min in liquid chromatography due to more efficient electrospray ionization\(^2,3\). The trace analysis is also facilitated by the new generation of high speed, high resolution tandem mass spectrometry such as the Q Exactive\(^4,5\).

Another aspect of trace analysis is the handling of biological samples. The sample preparation involves sample isolation (e.g. laser capture microdissection)\(^6,7\), sample purification (such as affinity purification)\(^8\) and sample digestion. The preparation approaches need to be optimized to minimize the sample loss in each step. Optimized sample preparation of limited of amounts biological samples ensures the maximum sample recovery for the following detection.

Development of high sensitivity approaches for trace biological sample analysis has many applications in biomarker discovery\(^9\), drug treatment\(^10\) and disease screening\(^11\). This dissertation involves the applications of proteomic analysis of limited human cells and pharmacodynamic study of siRNA drug.
1.1 High-performance liquid chromatography and Mass Spectrometry

1.1.1 High-Performance Liquid Chromatography (HPLC)

HPLC is the predominantly used chromatographic technologies to separate complex mixtures for purification, characterization or quantification of targeted analytes. Compared to very early liquid chromatograph techniques where solvent flow was driven by gravity, modern HPLC is considered a reproducible, robust instrumental technique consisting of pressure pump, dual solvent deliver system and separation column\(^{12}\).

1.1.1.1 HPLC Separation Mode

The fundamental separation mechanism is determined by the choice of various stationary and mobile phases. For trace sample analysis, the predominant separation mode is reversed phase; however, other separation modes are widely used to either assist the reversed phase separation or used in the sample preparation.

**Normal Phase Chromatography (NPC)** In NPC, the stationary is packed with polar, inorganic particles such as alumina and bare silica. The mobile phase is a non-aqueous solvent. The separation is based on polar-polar interaction between analytes and stationary phase. NPC is mainly used in the separation of small polar molecules\(^{13}\). Hydrophilic interaction chromatography (HILIC) is a variant of NPC, in which the hydrophilic stationary phase is surrounded by a water-enriched layer. The retention of the analyte occurs under the mechanism of liquid-liquid partitioning between the water-enriched layer and the hydrophobic aqueous elute. The use of water as the strong elute solvent gives HILIC advantage over NPC that 1) polar analytes have higher solubility in water and 2) the interface with electrospray MS is easier\(^{14}\).
**Ion Exchange Chromatography (IEC)** The particle of stationary phase is modified with charged groups that can bind ions of opposite charges, and the mobile phase is an aqueous buffer with salt at a certain pH. According to the charges on the stationary phase, ion exchange chromatography can be classified into cation exchange or anion exchange. For the separation of an ionized molecule, IEC is a powerful tool, especially for the separation of large biological samples such as peptides and proteins. IEC is widely used in multidimensional separation coupled with reversed phase separation to provide additional separation mechanism and increase the peak capacity of complex mixtures in proteomics study.

**Affinity Chromatography (AC)** is a method to separate a specific biomolecule or biological complex using highly specific interactions between, e.g. antibodies and antigens, enzyme and substrate, and receptor and ligand. Because of the high specificity, affinity chromatography is also used in depleting high abundant proteins (albumin and IgG) in blood and pulling out specific target proteins by binding its antibody on the stationary phase.

**Reversed Phase Chromatography (RPC)** is the dominant separation mode in analysis of biological samples (peptides and proteins). RPC includes a hydrophobic stationary which is generally made up of hydrophobic alkyl chains (C4, C8 or C18) (on porous silica particles) that interact with biomolecules by hydrophobic interaction. A short alkyl chain such as C4 is more appropriate to separate large proteins, and a long alkyl chain C18 is more suitable for the separation of peptides and small molecules. An alternative stationary phase is a synthesized organic polymer, and the long polymer chain can provide hydrophobic surface for the separation of biomolecules. Because of the high efficiency, versatility, flexibility and robustness, RPC is one of the dominant approaches used for the analysis of peptides and proteins.
1.1.1.2 Isocratic and Gradient Separation

**Isocratic Separation**

Isocratic separation is the earliest developed and simplest LC separation mode, in which the composition of the mobile liquid phase remains constant during the separation. For a given solvent, different analytes have different migrations in the separation column and exit the column as a peak band in the elution. To evaluate the efficiency of separation, isocratic elution theory has been established\(^2\).

Retention time \(- t_R\) The retention time \(t_R\) is the time from sample injection to the appearance of the peak maximum of individual analytes. The retention time of solvent or unretained analyte is referred to as dead time \(t_0\),

Retention factor \(- k\). The retention factor \(k\) is an important term in HPLC separation as it is defined as the migration rate of an analyte through a column. Usually, \(k\) is maintained between 2 < \(k\) < 10 for good separation. Small \(k\) < 1 will result in a fast elution but poor resolution; very large \(k\) > 10 will lead to long separation times and wide peak shape\(^1\). The retention factor \(k\) of an analyte can be calculated by Equation (1.1) from the retention time \(t_R\) and dead time \(t_0\),

\[
k = \frac{t_R - t_0}{t_0}
\]

(1.1)

Theoretical plate \(- N\). To improve separation and increase sensitivity, peak band broadening should be minimized to obtain sharp and narrow peaks. Theoretical plate numbers are used to evaluate the efficiency of the separation. The theoretical plate theory assumes that the equilibrium of analytes between the mobile phase and stationary phase is established for the
hypothetical stages. The performance of separation depends on the equilibrium of a series of these hypothetical stages; the higher the plate number \( N \); leads to a better efficiency in a separation. Theoretical plate \( N \) is given as \(^{22} \).

\[
N = 5.54 \left( \frac{t_R}{W_{0.5}} \right)^2 \quad \text{or} \quad N = 16 \left( \frac{t_R}{W} \right)^2 \quad (1.2)
\]

where \( W \), the peak width in time units, can be measured as \( 4\sigma \) of a Gaussian peak or using a peak with at half height, \( W_{0.5} \).

**Gradient Separation**

In many separations of complex samples, isocratic elution could not easily separate all the analytes in a good resolution. The reason is that in isocratic separation, the strength of solvent is constant; therefore, not all the analytes can be separated in an optimum \( k \) range \(^{23} \). The analytes which have weak interaction with the stationary phase will result in poor resolution (peaks 1-3 in Fig. 1.1A) and those having strong interaction will lead to a better resolution but wide peak shape and low intensity (peak 8,9 in Fig. 1.1A). In gradient elution, the composition of the mobile phase is changed with the separation time and the strength of the solvent is thereby increased, resulting in an optimal \( k \) range of all the analytes compared to isocratic separation (Fig. 1.1B) \(^{23} \).
Figure 1.1. General elution problem in isocratic condition (A) and the improved separation by gradient elution (B) (adapted from\textsuperscript{22})

To predict the retention behavior in gradient elution, a linear-solvent-strength model has been established\textsuperscript{24}. In this model, the retention factor $k$ varies with the percentage of strong solvent and can be described as\textsuperscript{24}:

$$\log k = \log k_w - S\phi$$

(1.3)

where $k_w$ is the $k$ value in water or the weakest solvent, $\phi$ is the volume fraction of organic or stronger solvent and $S$ is a constant for a given molecule and condition.

The overall $k$ during gradient separation can be defined as the gradient retention factor, $k^*$, which is the retention factor of the solute when it reaches the middle point of the column. The
value of $k^*$ can be obtained using Equation (1.4). $\Delta\phi$ is the volume fraction change of stronger solvent and $t_G$ is the gradient time which is the retention time window from the first to the last peak.

$$k^* = \frac{t_G}{(1.15 t_0 \Delta\phi S)}$$  \hspace{1cm} (1.4)

The plate number $N$ in gradient elution cannot be calculated directly from Equation 1.2 for isocratic separation. The resulting values of $N$ for gradient elution using Equation 1.2 will be much too large compared to the real plate number. Values of $N$ in gradient elution, $N^*$, should instead be determined from Equation (1.5),

$$N^* = 5.54 \left[ \frac{t_0 \left(1 + 0.5k^*\right)}{W_{0.5}} \right]^2 \quad \text{or} \quad N^* = 16 \left[ \frac{t_0 \left(1 + 0.5k^*\right)}{W} \right]^2$$  \hspace{1cm} (1.5)

Another term to evaluate the overall separation efficiency of complex mixtures is called peak capacity ($PC$), defined as “the maximum number of adjacent peaks with resolution of 1.0 that can be separated in a chromatography”\(^{26}\). The $PC$ provides a more direct comparison of efficiencies between different column formats. Generally, $PC$ is proportional to the plate number and is the upper limit of resolvable components for a given technique under prescribed conditions\(^{27}\). The gradient elution peak capacity can be determined from\(^{25}\):

$$PC = 1 + \frac{t_G}{W}$$  \hspace{1cm} (1.6)
**PC** is an essential measurement to evaluate the gradient elution performance. A gradient elution separation with higher peak capacity represents a higher potential to separate analytes in complex samples, such as proteomic samples. Increased peak capacity in gradient elution can be achieved by using longer columns, increasing gradient time, and elevating the temperature. The details of optimization of peak capacity will be Chapter 2.

### 1.1.1.3 Nanoflow Liquid Chromatography

In analysis for limited amounts of sample, the main concern is if the separation and detection technologies will be sensitive enough to detect specific targets. The mass spectrometer is the most widely coupled detector to liquid chromatography. The flow rate through the column has a significant impact on the signal intensity of a mass spectrometer when interfacing by electrospray ionization. Analytes in the liquid phase are required to be converted into gas phase before entering the MS. The smaller droplet created at very low flow rates results in increased sensitivity in mass spectrometry (MS) detection due to more efficient ionization, reduced ion suppression and improved ion transmission. This phenomenon is observed significantly at a very low flow rate < 20 nL/min (Fig. 1.2). The detailed mechanism of increased MS intensity at low flow rate is discussed in the Nano-ESI section.
To down-scale the flow rate to a nanoliter/min flow rate, the diameter of the analytical column needs to be decreased to keep the efficiency of separation. This can be seen from the van Deemter which shows the relationship between the linear velocity and column efficiency. The equation can be expressed as\(^{32,33}\):

\[
H = A + \frac{B}{u} + Cu \quad \text{or} \quad H = A + \frac{b \cdot D_M}{u} + \frac{c}{D_M} u \tag{1.7}
\]

where \(H\) stands for Height Equivalent to a Theoretical Plate (HETP) and can be expressed as \(H = N / L\); \(N\) is the plate number of column; \(u\) is the velocity of the mobile phase; \(D_M\) is the diffusion coefficient of the sample component in the liquid phase and \(L\) is the length of the column. In the van Deemter equation, the \(A, B\) and \(C\) terms represent three main band
broadening sources - eddy diffusion, longitudinal diffusion and mass transfer, respectively, that cause band broadening in the column. If the column is a packed bead column, the equation can be expressed as a function of particle diameter \( d_p \) \(^{34}\):

\[
H = 1.5d_p + \frac{D_M}{u} + \frac{d_p^2}{6 \cdot D_M} \cdot \frac{65.1}{u}
\]  

For open tubular (OT) columns without packing materials, the HETP can be expressed using the Golay Equation, \(^{35}\) as:

\[
H = \frac{2 \cdot D_M}{u} + \left( \frac{d_c^2}{D_M} \cdot \frac{k^2}{16(1+k)^2} + \frac{d_f^2}{D_s} \cdot \frac{k}{(1+k)^2} \right)
\]  

where \( d_c \) is the i.d. for OT column and \( d_f \) is the thickness of the stationary phase; \( D_M \) and \( D_s \) are the diffusion coefficients in the mobile phase and stationary phase, respectively.

Eddy diffusion, the A term in the van Deemter equation, is an independent factor and unaffected by linear velocity. Different analytes traveling in the column can take multiple pathways which results in band broadening due to the difference in elution time (Fig. 1.3A). Homogeneous packing and reduced particle size can minimize the length of a component’s path and therefore reduce the eddy dispersion effect \(^{36}\) (Fig. 1.3B). The reduced A term with particle size \( d_p \) is shown in Equation 1.8. Interestingly, the eddy diffusion is eliminated in the narrow open tubular column because no packing material exists in this type of column \(^{37,38}\) (Fig. 1.3C).
Figure 1.3. Eddy diffusion in the van Deemter equation for A. packed column with large particle and random packing, B> packed column with small particle and regular packing and C. open tubular column with open tubular structure (from top to bottom).

Table 1.1. Classification of different diameters of HPLC columns (Data adapted from Forster 44)

<table>
<thead>
<tr>
<th>Description</th>
<th>Dimension</th>
<th>Typical flow rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub-nanobore column</td>
<td>&lt; 25 µm i.d.</td>
<td>&lt; 20 nL/min</td>
</tr>
<tr>
<td>Nanobore column</td>
<td>25 µm ≤ i.d. ≤ 100 µm</td>
<td>25–4000 nL/min</td>
</tr>
<tr>
<td>Capillary column</td>
<td>100 µm &lt; i.d. &lt; 1 mm</td>
<td>50–1000 µL/min</td>
</tr>
<tr>
<td>Microbore column</td>
<td>1 mm ≤ i.d. ≤ 2.1 mm</td>
<td>0.4–200 µL/min</td>
</tr>
<tr>
<td>Narrow(small)-bore column</td>
<td>2.1 mm &lt; i.d. &lt; 4 mm</td>
<td>0.3–3.0 mL/min</td>
</tr>
<tr>
<td>Normal-bore column</td>
<td>4 mm ≤ i.d. ≤ 5 mm</td>
<td>1.0–10.0 mL/min</td>
</tr>
</tbody>
</table>
Longitudinal diffusion, the $B$ term, describes the diffusion of analytes along the axial direction of the column. Molecules naturally diffuse from high concentration zone to low concentration zone\(^{39}\). During a peak band moving through the column, the center of the band diffuses to the front and tail end of the peak where the concentration of the analyte is low, resulting in a peak broadening\(^{40}\). The longitudinal diffusion will increase with the residence time of molecules in the columns; therefore the peak broadening is inversely proportional to the velocity or flow rate.

Resistance to mass transfer, the $C$ term, relates to the mass transfer in the mobile phase and stationary phase. The $C$ term consists of two terms of resistance of mass transfer, resistances of mass transfer from mobile phase to stationary phase ($C_M$) and from stationary phase to mobile phase ($C_S$)\(^{41}\). Lower velocity enables a longer equilibration time for analytes between mobile phase and stationary phase, thereby increasing the efficiency of separation and decreasing the peak broadening up to a point beyond which longitudinal diffusion ($B$ term) takes over. The use of smaller particle size for packed columns and smaller i.d. for OT columns can significantly reduce the HETP in a separation due to the increased mobile phase mass transfer. With decreased diameter in particle size or open tube, the distance over which analyte molecules must diffuse to reach the stationary phase surface decreases, resulting in faster mass transfer between analytes to stationary phase\(^{42}\). A preparation of the packing material in a core-shell structure can also improve the column efficiency. The core-shell particles consist of a non-porous hard core and the porous thin layer outside the core. The thin layer decreases the distance of mass transfer in the stationary phase and hence $C_S$ decreases, resulting in an improved peak shape\(^{43}\).
Plotting the plate height as a function of linear velocity generates an $H-u$ van Deemter plot (Fig. 1.4). The van Deemter plot shows that the efficiency is significantly decreased at extremely high or low linear velocity. Simply decreasing flow rate on the analytical column to achieve ultralow flow rate will result in the velocity $u$ at the $B$ term region and therefore the column efficiency $H$ will significantly decrease. To achieve the ultralow flow rate <20 nL/min and keep the separation efficiency, the diameter of the column needs to be scaled down with the flow rate to maintain the velocity in the optimal range. Table 1.1 describes formats of separation column from relatively high mL/min flow rate to nL/min flow rate. Trends in developing narrower bore columns represent the demand for higher sensitivity in trace biological analysis and the consumption of less solvent\textsuperscript{44}. 


Figure 1.4. van Deemter plot of HETP as function of linear velocity. Parameters causing band broadening: $A$ term, eddy diffusion; $B$ term, longitudinal diffusion; $C$ term, Resistance to mass transfer.
1.1.1.4 Sub-nanobore Liquid Chromatography Columns

In the discussion above, we realize that achieving ultralow flow rates of $< 20 \text{nL/min}$ requires nanobore columns of diameter $< 25 \mu\text{m i.d.}$. Only a few formats of columns in reversed phase separation can generate this ultralow flow rate.

**Packed beads column in** $< 30 \mu\text{m i.d.}$ capillary can improve the performance of HPLC separation operated at ultralow flow rate$^{42}$. However, the problem in using this narrow bore packed column is that packing a narrow diameter column with sub-2 micro beads requires a special packing device$^{42}$. The pressure generated by this column is over 4000 bar, which is over the limit of most of the commercial HPLC and UHPLC pumps$^{45}$. To achieve the ultralow flow rate at moderate backpressure, some groups had to use a larger $3\mu\text{m or } 5\mu\text{m particle size to sacrifice the efficiency of separation. The results showed that the gain of sensitivity is more than the loss of efficiency}^{46}$.

**Monolithic column** is an alternative separation column to a packed beads column. The stationary phase in a monolithic column, in contrast to a slurry filled packed column, is in-situ polymerized with a composition of monomers and porogens to form a continuous piece of the stationary phase with porous structure$^{47}$. The porous structure of the polymer in a monolithic column enables high permeability of the mobile phase and reduces backpressure on the column. Our group has demonstrated the preparation of a poly(styrene-divinylbezene) (PS-DVB) based monolithic column in $20 \mu\text{m i.d.}$ fused silica capillary column, and the performance of separation in low amole detection$^{48}$.
**Porous layer open tubular (PLOT) column** is an open tubular structural column with a layer of porous packing stationary phase coated at the inner wall of fused silica capillary column. The PLOT column comes from open tubular columns in gas chromatography (GC) which were introduced in 1950s\textsuperscript{49}. In GC separation, gas is the mobile phase carrier for the analytes\textsuperscript{50} and GC has demonstrated high efficiency in separation of volatile compounds with over one million plates using long columns\textsuperscript{51,52}. After decades of GC, many approaches were evaluated to implement the OT columns (50-60 µm i.d.) into LC separation\textsuperscript{53–56}. Our group first developed a 10 µm i.d. poly-(styrene-divinylbezene) (PS-DVB) open tubular column with 1 µm thickness of porous PS-DVB layer on the inner wall for the use in HPLC (Fig. 1.5). In the preparation of the 10 µm i.d. PLOT columns, a single porogen (anhydrous ethanol) was used to precipitate the highly cross-linked polymer in the early stage of phase separation, resulting in an open tubular structure\textsuperscript{57,58}. With 8 µm open diameter, the PLOT column has very high permeability and produces only 200 bar backpressure at the length of 4 meters. The use of a long column at low backpressure demonstrated ultra-performance compared to packed beads column and ultra-sensitivity at low flow rate < 20 nL/min\textsuperscript{1,59}.

![Figure 1.5. Scanning electron micrographs (SEM) of the 10-µm-i.d. PS- DVB PLOT column](image-url)
1.1.2 Mass Spectrometry

Mass spectrometry has been widely coupled with HPLC as a detector to analyze complex biological samples. Mass spectrometry is a powerful tool to identify unknown sequences and structures in biological samples. The complete process involves an ion source that converts analytes eluted in a liquid mobile phase into gaseous ions, a mass analyzer that differentiates ionized analytes (with or without fragmentation) based on their different mass-to-charge ratios (m/z), and a detector that records and transfers the detected signal into their m/z values\(^60\).

1.1.2.1 Ionization

The first problem chemists faced when analyzing biomolecules such as peptides, glycopeptides and proteins was how to convert the polar, nonvolatile and thermally unstable analytes into the gas phase without the loss and degradation of biological samples. The development of two soft ionization techniques, matrix-assisted laser desorption ionization (MALDI)\(^61\) and electrospray ionization (ESI)\(^62\), brought a revolution to the analysis of biomolecules by MS.

MALDI

In MALDI analysis, the sample, other than injected into an analytical column, is first co-crystallized with a UV-absorbing organic acid matrix compound as a matrix\(^63\). After absorption of laser radiation, the heat generated by the laser causes desorption of the acidified matrix into the gas phase\(^63\). The ionization of neutral peptides or proteins were undergo a proton transfer from the primary protonated matrix\(^64\). One drawback of MALDI is that the solid matrix sample
preparation cannot provide separation of the biological samples, and hence LC-MALDI was developed. LC-MALDI methods utilize discontinuous separation on LC and drop the fraction of mobile phase onto the MALDI multiple-well plate followed by crystallization and ionization of matrix. Thus, LC-MALDI MS methods have some issues for high-throughput and complex sample analysis.

**ESI**

In contrast to MALDI, ESI produces vaporized ions continuously from HPLC elution. In ESI, a continuous stream of biomolecules eluted with the mobile phase is sprayed from a fine tip, which is maintained at a high voltage (e.g. 2.0 - 6.0 kV) relative to the wall of the surrounding chamber. Coulombic repulsion is accumulated between the charged analytes with the evaporation of a volatile solvent (e.g. water and acetonitrile) resulting in increased repulsion between the positive charges, and eventually the initial droplet generates smaller droplets. With a heated capillary in the ESI source and in some cases a stream of drying nitrogen gas, the ions on the surface of the droplet are ejected into the gas phase and transferred into the mass analyzer. Unlike MALDI, ions produced by ESI are usually multiply charged which benefits the MS to analyze large mass peptides or proteins in the low m/z value range.

**Nano-ESI**

An important advancement of ESI was the nanoelectrospray technique developed by Mann. Flow rates at the nanoliter per min range can reduce the initially size of thedroplets produced in electrospray and, as a result, a larger portion of analytes compared to solvent and other
contaminants is made accessible to entering the gas phase in the initial droplet (Fig. 1.6 bottom). This process can increase the ionization efficiency as well as reduce ion suppression to improve sensitivity in MS detection. Smith, for example, demonstrated a significant gain in sensitivity by packed columns with 15 µm i.d. at 20 nL/min flow rate in comparison to a 75 µm i.d. column at 400 nL/min. Furthermore, nano-ESI was shown as more tolerant towards salts than conventional ESI, which can benefit the multiple dimensional separation where a high concentration of salt is used in SCX. Nano-ESI is the primary ionization technique used in this dissertation.

Figure 1.6. Ionization of Electrospray in conventional flow and nanoflow (adapted from)
1.1.2.2 Mass analyzers

A mass analyzer, an important component of the mass spectrometer, separates the ionized analytes in the gas phase based on their mass to charge ratios and outputs the signal responses; the responses are converted to a digital output which represents the m/z value. There are four general types of mass analyzers with different ion separation mechanisms widely used in the detection of ions as described as below.

**Quadrupole**

The quadrupole mass analyzer, as the name implies, consists of four or more parallel metal rods. Applied direct current (DC) and radio-frequency (RF) voltages are used in a chamber of the parallel rods to control the ions traveling down through the quadrupole. DC and RF potential can affect the trajectories of a beam of ions in the X-Z planes and Y-Z planes separately. For a given DC and RF combination, only a specific pre-defined m/z range will be allowed to pass through the quadrupole and reach the detector; other ions will be filtered onto the walls by collisions because of their unstable trajectories. This feature of the quadrupole analyzer allows mass filtering of an ion with a particular m/z or scanning of an m/z range by continuously altering the combination of DC/RF voltage.

**Time-of-Flight (TOF)**

The TOF analyzer, the most straightforward mass spectrometer, was discovered in the middle 20th century. TOF analyzers separate ions simply based on their traveling time from the ion source to detector in a long tube. In the long tube where a given strength of the electric field is applied, ions are accelerated and obtain kinetic energy from the electric field as.
where \( U \) is the voltage of electric field, \( z \) is the charge of the ion, \( m \) is the molecular weight of the ion and \( v \) is the velocity. In given length \( L \) of tube and strength of electric field, the time of flight is determined by the \( m/z \) value of ions:

\[
t = \frac{L}{v} = \frac{L}{\sqrt{U}} \cdot \sqrt{\frac{m}{z}}
\]

The advantage of TOF is that the detection has a high mass range and fast speed and therefore TOF has been widely coupled with MALDI ion source for peptides and intact proteins analysis with high \( m/z \) value.

**Ion Trap**

An ion trap mass analyzer, as its name implies, is used to store and eject ions. In an ion trap, ions are captured in a region of a vacuum system by a combination of electric fields between electrodes. Variants of ion traps were developed based on different geometries of electrodes, three-dimensional quadrupole ion trap (3D QIT) and linear ion trap (LIT) being the most widely used ion traps. The conventional 3D QIT is composed by a ring electrode and two endcap electrodes. The invention of 3D QIT provides the capabilities of a low energy collision induced dissociation and multiple stages fragmentation (MS\(^n\)). The versatile QIE analyzer facilitates relatively high resolution, \( m/z \) measurement and multiple stages mass spectrometry analysis of biomolecules. The LIT is installed in series of QIT to manipulate the ions by a 2D RF filed in the radial dimension and static electrical potential on the end of the rods axially. In LIT, the Q1 ion trap is used to eject unwanted ions; the Q2 ion trap is presented as a pressurized collision cell...
where the fragmentations of precursor ions occur and the Q3 generates a mass spectrum (Fig. 1.7). The advantage of LIT is that it overcomes the low trapping capacity issue in 3D QIT due to the small trapping volume; thus, LIT is more sensitive and suitable for quantitative analysis. 

The Q3 in LIT could be also replaced with other front end analyzers which will be discussed later in this section. The hybrid LIT instruments, such as LTQ-Orbitrap and LTQ-FTICR, provide ultra-high resolution and versatile peptide fragmentation modes in the analysis of biological samples.

Figure 1.7. Schematic diagram of a LIT consisting of three quadrupole ion traps (Adapted from 86).
Fourier Transform Ion Cyclotron Resonance (FTICR)

In the FTICR mass analyzer, determination of m/z of ions is based on their cyclotron motion in a uniform magnetic field. FTICR has been developed based on the Penning trap, a device for storage of ions using a homogenous axial magnetic field and a quadrupole electric field\textsuperscript{89}. The ions trapped in a Penning trap resonate at their frequency in the magnetic field and then are “excited into a larger cyclotron radius by an electric field”\textsuperscript{90}. The induced current of charged ions can be detected on a pair of electrode plates when the ions pass through the plates\textsuperscript{90}. The result of the detectable signal is then transferred to a mass spectrum by performing a Fourier transform. The m/z value is proportional to the induced current, and the signal increases linearly with ion charge\textsuperscript{91}; hence, the FTICR is more sensitive to multiple charged analytes such as peptides and proteins. FTICR is developed for the needs of high resolution and high mass accuracy. The mass resolution is defined as the resolving power to distinguish two adjacent m/z values. For a single m/z peak, a standard definition of the resolution is based on the mass difference $\Delta M$ being the full width at half maximum height (FWHM)\textsuperscript{92}:

$$\text{FWHM} = \frac{M}{\Delta M}$$

In FTICR, the quality of the signal including the mass resolution increases significantly with the magnetic field\textsuperscript{93}. For example, resolution $> 300,000$ could be obtained at a magnet of superconducting magnet 9.4 Tesla for m/z value in the range of 300-1000\textsuperscript{94}. Because of the high resolution, wide dynamic range and high sensitivity of FTICR MS, it has been widely used in the characterization of biomolecules, top-down and bottom-up proteomics with mass accuracy of ppm $< 2$ for peptides and $< 10$ than proteins\textsuperscript{95,96}. 

24
**Orbitrap**

The recent novel FT-like mass spectrometer, Orbitrap, was developed by Makarov in 1999\(^9\) \(^7\) based on the Kingdon ion trap and put into practice in 2005\(^9\) \(^8\). The Orbitrap mass analyzer consists of a “spindle-like central electrode” and two outer cylinder electrodes facing each other\(^9\) \(^9\), as shown in Fig. 1.8. When voltage is applied in the chamber, ions are trapped because their movements towards the central electrode by the radial electric field are balanced by an opposing centrifugal force created by a tangential velocity\(^9\) \(^8\). With a correct potential of the electric field, ions oscillate around the central electrode, while an axial electric field excites ions into a larger trajectory. The axial oscillation of ions is harmonic, and the frequency of ion axial oscillation is simply related to the m/z. Similar to FTICR, the induced ion current is detected by outer electrodes, and then converted into frequency by Fourier transform, finally output into a mass spectrum\(^10\) \(^0\).

The innovation of the Orbitrap analyzer with high resolution, high mass accuracy and wide dynamic range overcomes the disadvantage of low mass resolution (~10,000) and mass accuracy (10 – 15 ppm) in the LIT\(^7\) \(^5\) \(^8\) \(^0\). Compared to the other ultrahigh resolution FTICR, the mass resolution in Orbitrap decreases only with the \(\sqrt{m/z}\); however, it decreases linearly with m/z in FTICR\(^10\) \(^1\) \(^0\) \(^2\). Therefore the orbitrap is more suited for the analysis of intact proteins within large m/z range. Additionally, the orbitrap analyzers have a faster acquisition time than FTICR, which improves the sampling rate for LC/MS analysis\(^10\) \(^3\). The advanced features as well as smaller size and affordability of the Orbitrap analyzer make it a dominant component in novel developed mass spectrometry such as Orbitrap Elite and Q Exactive instruments. In the Orbitrap Elite, resolving power up to 240,000 at m/z 400 combined with multiple fragmentation techniques
(HCD, CID and ETD discussed in section 1.1.2.3) and analyzer (ion trap, Orbitrap) make it a powerful tool for protein and peptide identification\textsuperscript{104}. The hybrid Q Exactive coupling a quadrupole analyzer and the Orbitrap with fast HCD fragmentation have become very popular for proteomics analysis and high-throughput screening\textsuperscript{4,105,106}.

Figure 1.8. Three-dimensional schematic of an Orbitrap cell. (Copyright from ACS publications\textsuperscript{99})
1.1.2.3 Fragmentation Methods

To obtain more information on the structure of unknown peptides or proteins, ionized molecules are often cleaved into smaller fragment ions. The power of this approach can be seen in that the first mass analyzer allows selective ions to pass into a reactor for fragmentation, and the second mass analyzer records the m/z values of fragmented ions.

**Collision Induced Dissociation** (CID) is the most commonly used fragmentation method to cleave peptides and proteins. In CID, precursor ions are accelerated in an electric field and collide with slow neutral gas molecules (often helium or nitrogen)\(^\text{107}\). Collisions between the accelerated ions and a neutral target gas are accompanied by an increase in internal energy which results in backbond cleavage and dissociation of large molecules into small fragments\(^\text{108}\). High energy CID (>14 keV) in four sector and TOF/TOF instruments generates all y, z, a and b ions\(^\text{109}\) in Fig. 1.9. However, in low energy CID (1-100 eV) in triple quadrupole or ion trap MS, positively charged peptide fragment mainly along its backbone, generating predominantly b and y-types ions and neutral loss of water, ammonia or carbon monoxide are often observed\(^\text{110}\). Limitations of CID are that 1) for peptides with posttranslational modifications (PTMs) such as phosphorylation and glycosylation, CID is preferred to cleave the labile PTMs other than the backbone and results in less peptide sequence information and 2) to achieve a reasonable fragmentation efficiency, the trapping voltage level is selected such that some low m/z product ion will not be detected\(^\text{111}\).

**Electron Transfer Dissociation** (ETD), as well as the similar electron capture dissociation (ECD), is a radical induced fragmentation method at low energy. In ETD fragments, initial
radical anion (e.g. anthracene or azobenzene) transfers an electron a protonated peptide by electron transfer reaction. The odd-protonated peptide undergoes rearrangement and induces the fragmentation of the peptide backbone, resulting in a c-z type cleavage of the Cα-N bond\textsuperscript{112} (Fig. 1.9). This process generates predominantly c and z-type ions rather than the typical b and y ions generated in CID\textsuperscript{113}. Unlike CID, ETD randomly cleaves the peptide backbone by electron transfer other than PTMs on the peptide\textsuperscript{112}. This feature of ETD is widely used in the analysis of PTMs such as phosphorylation\textsuperscript{114} and glycosylation\textsuperscript{115}. Another unique feature of ETD is that the electron transfer reaction can induce the breakage of disulfide bond, which is not available in other fragmentations, making ETD a powerful tool to determine the disulfide linkage in protein structure\textsuperscript{116,117}.

Figure 1.9. Chemical bond cleavages in MS/MS Fragmentations. (Adapted from\textsuperscript{118}).
**Higher Energy Collision Dissociation** HCD, formerly called higher energy c-trap dissociation is a recently introduced fragmentation method in the Orbitrap\(^{119}\). Initially, ions were fragmented in the C-trap by varying radiofrequency voltages without changing hardware to retain the maximum efficiency in trapping fragment ions\(^{119}\) (Fig. 1.10.a). To maintain the higher trapping efficiency of the C-trap at lower radiofrequency voltages, an octopole collision cell was placed after the C-trap for fragmentation\(^{119}\) (Fig. 1.10.b). The brief process of HCD involves ions with wide mass range being trapped in the C-trap with high trapping efficiency at low voltage, ejected into octopole collision cell for fragmentation and sent back through the C-trap into the Orbitrap for detection\(^{119}\). Within this process, the C-trap is always maintained at high trapping efficiency, thus overcoming the low m/z cut-off rule in CID. Therefore, HCD is useful for quantitative proteomics analysis based on iTRAQ\(^{120}\) and TMT\(^{121}\) isobaric tags.
1.2 Proteomics

1.2.1 Overview of Proteomics

Proteomics is a term defined as the study of the proteome which describes the whole set of proteins expressed by the transcriptions of genome. Proteomics involves not only all the proteins in a given cell, tissue or organism, but also protein isoforms and post-translational modifications and their biological network in higher-order.\textsuperscript{122}

Proteomics is challenging due to the complicated information in the proteome and the low abundance of many proteins. The mass spectrometry based proteomics approach fulfills the complexity and sensitivity requirements in proteomics studies, exhibiting an even more important role with the development of new instruments. Generally, for most proteomic workflows, proteins are extracted from cells or tissues, digested into peptides, separated by HPLC and then subjected to MS analysis using high speed, high resolution and a wide dynamic range\textsuperscript{123,124}. Mass spectra obtained in MS are searched in a database for possible matching, which is the output of MS-based proteomic approach.

1.2.2 Protein Quantitation

The proteome differs for the different types of cells and different phases of the cell cycle, where the same gene may express a significantly distinct set of proteins\textsuperscript{125}. Some proteins perform the regulation in a specific cell state, particularly a state relevant to the presence of the disease and can thus be considered as potential biomarkers for clinical diagnosis\textsuperscript{126}. MS-based proteomics is a method to screen biomarkers and study protein expression in a specific biological and disease state. Adding a quantitation dimension into the proteomic method is therefore
necessary. Generally, quantitation approaches can be divided into labeled and label-free quantitation.

1.2.2.1 Labeled Quantitation

The labeled quantitative approach is based on the fact that isotopically labeled analytes with identical chemical behavior can be differentiated by their mass differences in MS. The signal intensity ratio of isotopically labeled pairs indicates the abundant ratio of the two analytes. Several stable isotope labels have been introduced and include $^{18}$O/$^{16}$O enzymatic labeling, tandem mass tags (TMT), isobaric tags for relative and absolute quantification (iTRAQ) and stable isotope labeling by amino acids in cell culture (SILAC).

$^{18}$O/$^{16}$O

Peptides are isotopically labeled with $^{18}$O during enzymatic digestion in the presence of H$_2^{18}$O. Incorporation of $^{18}$O occurs during hydrolysis with H$_2^{18}$O at the carboxyl terminal of each peptide and results in a mass shift 2 Da or 4 Da (two $^{18}$O atoms). Peptide abundance is quantitated by calculating the $^{16}$O/$^{18}$O ratio between samples from high resolution scan.$^{127}$

TMT/iTRAQ

A tandem mass tag (TMT) reagent is designed for quantitation based on MS/MS analysis. The TMT reagent consists of a reactive group toward free N-terminal peptides and epsilon-amino functions of lysine residues, a mass normalization group and a fragment ion reporter (Fig. 1.11)$^{128}$. The reporter group is isotopically labeled and the normalization group is for balance of total mass of TMT reagent. Therefore, pairs of peptides or proteins labeled with TMT tags
remain isobaric and chromatographic identities\textsuperscript{129}. Pairs of TMT labeled peptides are differentiated by different isotopic reporter under CID MS/MS fragmentation. A 8-plex iTRAQ tag with 8-plex (6-plex in TMT) sharing the similar principle is capable of comparing of 8 samples in a single LC-MS analysis\textsuperscript{130}. Recently, the HCD fragment available in Orbitrap was reported useful for quantitative proteomics using TMT or iTRAQ due to its feature of no low m/z cut-off for small reporters as seen in the ion trap\textsuperscript{121}.

![Figure 1.11. Design of a tandem mass tag (TMT). (Copyright from ACS publication)](image)

**SILAC**

The stable isotope labeling by amino acid in cell culture (SILAC) method is a metabolic labeling method which grows cells in media with natural light amino acid and heavy isotope labeled amino acids or media with light or heavy isotope labeled ammonium salts separately for at least five cell doublings in order to force incorporation of the light or heavy labeled amino acids in cellular synthesized proteins. The light or heavy labeled amino acids in proteins introduces mass differences which can be detected by mass spectrometry. Isotope labeled lysine
and arginine (e.g. $^{13}\text{C}_6$, $^{15}\text{N}_4$ L-arginine and $^{13}\text{C}_6$, $^{15}\text{N}_2$ L-lysine) are most commonly used in the SILAC experiment could incorporate the light or heavy labeled lysine and arginine residues from the growth medium which could improve the coverage of quantitation$^{131}$. Compared to most chemical labeling approaches which usually combine labeled and unlabeled samples after digestion on the peptide level, the SILAC method enables mixing samples prior to the digestion process to eliminate bias and variations$^{132}$.

1.2.2.2 Label-Free Quantitation

1.2.2.2.1 Quantitation by Peak Intensity or Peak Area

Label-free quantitation by peak intensity or area is based on the observation that the signal intensity from ESI is related to the concentration in the sample. A quantitative study of tryptic digests of myoglobin in a complex matrix with a wide range between 10 fmol and 100,000 fmol resulted in a linear correlation of 0.991 between peak area and the concentration of the protein$^{133}$. To reduce variations in run-to-run analysis and sample preparation, measurements of peak areas of many proteins are averaged in replicates by computer algorithms with normalization of peak areas to a reference sample. Finally, statistical tests (e.g., a Student t-test) are performed to test if there are significant differences in protein expression between multiple samples, and the probability thresholds for statistical significance are calculated by p-value algorithm to reduce the rate of positive results$^{134}$.

Recently, an absolute label-free quantitation approach was carried out in the quantitation of large sets of proteomics data, called intensity-based absolute quantification (iBAQ)$^{135}$. In the iBAQ approach, the maximum observed peak intensities (including the isotope ions) of all
identified peptides belonging to one protein were summed as a total protein intensity. The sum value was divided by the theoretical number of observed peptides within the length of 6-30 amino acids in a fully tryptic digestion without any missed cleavages, resulting an iBAQ intensity for the protein\textsuperscript{136}. As the PERL script is available for MaxQuant\textsuperscript{137}, a software using algorithms specifically designed for high resolution proteomics, quantitation of large sets of proteomics data can be easily generated from the identified peptide results from MaxQuant. Using the iBAQ approach, over 10,000 proteins from a human cancer cell line could be quantified\textsuperscript{138} (Fig. 1.12). Compared to other label-free quantitation approaches, e.g. exponentially modified protein abundance index (emPAI) and absolute protein expression index (APEX), the iBAQ approach demonstrated a better correlation between biological replicates\textsuperscript{139}. The accuracy of the iBAQ approach will be examined in Chapter 3 for the proteomic analysis from limited amount of cells.

Figure 1.12. Ranked protein abundances calculated by iBAQ intensities\textsuperscript{138}. (Copyright under CC BY-NC-SA 3.0 License)
1.2.2.2 Quantitation by Spectral Count

A viable label-free quantitative strategy is spectral counting where the quantitation of a protein is performed by counting the total number of MS/MS spectra matched to the identified peptides from the same protein. Higher abundant proteins typically result in an increased number of peptides, and hence, the total number of MS/MS spectra is increased\textsuperscript{140}. Spectral count quantitation is demonstrated as being sensitive and quantitative as quantitation by signal intensity\textsuperscript{141}. Compared to the peak intensity approach, spectral counting requires no computer algorithms. A normalization of spectral count (SpC) divided by length of protein (L) to sum of all SpC/L is performed to compare the abundance of individual protein expression in various complexes. This normalization process is defined as a normalized spectral abundance factor (NSAF)\textsuperscript{142}.

1.2.2.2.3 Quantitation by Normalized Spectral Index (SI\textsubscript{N})

SI\textsubscript{N} is a recently developed quantitative method represents for the sum of intensities and SpC of all fragment ion spectra identified for all the peptides to a given protein\textsuperscript{143}

Normalization is performed by correcting the individual SI value of each protein for the total SI value of all proteins identified in the data set and dividing by the protein length (length of amino acid sequence). Compared to the SpC based method, the SI\textsubscript{N} method showed its superiority in reducing the variability between several replicate MS measurements of the same liver endothelia plasma membrane sample\textsuperscript{143}. The SI\textsubscript{N} value also enables the estimation of the protein expression level in complex samples since a correlation analysis of 2,660 shared proteins between 40 and 150 µg sample loading amount results in a linear fit with a slope of 3.72 which is
very close to the protein abundance ratio 3.75 (150 µg: 40 µg). This result exhibits the capability of SIₙ to quantitate thousands of proteins in complex biological samples\textsuperscript{143}.

1.2.3 Bioinformatics

MS-based proteomics usually generates a large number of MS/MS spectra, and this spectral information needs to be translated into peptides and their corresponding proteins for protein profiling, protein interaction and system biology\textsuperscript{144} (Fig. 1.13). Analysis of the large data generated by MS-based proteomics is challenging, and computer-based bioinformatics has been developed as a tool for interpretation of the raw data.

Figure 1.13. Bioinformatics analysis paths for MS-based proteomics (Adapted from \textsuperscript{144}).
1.2.3.1 Peptide Identification

For large scale CID or HCD MS/MS spectra, manual assignments of all peptide MS/MS spectra are not practical. Protein identification based on MS/MS spectral matching with a comprehensive peptide sequence database is available for high-throughput MS-based proteomic analysis. The most used matching softwares of these databases are MASCOT and SEQUEST. SEQUEST uses “a cross-correlation scoring function”, Xcorr score, to evaluate the matching between the MS/MS spectrum and the theoretical fragmentation of a peptide sequence\textsuperscript{145}. SEQUEST uses the intact peptide mass value to determine a set of peptide candidates in the database. Theoretical peptide MS/MS spectra are compared to the observed spectrum, and the correlation is evaluated by an Xcorr score\textsuperscript{145}. MASCOT uses a “probability-based scoring algorithm based upon the MOWSE scoring” method\textsuperscript{146}. For a peptide spectrum, the probability match in the database is randomly calculated, where the MOWSE score indicates the significance of match\textsuperscript{147}. Validation of peptide identification by determining false discovery rates (FDR) using a decoy database is recommended. A simple decoy database is created to keep the information in the target database but with reversed or random sequence\textsuperscript{148}. The false positive results from the same searching against the decoy database are supposed to be in the identification. Other database searching algorithms are also available, such as X!Tandem\textsuperscript{149} and PEAKS\textsuperscript{150}. Cross-validation using two different algorithms can improve the confidence and probability for peptide identification\textsuperscript{151}, and the combined search engine is available in the commercial software such as Proteome Discover and MaxQuant\textsuperscript{137}. 
1.2.3.2 Bioinformatic Methodologies for System Biology

Large peptide/protein identification and quantitative datasets from a database search need to be sorted to solve system biological questions of protein interaction, protein expression and protein networks. Several common bioinformatic tools below are available for system biology using data sets from the database search.

1.2.3.2.1 Gene Ontology (GO)

Gene Ontology (GO) is a major bioinformatics methodology to standardize the classification of gene and gene product protein across different species and databases with a systematic vocabulary of phrases. There are three overlapping categories of GO, including the molecular function which stands for the biological activity of gene or gene production on molecule level (e.g., alcohol dehydrogenase activity), biological process which stands for the transformation that the gene or gene product participated in (e.g., heart development) and cell component which describes the cellular locations where a gene product works (e.g., plasma membrane)\textsuperscript{152,153}.

GO could be used as a database to analyze high-throughput MS-based proteomics data efficiently; for instance, it could help to identify the cellular location or biological function of proteins at different expression levels among multiple MS data sets. Additionally, GO could also be applied to predict the large-scale interactions between proteins in an organism based on the knowledge on protein-protein interactions of other source organisms which have evolutionary relationships with the target organism\textsuperscript{154}.
1.2.3.2.2 Gene Set Enrichment Analysis (GSEA)

After individual gene is annotated by GO database, Gene Set Enrichment Analysis (GSEA) can be applied to evaluate large-scale of expression profile data of two biological states on gene sets level other than single gene level due to some important pathways are more likely to be regulated by groups of genes other than individual gene. The gene set is a group of genes, either with the same chromosomal locations or have the same GO terms.

GSEA aims at determining whether members in a gene set are enriched in the gene list which ranked according to the their expression differences between two datasets, generally by resulting $p$-values of multiple hypotheses testing. GSEA demonstrated the power of insight into several cancer-related data sets. GSEA can be applied to protein expression data from MS-based proteomic analysis.

1.2.3.2.3 Pathway and Network Analysis

The biological system, the system function is not regulated by single molecule; it is the result of interaction between numerous cellular components such as DNA, RNA and protein. Many tools are available for the pathway and network analysis, such as KEGG (Kyoto Encyclopedia of Genes and Genomes) and IPA (Ingenuity Pathways Analysis). KEGG was initiated in 1995 by the Japanese Human Genome Program and is an online database which links the genomic information with the functional information based on the published literatures. Three network databases on pathway analysis, Gene network, ligand network are available in KEGG and the databases are still continuously updated with the growing literature. IPA is commercially available software.
1.3 Sample Preparation of Limited Biological Samples

In recent biomedical research, interest has increased in the analysis against subpopulation of disease related samples, e.g. circulating tumor cells (CTCs) in peripheral blood\textsuperscript{159}, stem cells (SC)\textsuperscript{160}, specific tissues derived from living biological specimens by microbiopsy\textsuperscript{161}. However, those samples related to the specific diseases are always present in very small amount, e.g. CTCs are found in the order of 1-10 cells per mL of human blood\textsuperscript{162}. A successful LC/MS approach on trace biological sample analysis should provide sensitivity, reproducibility and robustness. The needs of appropriate sample preparations prior to LC/MS, such as sample collection, purification and digestion, have been realized\textsuperscript{163}. To improve the sensitivity and reproducibility, impurities other than target analytes are necessarily removed to decrease ion suppression and matrix effects\textsuperscript{164}.

1.3.1 Sample Collection

To study biomarkers or targeted analytes at the level of protein expression in certain clinical diseases, focusing on a specific subpopulation tumorigenic cells or tissues is necessary\textsuperscript{165,166}. However, in most cases, this subpopulation is not present in a homogenous collection. For example, tumor samples usually contain a majority of normal cells and the highly proliferative cells manipulating the cancer are minor\textsuperscript{167}. Technologies of sample collection are required to isolate the specific cells from the heterogeneous samples.

1.3.1.1 Centrifugation

Centrifugation is a common method used in blood samples to collect plasma or serum from whole blood. Blood cells and clots are removed by centrifuging at 1000-2000 g in a refrigerated
centrifuge. The difference in serum and plasma is that serum is centrifuged after a whole blood specimen is allowed to clot, and plasma is centrifuged of whole blood after anticoagulant-treated such as EDTA\textsuperscript{6}.

1.3.1.2 Microbiopsy

Microbiopsy is proposed to enable a fine needle with a micro capsule endoscope to target a small region of animal or human bodies and obtain a tiny tissue sample without major damage to the animal or human\textsuperscript{168}. Preservation of biological samples presents a major challenge to maintain the protein structure in living tissues during sample preparations\textsuperscript{169}. Cryofixation by high pressure freezing in microbiopsy has been shown to preserve high resolution detail and protein structure in living tissue over conventional formalin-fixed or paraffin-embedded methods. Microbiopsy with high pressure freezing is widely used in the sampling of fresh cells or tissues from muscle, liver and brain\textsuperscript{170–172}.

1.3.1.3 Laser Capture Microdissection (LCM)

Unlike RNA and DNA profiles which are unique in all types of cells, protein expression can vary significantly in different cells and even different phases of the cell cycle. Analysis of heterogeneous tissue is necessary and required. LCM is a technique to isolate specific types of cells from microscopic regions of the cells, tissues and organism. In LCM (Fig. 1.14), a transparent polymer is attached to a slide of the tissue section\textsuperscript{173}. Via direct visualization of tissues under electron microscopy, an IR/UV laser pulse can be applied to the polymer to cut the specific subpopulation of cells with adhesion to the polymer\textsuperscript{7,174}. The LCM method has been widely used in the isolating of specific or cancer cell in tumor tissue for proteomic discovery of biomarkers in diseases such as prostate cancer and colon cancer\textsuperscript{175,176}. 

41
Microfluidic devices have proved as an advance in the capture of rare cells from complex suspensions such as whole blood. Microfluidics enables control and manipulation of fluids at the µm scale while implementing multiple chemical or electrokinetic functionalizations for cell capture\(^{177}\). Major microfluidic cell capture approaches include: 1) immunocapture by

---

**Figure 1.14.** Examples of CTC capture technologies that combine multiple enrichment principles. (Adapted from \(^{159}\)).

**Figure 1.15.** Schematic of a laser capture microdissection process (Copyright from NPG\(^{173}\)).

1.3.1.4 Microfluidic Cell Capture

Microfluidic devices have proved as an advance in the capture of rare cells from complex suspensions such as whole blood. Microfluidics enables control and manipulation of fluids at the µm scale while implementing multiple chemical or electrokinetic functionalizations for cell capture\(^{177}\). Major microfluidic cell capture approaches include: 1) immunocapture by
immobilized antibody or aptamer of cancer cells\textsuperscript{178}; 2) size-based sorting in open obstacle arrays\textsuperscript{179} (as Fig. 1.15b); 3) electrophoresis and dielectrophoresis to manipulate cells based on the action of electric fields on the net free charges of the cell surface\textsuperscript{180} as similar as Fig. 1.15a. One impressive application of microfluidics is the capture of circulating tumor cells (CTCs) in whole blood. CTCs are extremely rare cells (1-10 CTCs/mL blood) that shed into blood from prior cancer tumors and migrate into other organisms; therefore, the concentration and isolation of CTCs by microfluidics are essential for understanding how CTCs regulate the metastatic processes\textsuperscript{159}.

### 1.3.2 Sample Purification/Isolation

A critical issue for biological sample analysis is that the detection of a trace amount of targeted analytes is suppressed by the high abundant proteins. Protein expression exhibits a wide dynamic range of 12 orders of magnitude in human blood\textsuperscript{181} and 7-8 orders in human cells\textsuperscript{182}. Sample simplification from wide dynamic range complex samples is necessary prior to LC/MS analysis. Such simplification often involves some separation of the complex protein mixture.

#### 1.3.2.1 Gel Electrophoresis

A well-established protein separation method is polyacrylamide gel electrophoresis (PAGE), based on the size (molecular weight, MW) or charge differences of proteins. Although resolution of separation is poor, PAGE provides MW information and visualization of the separated proteins by Coomassie brilliant blue (CBB) or more sensitive silver nitrate staining. Cutting
specific sections on PAGE gel followed by in-gel digestion is a convenient approach to simplify sample complexity and analyze trace amount of targeted analytes\textsuperscript{183}.

1.3.2.2 Solid Phase Extraction (SPE)

SPE is a method used for the isolation and concentration of selected analytes from a wide variety of matrices, including urine, blood, and water\textsuperscript{8}. Recent adaptions of SPE have demonstrated suitability for high-throughput preparation of limited amounts of sample in a 96-well plate format with low solvent consumption\textsuperscript{184}. SPE is a main purification method used in Chapter 4 to extract hepcidin from serum or plasma sample.

1.3.2.3 Immunoprecipitation (IP)

IP is a protein selective separation technique using antibody binding to a solid support (usually agarose resin or magnetic bead) to pull a specific target antigen out of a complex biological sample. An example of IP is to remove highly abundant proteins such as albumin and IgG, which account for approximately 60-80\% in plasma or serum sample\textsuperscript{185}. Considering the cost and availability of the antibody, isolation of the target protein by its antibody is desirable to reduce the matrix effect and improve the sensitivity of detection. For example, haptoglobin, the potential biomarker in lung cancer, was IP pulled down by a specific monoclonal antibody (mAb) followed by depletion of albumin and IgG in the plasma sample\textsuperscript{186}. 
1.3.2.4 Cell Lysis

Protein extraction from cell or tissue samples with high yield is important for protein analysis, especially when the availability of the cell number is limited. The lysis of a cell to disrupt the cell wall and release proteins can be achieved by detergent, denaturants and/or acoustic means.\textsuperscript{187,188} Detergent-based lysis incorporates detergents such as sodium dodecyl sulphate (SDS) into the cell membrane. The detergent solubilizes lipids and proteins in the membrane, disrupts the cell membrane structure and releases the proteins. Cells in a high concentration of denaturants such as urea or guanidine buffer can result in high osmotic pressure gradient resulting in a cell breakage. High-frequency oscillation generated by sonication causes localized high pressure areas resulting in cavitation and impaction that can break apart cells\textsuperscript{189}. The advantage of sonication cell lysis over chemical lysis is that less sample clean-up steps are required.

1.3.3 Sample Digestion

An important step in bottom-up protein analysis is to digest a protein into peptide fragments with the preferred mass range (m/z 400-1600) for MS detection. To breakdown the tertiary structure of proteins, disulfide bond have to be reduced using reducing agents such as dithiothretil (DTT) or tris(2-carboxyethyl)phohpine hydrochloride (TCEP)\textsuperscript{190}. Following reduction, free thiol groups must be blocked by iodoacetamide (IAM) and iodoacetic acid (IAA) to quench the thiol-disulfide reaction\textsuperscript{191}. Protein digestion with a protease is the key step to cleave protein into peptides. Trypsin is by far the most efficient and reliable protease which cleaves the peptide bonds, specifically after arginine (Arg) or lysine (Lys) residue\textsuperscript{192}. Other
proteases are also available to cleave protein on other sites, such as endoproteinase Lys-C on the C-terminal side of Lys and endoproteinase GluC on the C-terminal side of glutamic acid (Glu). Combination of multiple digestion steps such as Lys-C followed by trypsin can improve the efficiency of digestion⁹.

1.4 Pharmacokinetics (PK) and Pharmacodynamics (PD) of Biotech Drug

1.4.1 Overview of PK and PD

In the biotech drug development and evaluation process, the general paradigm is the comparison between the selected dose of biotech drug and effect of therapy. The relationship between the concentration of dose and the efficacy is defined by pharmacokinetics (PK) and pharmacodynamics (PD) of a biotech drug. PK includes the study of the mechanisms of drug absorption, distribution, metabolism and excretion vs. the time course. PD studies how the drug concentration affects the efficacy against the drug target¹⁹³.

The PK/PD modeling facilitates biotech drug development and has been widely performed in preclinical and clinical Phases I to III studies, in order to select the drug candidates for further development and study the appropriate dose amount¹⁹⁴.

1.4.2 Quantitation Method in PK and PD

PK and PD are often analyzed using MS because of the need of high selectivity in the highly complex nature matrix (e.g. plasma) and the demand for high sensitivity to monitor the drug
effect at a low dose or after a long time period. Multiple reaction monitoring (MRM) using a triple quadrupole (QQQ) MS is a powerful tool for quantitative measurement of the target biotech drug in a complex matrix. MS-based quantitative methods have the necessary characteristics such as high specificity, sensitivity and precision required for PK/PD study.

In the MRM mode, two quadrupoles, functioning as mass filtering, are involved in a QQQ MS to reduce the matrix effect. In the quadrupole Q1, a precursor ion is preselected, whereas other ions are filtered. The selected precursor ion travels to the second quadrupole Q2 and is fragmented by collision. In the third quadrupole Q3, instead of scanning all MS/MS fragmented ions from precursor ion (Fig. 1.16a), only a few defined product ions are allowed to be detected by Q3 (Fig. 1.16b). This specific precursor ion-product ions analysis using MRM can improve the sensitivity for lower limited of detection by filtering interference ions which have different m/z values from the interested target, thus allowing high-throughput monitoring of the interest ions.\textsuperscript{195}

MRM methods provide both absolute and relative quantitation of target concentration when coupled with stable isotopes. A synthetic peptide that has an identical sequence of the target of interest is labeled with a stable isotope and then spiked into the sample as an internal standard. The concentration of the target can be measured by the ratio of the signals from the spiked-in labeled and endogenous unlabeled species. Using the internal standard in quantitative MRM analysis can also correct for variations during sample preparation and sample injection because labeled and unlabeled targets have the same physicochemical properties and differ only by mass\textsuperscript{196}. The labeled internal standard approach quantified by MRM using QQQ was used in the
quantitation of hepcidin level dosing by interference RNA (siRNA) in animal plasma as described in Chapter 4.

Figure 1.16. Illustration of two targeted analysis approaches. (a) conventional data-dependent analysis and (b) targeted MRM analysis on QQQ MS. (Adapted from Carr\textsuperscript{195}, copyright from NPG)

1.5 References


40. Knox, J. & Scott, H. B and C terms in the Van Deemter equation for liquid

41. Causon, T. J., Shellie, R. a, Hilder, E. F., Desmet, G. & Eeltink, S. Kinetic optimisation of
    open-tubular liquid-chromatography capillaries coated with thick porous layers for

42. MacNair, J. E., Opiteck, G. J., Jorgenson, J. W. & Moseley, M. a. Rapid separation and
    characterization of protein and peptide mixtures using 1.5 microns diameter non-porous
    silica in packed capillary liquid chromatography/mass spectrometry. Rapid

43. Fekete, S., Ganzler, K. & Fekete, J. Efficiency of the new sub-2 μm core-shell (Kinetex™
    column in practice, applied for small and large molecule separation. Journal of
    pharmaceutical and biomedical analysis 54, 482–90 (2011).

44. Forster, S., Kolmar, H. & Altmaier, S. Synthesis and characterization of new generation
    open tubular silica capillaries for liquid chromatography. Journal of chromatography. A
    1265, 88–94 (2012).

45. MacNair, J. E., Lewis, K. C. & Jorgenson, J. W. Ultrahigh-pressure reversed-phase liquid

    peak capacity single dimension LC-MS/MS platform for high-throughput, in-depth

47. Xie, S., Svec, F. & Fréchet, J. polymer monoliths: preparation of sorbent materials with
    high-surface areas and controlled surface chemistry for high-throughput, online, solid-
    <http://pubs.acs.org/doi/abs/10.1021/cm9804867>

48. Ivanov, A. R., Zang, L. & Karger, B. L. Low-attomole electrospray ionization MS and
    MS/MS analysis of protein tryptic digests using 20-microm-i.d. polystyrenedivinylbenzene

    1958).

    chromatography: an overview of various concepts. Journal of chromatography. A 856,


126. Rifai, N., Gillette, M. a & Carr, S. a. Protein biomarker discovery and validation: the long

Vázquez, J. Quantitative proteomics using 16O/18O labeling and linear ion trap mass

128. Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative
analysis of complex protein mixtures by MS/MS. *Analytical chemistry* 75, 1895–904
(2003).

129. Dayon, L. *et al.* Relative quantification of proteins in human cerebrospinal fluids by

130. Li, Z. *et al.* Systematic comparison of label-free, metabolic labeling, and isobaric
chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. *Journal of
proteome research* 11, 1582–90 (2012).


132. Ong, S.-E., Foster, L. J. & Mann, M. Mass spectrometric-based approaches in quantitative

133. Chelius, D. & Bondarenko, P. Quantitative profiling of proteins in complex mixtures
using liquid chromatography and mass spectrometry. *Journal of proteome research*

quantification using LC-coupled ion trap or FT mass spectrometry: Reproducibility,
linearity, and application with complex proteomes. *Journal of Proteome Research* 5,

135. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control.

136. Schwanhäußer, B. Global analysis of cellular protein dynamics by pulse-labeling and

137. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized


Chapter 2

The Fundamental Theory of Porous Layer Open Tubular Column and Optimization of LC Separation
Abstract

In this study, the gain in ESI-MS sensitivity was first demonstrated at ultralow flow rates < 20 nL/min for peptides separated by 10 µm i.d. PLOT columns compared to the sensitivity at conventional flow rate > 100 nL/min. Up to a 10-fold increase in sensitivity was observed at 2.5 nL/min in comparison to 80 nL/min. The van Deemter plot of a 3 meter PLOT column was established, and the optimal flow rate was found to be 5 nL/min. A complete equation, including the parameters of gradient time, column length, column diameter, flow rate and temperature, was derived from Snyder’s gradient elution theory. Based on the established van Deemter plot, the performance of different PLOT columns could be extrapolated, and the optimal conditions for best separation of a limited amount of a complex sample was shown to be simply 5 nL/min flow, 8-10 m column length and 6 h gradient. Further potential improvements could be obtained at elevated column temperatures and use of smaller i.d. PLOT column. In the 1D analysis of 500 cells of an MCF-7 lysate, a total of 3679 proteins could be identified in a single analysis and 4720 proteins in a combination of three runs. A 2D SCX-PLOT analysis of 500 MCF-7 cells generated a 13% gain in protein IDs but 6-fold increase in analysis time. The attempts of modifying PLOT column stationary phase showed that the 0.5% incorporated carbon nanotubes (CNTs) and 2nd thin layer octadecene-divinylbenzene modified PLOT column exhibited increased hydrophobicity and improved performance. Modifications of monolithic SPE columns with alky monomer also presented improved loading capacity, hydrophobicity and separation. In additional, a double layer PLOT column in 10 µm i.d. capillary and single layer PLOT column in 5 µm i.d. capillary were prepared the demand of narrower bore PLOT column.
2.1 Introduction

The trend of miniaturization in high performance liquid chromatography (HPLC) has increased during recent years to satisfy the demand of analysis of limited amounts of samples and low consumption of solvent\(^1\). The packed bed columns, as the dominant format in conventional HPLC columns, were reduced in column diameters and particle size to meet such miniaturization, to generate considerably higher chromatographic performance than conventional\(^2\)\(^-\)\(^3\). For packed columns, the preparation of microbore or nanobore columns were highly dependent on the packing conditions, and the performance was limited by the produced high column backpressure\(^3\)\(^-\)\(^4\). For the potential advantage of high permeability, open tubular (OT) columns have reawakened interest in the application in micro-scale HPLC after its big success in gas chromatography (GC). OT columns were introduced by Golay in GC separation in 1950s\(^5\), demonstrating high efficiency in the separation of volatile compounds with over one million theoretical plates using a long column\(^6\)\(^,\)\(^7\). Tsuda and other co-workers introduced open tubular columns for microbore (50-60 \(\mu\)m i.d.) LC separation\(^8\)\(^-\)\(^11\). However, the theoretical prediction indicated that the power of OT column would be limited unless a diameter of 5-10 \(\mu\)m i.d. could be operated routinely\(^12\). For this purpose, our group developed a polymer-based 10 \(\mu\)m i.d. porous layer open tubular (PLOT) column with high resolving power, good column-to-column reproducibility and high permeability\(^13\).

10 \(\mu\)m i.d. PLOT columns in microscale separation have several advantages over packed columns. One advantage is that the tremendously increased permeability, due to the porous polymer and open tubular structure, produces much lower backpressure compared to packed columns\(^14\). This feature allows running the PLOT columns at ultralow flow rate (e.g. 20 nL/min)
with moderate backpressure < 350 bar, whereas the packed column of comparable inner diameter of 20-25 µm packed with sub-2 µm particles usually produces over 1000 bar backpressure at the same flow rate\textsuperscript{15,16}. Compared to conventional flow technique, the use of nanoelectrospray ionization (nESI) in miniaturized HPLC, resulted in increased MS sensitivity due to increased ionization efficiency, reduced ion suppression and improved ion transmission\textsuperscript{17}. Especially at ultralow flow rate < 20 nL/min, significant gain in MS response and nearly eliminated ion suppression were observed\textsuperscript{18,19,2}.

Another potential advantage is the improved HPLC performance of PLOT columns. In packed columns, eddy diffusion, which is the A-term in the van Deemter equation caused by multipath of the analytes, accounts for a major contribution to the band broadening\textsuperscript{14}. However, the eddy diffusion is theoretically eliminated in the PLOT column because no multipath process exists in the open tubular capillary\textsuperscript{20,21}. An improved mass transfer coefficient in nanobore OT column is also considered as an important factor to decrease band broadening. In PLOT columns, the improve the mass transfer rate is improved with decreased diameter, and therefore the C-term in the van Deemter equation (resistance to mass transfer coefficient) is smaller\textsuperscript{22}. With the above characteristics, the PLOT column with nanobore i.d. would exhibit excellent HPLC performance if operating the system at an ultralow flow rate is practical.

In this study, we demonstrate the routine operation of 10 µm i.d. PS-DVB PLOT column at ultralow flow rate < 20 nL/min with the expected gain in MS sensitivity. Based on the established van Deemter plot, we used Snyder’s gradient theory\textsuperscript{23} to predict the efficiency of the separation and optimized parameters such as the flow rate, gradient time, column length and
temperature. In the application of PLOT columns for analysis of limited sample amounts by 1D or 2D analysis, coupling to a high resolution mass spectrometer showed high efficiency and sensitivity. Several attempts to increase the hydrophobicity of PLOT column were also studied to improve the efficiency of the column. These studies are detailed below.

2.2 Experimental

2.2.1 Materials and Reagents

Styrene divinylbenzene, ethanol, tetrahydrofuran, 1-decanol, 3-(trimethoxysilyl) propyl methacrylate (TMSPMA), 2,2-diphenyl-1-picrylhydrazyl (DPPH), dimethylformamide (DMF), 2,2′-azobis(2-methylpropionitrile) (AIBN), decanol, THF, decene (analytical), octadecene (analytical), sodium hydroxide (1 M), and standard peptides were purchased from Sigma-Aldrich (St. Louis, MO). Formic acid, HPLC grade water and acetonitrile were obtained from Thermo Fisher Scientific (Waltham, MA). Fused silica capillary was from Polymicro Technologies (Phoenix, AZ). Six bovine digest mixture was from Bruker-Michrom (Auburn, CA). The carbon nanotubes were synthesized and cut by Harvard University.

2.2.2 PLOT Column Preparation

2.2.2.1 Single Layer PLOT Column

A 10 µm i.d. fused-silica capillary (~ 20 meters) was put in a stainless steel vessel and first flushed 2 h with 1.0 mol/L NaOH at 1000 psi. The column was then washed with water for 4 h and then flushed with acetonitrile for 4 h. The capillary was dried with nitrogen to remove
residual water and acetonitrile for 4 h each. 30% (v/v) TMSPMA and 0.5% (wt/v, g/mL) DPPH in DMF was freshly prepared. The solution was filled into the 10 µm i.d. pretreated capillary at 1000 psi for 2h. Both ends of the capillary were sealed with a septum, and the capillary was placed in an oven at 110 °C. After 6 h, the capillary was washed with acetonitrile and blown dry with nitrogen. A polymerization solution containing 5 mg of AIBN, 0.1808 g styrene, 0.1823 g divinylbenzene (DVB) and 0.4523 g ethanol was filled into the silanized capillary at 1000 psi for 2h. It is necessary to keep four significant numbers in the measurement to produce the appropriate pore size in the monolith. Both ends of the capillary were immersed into the polymerization solution and the vial was pressurized by a syringe. The capillary with the vial was heated at 74 °C for ~16 h in a water bath. The PLOT column was washed with acetonitrile and water alternatively overnight to remove polymer residues. Both ends were examined under the microscope and 20-30 cm long pieces were cut from both ends.

2.2.2.2 Double Layer PLOT Column

The double layer PLOT column was separately prepared by filling the same polymerization solution as in preparation of single layer PLOT column into a 5 m long PLOT column and reacted at 74 °C for ~16 h. The double layer PLOT column was washed with acetonitrile and water alternatively overnight to remove polymer residues not anchored to the first layer.

2.2.2.3 Carbon Nanotube PLOT Column

The carbon nanotube (CNT) modified PLOT column was prepared by incorporating well-cut CNTs in the polymerization. The multiwalled CNTs were synthesized and cut in acid by with an average diameter < 100 nm as described by Svec24. Two PLOT columns were modified by adding 0.25% and 0.5% CNTs (to the total weight of monomers) into the polymerization solution,
respectively and sonicated in a water bath for 30 mins to generate a fine suspension. The stability of the CNTs suspension was examined for 24 h without observing any precipitation. The solution was filled into the capillary and heated as the preparation for the PLOT column.

2.3.2.4 Second Layer OD Grafted PLOT Column

The octadecene (OD) grafted PLOT column was a modification of the layer of the PLOT column. A mixture of 60/20/20 (v/v) of OD/ethanol/decanol was freshly prepared with 1% (w/w) of AIBN/monomer solution. The solution was sonicated for 5 min and pumped into a 5 m PLOT column dried by nitrogen at a pressure of 1000 psi. The column was immersed in the water bath at 80 °C with the solution pumping into the PLOT column for 16 hr.

2.3.2.5 Second Thin Layer OD-DVB PLOT Column

A second thin layer OD-DVB approach involved coating a thin layer of cross-linked polymer with alky groups exposed on the surface. A mixture of 25/20/30/30 (v/v) of OD/DVB/ethanol/decanol with 1% (w/w) of AIBN to total monomers was sonicated for 5 min and filled into a well prepared PLOT column. Both ends of the capillary were immersed into the polymerization solution, and the capillary was heated at 74 °C for ~16 h in a water bath. The column was cleaned using the same protocol as in preparation of the PLOT column.

2.2.3 Monolithic Column (SPE) Preparation

2.2.3.1 PS-DVB Monolithic Column

The 50 µm i.d. fused-silica capillary (30 cm) was silanized in a similar manner as for the PLOT column, except using 30 min for each wash and 2 min for filling at 100 psi. For PS-DVB based monolithic column,
a polymerization mixture of 0.1808 g styrene, 0.1826 g DVB, 0.434 g decanol, 0.0622 g THF and 5 mg AIBN was filtered, degassed by ultrasonication for 5 min and then filled into the silanized capillary. Both ends of the capillary were immersed into the polymerization solution, and the capillary was heated at 70 °C for ~16 h in a water bath. The monolithic column was ready to use after washing by acetonitrile and water overnight.

2.2.3.2 Alkyl Monomer Modified PS-DVB Monolithic Column

Decene (C\textsubscript{10}) and octadecene (C\textsubscript{18}) modified columns were prepared by changing the decanol to THF ratio to obtain the rigid monolithic structure. The appropriate recipe for C\textsubscript{10} modified column was 5 mg AIBN, 0.453 g styrene, 0.1365 g decene, 0.1827 g divinylbenzene, 0.1010 g THF and 0.3987 g decanol and for C\textsubscript{18} modified column was 5 mg AIBN, 0.4530 g styrene, 0.1365 g octadecene, 0.1827 g divinylbenzene, 0.1331 THF and 0.3739 g decanol.

2.2.4 PLOT Columns Platforms

2.2.4.1 Split-injection PLOT Design

The split-injection system is the simplest system to test the performance of a PLOT column at ultralow flow rate. All ends of capillaries and PLOT columns used were precisely 90° cut using a diamond cutter (New Objective, Woburn, MA or Sigma, St. Louis, MO). Further polish was required using capillary polish station (ESI solutions, Wobrun MA) to minimize the dead volume between the connections.

The schematic diagram of the split-injection PLOT system is shown in Figure 2.1. The gradient delivery capillary, splitter and PLOT column were connected by a PEEK tee (IDEX, P-775, Lake Forest, IL) with 29 nL swept volume. The split ratio was set to 1:2000 at a given
length of splitter. A distal coated silica tip (New Objective, #FS360-20-10-D) was cut to 3 cm. To make a zero dead volume connection between the PLOT column and emitter, a PicoClear union (New Objective, PCU-360) was used. The PLOT column and spray tip were well aligned under the microscope and inspected without any scratched particles in the connection.

Figure 2.1. Schematic diagram of split-injection PLOT design.
2.2.4.2 SPE-PLOT 1D System

To improve the loading capacity, a 4 cm, 50 µm PS-DVB monolithic column was placed before the PLOT column as a trapping column. In order to achieve fast loading, a three way PicoClear tee (New Objective) was used to connect the SPE, PLOT column and waste capillary (Fig. 2.2). The three capillaries were well aligned so that the gap between the SPE and PLOT columns would be in the middle of the 50 µm i.d. waste line (Fig. 2.2). The splitter and waste were connected to an automatic 6 port valve with a plug to stop the flow. The gradient flow was delivered by the micropump at 400 nL/min and split at 1:20 ratio to obtain 20 nL/min flow rate on the PLOT column. By switching the valve, the system could change from the “injection” mode to the “separation” mode. In the injection mode (the blue lines shown in Fig. 2.2), the sample was directly loaded onto the SPE column at 200 nL/min, and the waste went through the waste line in the PicoClear tee. In the separation mode, the waste line in the PicoClear tee was closed, and therefore, the sample was eluted from the SPE and separated on the PLOT column at ultralow flow rate.
Figure 2.2. Schematic diagram of SPE-PLOT 1D system. The valve position (in blue) represents the “injection mode” of the system and zero dead volume connection of SPE, waste line and PLOT column was implemented by using a transparent PicoClear tee.
2.2.4.3 SCX-SPE-PLOT 2D System

In the two dimensional PLOT separation, an orthogonal separation column was introduced into the SPE-PLOT 1D system. In our 2D separation, an SCX (strong cation exchange) first dimension column was prepared in a 75 µm i.d. capillary with a frit\textsuperscript{25} by sequentially packing 2 cm SCX particle (PolySulfoethyl A, 300Å, 5 µm, Nest Group, Southborough, MA) and 2 cm C\textsubscript{18} particle (Magic, 200Å, 3 µm, Nest Group). Here, the RP packing in front of the SCX column is used to clean-up the salt and other buffers used in the sample preparation, and the salt or buffers in the prepared sample could cause the sample unretained on the SCX column. An additional valve was introduced into the 1D system so that the SCX column could automatically turn off-
line and on-line. In the “separation mode” with the RP-SCX column on-line (flow direction shown in Fig. 2.3), the sample was loaded on the RP packing and desalted at the high flow rate of 300 nL/min. The desalted sample was then transferred to the SCX packing with high organic solvent (90% ACN). After balancing the SCX and SPE column with 0.1% FA in water, the system was switched to the “injection mode” and different salt plugs were injected on SCX column and the eluted peptides trapped and desalted on the SPE column at 200 nL/min. Finally, with the RP-SCX off-line and system in the “separation mode”, the fractions were separated by the PLOT column at 20 nL/min split from 400 nL/min flow rate delivered by micropump.

2.2.4 LC and Mass Spectrometry

For performance testing of PLOT columns, the split-injection system was used. 40 µL/min flow from the loading pump of the Dionex Ultimate 3000 (Dionex, CA) was split in a 1:2000 ratio to obtain 20 nL/min flow in the PLOT column. For 1D analysis using the SPE-PLOT system, a 4 cm, 50 µm i.d. decene modified PS-DVB monolithic SPE precolumn was connected with a 4.2 m PS-DVB PLOT column. Digested peptides were first loaded on the SPE column at 200 nL/min flow using a Dionex NCS 3500 RS pump. Then, trapped samples were separated on the PLOT column with a 20 nL/min flow rate split from 400 nL/min. The separation was developed using a 4 hour gradient of 0%-27% mobile phase B (mobile phase A, 0.1% FA in water; mobile phase B, 0.1% FA in ACN). The SPE and PLOT columns were washed with 90% B for 20 min and re-equilibrated with mobile phase A. In 2D analysis, peptides trapped on the RP packing were transferred to the SCX packing by 90% B. The peptide fractions were eluted sequentially by injecting 2 µL of 0, 5, 20, 50, 100 and 500 mM ammonium acetate buffer (pH
3.7). Nano ESI spray was generated from a coated tip by an electrospray voltage of 1.1 eV. The ion transfer tube temperature was set at 275 °C.

For monolithic column tests, 20 cm PS-DVB and modified monolithic columns were connected to the micropump of the Dionex Ultimate 3000 system and operated at 100 nL/min flow rate. 50 fmol of the six bovine protein digest was directly loaded on the monolithic column and separated using a 30 min gradient from 0%-30% mobile phase B.

MS detection was accomplished using MS/MS data-dependent scan of the top 12 most intense precursor ions on the Q Exactive (Thermo Fisher, Waltham, MA) mass spectrometer. Full MS spectra were acquired over the range of m/z 380-1600 with a resolution of 70,000 (at m/z 200) and an automatic gain control (AGC) target of 3e6. The 12 most intense ions with charge state exclusion of single and unassigned charges were selected for HCD fragmentation with a normalized collision energy (NCE) of 28%. The MS/MS spectra were analyzed in the Orbitrap mass analyzer with a resolution of 17,500 and AGC of 1e5. The isolation window was set to 2 m/z and dynamic exclusion was 60 s for 4 hr separation or 15 s for 30 min gradient. The maximum injection for acquisition was 20 ms for full MS scan and 120 ms for MS/MS.
2.3 Results and Discussion

2.3.1 ESI-MS Sensitivity at Ultralow nL/min Flow Rates

Compared to conventional nanoflow, the use of nanoelectrospray ionization at ultralow flow rates, e.g., <20 nL/min, results in increased MS sensitivity due to increased ionization efficiency, reduced ion suppression and improved ion transmission\(^{17}\). Previous studies of MS response with flow rate were performed by direct infusion, and the MS efficiency was calculated using relative signal intensity ratios between analyte and suppressed background ions\(^{19}\). In the present study, we split-injected identical 800 amoles of standard peptides (angiotensin I and bradykinin 2-9) and monitored the signal intensity changes in the separation using a 3 m long PLOT column at flow rates ranging from 2.5 nL/min to 80 nL/min. Due to the pressure limit (800 bar) of the HPLC pump, the maximum flow rate obtained on the 3 m PLOT column was 80 nL/min. The absolute signal intensity \((2.02 \times E7)\) for angiotensin I at 2.5 nL/min flow rate was almost 10-fold higher than the intensity \((2.36 \times E6)\) at 80 nL/min (Fig. 2.4). Even at 20 nL/min which was a normal operational flow rate for PLOT column, the signal intensity of angiotensin I \((5.51 \times E6)\) was 2 to 3-fold higher than that at 80 nL/min (chromatogram not shown). A similar significant gain in ESI sensitivity was observed for the peptide standard bradykinin 2-9 fragment (Fig. 2.4). The plot of the absolute signal intensity of a given amount peptide vs. flow rate on the PLOT column agreed with our previous direct infusion study on PLOT using relative signal intensity ratios, which revealed that relative signal intensity could gain ~5-fold at 20 nL/min than conventional 200 nL/min nanoflow\(^{26}\). The results of Fig. 2.4 demonstrate that nanoESI/MS at 20 nL/min or lower can significantly increase the sensitivity of MS detection, which facilitates the analysis of limited amounts of biological samples.
2.3.2 PLOT Column Efficiency at Different Flow Rates

Evaluation of column efficiency has usually been measured at various flow rates under isocratic conditions\textsuperscript{27}. Due to the open tubular structure of PLOT column, the focus of analytes at the head of PLOT column was not as good as that in packed column. Evaluation of PLOT column under isocratic condition resulted in low efficiency during our study and could not reflect the true power of PLOT column in separating peptides using gradient elution. Therefore, evaluation of both the PLOT and packed columns were conducted under gradient conditions.
Theoretical plates $N^*$ in gradient condition can be calculated from Snyder’s linear strength model\textsuperscript{28}. $N^*$ is described as:

$$N^* = 5.54 \left[ \frac{t_0 \left(1 + 0.5k^*\right)}{W_{0.5}} \right]^2$$

or

$$N^* = 16 \left[ \frac{t_0 \left(1 + 0.5k^*\right)}{W} \right]^2$$

(2.1)

where $t_0$ is the column dead time, $W$ the $4\sigma$ peak width at 13.3% height, $W_{0.5}$ the peak width at half-height and $k^*$ the gradient retention factor, defined as the retention factor $k$ at the middle point of the column. The gradient retention factor $k^*$ can be expressed in the following equation from Snyder’s theory\textsuperscript{23}:

$$k^* = \frac{t_G}{1.15t_0 \Delta \phi S}$$

(2.2)

where $t_G$ is the gradient separation window from the first to the last peak, $\Delta \phi$ the mobile phase volume fraction change in the gradient and $S$ a constant value for a given analyte (slope of the $\ln k$ vs. volume fraction of mobile phase).

Recently, Snyder introduced the concept of gradient compression in which the compression of gradient delivery accounted for a narrower peak observed than in theory. The compression of peak band by the factor $G$ is introduced into the Equation 2.1 as\textsuperscript{29}:

$$N^* = 5.54 \left[ \frac{G \cdot t_0 \left(1 + 0.5k^*\right)}{W_{0.5}} \right]^2$$

or

$$N^* = 16 \left[ \frac{G \cdot t_0 \left(1 + 0.5k^*\right)}{W} \right]^2$$

(2.3)

The gradient compression factor $G$ is defined by the quantity $p$.
\[
G = \sqrt{\frac{1 + p + p^2}{3}} \quad (2.4)
\]

\[p \approx \frac{2}{k^*} \quad (2.5)\]

In isocratic separation, solvent strength and retention factor \( k \) are constant, regardless of the change in flow rate. However, the gradient strength will change with the flow rate\(^{28}\). To compare the efficiency of columns at different flow rates under the same gradient strength, the gradient retention factor \( k^* \) was maintained constant at approximately 3 by varying the gradient time with flow rate for use in the classic isocratic van Deemter plot\(^{30}\). The van Deemter plot equation can be described as\(^{31}\):

\[
H = A + \frac{B}{u} + Cu \quad (2.6)
\]

where \( u \) is the linear velocity of the flow and \( H \) stands for height equivalent to a theoretical plate (HETP) which is defined as:

\[
H = \frac{L}{N^*} \quad (2.7)
\]

where \( L \) is the column length.

By measuring plate counts at different linear velocities, we could plot the \( H-u \) relationship as shown in Fig. 2.5. Both packed 20 μm i.d. C\(_{18}\) column and 10 μm i.d. PLOT column presented
very low HETP of 1.6 µm and 3.7 µm, respectively, at the optimum 5 ng/mL flow rate for peptide [Glu¹]-fibrinopeptide B. The 6 cm packed C₁₈ column could achieve ~33,000 theoretical plates. At similar column backpressure (~160 bar), 3 m PLOT column could generate ~800,000 theoretical plates (Table 2.1). Although the plate height of PLOT column was higher, the PLOT column benefited from the much longer column length and produced over 20-fold increased column efficiency compared to packed column. Packed bead columns with similar theoretical plates could be obtained using a 20 µm i.d. capillary, 1.7 µm particle size and up to 1m length; however, it resulted in extremely high backpressure at 20,000 psi (1379 bar)³². To minimize the extra column effect at ultralow flow and obtain sensitivity in MS detection in the practical operation, 20 nL/min flow rate other than the optimal flow rate was used with the PLOT column in general operation, and still ~270,000 theoretical plates could be achieved. The extremely high theoretical plates enabled us to obtain relatively high column efficiency at normal backpressure <200 bar. As a final point, the observed C term in the PLOT column was smaller than that in packed column. This was probably accounted for the small i.d. of PLOT that increased the diffusion efficiency³³.
Figure 2.5. van Deemter plot of 10 μm i.d. PLOT and 20 μm i.d. packed C\textsubscript{18} columns with 1.7 μm beads at ultra-nano flow.

2.3.3 Theoretical Study of Peak Capacity Optimization of PLOT Column

Plate numbers between different columns or in different operational conditions may have larger variations, so that it is difficult to have direct visual comparisons. Therefore, peak capacity (PC), defined as the maximum number of adjacent peaks that can be separated in a chromatographic run, is widely used to evaluate the overall separation efficiency of complex mixtures\textsuperscript{34}. The gradient elution peak capacity can be described as follows\textsuperscript{35}:  

84
The peak capacity is related to the gradient theoretical plates by substituting the Equations (2.2) and (2.3) into (2.8) resulting the peak capacity as:

\[
P C = \frac{t_G}{W} + 1
\]  

(2.8)

\[
PC = \frac{\sqrt{N^*}}{4G} \times \frac{t_G}{t_0 + \frac{t_G}{2.3\Delta \phi S}}
\]  

(2.9)

As the discussed above, the theoretical plates \(N^*\) of a column is also affected by the linear velocity \(u\) in the van Deemter Equation (2.6). The column dead time \(t_0\) can be also expressed as a function of \(u\) as:

\[
t_0 = \frac{L}{u}
\]  

(2.10)

Taking Equation (2.6) and (2.10) into the peak capacity, we can get a complete expression of peak capacity as a function of column length \(L\), linear velocity \(u\) and gradient time \(t_G\).

\[
PC = \frac{\sqrt{L}}{4G\sqrt{A + B/u + Cu}} \times \frac{1}{\frac{L}{u t_G} + \frac{1}{2.3\Delta \phi S}}
\]  

(2.11)

With Equation 2.11 and the established van Deemter plot in Figure 2.5, we can predict the peak capacity using various \(L, t_G\) and \(u\) parameters in order to optimize the peak capacity. To simplify the estimation using Equation 2.11, we neglected the effect of gradient compression here due to the fact that in practice the actual \(G\approx1\) for reasonable slow gradient steepness in
proteomic separation of complex samples. In the further, a programed software can be written
to involve more parameters to predict complex situation.

We first fixed the flow rate at 20 nL/min and varied the column length as well as the gradient
time. As seen in Fig. 2.6A, at a short gradient time, increasing the column length of a 10 µm i.d.
PLOT column does not improve the PC significantly and at 30 min gradient time, the PC was
even decreased above a 4 m long PLOT column. The benefit of column length in PC occurs at
longer gradient times. For a given length of column, the PC showed a logarithmic growth with
the gradient time. The PC could increase 3 to 4-fold using 4 h gradient relative to the 30 min
gradient, increasing slower at longer gradient. Generally, to achieve the maximum PC, a longer
column within limited backpressure range and a relatively long gradient time would improve
the overall separation of a complex mixture. For the 3 m PLOT column at 20 nL/min in van
Deemter plot study, for example, a PC of 455 (90% of the maximum PC) would be reached at
240 min gradient for peptide [Glu^1]-fibrinopeptide B (all PCs below are extrapolated for
[Glu^1]-fibrinopeptide B) and 600 min gradient would generate PC of 506 which only increased
10% over the earlier one.

In the van Deemter plot of the PLOT column (Fig. 2.5), the optimal flow rate converted was
around 5 nL/min. For a given length of a 3 m PLOT column, we plotted the relationship
between PC, gradient time and flow rate. As observed in Fig. 2.6B, the PC increased from 86
for a gradient time of 30 min. at 2.5 nL/min to reach the maximum PC of 337 at 40 nL/min and
then slowly decreased with the flow rate. With a 240 min gradient, the optimal flow rate for
best peak capacity decreased to 15 nL/min and slope before and after the optimal flow rate
increased. Considering all effects, increasing the gradient time and maintaining the flow rate
close to 5 nL/min is predicted to improve the PC to 773 for a 3 meter column. The PC could be further improved to 1050 by using a 8 meter PLOT column.

Preparing a smaller i.d. PLOT column, e.g. 5 µm, is another approach to improve the efficiency of the PLOT since with small diameter of open tubular, the diffusion of analytes in mobile phase are increased (See Equation 1.9 in Chapter 1, \( H \propto d_c^2 \)). However, since the van Deemter plot of a 5 µm i.d. is not available, we can only make a rough assumption that the column efficiencies of different i.d. at the same flow rate are identical, because the main band broadening caused by \( C_m \) term - \( \frac{d_c^2 u}{D_M} \) will be constant (Equation 1.9 in Chapter 1). Therefore, for a given length, gradient time and flow rate, diameter will only change the linear velocity.

Fig. 2.6C shows several PC- column diameter plots under different conditions. One curve, which is for a 10 h gradient, with 8 m column and at 5 nL/min, shows that the PC ramps up from 964 to 1280 by decreasing the column diameter from 10 µm to 5 µm i.d.. The real performance of the 5 µm i.d. PLOT requires further validation on evaluation of a well-prepared 5 µm i.d. PLOT column.

Recently, a theoretical kinetic study on open tubular columns revealed the large benefits from elevated temperatures for column optimization\textsuperscript{36}. The \( H-u \) curve can be described in a more detailed equation\textsuperscript{37}:

\[
H = A + \frac{b \cdot D_M}{u} + \frac{c}{D_M} u
\]

(2.12)
where $D_M$ is the diffusion coefficient which is related to the absolute temperature $T$ and viscosity $\eta$ of the mobile phase:

$$D_M = \gamma \cdot \frac{T}{\eta}$$

(2.13)

where $\gamma$ is a simplified constant which is related to the molecular weight and molar volume of the analyte. At elevated temperature, the viscosity of mobile phase decreases and the value of $T$ in the numerator of Equation 2.13 increases; hence the diffusion coefficient increases, resulting in a higher $B$ term and a lower $C$ term. Therefore the HETP will be affected by varying the temperature and linear velocity. Here, we used an experimental viscosity of 20% ACN mobile phase at different temperature$^{38}$ and plotted the estimated $PC$ vs. flow rate and temperature. The 3D plot in Fig. 2.6.D illustrates that at flow rate $> 10$ nL/min, the $PC$ is increased with the elevated temperature; however, the $PC$ started to drop with elevated temperature at the flow rate lower than 10 nL/min. The explanation is that at higher linear velocity, the contribution of \(\frac{b \cdot D_M}{u}\) in the van Deemter is minor and \(\frac{c}{D_M} u\) is the main contribution to the HETP. Therefore, at elevated temperatures, the increased diffusion coefficient will improve the HETP. In contrast, at lower linear velocity, elevated temperature will not improve the separation. Another explanation easier to understand is that the optimal linear velocity (\(u_{opt} = \sqrt[3]{\frac{B}{C}}\)) in the van Deemter plot will be higher because of the higher $B$ term and lower $C$ term at elevated temperature. In this case, a 10–15 nL/min flow rate is recommended on the PLOT column with elevated temperature at 60 °C and a $PC$ of 1054 is predicted to be obtained for a 8 meter PLOT column using 600 min gradient.
The above discussion can help us better interpret the fundamental behavior of PLOT columns in gradient HPLC separation. More importantly, the established van Deemter plot of PLOT column combined with the Snyder’s linear gradient model can be used to predict the peak capacity and optimize the separation using PLOT column. The peak capacities measured by peak width of 4σ in the van Deemter plot experiment are listed in Table 2.1. Using the van Deemter plot and Equation 2.11, the measured peak capacities were very similar to the predicted values. In fact, the $PC$ of a separation of complex was usually lower than what we predicted using $[\text{Glu}^1]$-fibrinopeptide as a model. Results of Table 2.1 also showed that the average $PC$ of three peptides was 27% less than the $PC$ obtained by extrapolated $[\text{Glu}^1]$-fibrinopeptide, for the observed wider band broadening of other two peptides, bradykinin and angiotensin. In a complex peptide mixture, not all the peptides behave similar and some may have more band broadening due to their physical characteristics, such as higher diffusion in the mobile phase and slower diffusion coefficient in the stationary phase. Using the 73% extrapolated $PC$ from $[\text{Glu}^1]$-fibrinopeptide can give a better estimation of a complex mixture separation. For a 400 min separation with 8 m PLOT prepared by another group$^{39}$, we estimated a $PC$ of 564 using $[\text{Glu}^1]$-fibrinopeptide. The 27% less of extrapolated $PC$ (411) is very similar to their measured $PC$ of 415.
Figure 2.6. Prediction of effects of different parameters on peak capacity for (A) column length and gradient time; (B) flow rate and gradient time; (C) column diameter and (D) temperature.
Table 2.1. Comparisons of theoretical plates and peak capacities at different flow rate

### 3 m PS-PVB PLOT column (1.3 μm thickness)

<table>
<thead>
<tr>
<th>$F$ (nL/min)</th>
<th>$t_G$</th>
<th>$N^*$</th>
<th>$PC_{\text{measured}}$</th>
<th>$PC_{\text{extrapolated}}$</th>
<th>$PC_{\text{avg.}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5</td>
<td>240</td>
<td>853767</td>
<td>472</td>
<td>444</td>
<td>346</td>
</tr>
<tr>
<td>5</td>
<td>120</td>
<td>800770</td>
<td>457</td>
<td>460</td>
<td>322</td>
</tr>
<tr>
<td>7.5</td>
<td>80</td>
<td>613891</td>
<td>400</td>
<td>412</td>
<td>282</td>
</tr>
<tr>
<td>10</td>
<td>60</td>
<td>527715</td>
<td>371</td>
<td>368</td>
<td>246</td>
</tr>
<tr>
<td>15</td>
<td>48</td>
<td>452022</td>
<td>346</td>
<td>331</td>
<td>245</td>
</tr>
<tr>
<td>20</td>
<td>30</td>
<td>272883</td>
<td>262</td>
<td>264</td>
<td>203</td>
</tr>
<tr>
<td>25</td>
<td>24</td>
<td>205492</td>
<td>228</td>
<td>236</td>
<td>184</td>
</tr>
<tr>
<td>30</td>
<td>20</td>
<td>187292</td>
<td>217</td>
<td>215</td>
<td>145</td>
</tr>
</tbody>
</table>

### 8 m PS-PVB PLOT column (0.7 μm thickness)

<table>
<thead>
<tr>
<th>$F$ (nL/min)</th>
<th>$t_G$</th>
<th>$PC_{\text{measured}}$</th>
<th>$PC_{\text{extrapolated}}$</th>
<th>73% $PC_{\text{extrapolated}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>400</td>
<td>415</td>
<td>563</td>
<td>411</td>
</tr>
</tbody>
</table>

b. Extrapolated by peptides [Glu]$^1$-fibrinopeptide B
2.3.4 On-line 1D SPE-PLOT Analysis

Due to the thin monolithic layer and narrow column diameter, the loading capacity is not very high compared to a conventional 75 µm i.d. packed column. According to previous study in our group, the loading capacities were found ~100 fmol for angiotensin I and ~50 fmol for insulin. Also, it is not practical to load large volume sample (1~5 µL) using a 20 nL/min flow rate. To increase the loading capacity and reduce the sample loading time, a 4 cm 50 µm i.d. PS-DVB monolithic SPE trapping column was introduced into the 1D system by connecting the SPE and PLOT columns using a PicoClear tee (Fig. 2.2). To test the resolving power of SPE-PLOT system, 5 fmol of a mixture of equal mole standard peptides (1000 selected tryptic peptides conserved across Homo sapiens, labeled with stable isotope, obtained from Proteomics Standards Research Group) was first loaded on the SPE trapping column at 200 nL/min flow rate and then eluted onto a 4.2 m PLOT column for separation. In a 200 min gradient separation window, a high efficiency separation of the peptide mixture was observed, showing that 10 extracted ions with average peak width at 4σ of 0.6 min and a PC of 330, which is very similar to the estimated PC of 360 from the equation described above. Smith obtained a PC of 1150 estimated on one of the narrowest peaks using an 87 cm × 19.8 µm i.d. fused-silica capillary packed with 3 µm porous C18-bonded particles at a pressure of 20,000 psi (1379 bar). To use nanoflow in conventional HPLC pump, JA Marto packed a column (25 µm × 100 cm column) with larger 5 µm particle size to operate the column at 5 nL/min at 1,400 psi (97 bar) and obtained 755 PC in a 10 h gradient. If we perform a separation at 5nL/min flow on a 8 m PLOT column generated a similar backpressure, the predicted PC would be 704, and the efficiency is competitive to the packed column.
To test the SPE-PLOT-MS performance in the proteomic analysis of a biological sample with wide dynamic range, an MCF-7 cell lysate digest sample prepared from 10 million cultured cells was used. Cell lysate digest dilution corresponding to 500 MCF-7 cells was injected in this study to measure the sensitivity and dynamic range when the SPE-PLOT was coupled with a Q-Exactive mass spectrometer. The SPE-PLOT system showed a high efficiency and sensitivity in the separation of 500 MCF-7 cells (Fig. 2.8). An average of 3679 protein groups were identified in a single run and 4720 by combining three replicates (Fig. 2.9). The performance of the SPE-PLOT platform was compared with a 20 cm × 75 μm i.d. column packed with 2.7 μm C_{18} beads at conventional 200 nL/min flow rate. Injecting same 500 MCF-7 cell lysate, the packed C_{18} column resulted in only an average of 1152 proteins identified in a single run and 1496 proteins in all three replicates (Fig. 2.9). The SPE-PLOT resulted in more than 3-fold increase in protein IDs than the packed C_{18} column in detecting a limited amount of a biological sample.
Figure 2.7. Separation of standard peptide mixture by SEP-PLOT 1D system
Figure 2.8. Analysis of 500 MCF-7 cells using 4.2m PLOT column and 4 h long gradient

Figure 2.9. Advantage of SPE-PLOT system over 75 μm i.d. C_{18} column
Figure 2.10. 2D SCX-SPE-PLOT analysis of 500 MCF-7 cells and the comparison to 1D analysis.

2.3.5 On-Line 2D SCX-PLOT/LC-ESI-MS Analysis

Generally, one type of human cell can produce over 10,000 proteins (not considering proteoforms) from gene expression. No LC platform available has enough resolving power to separate the whole proteome in a single analysis. Combining a second orthogonal dimension of separation can significantly increase the peak capacity of separation\textsuperscript{41}. Using a pipette tip SAX protocol, Mann was able to identify 8,500 HeLa proteins per sample containing about 100,000 HeLa cells, which covered almost 80\% of the whole proteome\textsuperscript{42}.
As shown in Figure 2.3, we constructed an on-line 2D SCX-PLOT/LC-ESI-MS platform by adding an additional 6-port valve to the platform of Fig. 2.2. In 2D analysis, 500 MCF-7 cell lysate was first loaded onto the SCX column. The unretained peptides on the SCX column were trapped into the 50 µm i.d. PS-DVB monolithic SPE column, and then further separated by the PLOT column. 5 fractions of peptides from the SCX column eluted by step concentration of salt were analyzed by the SPE-PLOT system. The number of proteins identified was found highest in the salt concentration of 5 mM and 20 mM. In total, 4191 proteins were identified from 6 steps of 2D analysis (Fig. 2.10). A 13% gain in protein IDs was achieved in 2D analysis compared to 1D SPE-PLOT analysis at the cost of increased analysis time. In Mann’s research, protein groups identified in 2D analysis were almost 2-fold more than those in single dimensional analysis. The limited amount of digest peptides in our study was probably the reason for the lower gain of protein IDs in the 2D analysis. In the results of Fig. 2.9, the increase of accumulated protein groups became slower with the increase in the number of fractions. This indicated that some peptides were shared between different fractions. The second dimension separation by SCX fractions could provide additional separation but might also divide one peptide into several fractions. Some peptides whose levels were slightly above the limit of detection in a limited amount sample might be not below the level of detection after fractionation. Nevertheless, online 2D SCX-SPE-PLOT still provides additional identifications over 1D SPE-PLOT analysis.

2.3.6 Hydrophobicity Modified PLOT Columns

Because of the relative hydrophobicity of PS-DVB polymer, it has been used as a reversed stationary phase in the format of a monolith or cross-linked particle for the separation of proteins
or peptides\textsuperscript{43,44}. The PS-DVB PLOT columns developed by our group also demonstrated the ultraperformance of the PS-DVB stationary phase in peptide separation\textsuperscript{13,25,45,46}. However, the PS-DVB material is less hydrophobic compared to the most widely used commercial C\textsubscript{18} stationary phase\textsuperscript{47}, leading to poorer separation of small molecules and hydrophilic peptides.

From Equation 2.11, we could conclude that if the \( t_G \gg t_0 \), \( \frac{L}{u t_G} \) can be neglected and therefore the \( PC \) is proportional to \( \Delta \phi S \). The improved hydrophobicity of the PLOT column would lead to a larger \( S \) value hence an increase in \( PC \).\textsuperscript{48} Several approaches have proved useful to increase the hydrophobicity of PS-DVB monolithic columns such as surface alkylation, co-polymerization with alkylated monomers or incorporation of high hydrophobic carbon nanotubes\textsuperscript{24,49,50}. Improving the hydrophobicity of the PLOT column is challenging because it is difficult to maintain the open tubular structure when the content of polymerization mixture is changed.

2.3.6.1 Carbon Nanotubes (CNTs) Modified PLOT Columns

One strategy to increase the hydrophobicity of PLOT column is to incorporate hydrophobic particles—carbon nanotubes (CNTs) into our PLOT column. CNTs have previously been incorporated into the monolithic and open tubular columns to improve the hydrophobicity\textsuperscript{24,51,52}. In our study, well-cut carbon nanotubes were suspended in the porogen of a mixture of ethanol, styrene and divinylbenzene for 24 h. 0.25\% and 0.5\% (wt\%). CNTs were added into the standard polymerization mixture separately and further co-polymerized in the 10 \( \mu \)m i.d. capillary. The oxidative cutting of the CNTs was performed in the solution of strong sulfuric acid and nitric acid by sonication, resulting in an average size of 100 nm for CNTs\textsuperscript{24}. The cut CNTs were
observed to form a homogenous suspension in the porogen solution, after sonication for 10 min. The unmodified and the two CNTs modified PLOT columns were compared by using a 30 min gradient at a constant linear velocity. In the separations of six bovine protein digest (Fig. 2.11), we observed that in 0.25% CNTs PLOT column, the retention times of peptide 1 and 2 were not changed. The peak shape of peptide 3 was much sharper with a $W_{0.5}$ of 0.076 min compared to 0.55 min in the unmodified PLOT column and peptides 8-10 had a retention shift about 1.5 min. The improvement of peak shape of peptide 3 and retention of peptides 8-10 demonstrated that 0.25% wt CNTs could slightly improve the hydrophobicity. In the chromatogram of 0.5% CNTs PLOT, unretained peptides 1 and 2 showed broad peaks in former chromatogram and co-eluted with peptide 3. The retention times of peptides 3-10 were delayed about 2 min. The average peak width of peptides across the separation window was improved from 0.17 min to 0.91 min in 0.5% CNTs PLOT column and remained similar to the average peak width of the sharp peaks 4-10 in unmodified PLOT (Table 2.2).
Figure 2.11. Incorporation of carbon nanotubes into PLOT column. 500 amol 6 bovine protein digest was split-injected on each 3.7 m PLOT column. The gradient was 0-27% B (0.1% FA in ACN) in 30 min. Flow rate from loading pump was 40 μL/min and split by 1:2000 ratio.
2.3.6.2 Alkyl Monomer Modified PLOT Columns

Incorporating CNTs into PS-DVB stationary phase of the PLOT column increased the hydrophobicity. However, cutting the CNTs into homogenous length around 100 nm was difficult. Co-polymerization of an alkyl monomer or grafting an alkyl group on the surface of the PS-DVB by the Friedel-Crafts reaction has been shown to lead to improved resolution for HPLC separations\textsuperscript{49,50}. Modifying PLOT columns with the more hydrophobic alkyl groups is another potential approach to improve the hydrophobicity. For this purpose, co-polymerizing with octadecene (OD) and grafting 1-chlorooctadecane onto the surface of PS-DVB were attempted; however, neither of the approaches worked. The co-polymerization with the introduced monomer disturbed the original mixture and it was found hard to maintain the open tubular structure in the 10 µm i.d. capillary. For surface grafting, the solution of 1-chlorooctadecane and aluminum chloride produced a high backpressure when filling the PLOT column.

Based on the prepared PS-DVB PLOT column, two modifications were made on the surface of the existing polymer layer. One modification was to graft OD onto the PS-DVB layer by a free radical reaction between the double bonds in OD and PS-DVB. Another modification was to make a very thin layer of OD-DVB cross-linked polymer onto the existing layer by using a 30:70 ratio of monomers: decanol. For OD grafted column, the peak shape of peptide 3 improved and retention of peptides 7-10 shifted to the right about 2 min (Fig. 2.12, middle). For the 2\textsuperscript{nd} thin layer OD-DVB, peptides 1,2 were co-eluted with much shaper peaks (Fig. 2.12, bottom). The 2\textsuperscript{nd} layer OD-DVB PLOT column also exhibited a better trapping ability for the peptide mixture. After split-loading a peptide mixture on the 2\textsuperscript{nd} thin layer OD-DVB column for 1 h, the
hydrophilic peptides still appeared with good peak shapes with gradient separation (data not shown).

Figure 2.12. PS-DVB surface modified PLOT columns with alkyl-alkene. 500 amol 6 bovine protein digest was split-injected on each 3.7 m PLOT column. The gradient was 0-27%B (0.1% FA in ACN) in 30 min. Flow rate from loading pump was 40 μL/min and split by 1:2000 ratio.
2.3.7 Narrower Bore PLOT Columns

As the Golay Equation (1.9 in Chapter 1) implies, the $C_m$ term in HETP is proportional to the square of the i.d. of PLOT column. The preparation of the porous layer in a smaller i.d. capillary will improve the column efficiency. Another advantage of smaller i.d. PLOT column is that the column can be operated at a higher linear velocity as well as maintained the efficiency. This benefit will facilitate the fast separation.

2.3.7.1 Double PS-DVB Layer PLOT Columns

To reduce the i.d. of the open tube, we first tried to increase the thickness of polymer layer on inner wall of a 10 μm i.d. fused silica capillary by polymerizing another layer of PS-DVB based on a prepared PS-DVB PLOT column. In the scanning electron microscopy (SEM) imaging of the cross-section of the double layer PLOT column, the thickness of the polymer layer was increased from 1.2 μm to 2.1 μm and a boundary between the 1st layer and 2nd layer could be clearly observed (Fig. 2.13). As a result of decreased open tubular i.d., a 3.7 m double layer PLOT column generated 2-fold higher backpressure for a given flow rate in comparison to a single layer PLOT column with the same length.

For separation of 500 amole of digested peptides from a 6 bovine protein mix by the double layer PLOT column, the linear velocity was maintained the same to keep the $k^*$ constant. Ten selected peptides (sequences listed in Table 2.2) spanning the separation window were extracted and labeled with 1-10 in both chromatograms of the single and double layer PLOT columns (Fig. 2.13). As shown in Fig. 2.13, the retention times of peptides 1-10 were shifted to the right about 7 mins. The thicker polymer layer can cause the shift of the retention due to the phase ratio.
change; however, the peak widths of peptides 1 and 2 were still broad indicating the hydrophobicity of double layer PLOT column was not increased. But the double layer PLOT column still provides us an option to prepare an open tubular column with smaller open tube i.d. of 7-8 µm and higher capacity due to the thicker layer.

Figure 2.13. 2nd layer PS-DVB PLOT column. 500 amol 6 bovine protein digest was split-injected on each 3.7 m PLOT column. The gradient was 0-27%B (0.1% FA in ACN) in 30 min. Flow rate from loading pump was 40 µL/min and split by 1:2000 ratio.
2.3.7.2 Preparation of PS-DVB PLOT Column in 5 μm i.d. Capillary

The 2nd PS-DVB layer decreased the open i.d., however, the double layer PLOT did not exhibit higher column efficiency as expected. The reason is the $C_s$ term for resistance of diffusion in stationary phase cannot be ignored, as $C_s$ is proportional to the square of the thickness of the stationary phase on the wall (Equation 1.9 in Chapter 1). With the thickness increased from 1 μm to 2 μm, the $C_s$ increased 4-fold relevantly. Polymerization of a single PS-DVB layer in a narrower bore capillary is preferred.

Initially, the recipe for preparation of 10 μm i.d. PLOT was adapted in the preparation of 50 cm long PLOT column and produced a thickness of 0.8 μm layer on the inner wall of 5 μm i.d. capillary; however, large globule was also observed (Fig. 2.14). This large globule could cause the clogging of the PLOT column. Decreased layer thickness and size of the globules were achieved by increasing the ratio of porogen in the polymerization mixture. A mixture of 15% ST, 15% DVB and 70% ethanol could generated a 0.5 μm layer and 0.3 μm thickness was produced by using a mixture of monomers to porogen ratio of 20% (Fig. 2.14), both with smaller globules. The 30% monomer recipe is a potential candidate for preparation of 10 μm i.d. PLOT. According to the Golay Equation, both the $C_m$ term and $C_s$ term will improve by a factor of 4, due to the 2-fold decrease in the open diameter and thickness of the stationary phase. Therefore the $H_{\text{min}} \propto \sqrt{BC}$ could improve 2-fold$^{53}$.

The HPLC behavior of the 5 μm i.d. PLOT still needs to be evaluated in a gradient separation by preparing a long column ~4 m in the further. Operating this narrow bore column requires an ultra-high pressure LC pump to handle the high column backpressure or elevated temperature to reduce the backpressure.
Figure 2.14. Changes in the thickness of the porous PS-DVB layer in a 5 μm i.d. capillary by varying the ratio of monomers to porogen.

Table 2.2. Summary of the peptide separation by modified PLOT columns

<table>
<thead>
<tr>
<th>label</th>
<th>peptide sequence</th>
<th>m/z</th>
<th>W ½ (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>PS-DVB</td>
</tr>
<tr>
<td>1</td>
<td>HIQKEDVPSER</td>
<td>446.5648</td>
<td>0.219</td>
</tr>
<tr>
<td>2</td>
<td>VGDANPALQK</td>
<td>507.2603</td>
<td>0.322</td>
</tr>
<tr>
<td>3</td>
<td>NYTDNELEK</td>
<td>563.2293</td>
<td>0.55</td>
</tr>
<tr>
<td>4</td>
<td>TPEVDDEALEK</td>
<td>623.2944</td>
<td>0.085</td>
</tr>
<tr>
<td>5</td>
<td>LVEDLKR</td>
<td>487.2818</td>
<td>0.115</td>
</tr>
<tr>
<td>6</td>
<td>TPEVDDEALEKFDK</td>
<td>545.9304</td>
<td>0.061</td>
</tr>
<tr>
<td>7</td>
<td>SLHTLFDELCK</td>
<td>474.2310</td>
<td>0.086</td>
</tr>
<tr>
<td>8</td>
<td>LVEDLKTRETEEQK</td>
<td>858.9003</td>
<td>0.091</td>
</tr>
<tr>
<td>9</td>
<td>DHGMPGYN5WRGFCGLSQPKTLK</td>
<td>879.6690</td>
<td>0.075</td>
</tr>
<tr>
<td>10</td>
<td>FFVAPFPEVFGK</td>
<td>692.8682</td>
<td>0.148</td>
</tr>
</tbody>
</table>

Average peak width at half height (1-10) 0.175 0.171 0.091 0.162 0.142 0.093
Average peak width at half height (4-10) 0.094 0.137 0.109 0.117 0.130 0.118
2.3.7 Alkyl Monomers Modified Monolithic Columns

From the results of the modified PLOT columns, the hydrophobicity of the PS-DVB polymer was slightly improved by incorporation of CNTs or co-polymerization with alkyl monomer. Hydrophobicity improved in the 50 µm i.d. SPE monolithic column for capturing more hydrophilic peptides. Co-polymerization with alkyl monomer in the preparation of the monolithic column is much easier to control because two porogens, macroporogen and microporogen (decanol and THF, respectively, in our study), are used in the polymerization mixture system. The higher macroporogen to microporogen ratio can result in a larger macropore size and smaller mesopore size, while a lower ratio would generate a polymer with smaller macropore size and larger mesopore size\textsuperscript{54}. Polymerization of styrene (ST) and DVB with octadecene has been prepared in the 320 µm i.d. capillary with porogens of DMF and decanol\textsuperscript{50}.

In our study, polymerization of ST and DVB with the more hydrophobic alkyl monomers, decene and octadecene, was prepared in 50 µm i.d. based on the porogens of decanol and THF. We noticed that replacing 15% wt of ST with alkyl monomers in the mixture resulted in very low column backpressure and discontinuous monolithic bed, indicating a monolith with a large macropore size formed in an early stage of phase separation. Therefore, lower decanol/THF ratio was optimized for either decene or octadecene modified monolithic column to obtain a rigid monolith as the PS-DVB monolithic column. Finally, porogen compositions of decanol/THF ratio of 480/20 (v/v) and 450/150 were optimized for decene (C\textsubscript{10}) and octadecene (C\textsubscript{18}) modified columns, respectively. The evaluation of the PS-DVB, PS-C\textsubscript{10}-DVB and PS-C\textsubscript{18}-DVB monolithic columns was carried out by direct-loading 50 fmol, 6 bovine protein digest at 100 nL/min using 20 cm column length, with similar backpressure of 100 bar (1,500 psi). The
chromatograms of the 6 bovine protein digest separation by the three columns shown in Fig. 2.15A demonstrates the increase in hydrophobicity resulting from the co-polymerization of the alkyl monomers. In the separation by PS-C$_10$-DVB SPE column, a 10% more peptides (based on the retention window were trapped and detected after the PLOT separation compared to the unmodified PS-DVB column and another 5% more peptides were detected in the separation by the octadecene modified SPE column. Inspection of extracted ion chromatograms of selected peptides among the three separations revealed a 2-3 min retention shift in decene and octadecene modified columns (Fig. 2.14B); however, retention times of peptides 2-6 showed only slight changes between the two modified columns.

During the optimization in the preparation of alkyl monomer modified SPE columns, monolithic structures with higher or lower pore size were also generated. For the PS-C18-DVB monolithic column, pressure of 20 cm long column prepared in decanol/THF ratio of 455/145 was 65 bar (942 psi) and was 175 bar (2538 psi) in the porogens with 435/165 ratio. The two columns presented similar hydrophobicity in trapping the peptides, but sharper peaks were observed in the separation by the monolithic column with higher pressure (Fig. 2.16). The lower decanol/THF ratio represented less macroporogen and more microporogen; hence, a monolithic structure with smaller macropore size and mesopore size was produced. The smaller size resulted in a higher resistance of flow increased the mass transfer which led to the higher column efficiency in the separation.
Figure 2.15. PS-DVB monolithic column modified with alkyl-alkene. (A. Base peak chromatograph of 6 bovine mix digest, B. extracted ion chromatograph of selected peptides). 50
110 fmol 6 bovine protein digest was direct-injected on each 20 cm monolithic column. The gradient was 0-27%B (0.1% FA in ACN) in 30 min. Flow rate from micropump was 100 nL/min.

Figure 2.16. Effect of pore size of monolithic polymer on the column efficiency. 50 fmol 6 bovine protein digest was direct-injected on each 20 cm monolithic column. The gradient was 0-27%B (0.1% FA in ACN) in 30 min. Flow rate from micropump was 100 nL/min.
2.4 Conclusion

In this research, we first demonstrated the influence of flow rate on the nanoESI signal intensity. The absolute gain in sensitivity of the nanoESI signal at ultralow flow $< 20 \text{ nL/min}$ was shown under gradient elution separation. This experiment indicated the necessity of using ultralow flow $< 20 \text{ nL/min}$ in the analysis of limited amounts of a sample with biological and clinical importance. For such small sample sizes, our group developed the 10 $\mu\text{m}$ i.d. PLOT column suitable for operating at ultralow flow rate $< 20 \text{ nL/min}$. In the $H-u$ relationship, we found an efficiency of 90,000 plates/m for the PLOT column operating at 20 nL/min flow rate and further more plates about 260,000 plates/m were achieved at 5 nL/min. Using the established van Deemter plot, we could estimate the peak capacity of a gradient separation by Snyder’s gradient elution theory. Upon extrapolation, operating the PLOT column at 5 – 10 nL/min with a length of 8 – 10 m, we estimate a $PC$ of 600-700 in a complex mixture separation under moderate backpressure $< 4000 \text{ psi}$. Developing a PLOT column with narrow i.d. or operating the column at elevated temperature will further improve the efficiency.

In the application of PLOT column for the analysis of limited amounts of sample, a monolithic SPE column was introduced in front of PLOT to obtain improved loading capacity. As a result, proteomic analysis of a lysate of 500 MCF-7 cells identified 3679 proteins in a single 4 h long gradient separation. The number of identified protein groups by the PLOT column was 3-fold higher than that found in the analysis by packed 75 $\mu\text{m} \ C_{18}$ column at conventional 200 nL/min flow rate. The 2D analysis of the same amount of cell lysate only presented a 13% increase in identified proteins. The gain in 2D analysis of limited samples was not significant because the
potential sample loss of material on the first dimension. The result illustrates that at sensitivity available for now, 1D analysis is preferred and much shorter analysis time is achieved.

To obtain higher hydrophobicity for polymer based columns, we chemically modified the PLOT and monolithic columns. The comparison to non-modified PLOT columns showed that the 0.5% CNTs PLOT and 2\textsuperscript{nd} thin layer OD-DVB PLOT columns had better hydrophobicity for the hydrophilic peptides and demonstrated good resolution for overall separation. Increased hydrophobicity was also observed in decene and octadecene modified columns. Decene modified monolithic column is a good choice of trapping column when more hydrophobic PLOT column is available due to the higher loading capacity and improved efficiency.

The attempts to prepare narrower bore PLOT columns were carried out in preparing a double PS-DVB layer PLOT in 10 μm i.d. capillary and single layer PLOT in 5 μm i.d. capillary. The double layer PLOT had a 2 μm thickness layer and 6 μm open diameter. The double layer PLOT had higher capacity due to the thicker layer; however, the efficiency was not improved as expected for the possible reason of slower mass transfer in the thicker layer. The protocol for preparation of 5 μm i.d. PLOT column was optimized, and 30:70 of monomers to porogen was chosen. The performance of 5 μm i.d. PLOT column is still unknown and need to be evaluated in the future study.
2.5 References


Chapter 3

Ultrasensitive Microproteomic Profiling of Rare-cell Populations Using Microfluidic Magnethophoretic Isolation from Whole Blood Followed by Porous Layer Open Tubular Liquid Chromatography - Tandem Mass Spectrometer

* A manuscript based on the results presented in this chapter is currently in preparation.

Co-authors’ contribution to this work: Siyang Li (cell lysis, LC/MS, data analysis and interpretation, manuscript preparation), Dr. Brian D. Plouffe (cell capture, manuscript preparation), Somak Ray (data analysis for label-free quantitation); Arseniy Belov (SDS gel electrophoresis and western blot analysis); Dr. Shashi K. Murthy (experimental design, manuscript preparation), Dr. Alexander R. Ivanov (conceptual and experimental design, experimental assistance, data analysis, manuscript preparation, project oversight), Dr. Barry L. Karger (conceptual design, manuscript preparation, financial support).
Abstract

Clinical studies of a specific disease are often hampered by sample heterogeneity and limited availability of disease-relevant subpopulations of cells. Because of the lack of appropriate sample processing and analytical techniques, deep proteomic profiling of such limited samples by mass spectrometry (MS)-based proteomics is challenging. In this study, the 10 µm i.d. PS-DVB porous layer open tubular (PLOT) column coupled with high resolution/accurate mass (HR/AM) Q Exactive mass spectrometer exhibited ultralow sensitivity, as low as 5 zmol and ultra-performance in deep proteomic analysis of 500 MCF-7 cells, at the flow rate below 20 nL/min. A novel sample preparation approach, based on the Adaptive Focused Acoustics™ (AFA) technology, introduced by Covaris (Woburn, MA), was optimized with minimal sample dilution, manipulation and transfer. The methodology, including ultrasonication is shown to produce high recovery in the preparation of limited starting materials as low as 2,000 cells. To achieve clinical utility, a microfluidic device using magnetophoretic cell isolation was used for specific and quantitative capture of 1,000 to 10,000 MCF-7 cells from whole blood and integrated with downstream PLOT-nLC-MS. To demonstrate the applicability of the developed high sensitivity platform to “real-world” clinical and biological applications, endothelial progenitor cells (EPCs) and hematopoietic stem cells (HSCs) were isolated from unprocessed human blood drawn from healthy non-smoking and smoking donors, and the resulting proteomes were quantitatively characterized. The differences of proteomic profiles between the two groups of donors were assessed.
3.1 Introduction

Analysis of limited biological and clinical samples, e.g. circulating tumor cells (CTCs) in peripheral blood, stem cells (SC), tissue microbiopsies, has been increasingly recognized as important during the recent years. High sensitivity deep genomic and proteomic analysis of such scarce samples can facilitate the diagnosis of various diseases at early stages. However, relevant specimens are often available in limited amounts, e.g. CTCs are found in the order of 1-10 cells per mL of human blood. Genomic and transcriptomic approaches have become capable of profiling of few hundred cells or even a single cell, all enabled by amplification techniques. Mass spectrometry (MS)-based proteomics provides unique information about gene expression and post-translational modification levels. However, MS-based proteomics is not yet sensitive enough for deep analysis of limited clinical samples at the level of hundreds or few thousand cells due to the current limitations in the technology. Proteomic analysis of the limited clinical sample is challenging and hence requires specific and sensitive approaches to isolate, prepare and analyze samples as low as hundreds of rare cells.

Several protocols such as, acetone precipitation, pseudoshotgun (PSG), filter-aided sample preparation (FASP) and monolithic column based proteomic reactor have been applied in the preparation of as low as 500-1000 cultured cells, resulting in identifications of a few hundred proteins. These proteomic results were not capable of providing the profiling depth similar to the analysis of millions cultured cells which can result in over 10,000 protein identifications. A major issue in sample preparation of limited samples is the significant losses during the transfer or clean-up steps. Hence, the sample preparation requires a more specific and concise processing to minimize the sample losses.
Microfluidic devices provide the potential for the isolation of subpopulation of cells from heterogeneous suspensions such as whole blood and other unprocessed physiological fluids. Fluorescence-activated cell sorting (FACS) and magnet-activated cell sorting (MACS) are the two main approaches on sorting specific cells by attaching corresponding antibodies attached to fluorescent dyes or magnetic beads against specific antigens on the cell surface\textsuperscript{17,18}. The immunomagnetic separation method developed on a microfluidic device has been shown to detect and isolate as low as 2 CTCs in blood\textsuperscript{19}. To isolate a certain cell population from heterogeneous solid tissue, laser capture microdissection (LCM) is the dominant method using a laser to select and excise specific cell populations under direct microscopic visualization\textsuperscript{20}.

Other than isolation and preparation of rare cell samples, liquid chromatography-tandem mass spectrometry (LC-MS/MS) based proteomic analysis is also inefficient in detecting limited amount samples with a wide dynamic range. Compared to the conventional nanoflow technique, the use of nanoelectrospray ionization at ultralow flow rates, e.g. <20 nL/min, results in the increased MS sensitivity due to the increased efficient ionization efficiency, reduced ion suppression and improved ion transmission\textsuperscript{21}. By reducing the column diameter and flow rate, a significant gain in MS response was achieved using narrow-bore packed column or monolithic column (<20 \(\mu\)m i.d) at 20 nL/min flow rate in comparison to a 75 \(\mu\)m i.d. column at a conventional flow > 200 nL/min\textsuperscript{1,2}. To overcome the obstacle of the high backpressure on the packed columns, our group developed a polymer-based 10 \(\mu\)m i.d. porous layer open tubular (PLOT) column with high resolving power, good column-to-column reproducibility and high
permeability. Coupling PLOT column with high resolution mass spectrometer is a powerful tool for detection ultratrace level proteins in limited clinical samples.

In this work, we describe an integrated approach to perform the microproteomic analysis of rare cells from whole blood. The targeted cancer or stem cells in whole blood were captured using antibodies attached to the micromagnetic beads and isolated using a microfluidic magnetophoretic technique. Captured rare cells were further processed with Adaptive Focused Acoustics™ (AFA) ultrasonication and followed by enzymatic digestion in the same tube without sample clean-up and transfer. The resulting digested cell lysates were analyzed using the ultrahigh performance PLOT column coupled with high resolution/accurate mass (HR/AM) Q Exactive mass spectrometry. The integrated approach was demonstrated to be effective, reproducible and highly sensitive in analysis of as low as few hundreds cell resulting in deep coverage of the proteomes under study.
3.2 Methods

3.2.1 Cell Handling

MCF-7 human breast adenocarcinoma cells (ATCC, Manassas, VA) were cultured by Dr. Murthy’s lab in 75 cm$^2$ tissue culture flasks at 37 °C, 5% CO$_2$. MCF-7 cells were incubated in Eagle’s Minimum Essential Medium (EMEM; ATCC) supplemented with 10% fetal bovine serum, 100 U mL$^{-1}$ penicillin, 100 μg mL$^{-1}$ streptomycin, and 0.01 mg mL$^{-1}$ bovine insulin. Cells were grown to preconfluence and isolated for experiments by trypsinization using a 0.25% Trypsin-EDTA solution.

Prior to cell isolation or cell lysis, cells were centrifuged at 200 g for 10 min at 4 °C. The cell culture medium was removed and MCF-7 cells were washed twice with ice cold 1× Dulbecco’s Phosphate-buffered saline (DPBS, Sigma, St. Louis, MO containing 0.5% EDTA (v/v). The cells were resuspended at a concentration of approximately 1,000 cells/μL and the cell numbers were counted three times using hemacytometer and (or) flow cytometer (Cell Lab Quanta SC; Beckman Coulter, Brea, CA).

3.2.2 Microfluidic Device Design and Fabrication.

To validate the developed optimized device design, microfluidic channels were fabricated by Dr. Murthy’s lab as previously described.$^{23,24}$ Wire arrays were designed using PCB123® printed-circuit board design software and ordered from Sunstone Circuits (Mulino, OR). The wire dimensions were set to provide a gap encompassing the width of the device microfluidic channel; the height and width of the all of the wires were set to 35 μm and 178 μm, respectively. Teflon-insulated 18G copper wires were soldered to the ends of each of the printed circuit board arrays and the arrays were connected to a DC power supply (Elenco Electronics XP-4, Wheeling
IL) that provided three fixed-current settings of 0.25 A, 0.50 A and 1.00 A via standard alligator clip connectors. The PDMS channels and wire arrays were visually aligned.

3.2.3 Microparticle Modification.

DynaBeads® MyOne™ Carboxylic Acid particles (Life Technologies, Carlsbad, CA) were modified by Dr. Murthy’s lab with antibodies, either antibodies against the epithelial cell adhesion molecule (mouse monoclonal anti-human EpCAM; Santa Cruz Biotechnology, Santa Cruz, CA) or antibodies against CD133 (mouse anti-human CD133, Miltenyi Biotec Inc, Auburn, CA) using standard carbodiimide chemistry in ratios suggested by the reagent manufacturer (1:1 molar ratio of beads to protein; Pierce Biotechnology, Rockford, IL).

3.2.4 Microfluidic Capture of Spiked MCF-7 cells

In Dr. Murthy’s lab, several sample levels of MCF-7 cells (1000, 2000, 3000, 5000 and 10,000 cells) were spiked into 1 mL of whole human blood and mixed with the Dynal Myone EpCAM-functionalized magnetic microbeads. As negative controls, human blood without spiked-in MCF-7 cells was mixed with the EpCAM-functionalized microbeads. Cells functionalized with magnetic microbeads were isolated from human blood cells using a microfluidic magnetophoretic device as described in26. A flow cytometer was used to count MCF-7 cells. The cells were washed twice with 50 µL of DPBS and processed immediately as described below.
3.2.5 Whole Blood Cancer Cell Isolation.

In Dr. Murthy’s lab, whole blood was drawn from healthy volunteers and collected in EDTA-coated Vacutainer® tubes (Becton Dickinson, Franklin Lakes, USA). Approval from the Northeastern University Institutional Review Board was obtained for this purpose.

Prior to experiments, the location of the interface that forms between the injected blood and buffer was evaluated. As blood is a non-Newtonian, shear thinning fluid, it behaves differently from cells in buffer solutions and thus the required displacement for effective isolation is changed. The results of this evaluation influence the rational design optimization described previously.²⁶,²⁷

A Coulter counter/flow cytometer (Cell Lab Quanta™ SC; Beckman Coulter, Brea, CA) was used to count the number of target (MCF-7) cells versus native polymorphonuclear cell that were separated. A protocol, based on the distinct size difference of these two cells, was created to identify each cell population. The cells were gated by their electronic volume and granularity, and the total number of cells within the recovered suspension was assessed.

Various concentrations of MCF-7 cells (500-100,000 cells) were spiked into whole blood. Following the inoculation of the MCF-7 cells, 50 µL of modified anti-EpCAM magnetic microbeads was added to 5 mL of unprocessed blood samples and allowed to mix and incubate for 30 min. on a rotary mixer. This experiment was conducted at computationally and experimentally optimized flow rates as described in the theory section below. For all MCF-7 experiments the flow rate of the samples was fixed at 240 µL/min and a center stream of 1× RBC lysis buffer was flowed at 160 µL/mL. Target and non-target cells were collected in separate MeOH cleaned micro-centrifuge tubes.
3.2.6 Isolation of Hematopoietic Stem Cells and Endothelial Progenitor Cells from Whole Blood.

To illustrate the utility of the magnetophoretic rational design in cardiovascular disease, Dr. Murthy’s lab extracted hematopoietic stem cell (HSCs) and endothelial progenitor cells (EPCs) from whole blood using anti-CD133 functionalized microparticles. Again, whole blood was drawn from healthy volunteers and collected in EDTA-coated Vacutainer® tubes (Becton Dickinson, Franklin Lakes, USA).

Isolated cells were then labeled with antibodies to identify HSC and EPC populations. The HSCs were identified as labeling positive for mouse anti-human CD34 conjugated to fluorescein isothiocyanate (anti-CD34-FITC; Santa Cruz) and mouse anti-human CD45 conjugated to phycoerythrin (anti-CD45-PE; Santa Cruz), and negative for goat anti-human KDR (kinase insert domain receptor; Santa Cruz). The KDR was then conjugated to a secondary antibody donkey anti-goat peridinin chlorophyll protein (PerCP; R&D Systems, Minneapolis, MN). EPCs were identified as labeling positive for anti-CD34-FITC and anti-KDR-PerCP, and negative for anti-CD45-PE. Both cell populations were distinguished via a flow cytometer.

3.2.7 Sample Processing Prior to Lysis and LC/MS

After separation, the ~650uL samples (in RBC lysis buffer) was concentrated down using a NdFeB permanent magnet in Dr. Murthy’s lab. The samples were rinsed twice with 500uL of PBS and transferred to a lysis mini-tube. The cell losses due to concentrating, rinsing, and transfer were monitored to assess starting and ending cell numbers. See Table 3.1 for these results.
Table 3.1: Cells Losses During the Isolation and Purification Process

<table>
<thead>
<tr>
<th>Process</th>
<th>Cell Number [total]*</th>
<th>Percentage Loss [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Inoculation</td>
<td>5000 +/- 42</td>
<td>N/A</td>
</tr>
<tr>
<td>Microfluidic Separation</td>
<td>4854 +/- 122</td>
<td>97.1 +/- 2.5</td>
</tr>
<tr>
<td>Concentration and Rinse #1</td>
<td>3987 +/- 98</td>
<td>79.7 +/- 2.0</td>
</tr>
<tr>
<td>Concentration and Rinse #2</td>
<td>3454 +/- 101</td>
<td>69.1 +/- 2.0</td>
</tr>
<tr>
<td>Transfer to miniTube</td>
<td>2453 +/- 54</td>
<td>49.1 +/- 1.1</td>
</tr>
</tbody>
</table>

* All cells were counted via flow cytometry

3.2.8 Cell Lysis and Protein Digestion

In-Solution Sample Processing (conventional enzymatic digestion in large volume): For sample preparation using large amounts of cells, $10^7$ cells were rinsed twice with 1 mL volume of DPBS, spun down at 200 g for 10 min, and the cell pellet was resuspended in 100 µL of lysis buffer (8 M urea, 2 M thiourea, 5 mM TCEP in 25 mM ammonium bicarbonate (ABC), pH 8.4). To process limited cell amounts, 2 µL of a DPBS suspension containing approximately 2,000 rinsed cells was added into 8 µL of lysis buffer (10 M urea, 2.5 M thiourea, 6 mM TCEP in 30 mM ammonium bicarbonate, pH 8.4). Cells were sonicated on ice in a water bath for 15 min. Extracted proteins were reduced in 5 mM TCEP for 30 min at room temperature (~ 27 °C) and alkylated in 20 mM iodoacetamide (IAA) for 30 min in the darkness. Alkylation was quenched
with an addition of fresh TCEP to 5 mM with shaking for 5 min. Prior to digestion, the resulting lysate was diluted 10-fold with 25 mM ABC, pH 8.4 to bring the concentrations of urea and thiourea to 0.8 M and 0.2 M, respectively. Protein digestion was performed with endoproteinase Lys-C (Sequencing grade, Promega, Madison, WI) for 4 h at an enzyme/substrate (E:S) ratio of approximately 1:50, followed by addition of sequencing grade trypsin (Promega) at an E:S ratio of approximately 1:50 and overnight digestion in a shaker at 37 °C. The total protein abundance was estimated from literature that 1,000 cells contains ~200 ng total protein\textsuperscript{28}. The total volume of the resulting digests were 1000 µL and 100 µL for large (10\textsuperscript{7} cells) and small (2,000 cells) samples, respectively. Digested lysates were desalted using self-packed C\textsubscript{18} microSPE tips packed of five layers of porous membrane with embedded C\textsubscript{18} beads (3M, St. Paul, Minnesota), as described previously\textsuperscript{29,30}. Cell lysates were reconstituted in water/0.1% FA.

Micro SPE Tip Sample Processing: 2,000 MCF-7 cells were lysed in-solution, as described above, and the resulting lysates were loaded onto a C18 micro SPE tips. 1 µL of 50 mM TCEP in 25 mM ABC, pH 8.4 was loaded on the tip, and the reagent was partially pushed through to wet the tip layers of the packing. The micro SPE tips were kept exposed to TCEP for 30 min at room temperature. Then, the tips were rinsed with 10 µL of 25 mM ABC, pH 8.4 and exposed to 1 µL of 20 mM IAA alkylation reagent for 30 min at room temperature in the dark. The remaining chaotropic, reducing and alkylating reagents were removed by washing the micro SPE tip twice with 10 µL of 25 mM ABC, pH 8.4. Then, a solution of Lys-C in 25 mM ABC, pH 8.4 was loaded on the micro SPE column to an E:S ratio of approximately 1:10 and left in a shaker for 4 hr at 37 °C, followed by a displacement with a plug of trypsin in 25 mM ABC, pH 8.4 to an E:S ratio of 1:10 and shaking overnight at 37 °C. The digested peptides were eluted with 10 µL of 65%
acetonitrile (ACN) in 0.1% formic acid (FA), evaporated to dryness using a SpeedVac and reconstituted in 10 µL of water and 1% FA.

**AFA Ultrasonication:** Limited numbers (low hundreds to low thousands) of pure cultured MCF-7 cells or cells isolated magnetophoretically using the microfluidic platform were rinsed twice in PBS and resuspended in 15 µL of 5 mM TCEP, 25 mM ABC pH 8.4 immediately in Covaris microTUBE (SKU:520045, Covaris Inc., Woburn, MA). Cells were lysed in an Adaptive Focused Acoustics device (S220x, Covaris Inc., Woburn, MA) using ultrasonication (35 Watt, 10% duty cycle, 200 cycle/burst, 4°C) for 3 min. Reduction and alkylation were performed in the same sample tube by adding 1 µL of 80 mM TCEP to reach a concentration of 5 mM with incubation for 30 min in a shaker at room temperature. Alkylation was performed in 20 mM IAA for 30 min in the same sample tube, as described above. Alkylation was quenched with addition of 1 µL of 90 mM TCEP and incubation in a shaker for 5 min. A 1 µL aliquot of Lys-C was added to the lysate in an E:S ratio of approximately 1:50, and the sample was incubated in a shaker for 4 h at 37 °C. Trypsin was brought to the E:S ratio of approximately 1:50 and left for overnight digestion in a shaker. The total volume of the digest was approximately 18-20 µL (depending on the sample). Samples were concentrated to 10 µL using a SpeedVac and directly injected for LC-MS analysis. For captured cell samples, 5 fmol of a stable isotopically labeled (SIL) peptide standard (JPT, Acton, MA), containing 252 cancer related peptides with Lys and Arg labeled at the C-terminal, was added into the cell lysate before injection.
3.2.9 Preparation of PLOT and Monolithic Columns

Preparation of single layer PS-DVB PLOT column and 1-decene modified monolithic column was described in Chapter 2. Briefly, a 10 µm i.d. fused-silica capillary (~ 10 meters) was flushed with 1.0 M NaOH at 1000 psi for 2 hours using a pressure bomb. Then, the column was washed with water for 4 h and later with ACN for 4 h. The capillary was dried with nitrogen to remove residue water and acetonitrile for 8 h. 30% (v/v) TMSPMA and 0.5% (wt/v, g/mL) in DMF was freshly prepared. The solution was passed through the 10 µm i.d. pretreated capillary at 1000 psi for 2 h. Both ends of the capillary, filled with the solution, were sealed by inserting both ends in the vial, and the capillary was placed in an oven at 110°C. After 6 h, the silanized capillary was washed with ACN and blown dry with nitrogen. A polymerization solution containing 5 mg of AIBN, 0.1808 g styrene, 0.1823 g divinylbenzene and 0.4523 g ethanol was filled into the silanized capillary at 1000 psi for 2 h. Both ends of the capillary were immersed in the polymerization solution, and the capillary was heated at 74°C for 16 h in a water bath with forced water circulation. The PLOT column was washed first with acetonitrile and then water overnight to remove residual polymeric mixture.

A 50 µm i.d. fused-silica capillary (30 cm) was silanized using a similar protocol as provided above for the PLOT column, except using 30 min for each wash and 2 min for filling at 200 psi. A polymerization solution was prepared consisting of 5 mg of AIBN, 0.4530 g styrene, 0.1365 g 1-decene, 0.1844 g divinylbenzene, 0.0363 g THF and 0.4507 g 1-decanol. The solution was degassed by ultrasonication for 5 min and then filled into the silanized capillary. Both ends of the capillary were immersed in the polymerization solution, and the vial was pressurized through the sealed cap by a syringe. The filled capillary was heated at 70°C for ~16 h in a water bath. The monolithic column was ready to use after several washes with acetonitrile and water.
3.2.10 PLOT Nanoflow Liquid Chromatography-Tandem Mass Spectrometry (PLOT nanoLC-MS/MS)

For LC separation, 5 cm of 50 μm i.d. 1-decene modified PS-DVB monolithic SPE precolumn was connected to a 4.2 m PLOT by a PicoClear Tee (New Objective, Woburn MA). Digested lysates were first loaded on the monolithic SPE precolumn at a flow rate of 200 nL/min using a NCS 3500 RS pump (Dionex, Sunnyvale, CA). Then, entrapped and cleaned up digests were eluted off the precolumn and separated on the PLOT column using a linear solvent gradient at a 20 nL/min flow rate split from 400 nL/min using the NCS 3500 RS pump (Dionex). The separations were performed using 30-240 min gradient of 0%-27% mobile phase B (mobile phase A: 0.1% FA in water; mobile phase B, 0.1% FA in ACN). The SPE and PLOT columns were washed with 90% B for 20 min and re-equilibrated with mobile phase A for another 20 minutes. Nano ESI spray was conducted using an electrospray voltage of 1.1 kV and a distal coated tip (FS360-20-5-D-20, New Objective) butt-to-butt connected with an outlet of the PLOT column via a zero dead volume PicoClear union (New Objective). The ion transfer tube temperature was set for 275 °C.

MS detection was performed using a top 12 MS/MS data-dependent scans on the Q Exactive (Thermo Fisher Scientific) mass spectrometer. Full MS scans were acquired over the range of m/z 380-1600 Thompson units with resolution set to 70,000 (at m/z 200) and an automatic gain control (AGC) target set to 3 x 10^6. The 12 most intense parent ions, excluding singly charged ions and ions with unassigned charges, were selected for higher-energy collisional dissociation (HCD) fragmentation with the normalized collision energy (NCE) set at 28%. The MS/MS spectra were analyzed in the Orbitrap mass analyzer using resolution set at 17,500 and AGC set
at $1 \times 10^5$. The isolation window was set at 2 m/z and dynamic exclusion at 60 s. The maximum ion injection time was 20 ms for full MS scans and 120 ms for MS/MS scans.

### 3.2.11 Database Search

LC-MS/MS raw data files were analyzed using Proteome Discoverer 1.4 (Thermo Fisher Scientific) by two search engines Sequest HT (Thermo) and Mascot (Matrix Science) against the UniProt human database (2013 Jan version, containing 139905 sequences). MS/MS spectra were first searched against the NIST HCD and CID human spectral libraries (NIST human IT library, 2012-5-30). Unassigned spectra were then analyzed by Sequest HT and Mascot for fully tryptic and semitryptic peptide sequence matches. Carbamidomethylation (57.021 Da) was set as a fixed modification and N-terminal acetylation, methionine oxidation and deamidation (NQ) were set as variable modifications. The precursor peptide mass tolerance was 10 ppm and fragment tolerance 0.05 Da. The results of the searches were combined and validated using the Percolator module with filters set to high peptide identification confidence to achieve a false discovery rate (FDR) $\leq 1\%$ for SEQUEST and $\leq 0.5\%$ for Mascot. The protein identifications were reported as protein groups in case that a peptide sequence can be identified as a match for multiple proteins in the database.

### 3.2.12 Gene Ontology (GO) Annotation Analysis

GO annotation analysis was performed by a Software Tool for Rapid Annotation of Proteins (STRAP) developed by BU school of medicine. STRAP took the list of protein identifications
from database search as an input and generated protein annotation results from the on-line information on Uniprot.

3.2.13 Targeted Analysis

The peptide list from the database search was compared with the sequences of 252 SIL peptides. The endogenous peptides, which had matched SIL peptide pairs, were selected as targets for absolute quantifier. The abundance of the targeted peptides was presented as the ratio of peak area between endogenous and SIL peptides.

3.2.14 iBAQ Quantitation and IPA Network Analysis

The label-free quantification was performed by intensity-based absolute quantification (iBAQ), developed by Schwanhäusser\textsuperscript{34}. A PERL script available in MaxQuant software automatically calculated the iBAQ values for large sets of proteins identified from MaxQuant. Proteins with a 4-fold difference in iBAQ intensity were imported into IPA for the network analysis.
3.3 Results and Discussion

3.3.1 Performance of PLOT NanoLC Columns Coupled with Q-Exactive Mass Spectrometry

To capture the benefits of ultralow flow in nLC-MS, we developed 10 µm i.d. poly(styrene-divinylbenzene) (PS-DVB) porous layer open tubular (PLOT) columns with 1 µm thickness of layer with high permeability (Fig. 3.1a). The performance of the PLOC column was evaluated using split-injection system described in Chapter 2. A 3 m PLOT column exhibited a height equivalent to a theoretical plate (HETP) of 8 µm (corresponding to approximate 100,000 plates/m) at 20 nL/min ultralow flow rate (Fig. 2.5). To study the gain of the ESI sensitivity from the ultralow flow rate, the 3 m PLOT column was coupled to HR/AM mass spectrometry – Q Exactive. Analyzing an equimolar mixture of digested 6 bovine proteins with non-targeted data dependent method resulted in the detection limits down to 10 zmol level (Fig. 3.1b), based on an accurate m/z with 2 ppm, and 50 zmol (Fig. 3.1c), based on MS/MS fragment matching. A linear MS response was recorded over a wide dynamic range of four orders of magnitude in protein amounts spanning from 10 zmol to 100 amol, with a liner regression $r^2 > 0.99$.

The implementation of a quadrupole in the Q Exactive provides a new targeted analysis method in Orbitrap instruments, called parallel (or pseudo) reaction monitoring (PRM) mode\textsuperscript{35}. Unwanted ions were filtered in the quadrupole before entering the HCD collision cell, hence sensitivity of the targeted ions increased. In PRM mode, detection of peptides was confirmed at as low as 5 zmol, based on the MS/MS fragmentation, which was a 10-fold increase in sensitivity compared to the data dependent scanning mode (Fig. 3.1d right). The HCD spectrum in 5 zmol
detection was still of high quality and lower limits of detection (down to ymol) could be potentially achieved. Quantitation of targeted ions in PRM mode could be performed by extracting signal intensity for one or several most prominent product ions (e.g., y7⁺ in the case for peptide ALVYGEATSR) similar as in QQQ or QTOF. Excellent linear correlation between sample loads and experimental signal response with r² of 0.9998 for a range of 5 zmol to 500 zmol was obtained using PRM mode (Fig. 3.1.d left), demonstrating that the targeted mode in Q Exactive could be used as an alternative quantitative method other than QQQ. The power of target mode was also examined by spiking in the 6 bovine protein digest into the complex matrix. In 8 ng MCF-7 lysate as the background, we were able to reliably quantify peptides at such low levels as 10-50 zmol in one-dimensional separation of the unfractionated sample (Fig. 3.1.e and more examples of targeted peptide analysis in Fig. 3.2).
Figure 3.1. Performance characterization of a porous layer open tubular column coupled with a Q Exactive mass spectrometer. a, SEM of cross-section of 3 m×10 μm i.d. PLOT column. b, Extracted ion chromatography (EIC) of six bovine mix digest from 10 zmol to 100 amol injection by data-dependent analysis. c, MS/MS chromatography of peptide ALVYGEATSR (m/z 533.7800). d, Targeted-MS² analysis of peptide ALVYGEATSR. e, Examples of parallel product ion monitoring of targeted precursor ion of peptides ALVYEGATS, VLALDLSIK and FFVAPFPFVGK.
Figure 3.2. Parallel product ion monitoring of 8 targeted precursor ions from 6 bovine mix digest.
3.3.2 Sensitivity of Proteomic Analysis of Limited Samples

The PLOT column coupled with Q Exactive exhibited its great ability for the ultrasensitive deep proteomic analysis of limited sample amounts. To increase the loading capacity, a short (4-6 cm) 50 μm i.d. 1-decene modified PS-DVB monolithic microSPE pre-column was used (Fig. 2.2). As samples of high complexity, we selected a lysate prepared from 10 million human MCF-7 cultured cells by conventional lysis and digestion approach. Aliquots corresponding to 500 MCF-7 cells (~100 ng total protein) were loaded, concentrated and desalted on the microSPE column at a flow rate of 200 nL/min. The analytical separation was performed on a 4.2 m PLOT column using a 4 h long gradient at a flow rate of 20 nL/min coupled to Q Exactive mass spectrometer.

As a result, up to 3,700 unique protein groups and 22,645 unique peptides were identified in a single analysis of 500 MCF-7 cells on average and a combination of 5 replicate PLOT-nLC-MS runs resulted in identification of 5,183 proteins and 41,020 peptides (Fig. 3.3a), whereas the analysis using a conventional 75 μm i.d. x 20 cm packed column resulted in a significantly lower depth of only about one third protein IDs (Fig. 2.9), indicating the excellent fit and performance of the PLOT nLC-MS platform in ultrasensitive microproteomic profiling of limited availability samples. As expected, the number of identified proteins decreased when lower amounts of cells were injected (Fig. 3.3b). However, the platform demonstrated outstanding depth and performance in analysis of cell lysates corresponding to 50 and 100 cells, identifying 1,327 ± 143 and 2026 ± 98 protein groups respectively, n=3 (Fig. 3.3b). Very deep proteomic analysis of cell line resulting in identification of over 10,000 proteins, which covers almost a half of the whole expressed genome (excluding proteoforms) has been recently achieved, however, the
analysis required ~ $10^6$ cells and extensive sample fractionation\textsuperscript{15}. The PLOT nLC platform provided the tools to analyse few hundred cells in one-dimensional separation with a relatively high depth. The number of proteins (i.e., protein groups) identified by our approach was compared to other published data in analysis low number of cells. The recent analysis of 250-500 melanoma cells resulted in 629 and 867 proteins for 250 cells and 500 cells, respectively, using a pseudoshotgun (PSG) sample preparation method\textsuperscript{14}. The Matthias Mann group reported an identification of 905 proteins from formalin fixed paraffin embedded (FFPE) sample corresponding to 500 cells using filter-aided sample preparation (FASP) workflow\textsuperscript{36}. Our platform demonstrated a 4-5 times higher numbers of identified proteins for similar cell numbers and we also explored the level of 50 cells that was not previously assessed using alternative techniques.

Technical replicates in our analysis resulted in similar proteomic profiling data, with a significant overlap of 75\% in proteins identified by all three replicates (Fig. 3.3c). These results illustrated the high reproducibility of analysis enabled by the PLOT platform for such low sample levels. Over 95\% of identified proteins in the analysis of 50 cells and 93\% for profiling of 100 cells were also identified in the analysis of 500 cells (Fig. 3.3d), indicating that the reliability of our technique for such very low cell numbers.
Figure 3.3. Proteomic analysis of limited amount of MCF-7 cell lysate prepared by a conventional urea-based lysis and in-solution digestion. a, Increased protein and peptide IDs with multiple injection of 500 MCF-7 cells. MCF-7 lysate corresponding to 500 cells were injected 5 times. Database search results after each run was combined to the previous run(s) to gain identifications of protein groups (i) and peptides (ii). b, Sensitivity profile of proteomic analysis of MCF-7 lysate corresponding to 50, 100 and 500 cells. c, Venn diagram of protein IDs of three replicate analysis of 500 MCF-7 cells; d, venn diagram of total protein IDs of three replicate analysis of 50, 100 and 500 MCF-7 cells.
3.3.3 Sample Preparation for Limited Availability Samples

The sensitivity of the microSPE-PLOT platform was assessed injecting sample aliquots corresponding to low cell numbers while the sample was prepared from 10 million MCF-7 cells using conventional lysis and digestion protocols. Sample preparation for lower cell numbers was challenging and direct down-scaling of the protocol was found ineffective due to the more pronounced effect of sample losses on contact surfaces during sample manipulation. Thus, sample preparation starting from 500-5000 cultured MCF-7 cells using acetone precipitation approach resulted in only 167 ± 21 and 491 ± 63 most abundant proteins identified from 500 and 2,500 MCF-7 cells, correspondingly.

Our preliminary study on sample preparation using only 2,000 MCF-7 cells was performed by a conventional cell lysis in 8M urea-based lysis and in-solution digestion, resulting in identification of 1,286 ± 11 proteins from an injection of a aliquot of the sample corresponding to 500 cells, n=2 (Fig. 3.4). Therefore, the difference in identification results and hence, protein recovery, was substantial for sample preparations using limited numbers (2,000 cells) and a large cell number (10×10⁶) (Fig. 3.3b). Then we used a microSPE tip (StageTip) as the proteomic reactor³⁷, where on-line desalting and enzymatic digestion were performed on the solid phase support. This technique allowed us to double the number of identified proteins resulting in 2,582 ± 47 detected unique protein groups (Fig. 3.4). The decreased depth of the proteomic profiling when a significantly lower amount of cells was used in comparison to the above described experiments with the initial cell number equal to 10x10⁶ can be caused by the lower recovery of cells, proteins and peptides in all steps of sample processing, involving sample dilution, transfer and clean-up steps, where the sample was brought in contact with various surfaces at high
surface-to-volume ratios (i.e., adsorption in test tubes, pipet tips, stationary phase, etc.). Also, most likely the less concentrated lysates and digests resulted from low cell copy numbers, resulted in virtually eliminated “sample career” effect that was assisting to sample dissolution in highly concentrated high cell copy number samples. In order to alleviate these issues, we developed a single-tube sample preparation approach based on the AFA (Adaptive Focused Acoustics™) technology\(^{38}\) that did not require any sample clean-up and where the sample was processed with minimal dilution (hence, low surface-to-volume ratios). The AFA produces a focused acoustic beam of the 15-30-fold higher frequency than a conventional sonication bath or probe (Fig. 3.5a)\(^{39}\). The advantage of AFA is that the acoustic energy disrupts the tissue without adding common high concentration lysis buffers (e.g. urea- or guanidine-based), therefore avoiding the sample dilution and clean-up procedures and improving the sample recovery. Recently, the AFA sonication in the microscale chip analysis of 5,000 cells was applied and demonstrated high efficiency\(^{40}\).

Using 4 h long gradient on SPE-PLOT system, proteomic analysis of 500 MCF-7 cells prepared from 2,000 cells by AFA-aided approach resulted in identification of 3370 ± 119 protein IDs, which is similar to the profiling depth and sample recovery we achieved using 10 million of MCF-7 cells as the initial amount (over 90%). Surprisingly, AFA preparation exhibited higher sensitivity in the analysis of lower cell numbers. In profiling of 100 cells, 2,061 ± 39 proteins were identified (n=2), which is almost identical to the samples prepared from 10 million cells; whereas the analysis of 50 cells resulted in 1,802 ± 18 protein groups, a 36% increase in comparison to the bulk sample preparation. The increased sensitivity of the AFA-based single-tube preparation technique indicated an improved overall recovery of cells, proteins
and peptides. Hence, the AFA approach has been proven to be an attractive method to handle limited samples.

In addition, we investigated how the presence of magnetic beads, which were used for immunoaffinity magnetophoretic cell capture, affects the cell lysis. Approximately 2,000 cultured MCF-7 cells washed by DPBS buffer were counted by flow cytometry and analyzed using AFA approach in a volume of 50 µL (n = 4) with or without adding the beads. The cell lysis without microbeads followed by analysis of a sample aliquot corresponding to 500 cells resulted in identification of 1107 ± 85 protein groups, whereas the cell lysis performed with microbeads doubled the profiling depth (2311 ± 416 protein groups) most likely due to the bead milling effect which improves the cell disruption\textsuperscript{41}. Here, AFA- and bead-assisted cell lysis demonstrated a lower profiling depth due to the larger digestion volume used (50 µL), larger surface-to-volume ratio and correspondingly higher sample losses than in our previous AFA-assisted experiments without cell culture (10-15 µL, see Fig. 3.5b,c).
Figure 3.4. Improved protein IDs in different optimized sample preparation approaches for limited starting material analysis in comparison to conventional sample preparation from millions cells.
Figure 3.5. Optimization of sample preparation for analysis of limited samples. a. Workflows for sample preparation of low and high numbers of MCF-7 cells. b. Improved sample preparation methods of 2,000 cells using AFA lysis results in a comparable depth of proteomic profiling as the conventional in-solution protocol using 10 x 10^6 cells when sample aliquots equivalent to 500 cells were analyzed. c. Identification results of proteomic analysis of MCF-7 lysates corresponding to 50, 100 and 500 cells, prepared from either 2,000 cells using AFA-assisted or 10 million MCF-7 cells using conventional urea-based lysis and in-solution digestion.
Figure 3.6. Work flow of proteomic analysis of limited amount cells captured by microfluidic device.  

Microfluidic separation of 1,000-100,000 MCF-7 cells spiked-in whole blood.  

b, Microfluidic separation of EPCs and HSCs in blood draw from individuals.  

c, Sample preparation of isolated cells on micromagnetic beads.  

d, LC/MS analysis of cell lysate and bioinformatic analysis.
3.3.4 Performance Assessment of the Microfluidic Magnetophoretic System for Isolation and Analysis of Rare Cells

To isolate and purify specific subpopulations of cells, a microfluidic device based on immunoaffinity interaction followed by magnetophoresis developed by Dr. Murthy’s group was used in this study\(^2\) (Fig. 3.6a, b). To validate the cell capture platform in the proteomic aspect, 3000 MCF-7 cultured cells counted by flow cytometry were incubated with magnetic microbeads functionalized with anti-EpCAM antibodies, spiked in 1 mL of the whole human blood and captured by the microfluidic device (Fig. 3.6a). The functionalized anti-EpCAM was used to selectively interact with the selectins on the cell surface of MCF-7\(^4\). Approximately 60% (~ 1800 cells) were captured on the anti-EpCAM-functionalized magnetic microbeads as measured by flow cytometry. Two independent replicate isolations from two replicate samples were carried out for capture of MCF-7 cells at the investigated spiked cell concentration level. Isolated cells with bound magnetic microbeads were washed twice with DPBS buffer and lysed using the above discussed optimized AFA approach. Two kinds of quality control (QC) experiments were performed to assess the isolation efficiency and specificity: identical aliquots of either (1) negative control: the whole human blood without spiked-in MCF-7 cells or (2) positive control: 2,000 counted MCF-7 cells in DPBS were processed without passing through the device prior to analysis PLOT-nLC-MS. Each aliquot of cell lysate corresponding to 500 cells was loaded on SPE and separated by PLOT column.

Prior to analysis of the captured cells, the level of contamination and carryover of the SPE-PLOT-nLC-MS platform was examined by injecting an aliquot of the aqueous mobile phase (solvent A). Only 33 ± 20 proteins (protein groups) were identified in each run on average as
well as 54 proteins in all four performed injections (Fig. 3.7a). The negative control experiment where 1mL of blood without spiked-in MCF-7 cells was passed through the device resulted in identification of 103 ± 25 proteins in each injection and 298 proteins in all four analyses, indicating a relatively low level of non-specific binding. However, 3,126 ± 103 proteins were identified on average in each analysis of 500 cells captured from blood (from the 1800 captured cells) by the microfluidic device and 5,117 proteins were identified by combining the results of two technical and two biological replicates. These numbers as well as the constituents of the detected MCF-7 proteome were similar to the positive control experiment where the identical amount of cells in DPBS was processed without passing through the microfluidic device (3449 ± 204 protein groups on average, 5,223 IDs total, n=4). Based on our comparative analysis of the above described samples, we estimate that only approximately 3% of proteins identified in MCF-7 cells isolated from the whole blood correspond to nonspecifically bound blood proteins, which is in agreement with Dr. Murthy’s previous work where ≥95% purity of isolation was observed. The coefficient of variation (CV) of 3% in the number of protein groups identified in four analyses demonstrated the good reproducibility of both the cell capture and analysis.

Further investigation of the differences in the proteomes recovered by the above experiments was performed using comparative gene ontology (GO) enrichment analysis. The proteomic profiles of the captured MCF-7 cells from blood and the negative control isolates from blood demonstrated very significant differences in distributions of the corresponding cellular localization GO terms, especially for extracellular, cytosolic and nucleus proteins (e.g., 9% vs. 2% and 14% vs. 22% in MCF-7 isolates and negative controls, respectively; see Fig. 3.7b, total proteins). Such differences were especially pronounced when protein groups uniquely detected in
each sample type were compared (Fig. 3.7b). It is noteworthy that many high or moderate abundance plasma proteins, such as immunoglobulins and kallikrein (see Table S1), were unique for the negative control experiment. In the positive control experiment, where the cultured MCF-7 cells in DPBS were analyzed, the GO profiles were very similar to those in the analysis of the MCF-7 cells isolated from blood by the microfluidic device for both unique and total proteins (Figure 2.7c). This high similarity in GO profiles verified the high specificity of the cell capture microfluidic device.
Figure 3.7. Evaluation of the microfluidic cell capture approach using proteomic profiling of MCF-7 cells isolates from blood (“spiked-in MCF-7 cells in blood”) or positive control (“MCF-7 cells in DPBS buffer”) and proteomes recovered in negative control experiments. 
a. A comparison of the profiling depth. b. The results of comparative GO term enrichment analysis between proteomes of isolates from the whole blood with and without spiked-in MCF-7 cells. c. GO term enrichment analysis for proteomes of MCF-7 cells captured from blood and MCF-7 cells in DPBS buffer (positive control).
3.3.5 Quantitative Analysis of MCF-7 Cells from Blood Isolated by the Magnetophoretic Microfluidic Device

To evaluate the applicability of the microfluidic platform in efficient isolation of cells within a range of physiologically relevant concentrations, Dr. Murthy’s lab spiked 1000, 2000, 3000, 5000 and 10000 MCF-7 cells per mL of the whole blood and captured by the magnetophoretic device (Fig. 3.6a). Approximately 60-70% of the cells were captured by the device (Table 3.2), prepared by AFA sonication and ~20% of cell lysate was injected into the 4.2 m PLOT column. On average 2,512 ± 246 proteins were identified in each sample aliquot equivalent to 122 cells from the total of ~600 cells captured when 1,000 cells were spiked into1 mL of blood and 3,752 proteins in three technical replicates. The protein IDs increased with the number of cells injected and 3,402 ± 169 protein groups were identified in each injection of an aliquot of 1,440 cells corresponding to ~20% of ~7,200 cells captured from the experiment with 10,000 cells spiked-in 1 mL of blood. (Fig. 3.8a, Table 3.2). The increase of the identified peptides presented the similar trend (Fig. 3.8b). Quantitative analysis was performed using label-free relative and absolute quantitative techniques (MaxQuant) as well as synthetic stable isotope labeled (SIL) peptides for absolute quantitation of selected proteins.

Stable isotope labeling (SIL) peptides quantitation: prior to the MS analysis, 1 fmol of 252 proteotypic SIL peptides corresponding to proteins presumably related to cancer was spiked in each of the five samples corresponding to the tested levels of the lysed cell isolates (Fig. 3.6c, Table 3.2). Four SIL peptides corresponding to endogenous peptides were quantified across all the levels and selected as examples for demonstrating the capabilities of absolute quantitative
analysis. Peak intensities and peak areas of the 4 peptides increased with the number of cells spiked-in and captured (Fig. 3.8c, i-iv). For each peptide, the abundances at different levels were normalized to peak intensities of the corresponding SIL peptides, resulting in measured ‘light/heavy’ (i.e., endogenous/SIL peak area) ratios. Plots of cell number and ‘light/heavy’ ratio relationship demonstrated excellent linear correlation of 0.97-0.99 (Fig. 3.8c, v-viii), which confirms the ability of the microfluidic device to reflect the trends of target cell concentrations within the range of 1,000-10,000 cells/mL.

Label-free iBAQ: A label-free quantitative approach iBAQ\textsuperscript{34} available within the MaxQuant software platform was used in this study to access abundances of each of reliably identified proteins. The iBAQ approach estimates protein abundances by normalizing the accumulated peptide intensities to the theoretical number of peptides of a digested protein\textsuperscript{34}; it was validated by a good correlation in the quantitation of large sets of proteins\textsuperscript{43}. The iBAQ quantitative approach demonstrated significant differences between dynamic ranges recovered by the proteomic profiling of MCF-7 cells captured at the tested five levels of cell concentrations (Fig. 3.9a). A linear correlation with $R^2$ of 0.84 was determined for the median values of protein abundances determined at each concentration level (Fig. 3.9b), demonstrating of the applicability of the iBAQ approach in label-free quantitation of limited samples.
Table 3.2: MCF-7 Cells captured for the quantitative analysis. The numbers of captured MCF-7 cells were counted by flow cytometry.

<table>
<thead>
<tr>
<th></th>
<th>1000</th>
<th>2000</th>
<th>3000</th>
<th>5000</th>
<th>10000</th>
</tr>
</thead>
<tbody>
<tr>
<td># Spiked MCF-7 cells</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># Captured MCF-7</td>
<td>611</td>
<td>1196</td>
<td>2045</td>
<td>3434</td>
<td>7201</td>
</tr>
<tr>
<td># Injected MCF-7 cells</td>
<td>122</td>
<td>239</td>
<td>409</td>
<td>687</td>
<td>1440</td>
</tr>
</tbody>
</table>

Table 3.3: Numbers of hematopoietic stem cells (HSCs) and endothelial progenitor cells (EPCs) captured from the microfluidic device counted by flow cytometry.

<table>
<thead>
<tr>
<th></th>
<th>Non-smoker</th>
<th>Smoker</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td># HSCs</td>
<td>14645</td>
<td>12342</td>
</tr>
<tr>
<td># EPCs</td>
<td>8342</td>
<td>7782</td>
</tr>
<tr>
<td>Total</td>
<td>22987</td>
<td>20124</td>
</tr>
</tbody>
</table>
Figure 3.8. Titration analysis of spiked-in MCF-7 cells captured by the magnetophoretic microfluidic device. Identified proteins (a) and peptides (b) in the proteomic analysis of MCF-7 cells corresponding to 122, 239, 409, 687 and 1440 cells (~20% of total number of captured cells) injected in each level of LC-MS experiments. c, i-iv, Illustrative examples of extracted ion chromatograms corresponding to four endogenous peptides spiked at different levels and their corresponding SIL peptides; v-viii, linear correlation of the peak area ratios for endogenous/SIL (‘light/heavy’) ratios with the number of cells injected in the range of 100-1500 cells per analysis.
Figure 3.9. iBAQ Label-free Quantification. a, Ranked protein abundant (log_{10} value) from lowest to highest in the proteomic analysis in 122, 239, 409, 687 and 1440 cells. b, Box plot distribution of the protein abundances (log_{2} value) presented in the analysis from 122 to 1440 cells.
3.3.6 Proteomic Analysis of HSCs and EPCs Isolated from Whole Blood

To demonstrate the utility of the microfluidic device in the analysis of rare cells in biomedical applications, Dr. Murthy’s lab isolated endothelial progenitor cells (EPCs) and hematopoietic stem cells (HSCs) from human whole blood drawn from seven healthy donors that included five non-smokers and two smokers (Fig. 3.6b). The total cell numbers of combined EPCs and HSCs from non-smokers were approximately 20% more than in smokers (Table 3.3), which was in agreement with the clinical studies\textsuperscript{44,45}. However, proteomic analysis revealed an increase in the number of protein IDs in cell isolates from smokers (Fig. 3.10a), especially the increase was more pronounced when the number of protein IDs was normalized to the number of captured cells (Fig. 3.10b). The number of proteins identified in EPCs and HSCs (~2,000 proteins identified in ~1,000 cells) was lower than in case of MCF-7 cells. Similarly, the protein content expressed in stem cells was reported to be lower than in HeLa cancer cells\textsuperscript{46}.

The differences in the protein abundance were compared among the shared proteins identified in both non-smoker and smoker groups (Fig. 3.10c). The up- and down-regulated proteins were selected as criteria of 4-fold difference in iBAQ values (Fig. 3.10c), i.e., > 4 or < 0.25 in shared proteins between non-smoker and smoker groups (Fig. 3.10c). Five up-regulated proteins and 64 down-regulated proteins (see Supplementary Table S2) were imported into IPA for network analysis using the IPA Ingenuity Knowledge Base. A networks associated with dermatological diseases and conditions, organismal injury and abnormalities, inflammatory disease were generated with highest score of 52 (Fig. 3.10d and Fig. 3.11). The network revealed that two upstream proteins, NAMPT (nicotinamide phosphoribosyltransferase) and C3 (Complement component 3), which are more characteristic to the extracellular space, may
feedback regulate the cell functions by external stimulation. Several clinical studies demonstrated that NAMPT and C3 are relevant to the smoking$^{47,48}$. The presented results of the EPC/HSC capture experiments demonstrate the utility of the immunoaffinity magnetophoretic cell isolation coupled with the PLOT-nLC-MS-based proteomic profiling in clinical and biological studies.
Figure 3.10. Proteomic analysis of HSCs and ECPs isolated from healthy non-smoking and smoking donors. a, Identified proteins in ~1000 of HSCs and EPCs from individual non-smokers and smokers. b, Identified proteins normalized to the number of captured cells from individual non-smokers and smokers; c, Distribution of iBAQ ratio between smokers and control group; the cut-off for the up- and down-regulated proteins was a 4-fold difference in the iBAQ ratio. d, Associated network functions in dermatological diseases and conditions, organismal injury and abnormalities and inflammatory disease, generated by IPA. (See enlarged figure in Fig. 3.11)
Figure 3.11. IPA analysis of associated network functions in dermatological diseases and Conditions, organismal injury and abnormalities and inflammatory disease.
3.4 References


**Supplement**

**Table S1:** List of unique proteins identified from the control analysis without spiked MCF-7 cells in the whole blood.

<table>
<thead>
<tr>
<th>Uniprot</th>
<th>Description</th>
<th>Sum PSMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q5DT20</td>
<td>Hornerin OS=Homo sapiens GN=HRNR PE=2 SV=1 - [Q5DT20_HUMAN]</td>
<td>73</td>
</tr>
<tr>
<td>Q5D862</td>
<td>Filaggrin-2 OS=Homo sapiens GN=FLG2 PE=1 SV=1 - [FILA2_HUMAN]</td>
<td>22</td>
</tr>
<tr>
<td>P02814</td>
<td>Submaxillary gland androgen-regulated protein 3B OS=Homo sapiens GN=SMR3B PE=1 SV=2 - [SMR3B_HUMAN]</td>
<td>17</td>
</tr>
<tr>
<td>Q0EFA5</td>
<td>S protein OS=Homo sapiens GN=S PE=2 SV=1 - [Q0EFA5_HUMAN]</td>
<td>10</td>
</tr>
<tr>
<td>P31151</td>
<td>Protein S100-A7 OS=Homo sapiens GN=S100A7 PE=1 SV=4 - [S10A7_HUMAN]</td>
<td>10</td>
</tr>
<tr>
<td>:sp</td>
<td>&gt;sp[CASI_BOVIN]</td>
<td>10</td>
</tr>
<tr>
<td>F8VV13</td>
<td>Mucin-like protein 1 OS=Homo sapiens GN=MUCL1 PE=3 SV=1 - [F8VV13_HUMAN]</td>
<td>8</td>
</tr>
<tr>
<td>P01859</td>
<td>Ig gamma-2 chain C region OS=Homo sapiens GN=IGHG2 PE=1 SV=2 - [IGHG2_HUMAN]</td>
<td>7</td>
</tr>
<tr>
<td>B4DRV8</td>
<td>cDNA FLJ57574, highly similar to Protein-glutamine gamma-glutamyltransferase E (EC 2.3.3.13) OS=Homo sapiens PE=2 SV=1 - [B4DRV8_HUMAN]</td>
<td>7</td>
</tr>
<tr>
<td>A2TIK1</td>
<td>Collagen type I alpha 2 (Fragment) OS=Homo sapiens GN=COL1A2 PE=4 SV=1 - [A2TIK1_HUMAN]</td>
<td>7</td>
</tr>
<tr>
<td>C9JEV0</td>
<td>Zinc-alpha-2-glycoprotein OS=Homo sapiens GN=AZGP1 PE=3 SV=1 - [C9JEV0_HUMAN]</td>
<td>6</td>
</tr>
<tr>
<td>B7Z1E3</td>
<td>Neuregulin-1 OS=Homo sapiens GN=NRG1 PE=2 SV=1 - [B7Z1E3_HUMAN]</td>
<td>6</td>
</tr>
<tr>
<td>P22531</td>
<td>Small proline-rich protein 2E OS=Homo sapiens GN=SPRR2E PE=2 SV=2 - [SPR2E_HUMAN]</td>
<td>5</td>
</tr>
<tr>
<td>P31944</td>
<td>Caspase-14 OS=Homo sapiens GN=CASP14 PE=1 SV=2 - [CASPE_HUMAN]</td>
<td>4</td>
</tr>
<tr>
<td>P01877</td>
<td>Ig alpha-2 chain C region OS=Homo sapiens GN=IGHA2 PE=1 SV=3 - [IGHA2_HUMAN]</td>
<td>4</td>
</tr>
<tr>
<td>P01857</td>
<td>Ig gamma-1 chain C region OS=Homo sapiens GN=IGHG1 PE=1 SV=1 - [IGHG1_HUMAN]</td>
<td>4</td>
</tr>
<tr>
<td>Q5JR98</td>
<td>Tctex1 domain-containing protein 4 OS=Homo sapiens GN=TCTEX1D4 PE=1 SV=1 - [TC1D4_HUMAN]</td>
<td>3</td>
</tr>
<tr>
<td>Q59F16</td>
<td>Alpha 1 type VII collagen variant (Fragment) OS=Homo sapiens PE=2 SV=1 - [Q59F16_HUMAN]</td>
<td>3</td>
</tr>
<tr>
<td>F8VV32</td>
<td>Lysozyme C OS=Homo sapiens GN=LYZ PE=3 SV=1 - [F8VV32_HUMAN]</td>
<td>3</td>
</tr>
<tr>
<td>E9PBV3</td>
<td>Suprabasin OS=Homo sapiens GN=SBSN PE=4 SV=1 - [E9PBV3_HUMAN]</td>
<td>3</td>
</tr>
<tr>
<td>C9J9J2</td>
<td>Protein ZNF783 OS=Homo sapiens GN=ZNF783 PE=4 SV=1 -</td>
<td>3</td>
</tr>
<tr>
<td>Accession</td>
<td>ID</td>
<td>Description</td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>-----------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>B7ZMK6</td>
<td>C9J9J2</td>
<td>GOLGA8B protein OS=Homo sapiens GN=GOLGA8B PE=2 SV=1 - [B7ZMK6_HUMAN]</td>
</tr>
<tr>
<td>Q92681</td>
<td>S1A7B</td>
<td>Regulatory solute carrier protein family 1 member 1 OS=Homo sapiens GN=RSC1A1 PE=2 SV=1 - [RSCA1_HUMAN]</td>
</tr>
<tr>
<td>Q5SY68</td>
<td>S100A7</td>
<td>Protein S100-A7-like 2 OS=Homo sapiens GN=S100A7L2 PE=2 SV=1 - [S1A7B_HUMAN]</td>
</tr>
<tr>
<td>Q5BQ95</td>
<td>KLK13</td>
<td>Kallikrein 13 splice variant 7 OS=Homo sapiens GN=KLK13 PE=2 SV=1 - [Q5BQ95_HUMAN]</td>
</tr>
<tr>
<td>Q53RT3</td>
<td>ASPRV1</td>
<td>Retroviral-like aspartic protease 1 OS=Homo sapiens GN=ASPRV1 PE=1 SV=1 - [ASPRV1_HUMAN]</td>
</tr>
<tr>
<td>P23490</td>
<td>LOR</td>
<td>Loricrin OS=Homo sapiens GN=LOR PE=1 SV=2 - [LORI_HUMAN]</td>
</tr>
<tr>
<td>F8W9L1</td>
<td>SERPINB4</td>
<td>Serpin B4 OS=Homo sapiens GN=SERPINB4 PE=3 SV=1 - [F8W9L1_HUMAN]</td>
</tr>
<tr>
<td>B9EGH5</td>
<td>PLCB2</td>
<td>1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-2 OS=Homo sapiens GN=PLCB2 PE=2 SV=1 - [B9EGH5_HUMAN]</td>
</tr>
<tr>
<td>B2R4M6</td>
<td>S100A9</td>
<td>cDNA, FLJ92148, highly similar to Homo sapiens S100 calcium binding protein A9 (calgranulin B) (S100A9), mRNA OS=Homo sapiens PE=4 SV=1 - [B2R4M6_HUMAN]</td>
</tr>
<tr>
<td>Q9P173</td>
<td>PRO2275</td>
<td>Isoform 2 of Neutral ceramidase OS=Homo sapiens GN=ASAH2 - [ASAH2_HUMAN]</td>
</tr>
<tr>
<td>Q9NR71-2</td>
<td>ASAH2</td>
<td>Isoform 2 of Neutral ceramidase OS=Homo sapiens GN=ASAH2 - [ASAH2_HUMAN]</td>
</tr>
<tr>
<td>Q9HCN3-2</td>
<td>KIAA1549</td>
<td>Isoform 2 of UPF0606 protein KIAA1549 OS=Homo sapiens GN=KIAA1549 - [KIAA1549_HUMAN]</td>
</tr>
<tr>
<td>Q96B18</td>
<td>DACT3</td>
<td>Dapper homolog 3 OS=Homo sapiens GN=DACT3 PE=1 SV=2 - [DACT3_HUMAN]</td>
</tr>
<tr>
<td>Q8NFU5</td>
<td>IPMK</td>
<td>Inositol polyphosphate multikinase OS=Homo sapiens GN=IPMK PE=1 SV=1 - [IPMK_HUMAN]</td>
</tr>
<tr>
<td>Q8N3J5</td>
<td>PPM1K</td>
<td>Protein phosphatase 1K, mitochondrial OS=Homo sapiens GN=PPM1K PE=1 SV=1 - [PPM1K_HUMAN]</td>
</tr>
<tr>
<td>Q59EE7</td>
<td>Q59EE7</td>
<td>Pro-alpha-1 type V collagen variant (Fragment) OS=Homo sapiens PE=2 SV=1 - [Q59EE7_HUMAN]</td>
</tr>
<tr>
<td>P29322</td>
<td>EPHA8</td>
<td>Ephrin type-A receptor 8 OS=Homo sapiens GN=EPHA8 PE=1 SV=2 - [EPHA8_HUMAN]</td>
</tr>
<tr>
<td>P12273</td>
<td>PIP</td>
<td>Prolactin-inducible protein OS=Homo sapiens GN=PIP PE=1 SV=1 - [PIP_HUMAN]</td>
</tr>
<tr>
<td>I0BOK3</td>
<td>FLG</td>
<td>Truncated profilaggrin OS=Homo sapiens GN=FLG PE=4 SV=1 - [I0BOK3_HUMAN]</td>
</tr>
<tr>
<td>E7ESR8</td>
<td>MAML3</td>
<td>Mastermind-like protein 3 OS=Homo sapiens GN=MAML3 PE=4 SV=1 - [E7ESR8_HUMAN]</td>
</tr>
</tbody>
</table>
### Table S2: Up-regulated proteins and down-regulated proteins from HSCs and EPCs in smokers with 4-fold difference compared to non-smokers. (ratio > 4, up-regulation, red; ratio <0.25, down-regulation, blue)

<table>
<thead>
<tr>
<th>Uniprot ID</th>
<th>Fasta headers</th>
<th>iBAQs/iBAQc ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>B4DP27</td>
<td>&gt;tr</td>
<td>B4DP27</td>
</tr>
<tr>
<td>P26368-2</td>
<td>&gt;sp</td>
<td>P26368-2</td>
</tr>
<tr>
<td>Q9P265</td>
<td>&gt;sp</td>
<td>Q9P265</td>
</tr>
<tr>
<td>F5GWX5</td>
<td>&gt;tr</td>
<td>F5GWX5</td>
</tr>
<tr>
<td>Q8NE71-2</td>
<td>&gt;sp</td>
<td>Q8NE71-2</td>
</tr>
<tr>
<td>Accession</td>
<td>Description</td>
<td>Gene</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Q9NPD3</td>
<td>Exosome complex component RRP41</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>B4DU58</td>
<td>Macrophage-capping protein</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>B5MC98</td>
<td>Prolactin regulatory element-binding protein</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>O43583</td>
<td>Density-regulated protein</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>P01024</td>
<td>Complement C3 complement</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Q8TCE1</td>
<td>SERPINC1 protein</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>J3KRE2</td>
<td>Rho GDP-dissociation inhibitor 1</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>O60744</td>
<td>Thioredoxin delta 3 (Fragment)</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>P08253</td>
<td>Collagen alpha-1(VI) chain</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>P12109</td>
<td>Collagen alpha-1(VI) chain</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Q4QRK8</td>
<td>DNA-directed RNA polymerase (Fragment)</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Q02750-2</td>
<td>Isoform 2 of Dual specificity mitogen-activated protein kinase</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>P52926</td>
<td>HMGA2_HUMAN High mobility group protein</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>F5H7U0</td>
<td>6-phosphogluconate dehydrogenase, decarboxylating OS</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>F6SBX2</td>
<td>Isoleucine--tRNA ligase, mitochondrial OS</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Q5VY30</td>
<td>Plasma retinol-binding protein(1-182)</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Q9UHV9</td>
<td>Prefoldin subunit 2 OS</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>B4DY59</td>
<td>cDNA FLJ56267, highly similar to Transducin beta-like 2 protein</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>O75489</td>
<td>[ubiquinone] iron-sulfur protein 3, mitochondrial OS</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Q13136-2</td>
<td>Isoform 2 of Liprin-alpha-1 OS</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>P02748</td>
<td>Complement component C9 OS</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Accession</td>
<td>ID</td>
<td>Description</td>
</tr>
<tr>
<td>-----------</td>
<td>----</td>
<td>-------------</td>
</tr>
<tr>
<td>Q5T6W8</td>
<td>&gt;tr</td>
<td>Q5T6W8</td>
</tr>
<tr>
<td>P10619</td>
<td>&gt;sp</td>
<td>P10619</td>
</tr>
<tr>
<td>Q6NUR7</td>
<td>&gt;tr</td>
<td>Q6NUR7</td>
</tr>
<tr>
<td>Q96FQ6</td>
<td>&gt;sp</td>
<td>Q96FQ6</td>
</tr>
<tr>
<td>P62072</td>
<td>&gt;sp</td>
<td>P62072</td>
</tr>
<tr>
<td>P23142-3</td>
<td>&gt;sp</td>
<td>P23142-3</td>
</tr>
<tr>
<td>O60831</td>
<td>&gt;sp</td>
<td>O60831</td>
</tr>
<tr>
<td>P13796</td>
<td>&gt;sp</td>
<td>P13796</td>
</tr>
<tr>
<td>A8K6S3</td>
<td>&gt;tr</td>
<td>A8K6S3</td>
</tr>
<tr>
<td>D9YZV3</td>
<td>&gt;tr</td>
<td>D9YZV3</td>
</tr>
<tr>
<td>P68871</td>
<td>&gt;sp</td>
<td>P68871</td>
</tr>
<tr>
<td>B4DDL4</td>
<td>&gt;tr</td>
<td>B4DDL4</td>
</tr>
<tr>
<td>P02533</td>
<td>&gt;sp</td>
<td>P02533</td>
</tr>
<tr>
<td>Q5U086</td>
<td>&gt;tr</td>
<td>Q5U086</td>
</tr>
<tr>
<td>Q5VZ89</td>
<td>&gt;sp</td>
<td>Q5VZ89</td>
</tr>
<tr>
<td>E7ERZ4</td>
<td>&gt;tr</td>
<td>E7ERZ4</td>
</tr>
<tr>
<td>P04792</td>
<td>&gt;sp</td>
<td>P04792</td>
</tr>
<tr>
<td>B4DMX4</td>
<td>&gt;tr</td>
<td>B4DMX4</td>
</tr>
<tr>
<td>Q53FR6</td>
<td>&gt;tr</td>
<td>Q53FR6</td>
</tr>
<tr>
<td>Q9H4M9</td>
<td>&gt;sp</td>
<td>Q9H4M9</td>
</tr>
<tr>
<td>K7EMS3</td>
<td>&gt;tr</td>
<td>K7EMS3</td>
</tr>
<tr>
<td>Q9XB4</td>
<td>&gt;sp</td>
<td>Q9XB4</td>
</tr>
<tr>
<td>Accession</td>
<td>Description</td>
<td>Similarity Score</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>P14735</td>
<td>&gt;sp</td>
<td>P14735</td>
</tr>
<tr>
<td>Q06033-2</td>
<td>&gt;sp</td>
<td>Q06033-2</td>
</tr>
<tr>
<td>F8VZY9</td>
<td>&gt;tr</td>
<td>F8VZY9</td>
</tr>
<tr>
<td>P13647</td>
<td>&gt;sp</td>
<td>P13647</td>
</tr>
<tr>
<td>C9JQ45</td>
<td>&gt;tr</td>
<td>C9JQ45</td>
</tr>
<tr>
<td>P43490</td>
<td>&gt;sp</td>
<td>P43490</td>
</tr>
<tr>
<td>Q5TZZ9</td>
<td>&gt;tr</td>
<td>Q5TZZ9</td>
</tr>
<tr>
<td>F5GWP8</td>
<td>&gt;tr</td>
<td>F5GWP8</td>
</tr>
<tr>
<td>Q5D862</td>
<td>&gt;sp</td>
<td>Q5D862</td>
</tr>
<tr>
<td>Q8N1N4</td>
<td>&gt;sp</td>
<td>Q8N1N4</td>
</tr>
<tr>
<td>Q01469</td>
<td>&gt;sp</td>
<td>Q01469</td>
</tr>
<tr>
<td>P08779</td>
<td>&gt;sp</td>
<td>P08779</td>
</tr>
<tr>
<td>Q02413</td>
<td>&gt;sp</td>
<td>Q02413</td>
</tr>
<tr>
<td>Q5T749</td>
<td>&gt;sp</td>
<td>Q5T749</td>
</tr>
<tr>
<td>P15924</td>
<td>&gt;sp</td>
<td>P15924</td>
</tr>
<tr>
<td>P22531</td>
<td>&gt;sp</td>
<td>P22531</td>
</tr>
<tr>
<td>P31947</td>
<td>&gt;sp</td>
<td>P31947</td>
</tr>
<tr>
<td>Q9NTJ3-2</td>
<td>&gt;sp</td>
<td>Q9NTJ3-2</td>
</tr>
<tr>
<td>B4DRR0</td>
<td>&gt;tr</td>
<td>B4DRR0</td>
</tr>
<tr>
<td>Q92783-2</td>
<td>&gt;sp</td>
<td>Q92783-2</td>
</tr>
</tbody>
</table>
Chapter 4

Quantitative LC-MS Measurements on Mouse and Monkey hepcidin Using QQQ and Q Exactive Mass Spectrometry

* A paper based on this chapter is in press, in collaboration with Alnylam Pharmaceutical.

Co-authors’ contribution to this work: Shiaw-Lin Wu, idea contribution, experimental design, data analysis, manuscript writing and revision; Alnylam Pharmaceuticals, siRNA drug development, PD experiments; Barry L. Karger, idea contribution, manuscript writing and revision and grant support.
Abstract

This research describes the development of a fast and high throughput method with precise hepcidin extraction using solid phase extraction on an automated liquid handling platform. For quantitation, an LC-MS method, using either a QQQ or Q-Exactive mass spectrometer, was developed to measure hepcidin from monkey and mouse serum. The method was applied to determine the level of reduction of hepcidin in the sera of monkey and mice resulting from treatment with an siRNA drug that effectively bound the mRNA for production of the peptide hormone. SRM with QQQ method was able to quantitatively measure hepcidin in monkey serum in the range of 1 to 250 ng/mL with the limit of detection (LOD) at 0.5 ng/mL (S/N ≥ 3) and in mouse serum over the range from 0.5 to 250 ng/mL with the LOD at 0.2 ng/mL (S/N ≥ 3). Sensitivity and specificity were further improved using SRM with Q-Exactive method to the LOD at 0.1 ng/mL in monkey serum and LOD at 0.05 ng/mL in mouse serum. A more than 10 fold increase in sensitivity and specificity using the Q-Exactive method was achieved. With the improvement, less serum consumption per analysis (e.g. from 250 µL to 25 µL) could be required for preclinical studies in the future, and that should reduce the stress on small size animals (e.g. mouse).
4.1 Introduction

Certain chronic medical conditions, such as infectious or inflammatory diseases or cancer, can lead to a drop of red cell production and hemoglobin in spite of normal iron stores in macrophages\(^1\). The medical term is called anemia of chronic disease (ACD). Abnormal high levels of hepcidin can significantly cause reduction in the plasma iron concentration, leading to iron deficiency and thus anemia\(^2,3\). Treatment of ACD with iron supplements have limited effectiveness because of the over-expressed levels of hepcidin, which inhibit the activity of ferroportin, thus reducing the exportation of iron into plasma\(^4\).

Current therapies for ACD focused on the hepcidin-ferroportin interaction and related signaling pathways, such as anti-hepcidin antibodies and pathway inhibitors\(^5\). However, development of an effective monoclonal antibody is still challenging to obtain high specificity and affinity against hepcidin, a small peptide hormone. Alternatively, RNA interference (RNAi) has the potential to provide a useful approach for specific silencing of the hepcidin gene\(^6\). Briefly, a double-strand RNA molecule containing a 21-25 nucleotide sequence mRNA (small interfering RNA, [siRNA]) can be synthesized to degrade a target endogenous mRNA by base-pairing\(^6,7\). P siRNA against the mRNA for hepcidin thus has the potential to reduce the expression of the peptide hormone, leading to a potential treatment of ACD\(^8\).

To advance hepcidin-siRNA towards clinical studies, it is important to measure the pharmacological properties of the drug. In additional to studies of the drug with dose level, the effect of the drug target in blood (before and after different doses of hepcidin-siRNA drug) also needs to be determined in order to understand the pharmacokinetic (PK)/pharmacodynamic (PD) behavior\(^9\). For measurement of the drug target, quantitative detection of hepcidin using
enzyme-linked immunosorbent assays (ELISA) have been developed with the advantages of easy-use and high throughput\textsuperscript{10,11}. However, quantitative ELISA is often limited by cross-reactivity with pro-hepcidin or other hepcidin related forms, in addition to requiring potentially different ELISA formats for different animal species (e.g. hepcidin exhibits slightly different sequence variations in mouse, monkey, rabbit, and human, see discussion). On the other hand, mass spectrometry based quantitative approaches provide high specificity by differentiating the sequence variations with precise precursor ion and/or fragment ion selections of hepcidin from other interferences\textsuperscript{12}. Recently SELDI-TOF and MALDI-TOF assays have been published for quantitation of hepcidin in serum and urine with a limit of detection at 5-10 ng/mL based on the precursor ion selection\textsuperscript{13,14,15}. To further increase the sensitivity and improve the specificity, LC-MS using triple quadrupole mass spectrometry (QQQ) with selected reaction monitoring (SRM) has been developed (based on both the precursor ion and fragment ion selections) to quantitatively measure hepcidin in a model system (i.e. spike-in hepcidin in a complex serum matrix) with high sensitivity\textsuperscript{16,17}.

In this study, an SRM approach (using QQQ) has been developed to quantitate the endogenous levels of hepcidin in 40 monkey and 40 mice using isotopically labeled hepcidin as internal standard. Once we established the normal endogenous level of hepcidin in both species, we then compared hepcidin levels after the treatment with three different dosages of the hepcidin-siRNA drug in the monkey model. The dose effect of the drug was determined by measurement of hepcidin level at different time intervals before and after injection of the drug. With this approach, the effectiveness of the RNAi drug was indeed observed at a specific level of dosage by reduction in the level of hepcidin in these animals. The proper dosage and duration effect of the drug \textit{in vivo} can be determined using this method.
There are, however, problems with the approach, particularly in the analysis of limited sample amounts as would exist for mice. The hepcidin peptide hormone consists of 25 amino acids. A product ion with multiple charge states is likely to be selected for compound specificity and quantitation. However, for the current QQQ mass platform, a small product ion (m/z 120 derived from the cleavage of a single phenylalanine residue) was used because the large m/z with multiple charges is not suitable for QQQ (low sensitivity with inconsistent signals). However, this small m/z (immonium ion of Phe) could suffer specificity or lose sensitivity with high background noise (e.g. any peptide with Phe in serum matrix eluted at the similar retention time could contribute the noise). An LC-MS system with the capability to observe high m/z with high resolution to gain specificity is desirable.

We thus developed another mass spectrometry platform to quantitate hepcidin based on a quadrupole-Orbitrap hybrid mass spectrometry (Q-Exactive mass spectrometer)\(^\text{18}\). This instrument is capable of detecting the entire MS/MS spectrum rapidly with high mass accuracy.\(^\text{19}\). The Q-Exactive has been used in targeted proteomic quantitation and PK/PD study for small molecules\(^\text{20,21,22,23}\). Not only an accurate precursor can be selected but also product ions with high resolution and accuracy in the entire MS/MS spectrum can be obtained\(^\text{24}\). Since there is no need to select a particular product ion as used in SRM-QQQ mass spectrometer, the entire product ions are automatically acquired in Q-Exactive and any desired product ion can be selected after the data collection. Thus, there is no need to pre-define a particular product ion before analysis, which saves the development time in mass spectrometry.

The development of the current SRM with the QQQ platform is time-consuming, and the sensitivity may not be sufficient to study the drug effect in a small animal such as a mouse.
Specifically, a choice of a peptide product ion (with the desired charge state) could be problematic because the original chosen product ion for a particular m/z may be shifted to a different charge state or suppressed with the compound containing sample matrix at different concentration levels. For a larger peptide with multiple charge states, this problem could be even more pronounced. As for hepcidin, high charge precursor ions and large product ions were selected for quantitation with Q-Exactive mass spectrometer. Using this new platform, hepcidin was detected with much higher specificity and sensitivity (significant reduced background noise) than the SRM-QQQ platform (at least 10 x improvements). We can anticipate that the method can be extended to PK/PD studies for other large peptides or even intact protein targets (e.g. without the need for enzymatic digestion as often required in QQQ measurements).
4.2 Materials and methods

4.2.1 Sample

Mouse and monkey serum samples, with and without the drug treatment for PK/PD study, were supplied by Alnylam Pharmaceuticals (Cambridge, MA). Wild-type hepcidin of mouse and monkey along with their isotopic labeled counterparts (labeled with $^{13}$C$_9$ and $^{15}$N$_1$ of Phe$_9$ amino acid at the hepcidin sequence) were synthesized by Peptide Institute Inc. (Osaka, Japan). Stable isotopic labeled mouse/monkey hepcidin had +10 dalton mass as compared to wild-type mouse/monkey hepcidin.

4.2.2 Reagents

HPLC grade water, acetonitrile and methanol were purchased from EMD Chemicals (Darmstadt, Germany). LC/MS grade formic acid (FA) was from Fisher Chemical (San Jose, CA). Ammonium hydroxide and standard rabbit serum were obtained from Sigma Aldrich (St. Louis, MO). RP, S AM Bravo Cartridge was supplied by Agilent Technologies (Santa Clara, CA).

4.2.3 Preparation of calibration standards and quality control (QC) samples

0.1 mg mouse/monkey hepcidin or isotopically labeled hepcidin were dissolved in 1 mL 50% methanol/0.1% FA to prepare 100 μg/mL stock solution. The stock solution was aliquoted and stored at -80 °C prior to use. Serially diluted calibration standards were prepared from the stock solution. Preparation of all serum samples were performed on ice. A 20 μL hepcidin stock solution of 100 μg/mL standard mouse/monkey was added into 980 μL rabbit serum to prepare 2000 ng/mL mouse/monkey hepcidin in rabbit serum. Linear standards of mouse or monkey
hepcidin serum of 0.05, 0.1, 0.2, 0.5, 1, 2.5, 5, 10, 25, 50, 100 and 250 ng/mL were serially
diluted from the 2000 ng/mL mouse/monkey hepcidin with rabbit serum.

Monkey hepcidin QC samples were prepared in 10, 50 and 100 ng/mL, and mouse hepcidin
QC samples were prepared in 10, 25 and 60 ng/mL in rabbit serum.

4.2.4 Extraction hepcidin from serum samples

Protein purification and extraction were automatically performed using the Agilent Bravo
Automated Liquid Handling Platform G5409A and RP, S AM Bravo Cartridge (Agilent
Technologies, Santa Clara, CA). Purification and extraction procedure was developed based on
the method developed by Li et al. Briefly, 200 μL of each mouse/monkey serum sample was
added into a 500 μL 96-well plate containing 50 μL of 50 ng/mL mouse/monkey internal
standard and 100 μL of water/0.1% FA. Samples were gently vortex mixed and centrifuged at
3000 rpm for 2 min. By computer programming, the RP cartridge plate was conditioned by 200
μL of methanol and 200 μL of water. 350 μL of each mixed sample was transferred to the PR
cartridge and drawn through the cartridge at the speed 50 μl/min. To remove interferences, 200
μL of 5% methanol, 200 μL of 30% methanol/65% water/5% NH₄OH and 200 μL water were
sequentially used to wash the cartridge. Hepcidin was eluted with 75 μL of 90% methanol/0.1%
FA into a collection 96 well plate. 60 μL of water with 0.1% FA was added to each well to dilute
the sample to 50% methanol. The samples were vortex mixed and centrifuged. Finally, 15 μL of
each sample was injected for LC/MS analysis.
4.2.5 LC-MS for Q-Exactive study

The extracted samples (15 µL) were injected and separated by an Agilent Polaris C18-A column (2.0×50.0mm, 5 µm) at a flow rate of 200 µL/min by a Dionex 3500RS UHPLC system (Thermo Fisher Sci, Thousands Oaks, CA). Each sample was run three times. Mobile phase A was 0.1% FA in water, and mobile phase B was 0.1% FA in acetonitrile. The gradient was 0 to 2 min, maintained at 10% B; 2 min to 8 min, 10% B to 60% B; 8 min to 10 min, maintained at 90% B; 10 min to 11 min, 90% B to 10% B and 11 min to 15 min, maintained at 10% B. The column was kept in the oven compartment at 40 °C during the separation. Targeted MS/MS detection was acquired by Q-Exactive mass spectrometer (Thermo Fisher Sci, San Jose, CA) with an H-ESI II ion source. Ion source parameters were optimized as vaporizer temp. 350°C, capillary temp. 350°C, sheath gas 35, Aux gas 10, sweep gas 0 and spray voltage 3.8 kV. Targeted MS parameters were optimized at resolution 35,000, AGC target 2e5, max IT 200 ms, MSX count 1, isolation window 2.0 m/z, fixed first mass 100.0 m/z. The precursor ion selected for monkey hepcidin was [M+5H]^{5+} ion with m/z 564.414. The product ion y_{19}^{3+} (m/z 703.237) was selected for quantitation and b_3^{+} (m/z 354.139), b_4^{+} (m/z 501.208) ions were selected for validation. The transition selected for quantification of +10 Da labeled monkey hepcidin was m/z 566.417 → 703.236. The normalized HCD collision energy optimized for monkey hepcidin was 22. The precursor ion selected for mouse hepcidin was [M+3H]^{3+} ion with m/z 918.681. The product ion y_{21}^{2+} (m/z 1138.920) was selected for quantitation, and b_3^{+} (m/z 331.123), b_4^{+} (m/z 478.191) ions were selected for validation. Transition selected for internal standard +10 Da labeled monkey hepcidin was m/z 922.026→ 1138.916. The normalized HCD collision energy optimized for mouse hepcidin was 20.
4.2.6 LC-MS for QQQ study

The HPLC system was an Agilent 1290 Infinity LC system (Agilent Technologies, Santa Clara, CA) consisting of Infinity binary pump, autosampler, thermostated column compartment. The extracted sample (15 µL) was injected and separated by an Agilent Polaris C18-A column (2.0×50.0mm, 5 µm) at the flow rate of 200 µL/min. Each sample was run three times. The mobile phase A was 0.1% FA in water, and mobile phase B was 0.1% FA in acetonitrile. The gradient was 0 to 1.5 min, 15% B to 15% B; 1.5 min to 5 min, 15% B to 90% B; 5 min to 7 min, 90% B to 90% B; 7 min to 7.2 min, 90% B to 15% B and 7.2 min to 12 min, 15% B to 15% B. The column was kept in the oven at 40 ºC. MS/MS detection was acquired by an Agilent 6460 triple-quadrupole mass spectrometer with Agilent Jet stream technology. Agilent Masshunter Workstation Data Acquisition software was used to collect and process the data. Jet steam parameters were optimized as gas temperature: 300 ºC, gas flow: 5 L/min, nebulizer pressure: 45 psi, sheath gas temperature: 350 ºC, sheath gas flow: 11 L/min, capillary voltage: 3500 V and nozzle voltage: 500 V. The SRM transitions selected for quantification of monkey hepcidin were m/z 564.3 → 120.0 and m/z 566.2 → 120.0 for isotopically labeled monkey hepcidin (internal standard). Two SRM transitions m/z 564.3 → 110.0, m/z 564.3 → 353.7 were used for validation. The SRM transition selected for quantification of mouse hepcidin was m/z 918.7 → 1138.8 and isotopically labeled hepcidin (internal standard) was m/z 921.9 → 1138.8. The other two SRM transitions m/z 918.7 → 331.0, m/z 918.7 → 478.2 were used as validations for mouse hepcidin. SRM acquisition parameters were Delta EMV (+) value: 400, dwell time: 50 ms, fragmentor: 135 V, collision energy: 24 eV, polarity: positive, MS1 and MS2 resolutions: unit.
4.2.7 Data process for Q-Exactive method

Peaks are integrated by the Genesis algorithm function in the Thermo Xcalibur Quan Browser, and 10 ppm mass tolerance was applied for the extraction of target product ions. The relative amount of hepcidin was quantified by the peak area ratio between hepcidin and the internal standard. Blank and serial calibration standards from 0.1 to 250 ng/mL for monkey hepcidin and from 0.05 to 250 ng/mL mouse hepcidin were analyzed by Quan Browser. The calibration curve was generated by the software using linear regression and $1/x^2$ weight calibration curve setup. The regression equation from the calibration curve was used to back-calculate the concentration of each QC sample.

4.2.8 Data process for QQQ method

Peaks were integrated by the Agilent QQQ Quantitative Analysis software (Masshunter Quan software). The relative amount of hepcidin was quantified by the peak area ratio between hepcidin and the internal standard. Blank serum and serial dilutions from 0.2 to 250 ng/mL mouse/monkey hepcidin in rabbit serum were analyzed by Agilent’s Masshunter Quan software. The measured value from the blank was treated as background noise, and thus the value was subtracted from other sample measurements. The calibration curve was generated by using linear regression and $1/x^2$ weight calibration curve setup. The limit of quantitation (LOQ) was defined as the lowest concentration that could be measured with a precision better than 20% CV and accuracy within ± 20% of the theoretical concentration. The regression equation from the calibration curve without the intercept was used to back-calculate the concentration of each PK/PD sample.
4.3 Results and Discussion

In this section, methods for hepcidin extraction and quantitation from serum are described, followed on the measurements of hepcidin in monkeys and mice as well as the siRNA drug effect in monkey model study. An improved method for hepcidin quantitation using the newly developed Q-Exactive platform is described in the last section.

4.3.1 Development of hepcidin extraction from serum

Hepcidin was found to be a challenging peptide, which could bind to hydrophobic surface including tubing, containers, and plasma proteins\(^25\). For control of the recovery in the sample preparation, we added isotopically labeled hepcidin to the sample matrix at the beginning of sample preparation to account for any potential losses in subsequent steps (see the details in the method section). As for extraction of hepcidin, there were several reports discussing the extraction efficiency from serum or plasma samples. For example, a protein precipitation approach using trichloroacetic acid (TCA) recovered 34% hepcidin in human plasma\(^26\). An enrichment method based on weak cation exchange beads resulted to 75% recovery but required 1.5 hr for the enrichment process\(^27\). Li et al. published a high throughput method using a 96-well \(\mu\)Elution SPE plate with recovery of more than 60% hepcidin\(^16\). In our approach, we first used the extraction based on Li’s method and then modified it with automated procedures. As described in the methods section, the sample transfer and extraction of hepcidin by buffer addition and elution on 96-well RP cartridges were automatically performed by an Agilent Bravo Automated Liquid Handling Platform. The platform provided speed and precision for hepcidin extraction from large numbers of serum samples.
4.3.2 Development of LC-MS method

Measurement of hepcidin based on HPLC-chip/MS was reported to increase the detection sensitivity from 1-2 ng/mL to 0.25 ng/mL in human plasma sample\textsuperscript{28}. However, we found that the use of HPLC-chip or nano-column was not robust for the PK/PD study because the chips or nano-columns were frequently clogged from the serum samples, and the turnaround per analysis was slow due to the low flow (0.2 µL/min). To increase the robustness and speed, we adopted a large dimension column (2.0 mm i.d x 10 cm length) with high flow (200 µL/min) for the method. Loss of sensitivity on high flow was compensated by injection of larger sample volumes (10x more). Fast turnaround was achieved in 15 minutes (reduced from 60 minutes) for each sample analysis. Total analysis time included sample injection, gradient separation, MS detection, and column reequilibration.

Optimizations were required for hepcidin in different species because of their slight variation in sequence (Fig. 4.1). The mouse sequence contains only two charged lysine (Lys or K) residues, and 3+ charge state is thus the dominant ion in acid condition (Fig. 4.3A). In human and monkey sequences, more amino acids are replaced with charged histidine (His or H), lysine and arginine (Arg or R) (Fig. 4.1), resulting in a dominate 4+ charge state for human hepcidin (data not shown) and 5+ charge state for monkey hepcidin (Fig. 4.2A).

|       | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
|-------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Mouse | D | T | N | F | P | I | C | I | F | C | C | K | C | C | N | N | S | Q | C | G | I | C | C | K | T |
| Human | D | T | H | F | P | I | C | I | F | C | C | G | C | C | H | R | S | K | C | G | M | C | C | K | T |
| Monkey| D | T | H | F | P | I | C | I | F | C | C | G | C | C | H | R | S | K | C | G | M | C | C | R | T |

Figure 4.1. Differences in the hepcidin sequences between mouse, human and monkey species.

183
For the MS analysis, SRM using QQQ mass spectrometer was used to select m/z 564.3 (5+) as the precursor ion (Figure 4.2A) and m/z 120 (1+) as the product ion (Fig. 4.2B) for detection of monkey hepcidin. The fragmentation of the large peptide with high charge resulted in product ions with low abundance and also inconsistent signals for large b and y ions. As shown in Figure 4.2B, a small m/z at 120.0 (1+), corresponding to immounium ion of phenylalanine (Phe or F), could be found with sufficient signal and consistency. After optimization of the collision energy (to 24 eV), the SRM transitions of m/z 564.3 → 120.0 and m/z 566.2 → 120.0 were selected for measurement of the wild type and isotopically labeled monkey hepcidin, respectively. Using the same approach for mouse hepcidin (see Fig. 4.3) an optimized collision energy (at 24 eV) and SRM transitions of m/z 912.9 → 1138.8 and m/z 921.9 → 1138.8 were selected for measurement of the wild type and isotopically labeled hepcidin.

4.3.3 Establishment of linear curves, LOD and LOQ

In the SRM measurement for monkey hepcidin, the limit of detection (LOD) was observed at 0.5 ng/mL, with a signal to noise ratio (S/N = 3), as illustrated in Figure 4.4B. Although background noise at the same retention time was observed in the blank serum, as illustrated in Figure 4.4A, this noise was subtracted from each level of hepcidin measurement. A linear curve was obtained from samples at 0.5 ng/mL to 250 ng/mL concentration with a correlation coefficient \( r^2 \) of 0.9996, as seen in Figure 4.5. The limit of quantitation (LOQ) was determined at 1 ng/mL, defined as the lowest concentration with a mean accuracy (RE) to be less than 20% (Table 4.1). The coefficient of variation (CV) was between 0.2% and 14.2% in samples measured from all the range of concentrations (Table 4.1). Similarly, the LOD of mouse
hepcidin could be obtained as low as 0.2 ng/mL, with LOQ at 0.5 ng/mL. The linear curve was obtained from mouse standards at 0.2 ng/mL to 250 ng/mL with a correlation coefficient $r^2$ of 0.9972, as shown in Figure 4.6, and precision between 2% and 14% over all.

Figure 4.2. SRM method development using QQQ for detection of monkey hepcidin. (A) precursor ion scan, (B) product ion scan at three levels of collision energy.
Figure 4.3. SRM method development using QQQ for detection of mouse hepcidin. (A) precursor ion scan, (B) product ion scan at three levels of collision energy.
Figure 4.4. Using optimized SRM method developed in QQQ as m/z 564.3 (precursor ion) and m/z 120.0 (product ion) at 24 eV collision energy for detection of monkey hepcidin at low levels. (A) blank serum, (B) at 0.5 ng/mL, and (C) at 50 ng/mL.
Figure 4.5. Linear regression curve of monkey hepcidin standards measured from 0.5 ng/mL to 250 ng/mL by the optimized SRM method described in Figure 4.2.

Figure 4.6. Linear calibration curve for mouse standards in rabbit blood at 0.2 ng/mL to 250 ng/mL concentration measured using QQQ.
Table 4.1. Precision and accuracy of SRM using QQQ for quantitation of monkey hepcidin standards.

<table>
<thead>
<tr>
<th>Monkey hepcidin concentration (ng/mL)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theoretical conc.</td>
<td>0.5</td>
<td>1</td>
<td>2.5</td>
<td>5</td>
<td>10</td>
<td>25</td>
<td>50</td>
<td>100</td>
<td>250</td>
</tr>
<tr>
<td>Final mean conc. (n=3)</td>
<td>0.66</td>
<td>0.93</td>
<td>2.09</td>
<td>4.46</td>
<td>9.26</td>
<td>23.57</td>
<td>48.14</td>
<td>98.09</td>
<td>251.32</td>
</tr>
<tr>
<td>Precision % CV</td>
<td>14.2%</td>
<td>3.5%</td>
<td>0.2%</td>
<td>5.2%</td>
<td>5.9%</td>
<td>1.7%</td>
<td>1.5%</td>
<td>0.5%</td>
<td>1.1%</td>
</tr>
<tr>
<td>Accuracy % RE</td>
<td>31.7%</td>
<td>7.2%</td>
<td>16.2%</td>
<td>10.8%</td>
<td>7.4%</td>
<td>5.7%</td>
<td>3.7%</td>
<td>1.9%</td>
<td>0.5%</td>
</tr>
</tbody>
</table>

1. Standards from 0.5 to 250 ng/mL were spiked into serum (labeled as Theoretical conc).
2. Observed concentrations were the average of the three repeat measurements (labeled as Final mean conc).
3. Precision or percent of coefficient variation were derived from the variation of the three repeat measurements (labeled as % CV).
4. Accuracy or percent of relative accuracy were derived from the variation of the observed concentration vs. the theoretical concentration (labeled as %RE).
Table 4.2. Precision and accuracy of SRM using QQQ for quantitation of mouse hepcidin standards.

<table>
<thead>
<tr>
<th>Mouse hepcidin concentration (ng/mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Theoretical Conc.</td>
</tr>
<tr>
<td>Final mean conc. (n=3)</td>
</tr>
<tr>
<td>Precision % CV</td>
</tr>
<tr>
<td>Accuracy % RE</td>
</tr>
</tbody>
</table>

Figure 4.7. Endogenous hepcidin levels in 40 cynomolgus monkeys and 40 mice (measured by the optimized SRM method described in Figure 4.2 and Figure 4.3).
4.3.4 Detection of endogenous hepcidin in monkey and mouse species

Endogenous hepcidin levels in 40 cynomolgus monkeys and 40 mice were studied using the above SRM with the QQQ approach. Approximately 250 µL of serum was obtained from each animal. Distribution of hepcidin levels in 40 monkeys and 40 mice is illustrated in Figure 4.7. The average of hepcidin level was 50.0 ng/mL in 40 monkeys. The average of hepcidin concentration was 46 ng/mL among 40 mice with 75% of the population. Only 3 mice exhibited a relatively high hepcidin level (> 76 ng/mL). It should be noted that although the average concentration of hepcidin is similar between the monkey and mouse species, the requirement of blood volume of 250 µL for one measurement prevented the study of the drug dosage effect from the mouse model. The need to draw the same volume of blood multiple times (for different time intervals) would be a significant stress for such a small animal model. As mentioned in the Introduction, a more sensitive method (using less material) is needed, and that will be discussed later in the section using the Q-Exactive mass spectrometer. Nevertheless, using the SRM with the QQQ approach would be sufficient to study the drug dose effect in monkey.

4.3.5 Determination of the siRNA drug dosage

Each of low, medium and high doses of siRNA drugs were used to study the dose effect, with each dose repeated three times (applying each dose to three monkeys). For each monkey, blood was drawn by Alnylam on 9, 6, and 3 days before dosing as well as 24 and 48 hr after dosing. The baseline level of hepcidin for each monkey was determined from the samples drawn before dosing (averaging the hepcidin levels at 9, 6, and 3 days). Thus, the inhibition rate (reduction of hepcidin) could be determined by the difference between the baseline and the 24 or 48 hr
measurement divided by the baseline level (calculated as % inhibition). In the dose study, the high dose of the siRNA drug was observed to effectively reduce the level of hepcidin in monkey, achieving inhibition rates of 87.4% after 24 hr and 74.0% after 48 hr, as illustrated in Figure 4.8 and Table 4.2. However, medium and low doses represented inhibition rates of only 32.0% and 25.4% after 48 hr, respectively. The high dose of the siRNA drug was chosen for further study.

Figure 4.8. Effect of hepcidin reduction by three different doses of siRNA drug measured at 0, 24hr, and 48 hr after injection of the drugs (measured by the optimized SRM method described in Figure 4.2).
Table 4.3. Dose Effect of siRNA drug in reduction of hepcidin in monkey model

<table>
<thead>
<tr>
<th>siRNA dose level</th>
<th>Animal ID</th>
<th>Hepcidin level (ng/mL)</th>
<th>-9 days</th>
<th>-6 days</th>
<th>-3 days</th>
<th>24 hr</th>
<th>48 hr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>1001</td>
<td></td>
<td>13.79</td>
<td>13.10</td>
<td>17.57</td>
<td>10.01</td>
<td>10.67</td>
</tr>
<tr>
<td></td>
<td>1002</td>
<td></td>
<td>15.83</td>
<td>18.12</td>
<td>20.12</td>
<td>13.38</td>
<td>14.72</td>
</tr>
<tr>
<td></td>
<td>1003</td>
<td></td>
<td>9.23</td>
<td>24.62</td>
<td>15.77</td>
<td>13.54</td>
<td>14.93</td>
</tr>
<tr>
<td>Medium</td>
<td>2001</td>
<td></td>
<td>4.86</td>
<td>10.19</td>
<td>6.95</td>
<td>8.61</td>
<td>10.29</td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td></td>
<td>11.10</td>
<td>11.36</td>
<td>16.73</td>
<td>3.36</td>
<td>5.59</td>
</tr>
<tr>
<td></td>
<td>2003</td>
<td></td>
<td>19.33</td>
<td>12.16</td>
<td>13.12</td>
<td>9.06</td>
<td>8.25</td>
</tr>
<tr>
<td>High</td>
<td>3001</td>
<td></td>
<td>4.76</td>
<td>14.68</td>
<td>9.16</td>
<td>1.72</td>
<td>5.66</td>
</tr>
<tr>
<td></td>
<td>3002</td>
<td></td>
<td>26.40</td>
<td>37.46</td>
<td>20.29</td>
<td>5.56</td>
<td>4.61</td>
</tr>
<tr>
<td></td>
<td>3003</td>
<td></td>
<td>5.78</td>
<td>9.00</td>
<td>7.51</td>
<td>0.00</td>
<td>0.16</td>
</tr>
</tbody>
</table>

1. Three dose levels (labeled as low, medium and high) and three monkeys per dose (such as 1001, 1002, and 1003 for the low dose) were used for the time course study.

2. Hepcidin levels were measured before (at -9, -6, and -3 days) and after (at 24 and 48 hr) the drug injection.
4.3.6 Consistency of siRNA drug effect

The high dose of the siRNA drug was repeatedly prepared (as three different lots). These three different batches of siRNA drugs were injected into three different monkeys to test the consistency. A non-siRNA drug labeled as “PBS” buffer was also injected into another monkey as a side-by-side control. Hepcidin levels were determined on 9, 6, 3 days before dosing, with the average of the three measurements as the baseline level. The siRNA drugs exhibited consistent efficacy with a > 60% inhibition rate in all three animals after 2 days (48 hr), and remained >40% after 14 days of dosing, as illustrated in Figure 4.9. In contrast, the control group (injection with PBS buffer) showed that the hepcidin levels were varied with time (no correlation with the time as shown in the purple line of Figure 4.9). The statistical difference between the three siRNA administrations and the PBS control was evaluated using Student’s t-test. P values were 0.031, 0.038 and 0.049 for the three drugs (labeled as 48141, 52590 and 51707), respectively. All the p values were < 0.05 or > 95% confidence. The significance of the statistical difference for all the three lots demonstrated the consistency of the drug effect.
Figure 4.9. Effect of hepcidin reduction by three different batches of siRNA drugs, and a non-siRNA drug measured at 1, 2, 4, 7, 10, and 14 days after injection of the drugs (measured by the optimized SRM method described in Figure 4.2).
4.3.7 Development of LC-MS method using Q-Exactive

The Q-Exactive mass spectrometer was purchased after the monkey PK/PD studies. To study this new mass spectrometer in our lab, we first used monkey hepcidin for evaluation. For the Q-Exactive, similar to the QQQ, m/z 564.414 (5+) of monkey hepcidin had the highest intensity and thus is used as the precursor ion in the SRM method (Figure 4.10A). However, a major difference was found in the selection of the production ions. As described in the introduction section, the entire product ion spectrum is automatically acquired in Q-Exactive with high mass accuracy and fast speed. Thus, any of the fragment ions, such as Phe immounium ion, $b_3^+$, $b_4^+$, $y_{19}^{3+}$, $y_{21}^{2+}$, and $y_{22}^{2+}$ ions could all be used as the product ion for quantitation (see Figure 4.10B). After optimization with collision energy (defined as NCE at 20%), m/z 564.414 (precursor ion) and m/z 703.236 (product ion) ($y_{19}^{3+}$) were selected for quantitation of monkey hepcidin. Instead of using a small m/z (Phe immounium ion), a large product ion ($y_{19}^{3+}$) with consistent high intensity was used in the Q-Exactive, providing much higher specificity than the small m/z used in QQQ. Similar optimization was performed for mouse hepcidin (Figure 4.11).
Figure 4.10. SRM method development using Q-Exactive for monkey hepcidin detection. (A) precursor ion scan, (B) product ion scan at three levels of collision energy, and (C) Detection of monkey hepcidin using the optimized SRM parameters as m/z 564.4139 (precursor ion) and m/z 703.2360 (product ion) at 22 normalized collision energy.
Figure 4.11. SRM method development using Q-Exactive for mouse hepcidin detection. (A) precursor ion scan, (B) product ion scan at three levels of collision energy, and (C) Detection of monkey hepcidin using the optimized SRM parameters as m/z 918.6814 (precursor ion) and m/z 1138.9192 (product ion) at 20 normalized collision energy.
We then used the same QC samples as for QQQ (rabbit serum spiked in different levels of monkey hepcidin) to study the linearity, limit of detection (LOD), and limit of quantitation (LOQ). The performance of Q-Exactive measurement is illustrated in Table 4.4. Precision of measurements was < 20% in all the samples measured from 0.1 ng/mL to 250 ng/mL concentrations, and especially < 2% when the concentration of hepcidin was > 1 ng/mL. Accuracy of the methods was from 0.08% to 8.6 % (Table 4.4) with excellent correlation coefficient ($r^2 = 0.9986$) from 0.1 ng/mL to 250 ng/mL, as illustrated in the Figure 4.12. In addition, we used samples from one of the previous PK/PD studies for validation. Specifically, monkey hepcidin concentration identical to the PK/PD results of siRNA 52950 drug were prepared according to the time course study. A similar inhibitory trend could also be obtained as shown in Figure 4.14 and Table 4.5.

Table 4.4. Precision and accuracy of SRM using Q-Exactive for quantitation of monkey hepcidin standards.

<table>
<thead>
<tr>
<th>Monkey hepcidin concentration (ng/mL)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theoretical Conc.</td>
<td>0.1</td>
<td>0.2</td>
<td>0.5</td>
<td>1</td>
<td>2.5</td>
<td>5</td>
<td>10</td>
<td>25</td>
<td>50</td>
<td>100</td>
<td>250</td>
</tr>
<tr>
<td>Final mean conc. (n=3)</td>
<td>0.18</td>
<td>0.21</td>
<td>0.52</td>
<td>0.94</td>
<td>2.45</td>
<td>5.00</td>
<td>10.20</td>
<td>22.85</td>
<td>51.80</td>
<td>107.58</td>
<td>252.32</td>
</tr>
<tr>
<td>Precision % CV</td>
<td>20.6%</td>
<td>13.8%</td>
<td>4.9%</td>
<td>1.3%</td>
<td>1.6%</td>
<td>1.9%</td>
<td>2.1%</td>
<td>1.5%</td>
<td>2.1%</td>
<td>1.3%</td>
<td>2.1%</td>
</tr>
<tr>
<td>Accuracy % RE</td>
<td>83.5%</td>
<td>7.3%</td>
<td>3.4%</td>
<td>6.0%</td>
<td>1.9%</td>
<td>0.1%</td>
<td>2.0%</td>
<td>8.6%</td>
<td>3.6%</td>
<td>7.6%</td>
<td>0.9%</td>
</tr>
</tbody>
</table>

1. Definitions of standards, precision, accuracy, theoretical, and final mean con were the same as in Table 4.1.
Figure 4.12. Linear calibration curve for monkey and mouse standards in rabbit blood measured using target-MS/MS mode on Q Exactive.

Figure 4.13. Using optimized targeted-MS2 method described in Figure 4.10 for detection of monkey hepcidin at low levels. (A) blank serum, (B) at 0.1 ng/mL, and (C) at 50 ng/mL.
Figure 4.14. Effect of hepcidin reduction by one siRNA drugs, and a non-siRNA drug measured at 1, 2, 4, 7, 10, and 14 days after injection of the drugs (measured by the optimized targeted-MS\(^2\) method described in Figure 4.10).

Table 4.5. Quantitation of spiked-in monkey hepcidin level in rabbit serum by Q-Exactive

<table>
<thead>
<tr>
<th>days</th>
<th>Cal. (ng/mL)</th>
<th>Theo. (ng/mL)</th>
<th>% CV</th>
<th>% RE</th>
</tr>
</thead>
<tbody>
<tr>
<td>-9</td>
<td>43.21592</td>
<td>46.5</td>
<td>1.57%</td>
<td>7.06%</td>
</tr>
<tr>
<td>-6</td>
<td>40.75902</td>
<td>42.5</td>
<td>1.23%</td>
<td>4.10%</td>
</tr>
<tr>
<td>-3</td>
<td>35.43657</td>
<td>39</td>
<td>0.93%</td>
<td>9.14%</td>
</tr>
<tr>
<td>1</td>
<td>4.991023</td>
<td>5</td>
<td>1.32%</td>
<td>0.18%</td>
</tr>
<tr>
<td>2</td>
<td>12.22065</td>
<td>11.5</td>
<td>3.90%</td>
<td>6.27%</td>
</tr>
<tr>
<td>4</td>
<td>16.91287</td>
<td>17.5</td>
<td>3.11%</td>
<td>3.36%</td>
</tr>
<tr>
<td>7</td>
<td>17.69954</td>
<td>18</td>
<td>0.92%</td>
<td>1.67%</td>
</tr>
<tr>
<td>10</td>
<td>19.28126</td>
<td>22</td>
<td>0.67%</td>
<td>12.36%</td>
</tr>
<tr>
<td>14</td>
<td>33.25867</td>
<td>36.5</td>
<td>1.14%</td>
<td>8.88%</td>
</tr>
</tbody>
</table>
For comparison of the detection limit, the LOD of QQQ (at 0.5 ng/mL) was compared with the LOD of the Q-Exactive (at 0.1 ng/mL). As illustrated in Figure 4.13B, the LOD of Q-Exactive was detected without background noise (S/N > 100), as compared with the high background noise (S/N ≥ 3) of LOD in QQQ (Figure 4.4B). In addition, background noise in the blank run was detected by the QQQ but not by the Q-Exactive (compare Figure 4.4A with Figure 4.13A). The use of the large product ion in the Q-Exactive (high specificity) with high mass accuracy eliminated the contamination (or background noise) from the serum matrix. Although the LOD improved by 5× (from 0.5 to 0.1 ng/mL), the significant reduction of noise enabled the Q-Exactive to achieve the overall improvement with > 30× (from S/N ≥ 3 to S/N > 100). With the improvement, much less serum consumption per analysis using the Q-Exactive (e.g. from 250 µL to 25 µL) could be used to achieve the similar detection limit as in QQQ (data not shown). A similar improvement was also achieved by the Q-Exactive for detection of mouse hepcidin (at least > 10x increase in detection sensitivity, see Figures 11, 15, and Table 4.5.)
Figure 4.15. SRM-like validation of precursor ion by multiple product ions in the method development using Q-Exactive for mouse hepcidin detection. Product ions, \( y_{21}^{2+}, b_3^+ \) and \( b_4^+ \) of mouse hepcidin were extracted in the level of (A) 0.05 ng/mL and (B) 200 ng/mL.
4.4 Conclusions

In this study, LC-MS using either the QQQ or Q-Exactive mass spectrometers could measure monkey and mouse hepcidin, with at least 10x increases in sensitivity and specificity (noise reduction) using the Q-Exactive method. As a result, much less serum consumption could be achieved to study the effect of siRNA drug using the Q-Exactive mass spectrometer, which can be beneficial to study a small size of animal (i.e. mouse) in the future. The Q-Exactive method should also be more useful (sensitive) for larger peptides or even intact protein targets. The extension in detection of intact proteins could significantly reduce the sample preparation steps and times (e.g. without the need for reduction, alkylation, and enzymatic digestion). In addition, rapid SRM method development is possible on the Q Exactive instrument.
4.5 References


(8) Akin Akinc, Amy Chan-Daniels, Alfica Sehgal, Don Foster, Brian R Bettencourt, Julia Hettinger, Tim Racie, Justin Aubin, Satya Kuchimanchi, H. E.-B. and T. N. In 53rd American Society of Hematology Annual Meeting and Exposition; San Diego, 2011.


(24) Zubarev, R. a; Makarov, A. A. *Analytical chemistry* **2013**.


Copyrights

American Chemical Society’s Policy on Theses and Dissertations

If your university requires you to obtain permission, you must use the RightsLink permission system. See RightsLink instructions at http://pubs.acs.org/page/copyright/permissions.html.

This is regarding request for permission to include your paper(s) or portions of text from your paper(s) in your thesis. Permission is now automatically granted; please pay special attention to the implications paragraph below. The Copyright Subcommittee of the Joint Board/Council Committees on Publications approved the following:

Copyright permission for published and submitted material from theses and dissertations
ACS extends blanket permission to students to include in their theses and dissertations their own articles, or portions thereof, that have been published in ACS journals or submitted to ACS journals for publication, provided that the ACS copyright credit line is noted on the appropriate page(s).

Publishing implications of electronic publication of theses and dissertation material
Students and their mentors should be aware that posting of theses and dissertation material on the Web prior to submission of material from that thesis or dissertation to an ACS journal may affect publication in that journal. Whether Web posting is considered prior publication may be evaluated on a case-by-case basis by the journal’s editor. If an ACS journal editor considers Web posting to be “prior publication”, the paper will not be accepted for publication in that journal. If you intend to submit your unpublished paper to ACS for publication, check with the appropriate editor prior to posting your manuscript electronically.

Reuse/Republication of the Entire Work in Theses or Collections: Authors may reuse all or part of the Submitted, Accepted or Published Work in a thesis or dissertation that the author writes and is required to submit to satisfy the criteria of degree-granting institutions. Such reuse is permitted subject to the ACS “Ethical Guidelines to Publication of Chemical Research” (http://pubs.acs.org/page/policy/ethics/index.html); the author should secure written confirmation (via letter or email) from the respective ACS journal editor(s) to avoid potential conflicts with journal prior publication embargo policies. Appropriate citation of the Published Work must be made. If the thesis or dissertation to be published is in electronic format, a direct link to the Published Work must also be included using the ACS Articles on Request author-directed link — see http://pubs.acs.org/page/policy/articlesonrequest/index.html

* Prior publication policies of ACS journals are posted on the ACS website at http://pubs.acs.org/page/policy/prior/index.html

If your paper has not yet been published by ACS, please print the following credit line on the first page of your article: "Reproduced (or 'Reproduced in part') with permission from [JOURNAL NAME], in press (or 'submitted for publication'). Unpublished work copyright [CURRENT YEAR] American Chemical Society." Include appropriate information.

If your paper has already been published by ACS and you want to include the text or portions of the text in your thesis/dissertation, please print the following credit line on the first page of your article: “Reproduced (or 'Reproduced in part') with permission from [FULL REFERENCE CITATION.] Copyright [YEAR] American Chemical Society.” Include appropriate information.

Submission to a Dissertation Distributor: If you plan to submit your thesis to UMI or to another dissertation distributor, you should not include the unpublished ACS paper in your thesis if the thesis will be disseminated electronically, until ACS has published your paper. After publication of the paper by ACS, you may release the entire thesis (not the individual ACS article by itself) for electronic dissemination through the distributor; ACS’s copyright credit line should be printed on the first page of the ACS paper.

10/10/05, 01/15/04, 06/07/06, 04/07/10, 08/24/10, 02/28/11

207
This is a License Agreement between Siyang Li ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3258920202298</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Oct 30, 2013</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>John Wiley and Sons</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Rapid Communications in Mass Spectrometry</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>A new linear ion trap mass spectrometer</td>
</tr>
<tr>
<td>Licensed copyright line</td>
<td>Copyright © 2002 John Wiley &amp; Sons, Ltd.</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>James W. Hager</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>Feb 4, 2002</td>
</tr>
<tr>
<td>Start page</td>
<td>512</td>
</tr>
<tr>
<td>End page</td>
<td>526</td>
</tr>
<tr>
<td>Type of use</td>
<td>Dissertation/Thesis</td>
</tr>
<tr>
<td>Requestor type</td>
<td>University/Academic</td>
</tr>
<tr>
<td>Format</td>
<td>Print and electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>Figure/table</td>
</tr>
<tr>
<td>Number of figures/tables</td>
<td>1</td>
</tr>
<tr>
<td>Original Wiley figure/table number(s)</td>
<td>Figure 1</td>
</tr>
<tr>
<td>Will you be translating?</td>
<td>No</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
</tbody>
</table>
This is a License Agreement between Siyang Li ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

- **License Number**: 3258420954418
- **License date**: Oct 29, 2013
- **Licensed content publisher**: John Wiley and Sons
- **Licensed content publication**: Wiley Books
- **Licensed content title**: High-Performance Gradient Elution: The Practical Application of the Linear-Solvent-Strength Model
- **Book title**: High-Performance Gradient Elution: The Practical Application of the Linear-Solvent-Strength Model
- **Licensed copyright line**: Copyright © 2006, John Wiley and Sons
- **Licensed content author**: Lloyd R. Snyder, John W. Dolan
- **Licensed content date**: Dec 1, 2006
- **Type of use**: Dissertation/Thesis
- **Requestor type**: University/Academic
- **Format**: Print and electronic
- **Portion**: Figure/table
- **Number of figures/tables**: 1
- **Original Wiley figure/table number(s)**: Figure 1.1
- **Will you be translating?**: No
- **Total**: 0.00 USD
NATURE PUBLISHING GROUP LICENSE
TERMS AND CONDITIONS

Nov 21, 2013

This is a License Agreement between Siyang Li ("You") and Nature Publishing Group ("Nature Publishing Group") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3273830261906</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Nov 21, 2013</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>Nature Publishing Group</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Nature Methods</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>Michael A Gillette, Steven A Carr</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>Dec 27, 2012</td>
</tr>
<tr>
<td>Volume number</td>
<td>10</td>
</tr>
<tr>
<td>Issue number</td>
<td>1</td>
</tr>
<tr>
<td>Type of Use</td>
<td>reuse in a dissertation / thesis</td>
</tr>
<tr>
<td>Requestor type</td>
<td>academic/educational</td>
</tr>
<tr>
<td>Format</td>
<td>print and electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>figures/tables/illustrations</td>
</tr>
<tr>
<td>Number of figures/tables/illustrations</td>
<td>1</td>
</tr>
<tr>
<td>High-res required</td>
<td>no</td>
</tr>
<tr>
<td>Figures</td>
<td>figure 1</td>
</tr>
<tr>
<td>Author of this NPG article</td>
<td>no</td>
</tr>
<tr>
<td>Your reference number</td>
<td></td>
</tr>
<tr>
<td>Title of your thesis / dissertation</td>
<td>New Approaches to Trace LC/MS Analysis of Biological Samples</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Dec 2013</td>
</tr>
<tr>
<td>Estimated size (number of pages)</td>
<td>200</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
</tbody>
</table>
This is a License Agreement between Siyang Li ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3146710141357</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>May 12, 2013</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>John Wiley and Sons</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Rapid Communications in Mass Spectrometry</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data</td>
</tr>
<tr>
<td>Licensed copyright line</td>
<td>Copyright © 2008 John Wiley &amp; Sons, Ltd.</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>William R. Alley, Yehia Mechref, Milos V. Novotny</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>Dec 8, 2008</td>
</tr>
<tr>
<td>Start page</td>
<td>161</td>
</tr>
<tr>
<td>End page</td>
<td>170</td>
</tr>
<tr>
<td>Type of use</td>
<td>Dissertation/Thesis</td>
</tr>
<tr>
<td>Requestor type</td>
<td>University/Academic</td>
</tr>
<tr>
<td>Format</td>
<td>Electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>Figure/table</td>
</tr>
<tr>
<td>Number of figures/tables</td>
<td>5</td>
</tr>
<tr>
<td>Original Wiley figure/table</td>
<td>Figure 5</td>
</tr>
<tr>
<td>number(s)</td>
<td></td>
</tr>
<tr>
<td>Will you be translating?</td>
<td>No</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
</tbody>
</table>

211
This is a License Agreement between Siyang Li ("You") and Nature Publishing Group ("Nature Publishing Group") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number 3273830590733
License date Nov 21, 2013
Licensed content publisher Nature Publishing Group
Licensed content publication Nature Methods
Licensed content title Higher-energy C-trap dissociation for peptide modification analysis
Licensed content author Jesper V Olsen, Boris Macek, Oliver Lange, Alexander Makarov, Stevan Horning et al.
Licensed content date Sep 1, 2007
Type of Use reuse in a dissertation / thesis
Requestor type academic/educational
Format print and electronic
Portion figures/tables/illustrations
Number of figures/tables/illustrations 1
High-res required no
Figures figure 1
Author of this NPG article no
Your reference number
Title of your thesis / dissertation New Approaches to Trace LC/MS Analysis of Biological Samples
Expected completion date Dec 2013
Estimated size (number of pages) 200
Total 0.00 USD
This is a License Agreement between Siyang Li ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

<table>
<thead>
<tr>
<th>License number</th>
<th>3134300297335</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Apr 22, 2013</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>Elsevier</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>The Journal of Molecular Diagnostics</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Microfluidics and Circulating Tumor Cells</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>Yi Dong, Alison M. Skelley, Keith D. Merdek, Kam M. Sprott, Chunsheng Jiang, William E. Pierceall, Jessie Lin, Michael Stocum, Walter P. Carney, Denis A. Smirnov</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>March 2013</td>
</tr>
<tr>
<td>Type of Use</td>
<td>reuse in a thesis/dissertation</td>
</tr>
<tr>
<td>Intended publisher of new work</td>
<td>other</td>
</tr>
<tr>
<td>Portion</td>
<td>figures/tables/illustrations</td>
</tr>
<tr>
<td>Number of figures/tables/illustrations</td>
<td>2</td>
</tr>
<tr>
<td>Format</td>
<td>electronic</td>
</tr>
<tr>
<td>Are you the author of this Elsevier article?</td>
<td>No</td>
</tr>
<tr>
<td>Will you be translating?</td>
<td>No</td>
</tr>
<tr>
<td>Title of your thesis/dissertation</td>
<td>New Approaches to Trace LC/MS Analysis of Biological Samples</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Sep 2013</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
</tbody>
</table>
You are free:

- to Share — to copy, distribute and transmit the work
- to Remix — to adapt the work

Under the following conditions:

- Attribution — You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).
- Noncommercial — You may not use this work for commercial purposes.
- Share Alike — If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar license to this one.

With the understanding that:

- Waiver — Any of the above conditions can be waived if you get permission from the copyright holder.
- Public Domain — Where the work or any of its elements is in the public domain under applicable law, that status is in no way affected by the license.
- Other Rights — In no way are any of the following rights affected by the license:
  - Your fair dealing or fair use rights, or other applicable copyright exceptions and limitations;
  - The author's moral rights;
  - Rights other persons may have either in the work itself or in how the work is used, such as publicity or privacy rights.
- Notice — For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to this web page.