Development and Applications of Mass Spectrometric Methods for Phosphorylation Analysis

Hsin-Pin Ho

Graduate Center, City University of New York
DOCTORAL DISSERTATION

DEVELOPMENT AND APPLICATIONS OF MASS SPECTROMETRIC METHODS FOR PHOSPHORYLATION ANALYSIS

by

Hsin-Pin Ho

A dissertation submitted to the Graduate Faculty in Chemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2014
Approval Page

This manuscript has been read and accepted for the Graduate Faculty in Chemistry in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

09/14/2014 Dr. Emmanuel J. Chang
Date Chair of Examining Committee

09/16/2014 Dr. Brian Gibney
Date Executive Officer

Dr. Amy Ikui
Dr. Yi He
Dr. Adam Profit
Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK
Abstract

Development and Applications of Mass Spectrometric Methods for Phosphorylation Analysis

by

Hsin-Pin Ho

Advisor: Dr. Emmanuel J. Chang

Protein phosphorylation modification regulates numerous cellular functions by a reversible and selective control of kinases and phosphatases. To understand the entire dynamic network of phosphorylation requires sensitive and reliable quantification of phosphorylation, measurements that can be achieved by mass spectrometry. In this research, we established efficient MALDI-mass spectrometric methods as strategies for single- or multi-site phosphorylation quantification without the use of isotopes, chromatography and calibration curves. The methods were assessed by analyzing peptide standards with different single-multiple phosphorylation sites, showing a wide dynamic range, good accuracy and reproducibility. This is the first label-free MALDI method without using a calibration methodology proposed for quantification of in vitro phosphorylation in a kinase assay.

Moreover, advanced mass spectrometry empowers identification of a highly conserved Cdk2 phosphorylation site of HIV-1 reverse transcriptase (RT) at Thr 261 across thousands of HIV-1 strains. We demonstrated phosphorylation on HIV-1 RT peptides and protein in in vitro assays, and confirmed phosphorylation in vivo with antibodies and mutation studies. Blocking this phosphorylation by p21, a naturally occurring Cdk inhibitor, defines a potential Cdk2-mediated cell-intrinsic mechanism for restricting HIV-replication in a clinically significant way.
Acknowledgements

First and above all, I praise God, the almighty, for giving me the opportunity to grow with many blessing through the best and the toughest years of my life. Without your strength to endure, I might have escaped from the achievement.

I would like to forever thank my mentor Dr. Emmanuel J. Chang from the deepest bottom of my heart for his guidance in the fields of analytical and biological chemistry with mass spectrometry, for always being supportive of my projects in science and in life, and for being the best mentor with understanding, patients and encouragement. I would also like to thank Ms. Susan Chang for being so caring with emotional support.

I would like to show my appreciation to my committee members Prof. Ikui, Prof. Yi He, and Prof. Adam Profit for providing the necessary and helpful feedback to my work. I would also like to thank our collaborators, Dr. Mathias Lichterfeld, Dr. Xu G. Yu, and Dr. Jin Leng, for their financial support, experimental contribution and important insight into the project of phosphorylation of the HIV-1 RT.

There are a number of people whom I owe a debt of gratitude over my PhD years at York College. There may be some whom I have inadvertently omitted some names, but I am truly thankful to you. My coworkers Dr. Kevin J. Mark and Ms. Christine Tada, who helped me quickly overcome the language barrier in my first PhD year, provided me experimental suggestions, showed me biological laboratory skills, and boosted me positive energy in the workplace when I've struggled with experiments. Mr. Pratik Kumar as my labmate who always provides me substantial support whenever there is a need. Ms. Marissa Louis and Ms. Liz Wright, who always cheer me up and support me through the good and bad times. All
the undergraduate students in the lab, such as Mr. Abbas Nazir, Ms. Ai-Mei Chen, Mr. Emeka Nnaji, who bring lots of invaluable laughter into the work environment and constant support.

York College is like a big family – All the faculties, staffs and graduate students are always so supportive of each other. I would like to extend deep thanks to Maciej, Jayson, Robert, Gabriela for your lovely friendship, and all your help and assistance, making the time in the laboratories enjoyable and upbeat.

Finally I would like to thank my family – my mother Wan-Jen Lin, my father Ming-Hsin Ho, and my brother Tim Ho, for your endless support and understanding, without whom, my achievement would not be possible.

Lastly but especially, I would like to give my special thanks to my husband, Xiu Sheng Liang, for graphic design of the published paper on Cell Host & Microbe, and most importantly for your trust, selfless love, support, and encouragement.
Abstract

Acknowledgements

Table of Contents

List of Figures

List of Tables

Abbreviations

Chapter 1: Introduction

1.1 Objectives and rationale of the study
1.2 Protein Phosphorylation
1.2.1 Cdk phosphorylation
1.2.2 Significance and complexity of phosphoproteomics
1.3 Mass spectrometry
1.3.1 Desorption Ionization
1.3.2 Mass Analyzer
1.3.3 MS techniques applied to this thesis
1.3.4 Tandem Mass Spectrometry (MS/MS)
1.4 Protein identification and phosphorylation analysis in MS.
1.4.1. MS proteomic approaches.
1.4.2 Phosphorylation analysis in MS
1.4.3 MS-based quantitative analysis

Chapter 2: Development of label-free quantitative proteomic analytical method by using matrix-assisted laser desorption/ionization (MALDI) -tandem mass spectrometry

2.1 Introduction
2.2 Label-free MALDI-MS for single-site phosphorylation quantification
2.2.1 Materials and methods
2.2.2 Results and discussion
2.2.2.1 Mathematic approach for phosphorylation quantification
2.2.2.2 Consistency of instrument response factor, a
2.2.2.3 Effect of laser energy on phosphorylation quantification
2.2.2.4 Linearity, dynamic range and reproducibility
2.2.2.5 Validation of phosphorylation quantification by isotope dilution MALDI-MS
2.2.2.6 Application
2.2.3 Conclusion

2.3 Label-free MALDI-MS for multi-site phosphorylation quantification
2.3.1 Materials and methods
2.3.2 Results and discussion
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3.2.1 Mass spectra of phosphorylated AQP-2 isomers (pSer$<em>{256}$ and pSer$</em>{261}$)</td>
<td>60</td>
</tr>
<tr>
<td>2.3.2.2 Mathematic Approach for multiple phosphorylation site quantification</td>
<td>62</td>
</tr>
<tr>
<td>2.3.2.3 Consistency of MS$^3$ fragment ratios of p261, β</td>
<td>64</td>
</tr>
<tr>
<td>2.3.2.4 Linearity, dynamic range and accuracy</td>
<td>66</td>
</tr>
<tr>
<td>2.3.2.5 Effect of collision energy on multiple phosphorylation quantification</td>
<td>68</td>
</tr>
<tr>
<td>2.3.2.6 Applications to <em>in vitro</em> kinase assays</td>
<td>70</td>
</tr>
<tr>
<td>2.3.3 Conclusion</td>
<td>71</td>
</tr>
</tbody>
</table>

**Chapter 3: Investigation of CDK phosphorylation in HIV-1 reverse transcriptase by mass spectrometry**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Introduction</td>
<td>72</td>
</tr>
<tr>
<td>3.2 Materials and methods</td>
<td>74</td>
</tr>
<tr>
<td>3.3 Results</td>
<td>80</td>
</tr>
<tr>
<td>3.3.1 Host Cdks support HIV-1 reverse transcription in CD4$^+$ T cells</td>
<td>80</td>
</tr>
<tr>
<td>3.3.2 HIV-1 RT is a substrate for Cdk2-dependent phosphorylation</td>
<td>82</td>
</tr>
<tr>
<td>3.3.3 Substitution of T216 by Alanine reduces HIV-1 RT</td>
<td>91</td>
</tr>
<tr>
<td>3.3.4 Quantification of HIV-1 <em>in vitro</em> Cdk2-dependent phosphorylation</td>
<td>93</td>
</tr>
<tr>
<td>3.3.5 Cdk2-dependent phosphorylation increases function and stability of HIV-1 RT</td>
<td>95</td>
</tr>
<tr>
<td>3.3.6 p21 (<strong>waq-1/cip-1</strong>) suppresses Cdk2-dependent HIV-1 RT phosphorylation</td>
<td>96</td>
</tr>
<tr>
<td>3.4 Discussions and Conclusion</td>
<td>99</td>
</tr>
</tbody>
</table>

**Chapter 4: Summary and Future Directions**

**Reference**

110
List of Figures

Chapter 1

Figure 1.1 Schematic illustration of the MS instrument setup 8
Figure 1.2 Schematic illustration of electrospray ionization 10
Figure 1.3 Schematic illustration of matrix laser desorption ionization. 11
Figure 1.4 Electrospray ionization and matrix assisted laser desorption ionization. 19
Figure 1.5 Typically used matrices in MALDI-MS 19
Figure 1.6 Three-dimentional ion trap configuration with central ring electrode and two end-cap electrodes 20
Figure 1.7 Two-dimentional linear ion trap configuration with the quadrupole rods cut into three axial sections. 21
Figure 1.8 Schematic illustration of tandem mass spectrometry 22
Figure 1.9 Peptide sequence fragmentation 24
Figure 1.10 The structure of iTRAQ reagent composed of a reporter group balance group and reactive group 30
Figure 1.11 MALDI-MS workflow with iTRAQ labeling (m/z 114-117) 30
Figure 1.12 Labeling reactions for dimethyl labeling 31

Chapter 2

Figure 2.1 General workflow for proteomic quantification. 36
Figure 2.2 Mass spectra of the experiments in Table 1 45
Figure 2.3 Linearity of double reciprocal equation. 48
Figure 2.4 Mass spectra of the experiments for dynamic range test in Table 3. 50
Figure 2.5 Inter-day reproducibility. 50
Figure 2.6 Variability in MALDI-TOF MS signals. 54
Figure 2.7 In vitro inhibition of Cdk2 measured by DRIP-Q method. 55
Figure 2.8 MS\(^2\) fragmentation of pSer-256 and pSer-261. 61
Figure 2.9 MS\(^3\) fragmentation of pSer-256 and pSer-261. 61
Figure 2.10 The linearity between amount fraction (% pSer-261) and R\(_{\text{mix}}\). 66
Figure 2.11 The effect of collision energy on \(\alpha\), \(\beta\) and R\(_{\text{mix}}\). 69
Figure 2.12 Phosphorylation quantification of aquaporin-2 treated with PKA and CDK in a time course. 72

Chapter 3

Figure 3.1 The crystal structure of HIV-1 RT complexes with ds-DNA 74
Figure 3.2 Host CDK supports HIV-1 reverse transcription 81
Figure 3.3 HIV-1 reverse transcriptase sequence. 82
Figure 3.4 HIV-1 RT is a substrate for CDK2-dependent phosphorylation. 83
Figure 3.5 In vitro phosphorylation reaction of synthetic peptide [210-227] by cyclindependent kinases. 84
Figure 3.6 CDK2 supports HIV-1 reverse transcription.

Figure 3.7 Influence of CDK1, CDK4 and CDK9 on HIV-1 replication.

Figure 3.8 MS² spectrum of the phosphorylated synthetic peptide [210-227] analyzed using MALDI-LTQ-MS.

Figure 3.9 HIV-1 RT is a substrate for CDK2-dependent phosphorylation.

Figure 3.10 MS² and MS³ spectra of Thr216 [202-220] phosphorylated Lys-C proteolytic RT peptide.

Figure 3.11 Conservation of Thr 216 in RT. (A): Solvent-exposed position of Thr216 in crystal structures of HIV-1 RT. (B): Entropy of amino acid residues at position 214-219 in indicated clade A, B and C HIV-1 RT sequences obtained from the Los Alamos HIV-1 sequence database.

Figure 3.12 HIV-1 RT T216A mutation decreases HIV-1 reverse transcription.

Figure 3.13 Quantification of in vitro phosphorylation of HIV-1 RT by using stable isotope labeling in conjunction with mass spectrometry.

Figure 3.14 CDK2-dependent phosphorylation enhances activity and stability of HIV-1 RT.

Figure 3.15 Inhibition of p21 (waf-1/cip-1) Increases HIV-1 Reverse Transcription in CD4+ T Cells from Elite Controllers.
List of Tables

Chapter 2

Table 2.1 Instrument response factor (a) at different phosphorylation fractions (x) 44
Table 2.2 The effect of different MALDI laser energy on the instrument response factor (a) and accuracy (% errors). 46
Table 2.3 Effect of sample concentration on unphospho-to phosphor signal ratio (S_un/S_p) 49
Table 2.4 Comparison of the results obtained by the proposed DRIP-Q method and isotope dilution method. 52
Table 2.5 Normalized intensity of MS^3 fragments of m/z 1297 and m/z 1460 of two sample sets in different amounts. 63
Table 2.6 The effect of sample amounts on R_{261} consistency. 65
Table 2.7 Dynamic range of the constant (x). 67
Table 2.8 Determination of % pSer-261 by using the proposed single-point method and conventional calibration curve method. 67
Table 2.9 Effect of collision energies on the amount fractions, % pSer-261. 69
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D-LIT</td>
<td>Two-dimensional linear ion trap</td>
</tr>
<tr>
<td>3D-QIT</td>
<td>Three-dimensional quadrupole ion trap</td>
</tr>
<tr>
<td>AQP-2</td>
<td>Aquaporin-2</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>AZT</td>
<td>3’-azido-3’-deoxythymidine</td>
</tr>
<tr>
<td>CAD</td>
<td>Collisionally activated decomposition</td>
</tr>
<tr>
<td>CD4</td>
<td>Cluster of differentiation 4</td>
</tr>
<tr>
<td>Cdc25</td>
<td>Cell division cycle 25 homolog</td>
</tr>
<tr>
<td>CDK</td>
<td>Cyclin-dependent kinase</td>
</tr>
<tr>
<td>CHCA</td>
<td>α-cyano-4-hydroxycinnamic acid</td>
</tr>
<tr>
<td>CID</td>
<td>Collision-induced dissociation</td>
</tr>
<tr>
<td>Cks</td>
<td>Cdc kinase subunit</td>
</tr>
<tr>
<td>DC</td>
<td>Direct current</td>
</tr>
<tr>
<td>DHB</td>
<td>2,5-dihydroxybenzoic acid</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>ds-DNA</td>
<td>Double-stranded deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTPs</td>
<td>Deoxynucleotide triphosphates</td>
</tr>
<tr>
<td>DRIP-Q</td>
<td>Double reciprocal isotope-free phosphopeptide quantification</td>
</tr>
<tr>
<td>EC</td>
<td>Elite controller</td>
</tr>
<tr>
<td>ECD</td>
<td>Electron capture dissociation</td>
</tr>
<tr>
<td>ESI</td>
<td>Electrospray ionization</td>
</tr>
<tr>
<td>ETD</td>
<td>Electron-transfer dissociation</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>LC-MS</td>
<td>Liquid chromatography-mass spectrometry</td>
</tr>
<tr>
<td>FT-ICR</td>
<td>Fourier transform-ion cyclotron resonance</td>
</tr>
<tr>
<td>HIV-1</td>
<td>Human immunodeficiency virus-1</td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>iTRAQ</td>
<td>Isobaric tag for relative and absolute quantitation</td>
</tr>
<tr>
<td>MALDI</td>
<td>Matrix-assisted laser desorption ionization</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>MS/MS</td>
<td>Tandem mass spectrometry</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>m/z</td>
<td>Mass to charge</td>
</tr>
<tr>
<td>PKA</td>
<td>Protein kinase A</td>
</tr>
<tr>
<td>PMA</td>
<td>Phorbol myristate acetate</td>
</tr>
<tr>
<td>RF</td>
<td>Radio frequency</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNase H</td>
<td>Ribonuclease H</td>
</tr>
<tr>
<td>R.S.D.</td>
<td>Relative standard deviation</td>
</tr>
<tr>
<td>RT</td>
<td>Reverse transcriptase</td>
</tr>
<tr>
<td>SA</td>
<td>Sinapinic acid</td>
</tr>
<tr>
<td>SILAC</td>
<td>Stable isotope labeling by amino acids in cell culture</td>
</tr>
<tr>
<td>siRNA</td>
<td>Small interfering RNA</td>
</tr>
<tr>
<td>SRM</td>
<td>Selected reaction monitoring</td>
</tr>
<tr>
<td>TOF</td>
<td>Time of flight</td>
</tr>
<tr>
<td>VSV-G</td>
<td>Vesicular stomatitis virus</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction to Protein Phosphorylation and Mass Spectrometry

1.1 Objectives and rationale of the study

1.2 Protein phosphorylation
   1.2.1 Cdk phosphorylation
   1.2.2 Specific examples of phosphorylated Cdk substrates

1.3 Mass spectrometry
   1.3.1 Desorption Ionization
   1.3.2 Mass analyzer
   1.3.3 MS techniques applied to this thesis – MALDI-LIT MS
   1.3.4 Tandem Mass spectrometry (MS/MS)

1.4 Protein identification and phosphorylation analysis in MS
   1.4.1 MS protein identification approaches
   1.4.2 Phosphorylation analysis in MS
   1.4.3 MS-based quantitative analysis

1.1 Objectives and rationale of the study

The objectives of this study are to develop label-free mass spectrometric methods for quantification of single or double-site phosphorylation (chapter 2), and to investigate Cdk (cyclin-dependent kinase) phosphorylation in HIV-1 reverse transcriptase (RT) by mass spectrometry-based techniques (chapter 3). The organization of the dissertation is described below.

This chapter introduces the background of Cdk phosphorylation and gives specific examples of phosphorylated Cdk substrates. The following text depicts essential parts of an analytical biological mass spectrometers (including desorption ionization and mass analyzers but emphasizing the techniques applied to this thesis), and the state-of-the-art in tandem mass spectrometry. The approaches of mass spectrometry are described instrumentally and methodically for protein phosphorylation identification and quantification.

In Chapter 2 includes two projects with a mutual goal - to develop label-free mass spectrometric strategies as alternatives for quantitative phosphoproteomics due to its low cost and convenience. The first project is to establish an efficient MALDI-mass spectrometric method, called Double Reciprocal Isotope-free Phosphopeptide Quantification (DRIP-Q), as the first proposed
isotope-free and calibration-curve-free MALDI-MS method for phosphorylation quantification. The results as follows include 1) a mathematic approach for phosphorylation quantification; 2) the effect of laser energy on phosphorylation quantification; 3) linearity, dynamic range and reproducibility; 4) validation of phosphorylation quantification by comparison with isotope dilution MALDI-MS; 5) applications to Cdk inhibition.

The second project in Chapter 2 is to utilize high sensitivity and specificity of MS$^3$ to determine double-site phosphorylation of aquaporin, in which two different phosphorylation sites occur on Ser-256 and Ser-261 in response to vasopressin-dependent short-term regulation (2006 Hoffert, 2010 Xie). The MS$^3$ method proposed overcomes the difficulties in isomeric phosphopeptide stoichiometric analysis by CID based MS/MS without the use of isotopes. The results obtained of 1) MS$^2$ and MS$^3$ analyses of AQP-2 peptides; 2) a mathematic approach by using MS$^3$ fragmentation for double phosphorylation site quantification; 3) consistency of MS$^3$ fragment ratios; 4) the effect of collision energy on quantification; 5) quantitative dynamic range; and 6) applications to in vitro kinase assay will be discussed.

In Chapter 3, the project (in collaboration with Mathias Litchterfeld’s group at Massachusetts General Hospital) uncovered a cell-intrinsic mechanism of HIV-1 reverse transcription associated with Cdk2 phosphorylation. The results strongly suggest that p21 indirectly blocks HIV-1 reverse transcription by inhibiting Cdk2-dependent phosphorylation of RT, including the following parts: 1) Host Cdns support HIV-1 reverse transcription in CD4$^+$ T cells; 2) HIV-1 RT is a substrate for Cdk2-dependent phosphorylation; 3) substitution of T216 by Alanine reduces HIV-1 RT; 4) Quantification of HIV-1 in vitro Cdk2-dependent phosphorylation; 5) Cdk2-dependent phosphorylation increases function and stability of HIV-1 RT; 6) p21 suppresses Cdk2-dependent HIV-1 RT phosphorylation.
1.2  Protein Phosphorylation

Most eukaryotic proteins selectively and specifically undergo covalent modification to regulate protein functions after being translated from mRNA (2005 Yang; 1995 Newton). Phosphorylation is one of the most important post-translational modifications in the human cells, occurring on serine/threonine/tyrosine residues in a number of proteins via catalytic addition of phosphate group from adenosine triphosphate (ATP) (2003 Mann). Protein phosphorylation is controlled by a wide variety of kinase and phosphatases, which themselves may be regulated by phosphorylation. Once a protein is phosphorylated by a kinase, the phosphorylation event can cause conformational and functional changes of the protein to either activate or de-activate the protein – an effective way to regulate protein functions (1993 Johnson). Here, we review the literature on Cdk phosphorylation, a very important post-translational modification in cellular regulation, and describe specific examples of Cdk phosphorylation substrates.

1.2.1  CDK phosphorylation

In the 1970s and 1980s, Cyclin-dependent kinases (Cdks), a family of serine/threonine protein kinases, were discovered as key regulators of the cell cycle in eukaryotic cells via phosphorylation by Hartwell, Hunt, and Nurse who were awarded the Nobel Prize in Physiology or Medicine for 2001 (1995 Morgan; 1989 Hartwell; 1989 Gould; 1983 Evans). Cdk complexes are composed of a Cdk catalytic subunit and a cyclin subunit, i.e. Cdk2/cyclin A. The activation of cyclin-Cdk complexes requires the binding of the cyclin subunit, phosphorylation activation on a conserved threonine (Thr160) in the Cdk activation loop by Cdk-activating kinase (CAK) (1995 Morgan), and, in some cases (e.g. multisite phosphorylation), interaction with Cdc kinase subunit (Cks) protein (1996 Pines, 1996 Bourne, 2011 Balog).
Differences in cyclin-Cdk complexes give specificity of phosphorylation to particular substrates and sites in cells. For example, some cyclins contain a docking domain, a small hydrophobic patch for helping recruit their correct substrate to Cdns by interacting with an RXL motif on the specific substrates (2007 Ubersax, 2011 Kõivomägi). A slight change in the sequences of the hydrophobic patches of different cyclins can lead to proper substrate recognition by either increasing local concentration of the kinase around the substrate, or orientating the substrate in the kinase active site, even though it still remains unclear if the Cdk complex binds to the RXL motif and the phosphorylation site on the substrate synchronously, especially for multiple phosphorylation sites on a substrate (2007 Ubersax).

The Cdk’s sequences are complementary to the specific substrate’s sequences through hydrophobic interactions, charge-charge interactions or hydrogen bonding (1995 Jeffrey). Cdns have a canonical catalytic domain composed of \( \beta \)-sheets in a small N-terminal lobe and \( \alpha \)-helices in a larger C-terminal lobe. The ATP binding is found at a hydrophobic pocket between the two lobes where the adenosine moiety is buried with the phosphate backbone exposed to the solvent. The catalytic domain catalyzes the transfer of the terminal \( \gamma \)-phosphate moiety to the hydroxyl phosphorylation site on the substrate, which interacts to the kinase. The substrate consensus phosphorylation sites for Cdns 1, 2, 4, 5, and 6 (2010 Errico) are S/T-P-X-K/R, where S/T represents the phosphorylated residue, and X represents any amino acids followed by basic residues, K/R. The phosphorylated site “P” (S/T) is followed by a Pro residue, which contains a unique nitrogen atom among the 20 amino acids. Because the nitrogen atom is part of rigid pyrrolidone side chain ring; it cannot act as a hydrogen donor. Any other residue at P+1 is unfavorable owing to the existence of uncompensated hydrogen bonds from substrate to a main-chain nitrogen atom. The Pro residue at P+1 also constrain the conformation of the substrate, allowing a basic residue (K/R) at P+3 to interact with the phospho-Thr 160 in the kinase (1999 Brown).
Cks (cyclin-dependent kinase subunit), a subunit of Cdks, is a relatively small protein (9-18 kDa), interacting with Cdk complexes in the control of cell cycle, has been identified since 1980’s; however researchers have long been tantalized by the implication of Cks for modulation of specific Cdk activity (1996 Pines) and their precise function still remains elusive. The crystal structures of the Cks proteins as a part of a complex with Cdks have been solved (1996 Bourne), suggesting Cks binds CdkS as a monomer, although Cks also exists as a dimer. The exchange of the monomeric and dimeric states of Cks via its conserved $\beta$-hinge region regulates Cdks-Cks association and thereby regulates Cdk enzymatic activity (2006 Bader, 2001 Bousseau). Two highly conserved residues (Pro93 and Pro95) in the hinge region of Cks have been observed its role in tuning the dimerization equilibrium, and recently proposed to associate with formation of Cdk-Cks interface by stabilizing conformation for binding (2011 Balog). Besides, the structural data revealed a conserved cationic pocket that weakly binds to free phosphate as well as other anions, and interacts with the Cdk active site. Cks is considered as a phosphoadaptor subunit of the cyclin-Cdk-Cks complexes, positioning the kinase complex to phosphorylated substrates and promoting multiple phosphorylation, which occurs in clusters in majority of Cdk substrates (2013 Koivomagi). Cks binding sites have been predicted to preferentially occur in Cdk substrates that contain multiple phosphorylation sites, and more likely to be found in phosphorylated substrate on a Cdk consensus site (mostly Thr) nearby the phosphorylated Cdk consensus site (2013 McGrath).

1.2.2 Specific examples of phosphorylated Cdk substrate

Protein phosphorylation modification involves addition of phosphate to a protein. The bulky phosphate with a double negative charge, therefore, alters a protein’s physical and biochemical properties, and induces conformational changes that are responsible for the biological protein functions, such as activity, stability and interaction (2007 Singh; 2009 Li; 2010 Deribe). Several well-
studied Cdk phosphorylation substrates involved in regulation of cell cycle and other biological processes have been reported. Some specific examples are introduced below.

**Sic1** acts as one of the important regulators through phosphorylation in cell cycle, a process of cell duplication and division, in *Saccharomyces cerevisiae*. The cell cycle is an ordered bio-event, including four phases, G1 phase (gap1, a growth phase for the cell) - S phase (DNA synthesis) - G2 (gap2, a second growth phase for the next division) – M phase (mitosis, a process of cell division). Sic1 associates with Cdk1/cyclinB complexes to regulate the G1/S transition in yeast via proteolysis caused by phosphorylation as a part of the intrinsic cell cycle machinery. When Sic1 is unphosphorylated, the Cdk1/cyclinB stays inactive, whereas the inactive Cdk1 kinase complex is switched into active state by Sic1 degradation through its multiply phosphorylation by Cdk1 on the minimum six of nine Cdk consensus sites (1997 Verma, 1999 Hodge). One thing to note is that the multiple phosphorylation of Sic1 is stimulated by Cks in the early stage of phosphorylation, occurring after Cdk consensus site (T5 or T33) on Sic1 which is first phosphorylated by Cdk1. The phosphorylated T5 or T33 (Cks consensus sites) is recognized by Cks to induce multiple phosphorylation, allowing the binding of Cdc4, and consequently inducing Sic1 degradation. The Sic1 degradation is necessary for Cdk activation, leading to initiation of the following DNA synthesis (S phase) in the cell cycle (2011 Koivomagi).

*Schizosaccharomyces pombe* **Wee1** and **Cdc25** (cell division cycle 25 homolog A) are, respectively, negative and positive regulators of Cdk1 protein kinase activation and localization during the G2/M transition. During the period of late S and G2 phase, cyclin B synthesis increases, promoting the binding of cyclin B to Cdk1. The Cdk1/cyclinB complexes are normally held in an inactive state as a result of phosphorylation of Cdk1 on Tyr15 by Wee1 (1994 Coleman, 1996 Lew, 2007 Perry). Wee1 and Cdk1 show mutual regulation: Wee1 itself is first phosphorylated and
activated by Cdk1/cyclinB undergoing multiple site phosphorylation with opposing threshold control by Cdc55-mediated dephosphorylation. A switch-like inhibitory phosphorylation of Cdk1 by activated Wee1 is then induced by Cks proteins that interact with the Cdk1 active site and phosphorylated Cks consensus sites on Wee1. During the M phase, dephosphorylation of Cdk1 by phosphatase Cdc25 (a key regulator of mitotic entry) converts the kinase complexes into an active state, and leads to the progress of cell division. The activation of Cdk1 occurs in a stepwise manner controlled by a dynamic equilibrium between phosphorylation (de-activating) and dephosphorylation (activating) of Cdk1; phosphorylation of Cdk1 is regulated by the multiple-phosphorylation state of Wee1. Initially during M phase, Cdk1 is held in a state of low activity for spindle assembly, while later the activity is switched to a state of high activity for spindle elongation at the time of highest expression of cyclin B (2011 Harvey). The activities of Wee1 and Cdc25 appear to be oscillated in opposition to one another, similar to their balanced roles in Cdk1 inhibition and activation (2005 Harvey).

**Beside the cell cycle control, a host protein SAMHD1**, associated with restriction of HIV-1 replication in myeloid and quiescent CD4+ T cells, is regulated by Cdk phosphorylation (2013 Cribier). SAMHD1 is one of the identified HIV-1 restriction factors, and consists of a sterile alpha motif (SAM) and a HD domain. It has been reported that SAMHD1 acts as a host restriction nuclease that can convert nucleotide triphosphates to a nucleoside and triphospate. In the presence of SAMHD1, HIV-1 infection appears to be blocked by reducing the level of dNTPs to a level too low to be available to a complete reverse transcription step. It has been observed that Cyclin A2/Cdk1 and Cyclin A/Cdk2 complexes phosphorylate SAMHD1 at its Thr 592 residue both *in vitro* and *in vivo* (2013 Cribier, 2014 St Gelais), whereas the phosphorylated SAMHD1 was significantly reduced in the PMA treated cells without affecting the level of unphosphorylated SAMHD1. A mutagenesis experiment showed that a Cdk phosphorylation-defective mutant
(T592A) has sufficient ability to inhibit HIV-1 replication in both of the PMA treated and untreated cell. In other words, the phosphorylation of SAMHD1 may play a role in a negative regulator of its restriction activity toward HIV-1. The authors proposed that the restriction activity of SAMHD1 may be regulated by the equilibrium between Thr 592 phosphorylated and unphosphorylated forms favored in the proliferating cells (untreated cells), and non-proliferating cells (treated cells), respectively (2013 Cribier, 2014 St Gelais).

1.3 Mass spectrometry

Mass spectrometry is an analytical technique that determines mass-to-charge ratio ($m/z$) of ionized molecules. In MS analysis, analytes are ionized and volatilized in the source, separated by $m/z$ in the analyzer, and counted in the detector. MS not only provides $m/z$ of the molecules, but in some cases also disintegrate the molecular ions into particular fragmentation pattern for qualification and quantification through tandem mass spectrometry (MS/MS). Typically, a mass spectrometer consists of an inlet system, ion source, a mass-selective analyzer, an ion detector and data system (see Figure 1.1). There are a variety of combinations of the MS ion sources and analyzers designed for different analytical purposes.

**Figure 1.1** Schematic illustration of the MS instrument setup
1.3.1 Desorption Ionization

Mass spectrometers determine the mass-to-charge ratio ($m/z$) of gas-phase molecular ions generated in an ion source. Successful MS analyses are mainly dependent on the volatilization and ionization of analytes in a sample. In biological mass spectrometry, because the large polar biomolecules, such as proteins and polypeptides, are non-volatile, it has been challenging to desorb the macromolecules without decomposition into smaller charged particles until desorption ionization techniques was developed. Currently, the majority of biological MS is couple to one of the two desorption ionization techniques: Electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI).

ESI is an ionization process by which molecules was directly desolvated from solution-phase into gas-phase under atmospheric pressure. In the late 1980s, John Fenn, the 2002 Nobel Prize Laureate in Chemistry, applied ESI to convert liquid solution of large molecules into gas ions for MS. The ESI technique made a breakthrough for MS to be attributed to the applications of LC-MS in biochemistry. In ESI, the analytes are first dissolved in a solution of sufficient ionic strength, sprayed through an ESI needle with 2-5 kV voltages, positive or negative, that results in the accumulation of ions with identical charge polarity formed at the droplet surface at the tips of the capillary relative to vacuum orifice. To evaporate the neutral solvent molecules from the droplet surface, an inert gas flow is directed to the droplets to the spray exit where a “Taylor cone” shape is formed. As a result, the charge density increases while the droplet size decreases. When the coulomb repulsion between the like charges on the droplet surface exceeds the forces of surface tension (Rayleigh instability limit), the droplets break into smaller highly charged droplets. The same process occurs many times until only one charged analyte molecule is left in single droplet. The lone ions thus are ejected into the gas phase by electrostatic repulsion entering into the mass spectrometer to be analyzed (Figure 1.2).
MALDI-MS is based on the concept of laser desorption mass spectrometry proposed in 1976, in which pulsed laser has been employed to produce intact gas phase ions from solid samples (1976 Macfarlane). Figure 1.3 shows a scheme graph of MALDI. In the early investigation of laser desorption mass spectrometry, before the development of MALDI, only a few short peptides could be ionized and generate intact ions. Successful ionization was highly dependent on the specific physical properties of the peptides. In 1985, Hillenkamp and Karas first mentioned of the term “matrix-assisted laser desorption” in a publication, reporting the observations of intact protonated molecular ions of amino acids and dipeptides under different influence of UV- laser wavelength for desorption (1985 Karas). One of the experiments shows that the intact molecular ion of alanine was produced in the presence of tryptophan at an irradiation of about a tenth of that necessary for obtaining spectra of alanine alone. The authors proposed that the tryptophan acted as an “absorbing matrix” which absorbed excess of the laser energy and transfer appropriate energy to the non-absorbing to generate a soft ionization, yielding intact molecular ions. In 1988, Hillenkamp and
Karas further performed experiments to successfully analyze bovine albumin (67 kDa) in the presence of high concentration of nicotinic acid, and observed the intact ions (1988 Karas). In 1988, Tanaka introduced another sample preparation method into MALDI-TOF-MS for detection of intact ions of large proteins, such as lysozyme (14 kDa) and chymotrypsinogen (26 kDa), by mixing an ultra fine cobalt powder with glycerol in an organic solvent and dropped a few microliters of the mixed solution onto a sample holder where sample solution is placed (1988 Tanaka). Both methods were a breakthrough in the application of MALDI-TOF-MS extended to include analysis of high-mass molecules (proteins). Although today’s technique typically employs the use of a small organic acid similar to Hillenkamp and Karas’s method, Tanaka was awarded the Nobel Prize in chemistry in 2002.

Figure 1.3 Schematic illustration of matrix assisted laser desorption ionization.
1.3.2 Mass Analyzer

After being ionized to charged gas-phase particles ion an ionization source, sample ions are analyzed depended on their mass \((m)\) and charge \((z)\) in a mass-selective analyzer. In order to accurately separate and measure the \(m/z\) of the ions, the mass analyzer should be operated under high vacuum and calibrated by using ions of known \(m/z\). There are five major types of mass analyzers: (i) Sector (magnetic or electric), (ii) Time-of-flight (TOF), (iii) Fourier Transform-Ion-Cyclotron Resonance (FT-ICR), (iv) Quadrupole, (v) Ion trap. Different analyzers have different capabilities. In general, the analyzers are separated into two categories: high resolution (Sector, TOF and FT-ICR) and low resolution (Quadruple and ion-trap). High-resolution analyzers have better ability to distinguish between ions of high \(m/z\); i.e. intact protein ions, than low-resolution analyzers, whereas, the speed of data processing of high-resolution analyzers can be limited. In recent years, hybrid analyzers are designed to include more-than-one analyzer with compensatory advantages of individual analyzers; it therefore causes the use of hybrid analyzers to rapidly become a popular technique to achieve high performance.

**TOF analyzers** separate ions of different \(m/z\) traveling with their mass-dependent velocities in space when the same voltage is applied. The TOF analyzers are commonly combined with MALDI ionization technique. After being ionized in the source, all ions gain the same amount of kinetic energy from an acceleration voltage. Therefore, the energy allows the ions to drift through a field-free flight tube under a vacuum with different velocities \((K.E.=1/2mv^2 = zV);\) where KE, \(m, v, z\) are the kinetic energy, the mass of the ion, the velocity, the charge of the ion, and the applied voltage, respectively). In other words, for same distance, ions with larger \(m/z\) values fly more slowly, whereas ions with smaller \(m/z\) values fly faster to the detector; thus different drifting time can be recorded with respect to each \(m/z\) value. The TOF analyzers are advantageous of very high mass
range with high resolution and high sensitivity, although it requires high vacuum, and high voltage/pulsed/high precision electronics.

**Sector or magnetic analyzers** also separate ions in space according to their \( m/z \) values detected at different location. Most sector analyzers are of double focusing design, including an electrostatic analyzer and magnet sector. Ions formed in the ion source pass through the electrostatic analyzer, where the ions of different energy are trajected along different paths toward the magnet sector. The magnet field is set up perpendicular to the path of the ions, leading to ions of identical \( m/z \) but different energy are brought to same focus at a single point on the detector. This combination of the electrostatic analyzer and magnet sector creates a high-resolution sector mass spectrometer with typical accuracy of 1-5 ppm (1997 Russell). The double focusing magnetic sector mass analyzers provide reproducibility with the best quantitative performance of all mass spectrometer analyzers, such as high resolution, high sensitivity, high dynamic range, but it is not well-suited for pulsed ionization methods (e.g. MALDI), and the instrument itself is usually bulky and higher cost than other mass analyzers.

**FT-ICR analyzers** measure \( m/z \) values of ions based on the cyclotron motion of ions caused by cyclotron frequency in a strong magnetic field. It is similar to a sector analyzer in that a magnetic field combined with electric field is used, but an ICR analyzer traps all ions inside of an ion trap (a cell) into an orbit by a centripetal force counterbalanced by a Lorentz force (perpendicular to both the ions velocity and magnetic field lines) (2001 Schmid). In the cell, ions with the identical \( m/z \) have identical cyclotron frequency but differences in energies, orbital radii, and consequently velocities in their orbits. To obtain measurable signals, the ions are then excited by a pulse of a radio frequency, generating a larger cyclotron radius. As the ions pass by the detection plates, an “image current” is induced. The complex signals are converted by Fourier Transformation (FT) from the time domain to the frequency domain, which can be further converted into a \( m/z \) spectrum. In
contrast to the magnetic sector analyzers, ICR is well suited for use with pulsed ionization methods (e.g. MALDI), has the highest recorded mass resolution of all mass spectrometer, and has powerful capability for MS/MS experiments. However, it suffers from a limited dynamic range and high cost of high vacuum and cryogenic cooling to lower signal to noise ratios. Besides, the accuracy is subject to space charge effects and ion molecule reactions, once ions are overfilled in the cell.

**Quadrople analyzers** are resonance mass filters that consist of four parallel rods in a square cylinder configuration. The ions are passed through the quadrupole in a combination field of a direct current (DC) and a radio frequency (RF) voltage. Ions of each \( m/z \) value have their stable trajectory within a range of DC and RF values. In general, the DC and RF voltages are ramped together at a particular ratio, letting ions with a particular \( m/z \) value through the quadrupole to the detector at a time. The benefits of the quadrupole analyzer include good reproducibility, relatively small and low-cost systems, but it has limited resolution, and is not well suited for pulsed ionization methods.

**Ion trap analyzers**, analogous to the quadrupole, utilize ion path stability of ions for separating them by their \( m/z \) by adjusting a direct current voltage and radio frequency (RF) voltage. There are two configurations: three-dimensional quadrupole ion trap (3D-QIT), also called Paul traps) and a new generation two-dimensional linear ion trap (2D-LIT). Paul and Steinwedel invented the 3D-QIT in the early 1950’s after the development of electrodynamic quadrupole (1953 Pual). The 3D-QIT is a compact and inexpensive analyzer, which consist of a ring electrode with end caps on the top and bottom. The trap has high scan speed and high selectivity and sensitivity of tandem mass spectrometry (MS\(^2\)). In the last decade, 2D-LIT was induced to increase the capacity and loading efficiency of the trap and reduce space-charge problems (changes in ion motion), resulting from too many ions in the trap that causes the mutual coulombic interaction of ions, and consequently induce mass shift and lower the resolution and accuracy observed. The details of the ion trap analyzers will be described in the following section.
**Orbitrap analyzer** is a new type of high-resolution mass spectrometer, composed of an outer barrel-like electrode and a central spindle-like electrode along the axis (2008 Perry). As the name suggests, Orbitrap is considered an ion trap analyzer, but there is no RF or magnetic used to control ion stability. In Orbitrap, ions are trapped and oscillated into orbits between the two symmetric electrodes, where a DC voltage applied results in the electrostatic potential distribution—a combination of quadrupole and logarithmic potentials. The axial oscillation frequencies, which are exclusively dependent on \( m/z \) but not on initial velocities as well as coordinates of ions, can be detected by measuring the image current on the outer electrodes. The time-domain frequency signals are further converted into \( m/z \) values shown on a spectrum. Similar to FT-ICR, advantages of the Orbitrap include high resolving power (up to 150,000), multi-fragmentation capability and high sensitivity in the attomole to femtomole range for measurements of intact proteins, whereas Orbitrap may outperform FT-ICR because of increased space charge capacity at higher masses, larger trapping volume, and, most importantly, a cryogen-free detection system (2005 Hu, 2006 Macek).

**Hybrid mass spectrometers** combine more than one type of mass analyzer that each has their own pros and cons. The hybrid capability allows for the combination of the different desirable performance characteristics offered by various types of analyzers into one mass spectrometer (2008 Glish). Most importantly, the hybrid analyzer provides a platform to perform MS/MS experiments with high resolving power, accuracy and analysis speed; e. g., ions are mass-selected in the first analyzer, reacted and the product ions are then analyzed in the second analyzer. Although perhaps more than 30 configurations of hybrids have been explored, only a few have been commercialized and remain popular, such as Q-TOF, Ion trap-TOF, and LIT-Orbitrap (2012 Paulo). In early years, most research was devoted to MS/MS sector instrumentation. However, in 1978, Yost and Enke first published MS/MS results in a triple quadrupole system – a key development in hybrid MS/MS
mass spectrometry. The next year, upon the comparison of different performance characteristics of various analyzers, it was observed that the MS/MS collision in QqQ hybrid instruments required lower energy but was more efficient than the collision used in Sector instruments (2008 Glish). Later, in the 1980’s, Glish and Goeringer at Oak Ridge National Laboratory combined a quadrupole mass filter with a TOF analyzer (Q-TOF) (1984 Glish), which has later become one of the most popular mass spectrometers currently used due to the main advanced design of Orthogonal Acceleration that is placed between the quadrupole and TOF analyzer for minimizing the effect of different ion entrance velocities, leading to a high resolution. As the quadrupole is operated in RF-only mode, all ions are passed from the ion source as an ion beam to the orthogonal acceleration region, where a acceleration pulse is applied orthogonal to the direction of the entering ions to send the ions down the flight tube. This Q-TOF hybrid design exploits the higher MS/MS efficiency, which was demonstrated for QqQ in early years, and the high performance in sensitivity, resolution, and accuracy of a TOF analyzer. Subsequently, TOF was coupled to ion trap analyzers, very similar to Q-TOF, to offer high MS^n capability and high resolving power and mass accuracy at the same time. The most recent type of commercialized hybrid instrument is LIT-Orbitrap with an ion collector “C-trap” between the two analyzers. Instead of a continuous stream, the ions are accumulated in LIT while another packet of ions is collected by the C-trap and ejected into Orbitrap (2008 Perry).

1.3.3 MS techniques applied to this thesis – MALDI-LIT MS

Most of the research presented in the thesis is carried out by MALDI-linear ion trap (LIT) MS. In early years, the MALDI technique was typically combined with TOF because of the pulsed nature of MALDI that easily interfaces with pulsed TOF measurements, especially with implementation of delayed-ion extraction (1995 Brown; 1995 Vestal) and, which allow for the better reproducibility, accuracy, and resolution. Nowadays, MALDI is coupled to nearly every MS analyzer
available for various experiments, utilizing its high tolerance for contaminants, simple sample preparation, high sensitivity and wide analytical mass range. Although abundance of sample preparation methods and a wide range of matrices have been proposed and applied to MALDI analysis, the detailed mechanism is not totally understood and still being investigated. It is generally accepted that MALDI is a soft ionization with a combination of desorption and ionization (2006 Knochenmuss) divided into two steps. The first step is to form gas-phase primary ions, i.e. protonated matrix ([M+H⁺]); the second step is to involve proton transfer from the primary ions to an analyte (A) through gas phase ion-molecule reaction to produce protonated analyte secondary ions, i.e. [M+H⁺] + A \rightarrow M + [A+H⁺].

MALDI is an extremely complex ionization, involving many desorption/ionization events that occur in a micro to nanosecond during or shortly after laser ablation. Over the last decade, scientists have proposed differing hypotheses that attempt to explain the complicated ionization mechanism in a somehow unpredictable manner. The mechanism for the formation of ion-molecule secondary ions is less controversial than the formation of primary ions. The mechanism for the formation of primary ions proposed in the literature mainly encompasses two models, the gas-phase proton transfer model (2006 Knochenmuss; 1987 Karas) and the cluster (lucky survivor) model (2000 Karas; 2006 Alves; 2002 Fournier).

In the gas-phase proton transfer model, desorption is initiated by photoionization, where matrix absorbs photon energy from laser ablation and transfers it into excitation energy causing the formation of matrix primary ions. However, the mechanism is insufficient to explain how the high-mass protein molecules can be desorbed from the matrix and suspended into the gas phase. In the cluster model, similar to the mechanism of ESI proposed by John Fenn, upon the laser ablation, the gas-phase charged clusters may be produced by ion evaporation mechanism described by Iribarne and Thompson (1967 Iribarne). The most notable assumption in the cluster model is that all the
ions may exist in the sample before laser ablation; during ablation, the desorbed cluster ions are re-neutralized in gas phase, but only a few “lucky survivor” ions can survive as singly charged ions detected. Yet, a problem is why the formation of low charged ions is much more favorable than that of highly charged ions during desorption if proteins can carry several charges, which was commonly observed in ESI (Figure 1.4) (2012 Trimpin). Recently, both models were discussed by Jaskolla and Karas, reporting that both of the gas-phase proton transfer pathway and the lucky survivor pathway are true under multiple investigations, and both pathways may be part of one overall MALDI mechanism (2011 Jaskolla), but the details still remain problematic.

The typical MALDI sample preparation involves dissolving a small organic acid (mostly aromatic compounds used for efficient energy absorption during laser ablation) at a high concentration in an organic solution, mixing sample solution with excess matrix, spotting the mixture of sample and matrix on a MALDI plate, and allowing the sample to co-crystalize with the matrix prior to analysis. Several important factors determining a successful MALDI analysis in sample preparation steps are choice of matrix and matrix solvent, the ratio of matrix to sample, concentrations of samples, consideration of analyte nature (i.e. the compatible solubility of matrix and analyte), homogeneity of spots, and possible contamination that affects signal-to-noise ratios on spectra. A wide variety of matrices have been used in MALDI dependent on specific analytes. Three common matrices for peptides and proteins are α-cyano-4-hydroxycinnamic acid (CHCA), 2,5-dihydroxybenzoic acid (DHB), and sinapinic acid (SA) (Figure 1.5).
Figure 1.4 Electrospray Ionization (ESI) and Matrix Assisted Laser Desorption Ionization (MALDI). (A/B) an ESI/MALDI mass spectrum of a protein. (C) charge states and corresponding mass-to-charge ratios (m/z) of the protein.

Figure 1.5. Typically used matrices in MALDI-MS.
Linear Ion Trap analyzer (LIT)

Before the development of LIT, the ion-trap analyzers were designed as a 3D-cylindrical symmetry mass analyzer (3D-QIT), roughly the size of a tennis ball, made up of a central ring electrode and two end caps on the top and bottom (1990 Paul; 1991 Cooks; 1997 Jonscher) (Figure 1.6). In the 3D trap, ions of a particular range of \( m/z \) can be trapped in an oscillating electric field around the inner electrode with the damping gas (helium) by applying a RF oscillating potential to the ring electrode (the end cap electrodes are held at an offset DC potential to prevent ions from escaping axially) for a certain time interval. The presence of the helium gas in the trap is to stabilize the ion trajectories and, if MS² is employed, to collide with the molecule ions. After that, the ratio frequency amplitude is ramped (an additional auxiliary radio frequency voltage, 1/3 frequency of the main voltage, is applied) over a period of time to destabilize the ion trajectories and then eject the ions toward the detector.

![Figure 1.6 Three-dimentional ion trap configuration with central ring electrode and two end-cap electrodes.](image)
2D-LIT analyzers have evolved recently (2002 Schwartz; 2002 Hager) (Figure 1.7). The physical appearances of the LITs are more closely related to those of quadrupoles, but the LITs are still designed based on the principle of the QITs. As shown in Figure 1.7, the quadrupole was cut into three axial sections, including the front section, the end section, and the x rod of the central section. The three sections are connected a discrete DC potential and two different phases of RF are applied to each pair of two opposite electrodes, allowing the ion trajectories take the shape of sausage along the z-axis rather than at a single center point of 3D-QIT. This results in greatly increase the trap’s capacity and decrease the space charge effects. Another advantage of LIT is that ions will not be ejected from the front and the end section, but from the x rod of the central section to the two detectors placed axially on either side of the rods. Thus, a large amount of ions entering axially can be detected to reach maximum sensitivity. The LIT analyzer has been shown to successfully overcome two fundamental limitations of the traditional 3D-QITs, ion capacity and trap efficiency. The further combination of LTQ with other high resolution analyzers, such as LIT-TOF (qTOF) and LIT-Orbitrap, has provided powerful tools for biomolecule analysis.

![Figure 1.7 Two-dimentional linear ion trap configuration with the quadrupole rods cut into three axial sections.](image_url)
1.3.4 Tandem Mass Spectrometry (MS/MS)

MS/MS is a technique where ion fragmentation occurs in a collision chamber before the mass spectrum is acquired. The fragment information is useful for peptide/protein sequencing, structural elucidation, analyte identification and quantification through fragment fingerprinting. Several approaches to MS/MS experiments (product ion scan, precursor ion scan and neutral-loss scan) can be conceived in two ways: *in time* or *in space* with collision induced dissociation (CID), the most commonly used method of fragmentation, also called collisionally activated decomposition (CAD) (2002 McClellan). Tandem-in-space MS/MS generally is performed on magnetic sectors and quadruple based mass spectrometers. The processes occur sequentially in separate regions of the instruments. In the first region, which is most often a quadruple filter, specific ions (precursor ions) are induced into the subsequent part of the instrument, a gas-filled collision chamber, for fragmentation. All fragment ions (product ions) further enter into the third section of the instrument to be scanned (see Figure 1.8). Tandem-in-time MS/MS can be performed by an ion trap mass spectrometer with, for example, product ion scan which acquires a mass spectrum of products ions produced by retaining a precursor ion in the trap chamber for a period of time for and colliding the ion with a damping gas, such as helium.

![Figure 1.8 Schematic illustration of tandem mass spectrometry.](image-url)
MS/MS experiments can be classified into three types (precursor ion scans, neutral loss scans, and product ion scans), according to which these species are monitored. During MS/MS, specific ions (precursor ions) are selected based on $m/z$ values and collided with a collision gas into fragments (product ions). Precursor ion scanning is to detect a range of $m/z$ of precursor ions as a particular $m/z$ of the product ions is fixed. Neutral loss scanning is to synchronously scan both precursor and product ions, but offset by a constant $m/z$ value corresponding to a specific neutral loss fragment. Product ion scanning is to scan product ions (fragments) derived from an isolated precursor ion of only a selected $m/z$. Because analyte ions have been isolated from the background ions for detection, MS/MS, is able to strongly reduce the background, leading to a lower detection limit achieved and higher signal-to-noise ratios shown in MS/MS spectra, in most cases. Moreover, tandem-in time MS analyzers (traps) have capability of performing multi-fragmentation (MS$^n$), in which specific ions can be selected for further fragmentation sequentially from the previous MS/MS step. The controllable MS$^n$ experiments are usually applied to structure elucidation due to its power to continually break down any interested analyte fragments into smaller pieces to gain a better understanding of the structure or sequence of a peptide. Thus, tandem MS (MS/MS or MS$^n$) has been widely applied to qualitative and quantitative proteomic analysis with the advantages of high sensitivity and selectivity, providing abundant tandem mass spectra with specific fragmentation information needed for structural identification (1990 Johnson).

A peptide sequence is composed of three types of bonds that can be fragmented along the amino acid backbone: CH-CO, CO-NH, and NH-CH. A nomenclature for peptide sequencing by MS/MS was originally proposed by Roepstorff et al., and later modified by Biemann (1984, Roepstorff; 1988, Biemann). Figure 1.9 illustrates six possible fragment ions for each amino acid residues, caused by the three potential cleavage sites along the sequence backbone. The a, b, and c ions are formed due to charge retention on the N-terminal fragment, whereas the x, y, z ions
originated from the charge retention on the C-terminal fragment. Upon CID fragmentation, the cleavage occurs at the CO-NH bonds, resulting in dominant b and y ions shown on MS/MS spectra. The fragment ions are structurally useful for determining a peptide sequence.

\[
\text{b - y : CID cleavage site}
\]

![Peptide sequence fragmentation diagram](image)

**Figure 1.9** Peptide sequence fragmentation

1.4 Protein identification and phosphorylation analysis in MS

1.4.1 MS protein identification approaches

Two fundamental MS protein identification approaches are termed “top-down” and “bottom-up”. In top-down protein analysis, intact proteins are introduced into MS without proteolytic digestion and are fragmented into smaller peptide cations or neutral fragments. In bottom-up proteomics, purified proteins are treated with a protease to perform proteolytic digestion before being subjected to fragmentation for MS analysis.
Bottom-up proteomics measures exact mass of peptides from digests derived from proteins to identify the corresponding proteins; and determining fragments of the digest peptides to know structural information. The conventional bottom up approaches can be gel-based or gel free. Proteins are first extracted from a sample of cells, and the protein mixture is separated by one or two-dimensional gel or online liquid chromatography (2014 Magdeldin). The two-dimensional gel separates proteins by their isoelectric point and then by its masses using a standard one-dimensional gel. Once the proteins are isolated from one another, the individual protein bands of interest are cut out and digested enzymatically with a suitable enzyme (typically with trypsin) or chemically with a strong acid (such as 12% formic acid) into small peptides (digests) which are then extracted prior to MS analysis. Because the peptides have the better solubility and ionization efficiency, it is possible to readily detect the peptides by MS, compared to the parent proteins (2006 Chait). The mass spectra (peptide mass fingerprinting) from its proteolytic digestion are assigned to proteins by comparison with a sequence database. In addition, with the use of MS/MS, partial amino acid sequences of individual peptides can be obtained to further characterize and confirm the sequence.

With top-down analysis, the proteins remain intact without any digestion before being introduced into MS (2011 Zhang). To determine modifications in the protein sequence, the exact mass measurement of the intact protein is compared to with calculated values based on the DNA-predicted protein sequence; then, the target protein is isolated and fragmented in gas phase by tandem MS for sequence mapping. However, the top-down approach still suffers from several limitations currently. Because high-mass proteins yield complex isotopic distribution patterns shown on the mass spectra, the top-down MS requires a costly high-resolution mass analyzer at high-mass range, such as Orbitrap or FT-ICR, to avoid any overlapping of the distributions. In addition to the complex isotopic distribution, multiply charged proteins generated in ESI after chromatographic separation also results in complex spectra. For intact protein tandem MS, poor dissociation
efficiency (low sequence coverage) has been observed for very large proteins when many common
dissociation processes such as CID are used, and the mechanism of gas phase protein fragmentation
is not fully understood, even though many researchers have turned their attention to develop new
fragmentation methods and the instrumentation (one such method is electron-transfer dissociation
(ETD) that generates c- and z-type fragment ions and preserves unstable side-chain modification
during MS/MS). Moreover, the ionization efficiency for very large proteins is limited, leading to
difficulty of trace analysis of the large protein. For these reasons, most of the protein MS
identification and quantification still relies on bottom-up approaches.

1.4.2 Phosphorylation analysis in MS

Phosphorylation occurs at least 30% of proteins in human (2007 Ubersax). A new phosphate
group induces a conformational change of target proteins that can be activated, deactivated, or
changed in function. Numerous cellular functions are regulated by protein phosphorylation, a
reversible and selective reaction controlled by kinases and phosphatases. The kinases and
phosphatases have been found to be involved in a large number of human diseases, resulting from
their mutation, over expression or malfunction (2001 blume-Jensen). Phosphoproteomics with
recent advances in mass spectrometry (MS) has become an important approach to understand the
entire phosphorylation network, for example, which protein are phosphorylated; which sites are
phosphorylated on the proteins; what is the stoichiometry of these phosho-sites; which kinase
phosphorylates the site; what is the biochemical function of the phospho-event; what is the cell
biological function of the phosphor-event.

Traditionally, to trace phosphorylation, radioactive $^{32}$P is introduced into cellular proteins via
the transfer of $\alpha$-phosphate of labeled ATP by kinases (1986 Garton). Another method is to utilize
antibodies against phosphorylated proteins by western blotting (1999 Dimmeler). The probing methods are sensitive and highly selective; however, they provide no information of specific phosphorylation sites, unless specific antibodies are produced against each phosphorylation site on a protein, and phosphorylation on differing sites has to be determined individually, which can be prohibitively expensive. MS is a powerful tool for probe-free phosphorylation analysis in which the attached phosphate group can be observed on mass spectra with a mass shift of +80 Da. The specific mass difference between modified and unmodified proteins suggests the presence of protein modifications (2004 Chang). A number of MS strategies have been developed to perform phosphorylation quantification, yet the technique is being explored to find better and more efficient quantification approaches.

In phosphoproteomic analysis, the phosphorylated proteins first are isolated from tissues or cells, digested into peptides subjected to bottom-up analysis, or directly subjected to top-down analysis by mass spectrometry for site-specific identification and quantification. However, the dynamics and heterogeneity of protein phosphorylation increase complexity of the phosphoproteomic analysis, leading to the challenge of deciphering phosphorylation cellular signaling network. For instance, the same residue may be phosphorylated by several kinases through a variety of mechanisms at different time. Furthermore, the phosphorylation level of a substrate fluctuates dynamically in the cellular network, i.e., some proteins containing phosphorylated residues may be found to be strongly phosphorylated, weakly phosphorylated or un-phosphorylated as functional proteins in cells (2013 Tyanova). To construct the detailed phosphorylation signal pathways, it is therefore important to develop methodologies for identification and quantification of the dynamic and heterogeneous phosphorylation.
1.4.3 MS-based quantitative analysis

Measurement of phosphorylation stoichiometry is biologically important for understanding the functional dynamics of phosphorylation regulation in a cell. To do this, MS has been regarded as a powerful tool not only for protein identification but also protein quantification. MS-based phosphorylation quantification can be classified into two major strategies: absolute quantification, which determines the exact amount of proteins in the sample a quantified standard; and relative quantification, which evaluates phosphorylation changes with experiments. Unfortunately, MS is not inherently quantitative (2009 Doerr), because ionization efficiency of many peptides is different in a given sample and the instrument responses are unpredictable. Many quantitative MS approaches have also been developed to overcome this limitation. For instance, protein or peptide samples can be isotopically labeled and the abundance of the labeled target can be relatively or absolutely compared to the abundance of the nonlabeled or differently labeled target. Another quantitative approach similar to the labeling method is called isotope dilution, which is used by spiking $^{13}$C- or $^{15}$N-stable isotope analogs of peptide analytes in defined amount as internal standard before MS analysis. Besides, due to a strong demand for a lost cost and rapid analysis, MS label-free approach has been recently proposed to quantify protein or peptide amounts in different samples in response to the changes in signal intensities.

Isotope labeling methods

MS-based relative quantitative proteomics can be achieved with comparison of two or more samples by using stable isotope labeling methods. Some of the recently developed labeling methods are: Stable Isotope Labeling by Amino acids in Cell Culture (SILAC) (1999 Oda), Isobaric Tag for Relative and Absolute Quantitation (iTRAQ) (2004 Ross), Stable Isotope Dimethyl labeling methods (2009 Boersema). The SILAC method relies on the metabolic incorporation of a “heavy” or “light”
form of a given protein or peptide with substituted $^{13}$C$_6$-Lys and $^{13}$C$_6$, and $^{15}$N$_3$-Arg residues. In SILAC experiments, cells are grown in cell media where the control cell medium contains the natural (light) proteins; whereas, the reaction medium (i.e. kinases treated) is prepared where the light amino acids are replaced by heavy SILAC amino acids. The two cell populations are then mixed prior to MS analysis, where a difference in MS signals between light and heavy masses of peptides represents the quantitative difference in the target proteins between control and reaction medium. The ratio of the light-to-heavy mass response is utilized for the relative proteomic quantification (i.e. phosphorylation quantification).

The iTRAQ method uses tandem mass spectrometry (MS/MS) and isobaric labeling on the primary amines of peptides and proteins. The iTRAQ reagents are four isobaric tags, which have an identical total mass of 145 Da, consisting of three elements: a peptide reactive group (NHS ester) for N-terminal labeling of peptides, a balance group (28-31 Da) and a reporter group of variable mass ($m/\chi$ 114-117) with difference in the stable isotope composition of the reporter groups and the balance groups, such as $^{13}$C, $^{18}$O and $^{15}$N (Figure 1.10) (2004 Ross). The reagents tag samples individually by replacing the peptide reactive group with primary amines of peptides and proteins prior to MS/MS analysis, where the reporter group falls off from each tagged compounds, producing different reporter fragment ions shown on the same mass spectrum for quantification. The pattern of the reporter fragments shown as differing masses allows multiplexing four different samples, and all samples could be identified and quantified in one tandem mass spectrometric run to achieve a high throughput analysis (Figure 1.11).
Figure 1.10 The structure of iTRAQ reagent composed of a reporter group, balance group and reactive group.

Figure 1.11 MALDI-MS workflow with iTRAQ labeling \((m/z \ 114 – 117)\)
In dimethyl labeling experiments, protein samples are first treated with protease, and free amines on the derived peptides of different samples are labeled with normal or deuterated formaldehyde. The labeled samples are then mixed and analyzed by LC-MS whereby the change in mass of dimethyl labels is used to compare the peptide abundance in different samples. Demethyl labeling is based on dimethylation of primary amines (the N terminus and the side chain of Lysine residues) in a peptide mixtures with formaldehyde to convert the primary amines to dimethylamines. It is observed that the differential peak mass by 28 Da for each derivatized sites and by 4 Da for each isotopic dimethyl labels (Figure 1.12) (2009 Boersema). This method is rapid and economic to globally detect and quantify phosphorylation by comparing the signal response with a mass shift by labeling.

![Labeling reactions of dimethyl labeling.](image)

**Figure 1.12** Labeling reactions of dimethyl labeling.

**Isotope dilution methods**

While isotope labeling methods mostly provide relative quantification, MS-based absolute quantification is also possible for determining absolute abundance of phosphopeptides. Isotope dilution has long been for small molecule quantification by the addition of defined quantities of isotope standards as ideal internal standards, which have the chemical behaviors identical to the
analytes. In the past decade, this approach has been transferred to absolute quantification of proteins and peptides in biological samples. In 2003, Gygi et al. presented an absolute quantification (AQUA) strategy in quantitative proteomics where proteins are proteolyzed with trypsin in the presence of the synthesized isotope standards (AQUA peptides), followed by extraction, chromatographic separation, and analysis using selected reaction monitoring (SRM) on triple quadrupole mass spectrometers (2003 Gygi). The absolute phosphorylation quantification of a tryptic peptide containing a specific phosphorylation site can be achieved by utilizing the ratios of native and corresponding AQUA peptide chromatographic peak areas. A similar strategy was described by Ciccimaro et al. to absolutely quantify phosphorylation of kinase activation loop of cellular focal adhesion kinase (FAK) with addition of a labeled phosphorylated FAK standard and a labeled control peptide along with the edogenous FAK before cell lysis and protein digestion, respectively (2009 Ciccimaro). It is possible to calculate amount of phosphorylation of the edogenous FAK by comparing the observed amount of the edogenous FAK to the amount of the control peptide with corrections for immunoprecipitation efficiency during protein isolation from cell lysate. Overall, this labeling dilution approach is, however, not used for a large scale analysis and limited to the individual analyte, because of the difficulty of obtaining these internal standards, especially when a large number of proteins have to be quantified.

**Label-free methods**

There has been growing interest in the use of label-free approaches to quantitative mass spectrometry. Label-free phosphoproteomics allows samples to be directly and separately analyzed by MS without merging (mixing of unlabeled and labeled samples), resulting in a higher coverage of a dynamic range. It also has been reported that the measurements are linearly proportional to the concentrations of interest peptides in the range of 10 fmol to 100 pmol (2012 Nilsson). Most
importantly, a large amount of samples can be analyzed at the same time without the limit of the number of labeling reagents.

There are two ways of determining the abundance of a peptide or protein in label-free quantitative phosphoproteomics: (1) based on intensity changes in peptide peaks, (2) the spectral counting of analytes after LC-MS/MS analysis. The former methods are conducted by comparing ratios of intensities of phosphorylated to nonphosphorylated peptide ions. For example, Steen et al. developed a label-free phosphorylation quantification method by using the measurements of normalized ion intensities of phosphopeptide relative to nonphospho-tryptic peptides from the target protein (2005 Steen). This method was further modified by Yi et al. to normalize the peak-area of the reconstructed ion chromatogram for the target phosphopeptide against the peak areas of other nonphospho-tryptic peptides as endogenous internal standards (2007 Yi). The normalized peak areas for untreated and treated samples were compared for relative quantification of phosphopeptides. However, this method may be problematic for quantification of peptides containing adjacent phosphorylation sites regulated more than one kinases.

The same research group, therefore, reported the latter method by monitoring MS/MS product ions instead of parent ions for label-free phosphorylation quantification. The samples were first tryptically digested, chromatographically separated, and then analyzed in MS. The intensities of the MS/MS or even MS² fragments, in which b and y ions are unique to each phosphorylation isoform, were applied to detect phosphorylated tryptic peptides in a linear and reproducible manner (2010 Langlais). In general, the major limitation of label-free quantification is the requirement of extra care of intersample preparation and purification for reducing matrix effect prior mass spectrometric analysis; that explains why chromatographic separation coupled to ESI-MS has been considered to be the most desirable platform for the label-free analysis in spite of the complexity of its and time-consuming compared to MALDI-MS.
Chapter 2: Development of label-free quantitative proteomic analytical method by using matrix-assisted laser desorption/ionization (MALDI)-tandem mass spectrometry

2.1 Introduction

2.2 Label-free MALDI-MS for single-site phosphorylation quantification
   2.2.1 Materials and methods
   2.2.2 Results and discussion
      2.2.2.1 Mathematic approach for phosphorylation quantification
      2.2.2.2 Consistency of instrument response factor, \( a \)
      2.2.2.3 Effect of laser energy on phosphorylation quantification
      2.2.2.4 Linearity, dynamic range and reproducibility
      2.2.2.5 Validation of phosphorylation quantification by isotope dilution MALDI-MS
   2.2.2.6 Application
   2.2.3 Conclusion

2.3 Label-free MALDI-MS for double-site phosphorylation quantification
   2.3.1 Materials and methods
   2.3.2 Results and discussion
      2.3.2.1 Mass spectra of phosphorylated Aquaporin-2 isomers (pSer-256 and pSer-261)
      2.3.2.2 Mathematic Approach for double phosphorylation site quantification
      2.3.2.3 Consistency of MS\(^3\) fragment ratios of p261, \( \beta \)
      2.3.2.4 Linearity, dynamic range and accuracy
      2.3.2.5 Effect of collision energy on double phosphorylation quantification
      2.3.2.6 Applications to in vitro kinase assays
   2.3.3 Conclusion

2.1 Introduction

Phosphorylation is one of the most important post-translational modifications. Many proteins have been known to be phosphorylated at numerous sites by kinases, resulting in a change in function. To understand the special characteristics of phosphorylation requires analytical methods to quantify either single or multiple phosphorylation sites within a substrate. Many methodologies have been used to identify and quantify phosphorylation, such as western blotting, radioactive phosphate labeling, flow cytometry, and mass spectrometry (1999 Kreegipuu, 2002 Zhang, 2008 Ma). Mass spectrometry (MS) has become an irreplaceable technique in phosphorylation analysis, in which a signal corresponding to the phosphate moiety can be detected on the mass spectrum, because it has the capability to readily determine site-specificity, and does not require the generation
of sequence-specific antibodies. The general workflow for MS based quantitative proteomic analysis relies heavily on HPLC-ESI-MS technique in the combination with either labeling or labeling-free method (shown in Figure 2.1A). Calibration curve construction generally is required, especially when label-free methods are applied. Both the labeling and label-free methods for absolute and relative quantification have pros and cons, as mentioned above in chapter 1. For instance, the labeling methods are limited to high cost and less flexibility, but are generally considered more reliable. Recently, label-free methods have attracted growing attention due to their simplicity, reduced sample preparation and high analytical dynamic range; however the procedures of sample preparation and chromatographic separation are a requisite for ESI-MS or ESI-MS in the workflow to reduce its spectral complexity and ion suppression effect. To overcome this conflict, MALDI has great potential to replace the current ESI strategy to build up a simpler workflow for proteomic analysis (Figure 2.1B).

ESI and MALDI are the most common ionization techniques generally used in protein mass spectrometry (2011 Kuklenyik). ESI can produce multiple charged ions that could be problematic while analyzing protein or peptide mass spectra due to the spectral complexity. Thus, for most phosphoproteomic analyses, ESI-MS must be coupled to liquid chromatography to simplify the complex spectra (2011 Zhang). Compared to ESI, MALDI has more tolerance to buffer or impurities in the sample, so it requires only simple sample preparation, is generally fast and easy to use, and yields simpler spectra (singly charged peptide ions) that ordinarily do not require chromatographic separation. In spite of this, MALDI is inherently less quantitative due to unpredictable difference in ionization efficiency and in absolute ion intensity arising from heterogeneities in analyte/matrix crystallization.
Figure 2.1 General workflow for proteomic quantification. (A) Current workflow with ESI. (B) A simpler workflow proposed with MALDI in the study.
To minimize the quantitative problems, isotope derivative methods are generally required in quantitative proteomic MALDI by taking advantage of the fact that isotopomers respond identically under MS conditions, but are able to be separated by mass. It has been recently demonstrated that MALDI combined with isotope labeling methods is feasible to protein quantification in proteomic analysis (2009 Johnson). While effective, such isotopic labeling methods are limited in their multiplexing capacity, high cost, requirement for specialized reagents and a narrow dynamic range (2012 Gunasekera, 2010 Gouw, 2010 Boersema, 2009 Ye). In addition, the inherent limitations in reproducibility of MALDI-MS can at least in part be overcome by increasing the signal acquisition time. The signal from a single spot can be interrogated for several minutes to average the signal from thousands of laser shots, as compared to the smaller number of scans generally available when limited by a chromatographic time scale in ESI-LC-MS.

Label-free strategies have been increasingly prominent as alternatives for quantitative proteomics due to their lower cost and convenience. Liquid chromatography MS (LC-MS) has been one of the fundamental techniques in the label-free relative-quantitative proteomics (2007 Levin, 2008 Asara). Chromatographic retention peaks can be integrated with external calibration to measure peptide abundance (2010 Zhu). However, the chromatographic task is time-consuming and may suffer from interference problems. Mobile phase conditions must also be carefully controlled to avoid changing the solubility and ionization of the peptide analytes. As an alternative to chromatography based label-free methods, label-free MALDI-MS methods are in rising demand for high-throughput proteomic analysis. Kinumi et al. has successfully measured phosphopeptides derived from the phosrestin protein by using a label-free MALDI approach with calibration construction by plotting signal ratios versus peptide amount ratios, quantifying the ratio of phosphopeptide to total peptide (2005 Kinumi). The calibration curve showed a linear correlation between the peak ratios and amount ratios. In other cases, the unphosphorylated and
phosphorylated substrates have different ionization efficiencies, resulting from their differing physicochemical properties. For example, a study by Parker, et al. demonstrated non-linear relationships between signal ratios and amount ratios for seven different peptides using MALDI ionization (2008 Parker). This may be due to differential ionization effects between the different peptides. Therefore, in this chapter, to solve the problems, we aim to develop MALDI label-free mass spectrometric strategies as alternatives for quantitative single or double-site phosphoproteomics with the advantage of lower cost, simplicity and convenience.

2.2 Label-free MALDI-MS for single phosphorylation quantification

In this study, we present a new MALDI-MS method for quantification of phosphorylation stoichiometry, termed Double Reciprocal Isotope-free Phosphopeptide Quantification (DRIP-Q). This method does not require isotope labeling, chromatography, or construction of calibration curves. Instead, a double reciprocal transformation is applied that accounts for differences in ionization efficiencies of analytes, allowing for the direct comparison of unphospho- and phosphopeptide signals. The method is shown to be robust and reproducible over a wide range of analyte concentrations using three different peptide models encompassing phosphoserine, phosphothreonine and phosphotyrosine. The results obtained with this method are comparable to those obtained using traditional stable isotope labeling. To our knowledge, this is the first report of a calibration-curve-free MALDI-MS method for the determination of phosphorylation.

2.2.1 Materials and methods

Materials

HPLC grade water, acetonitrile, and formic acid were purchased as HPLC grade from Fisher Scientific (Pittsburg, PA, USA). Synthetic peptides, AAAAYRAAR, LRWGFTTPDKKHQKEPPF,
and RQSVELHSPQSLPR, were used for the quantitative method development. The peptide, AAAAYRAAR, and its phosphorylated form, AAAApYRAA, and their \([^{13}C_3]\)-alanine labeled forms, were synthesized and purified by University of Utah DNA/peptide facility. The HIV-RT peptides, LRGWFTTPDKKHQKEPPF, containing a Cdk phosphorylation consensus sequence (S/T)PXR/K, and its phosphorylated form, LRGWFTpTPDKKHQKEPPF, were synthesized in house using standard fmoc-based solid-phase peptide chemistry. The AQP2-peptide, RQSVELHSPQSLPR, and its two phosphorylated forms at pSer-256 and pSer-261, RQpSVELHSPQSLPR, and RQSVELHpSPQSLPR, were purchased from Anaspec (Fremont, CA).

**MALDI-MS quantitative analysis**

Peptide standards, AAAAYRAAR (peptide 1), LRGWFTTPDKKHQKEPPF (peptide 2), and RQSVELHSPQSLPR (peptide 3), and their phosphopeptides were selected as three different model systems to establish the label-free MALDI-MS quantitative method. 1 μL of the sample mixture of the unphospho- and phosphopeptide standards with various phosphorylation fractions was mixed with 9 μL of matrix solution. Then, 1 μL of the peptide mixture with the matrix was spotted on a MALDI plate prior to MALDI-MS analysis. MALDI-MS experiments were carried out on a Thermo LTQ XL linear ion trap mass spectrometer (Thermo Scientific) was equipped with a MALDI ion source. α-Cyano-4-hydroxycinnamic acid (CHCA) was purchased Sigma-Aldrich (St. Louis, MO, USA) from 2,5-Dihydroxybenzoic acid (DHB) was purchased from Acros Organics (Morris Plains, NJ, USA) and used without further purification. A saturated aqueous solution of α-Cyano-4-hydroxycinnamic acid (CHCA) (Sigma-Aldrich, St. Louis, MO, USA) in 0.1 % trifluoroacetic acid/acetonitrile (30/70, v/v) or 2,5-Dihydroxybenzoic acid (DHB) (Acros Organics Morris Plains, NJ, USA) mixed with water / acetonitrile (50/50, v/v) in a 1:4 ratio (v/v) was used as a matrix solution. Full scan mass spectra were acquired in positive mode by 400 scans with
automated gain control (AGC) set at 300,000 counts with automated spectrum filter (ASF) off, 3 microscan/step, and processed by Xcalibur™ 2.0.7 software.

**In vitro Phosphatase assay**

The phosphopeptide (AAAApYRAAR) was applied to the phosphatase assay, which was performed in a final volume of 50 μL, containing 25mM of Ammonium bicarbonate buffer, 4 Units of Alkaline Phosphatase (Roche), and 1.56 nmole of the phosphopeptide substrate at 30 °C for 1 h. The reaction was quenched by adding 20% acetic acid into the reaction sample, and heating it for at 90 °C for 10 min. A control (inactivated phosphatase) experiment was carried out by adding 20 μL of 20% acetic acid and heating at 90 °C for 10 min before adding the substrate, so that the phosphopeptide remained. The reaction and control samples were diluted and mixed in different phosphorylation fractions. 1 μL of 2.83 – 0.31 pmole peptide substrates was subjected to MALDI-MS analysis.

**Isotope dilution MALDI-MS**

[13C3]-alanine labeled unphosphopeptide U* (AAA*AYRAAR) and labeled phosphopeptide P* (AAA*ApYRAAR) phosphorylated forms served as internal standards. The labeled standards were mixed with the unlabeled standards (U and P) in 1:1 ratio with the amount ranging from 0.1 - 20 pmole/μL and analyzed by MALDI-LTQ. The amount of P and U in unknown samples was determined by using calibration methodology. Calibration curves were constructed for each labeled pair by plotting signal ratios of labeled to unlabeled peptides versus their amount ratios (i.e. the signal ratios of P* to P versus their amount ratios; the signal ratios of U* to U versus their amount ratios) (2011 Kuklenyik).

**In vitro kinase assays**
To investigate Cdk inhibitor’s activity, Cdk2/CyclinA was assayed using 1.4 μM HIV-1 RT peptide LRWGFTTPDKKHQKEPPF in the presence of 2 mM ATP and 1x NEBuffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl₂, and 1 mM DTT, New England BioLabs Inc.) at 30°C for 15 min with the addition of Cdk inhibitors, Roscovitine at a concentration of 0.1 μM to 5 μM or Olomoucine, at a concentration of 1 μM to 50 μM. A control experiment was performed with no inhibitor addition. Following the incubation, 1 μL of each sample was added into 9 μL of the matrix solution for MALDI analysis. IC₅₀ values are calculated by nonlinear regression analysis using GraphPad Prism (version 6 Prism Software Inc.).

2.2.2 Results and Discussion

MS is a fast, high-throughput, and probe-free method for phosphorylation quantification. Both HPLC-ESI-MS and MALDI-MS can be used for phosphorylation quantification. HPLC-ESI-MS is considered to be more reproducible for phosphorylation quantification than MALDI-MS, even though MALDI is advantageous due to its analytical speed and accommodation of more complex samples. When used with isotope dilution quantitative strategy, MALDI-MS shows improved precision and reproducibility for peptide quantification, whereas isotope-free MALDI methods still face some challenges. For instance, phosphorylation may cause a change of ionization efficiency, which will result in non-linearity between the signal ratios and actual mole ratios. This condition has been previously observed in calibration curves for isotope-free MALDI-MS. To solve the problem requires considering the difference of ionization efficiencies in quantification. Here, we propose a new mathematic approach, called “Double Reciprocal Isotope-free Phosphopeptide Quantification (DRIP-Q)” to find a linear correlation between signals and actual amounts for phosphorylation quantification.
2.2.2.1 Mathematic Approach for phosphorylation quantification

The DRIP-Q approach proposed here accounts for possible difference in ionization efficiencies for a phosphopeptide (P) and its unphosphorylated analog (U). It requires the assumption (which we later test experimentally) that signal intensity is directly proportional to sample concentration, where the proportionality constant is defined as the ionization efficiency, I. Therefore,

\[ S_P = I_P \times M_P \]  \hspace{1cm} (Eq. 1a)

\[ S_U = I_U \times M_U \]  \hspace{1cm} (Eq. 1b)

where \( S, I \) and \( M \) represent observed mass spectral signal intensity, ionization efficiency and amount in moles, and the subscript \( P \) or \( U \) represents the phosphorylated peptide or its unphosphorylated analog.

Eqs. 1a and 1b can be combined to give:

\[ \frac{S_P}{S_U} = \frac{I_P}{I_U} \frac{x}{(1-x)} \]  \hspace{1cm} (Eq. 2a)

where \( x \) is the mole fraction of phosphopeptide \( P \), and \( (1-x) \) is the mole fraction of \( U \). For example, if \( x = 0.1 \), 10% of the total peptide is phosphorylated.

Letting the signal ratio, \( \frac{S_P}{S_U} = y \), and the instrumental response factor ratio \( \frac{I_P}{I_U} = a \), the equation becomes:

\[ y = a \times \frac{x}{(1-x)} \]  \hspace{1cm} (Eq. 3b)

Converting \( y = f(x) \) into \( x = f(y) \), yields:
\[ x = \frac{y}{(a+y)} \quad \text{(Eq. 3c)} \]

Taking the reciprocal of both sides of equation gives

\[ \frac{1}{x} = a \times \frac{1}{y} + 1 \quad \text{(Eq. 3d)} \]

A plot of \( \frac{1}{x} \) versus \( \frac{1}{y} \) yields a straight line with a y-intercept of 1 and a slope of \( a \). Therefore, once the response factor ratio, \( a \), has been determined, the phosphorylation fraction \( x \) could be calculated by using the observed signal intensity ratio \( y \) of phospho- to unphospho-peptide. This method only requires determination of \( a \) and \( y \), and therefore avoids time-consuming calibration curve construction.

2.2.2.2 Consistency of instrument response factor, \( a \)

For Eq. 3d to hold, the response factor ratio, \( a \), must be constant. To investigate the linear correlation between the reciprocals of the phosphorylation percentage \( \frac{1}{x} \) and signal ratios \( \frac{1}{y} \) in Eq. 3d, the response factor ratio was tested for three peptide models over a range of phosphopeptide to unphosphopeptide concentration ratios. The three models (Table 2.1, data shown in Figure 2.2, ) include peptides that may be phosphorylated at tyrosine, threonine or serine.

For each model, the phosphopeptide and unphosphopeptide were mixed at nine different mole fractions of phosphopeptide \( (x) \), and the response factor ratio \( a \) was determined for each mixture from the observed mass spectrometric signal intensity ratio \( y \). The response factor ratio is found to be consistent across the range of concentrations for each model (Table 2), with RSDs for \( a \) varying from 2.72\% to 4.10\% for the three models, confirming the linear correlation between \( \frac{1}{x} \)
and \( \frac{1}{y} \) in Eq. 3d. These results demonstrate the potential for applying Eq. 3d to quantification of phosphorylation by isotope-free MALDI-MS.

### Table 2.1 Instrument response factor (\( a \)) at different phosphorylation fractions (\( x \))

<table>
<thead>
<tr>
<th>( x^a )</th>
<th>( 1^b )</th>
<th>( 2^c )</th>
<th>( x^d )</th>
<th>( 3a^d )</th>
<th>( 3b^d )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.670 ± 0.027</td>
<td>0.189 ± 0.016</td>
<td>0.32</td>
<td>0.723 ± 0.034</td>
<td>1.032 ± 0.073</td>
</tr>
<tr>
<td>0.20</td>
<td>0.671 ± 0.029</td>
<td>0.176 ± 0.014</td>
<td>0.38</td>
<td>0.716 ± 0.015</td>
<td>1.017 ± 0.074</td>
</tr>
<tr>
<td>0.30</td>
<td>0.672 ± 0.037</td>
<td>0.177 ± 0.014</td>
<td>0.43</td>
<td>0.691 ± 0.038</td>
<td>1.062 ± 0.105</td>
</tr>
<tr>
<td>0.40</td>
<td>0.634 ± 0.009</td>
<td>0.180 ± 0.006</td>
<td>0.48</td>
<td>0.731 ± 0.042</td>
<td>1.022 ± 0.045</td>
</tr>
<tr>
<td>0.50</td>
<td>0.669 ± 0.005</td>
<td>0.189 ± 0.014</td>
<td>0.54</td>
<td>0.763 ± 0.047</td>
<td>1.006 ± 0.095</td>
</tr>
<tr>
<td>0.60</td>
<td>0.639 ± 0.032</td>
<td>0.186 ± 0.017</td>
<td>0.63</td>
<td>0.732 ± 0.024</td>
<td>1.023 ± 0.086</td>
</tr>
<tr>
<td>0.70</td>
<td>0.658 ± 0.056</td>
<td>0.182 ± 0.008</td>
<td>0.74</td>
<td>0.711 ± 0.038</td>
<td>1.107 ± 0.034</td>
</tr>
<tr>
<td>0.80</td>
<td>0.600 ± 0.023</td>
<td>0.199 ± 0.017</td>
<td>0.91</td>
<td>0.787 ± 0.086</td>
<td>1.107 ± 0.131</td>
</tr>
<tr>
<td>Average</td>
<td>0.652</td>
<td>0.184</td>
<td>0.733</td>
<td>1.047</td>
<td></td>
</tr>
</tbody>
</table>

\( a \) phosphorylation fraction  
\( b \) peptide 1 and phosphopep.1 (AAAApYRAAR, AAAApYRAA) within the range of 0.31 to 2.83 pmole  
\( c \) peptide 2 and phosphopep.2 (LRWGFTpTPDKKHKLPFF, RWGFTpTPDKKHKLPFF) within the range of 0.14 to 1.26 pmole  
\( d \) peptide 3, phosphopeptide 3a, phosphopeptide 3b (RQSVELHpSPQSLPR, RQpSVELHpSPQSLPR, RQSVELHIpSPQSLPR) within the range of 10.0-90.0 fmole  
\( e \) \( n = 3 \)
Figure 2.2 Mass spectra of the experiments in Table 1, in which response factor ratio $a$ was determined for each phospho- and unphosphopeptide mixtures from the observed mass spectrometric signal intensity ratio $y$ for three models in table 1, pep.1 AAAAYRAAR / AAApYRAA, pep.2 LRWGFTTPDKKHQKEPPF / RWGFTTPDKKHQKEPPF, pep. 3a RQSVELHSPQSLFR / RQpSVELHSPQSLFR, and pep. 3b RQSVELHSPQSLFR / RQSVELHpSPQSLFR.
2.2.2.3 Effect of laser energy on phosphorylation quantification

MALDI mass spectrometric signal response may vary with laser intensity. Changes in the signal response might further affect the signal ratios (\(y\)) and the consistency of \(a\) values in Eq. 3d. For example, low laser intensity may lead to low signal response due to the deficiency in sample ionization, whereas excessively high laser intensity may produce a high signal background, lowering the signal to noise ratio. To investigate the effect of laser intensity on the constant \(a\) in Eq. 3d, peptide 2 in its unphosphorylated and phosphorylated forms were mixed at various phosphorylation fractions (\(x\)), and were analyzed by MALDI with laser energies of 8 and 13 µJ. For each mixture, the response factor ratio \(a\) value was calculated from the observed signal ratio (\(y\)) and the known phosphorylation fraction (\(x\)). Average \(a\) values were 0.23 ± 5.70% RSD and 0.16 ± 4.40% RSD for 8 and 13 µJ, respectively (Table 2.2). These results indicate that changes in laser intensity can differentially affect the ionization efficiencies of the un-phosphorylated and phosphorylated peptides, but the ionization efficiency ratio \(a\) is consistent with a small RSD at a constant laser energy across a wide range of phosphorylation fractions. As a result, care should be taken to ensure that the same laser intensity is applied to all samples in an experiment.

To verify that phosphorylation fraction in a sample can be readily determined with the proposed DRIP-Q strategy, the phosphorylation fraction (\(x\)) calculation (Eq 3d) was assessed using the data already acquired as a test case. To do this, an experimental phosphorylation fraction was calculated for each sample by applying the observed peak ratio (\(y\)) and the average \(a\) value from Table 2 to Eq 3d. This was done for both the 8 and 13 µJ laser energy samples. The experimental phosphorylation fractions were then compared to their known values, and accuracy determined as % error (Table 2), obtained on the basis of five experimental replicates per sample. The % errors for the experimental \(x\) values fall in the range of 0.07 % - 6.19 % (overall RMS error of 3.35%), and 0.52 % - 5.33 % (RMS error of 2.43%), for 8 and 13 µJ laser energy, respectively. These error values
suggest that the ionization efficiency ratio \( a \) can indeed be applied to the Eq. 3d for accurate isotope-free MALDI quantification of phosphorylation.

### Table 2.2 The effect of different MALDI laser energy on the instrument response factor (\( a \)) and accuracy (% errors).

<table>
<thead>
<tr>
<th>( x )</th>
<th>8 µJ</th>
<th>13 µJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x ) (observed)</td>
<td>( a ) (observed)</td>
<td>% error</td>
</tr>
<tr>
<td>0.090</td>
<td>0.226 ± 0.017</td>
<td>0.088</td>
</tr>
<tr>
<td>0.110</td>
<td>0.215 ± 0.015</td>
<td>0.103</td>
</tr>
<tr>
<td>0.140</td>
<td>0.217 ± 0.013</td>
<td>0.133</td>
</tr>
<tr>
<td>0.200</td>
<td>0.231 ± 0.010</td>
<td>0.200</td>
</tr>
<tr>
<td>0.330</td>
<td>0.231 ± 0.012</td>
<td>0.330</td>
</tr>
<tr>
<td>0.670</td>
<td>0.244 ± 0.016</td>
<td>0.682</td>
</tr>
<tr>
<td>0.800</td>
<td>0.255 ± 0.011</td>
<td>0.815</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>0.231</td>
<td><strong>RMS (%)</strong></td>
</tr>
<tr>
<td><strong>R.S.D.</strong></td>
<td>5.70%</td>
<td></td>
</tr>
</tbody>
</table>

* a Phosphorylation fraction of the Pep. 2 mixture of 0.13 - 1.12 pmole

#### 2.2.2.4 Linearity, dynamic range and reproducibility

After establishing consistency of the response factor \( a \), across a wide range of phosphorylation fractions, in three different peptide models (Table 1), the same model peptides are further subjected to MALDI-MS analysis to investigate the linearity of the equation. The reciprocals of observed signal ratios \( (1/y) \) and the known phosphorylation fractions \( (1/x) \) are plotted in Figure 2.3, and good linearity of Eq. 3d is shown among all peptide models with \( R^2 \) values above 0.996.

The dynamic range of the method is further determined by evaluating the consistency of the response factors \( (a) \) among pep 2 mixtures of unphosphorylated/ phosphorylated peptides \( (x = 0.5) \) with a serial dilution. A consistent signal ratio \( (1/y) \) was observed within a concentration range of 7 to 0.04 pmole, shown in Table 2.3 (data shown in Figure 2.4). This suggests that the proposed
DRIP-Q method is feasible for phosphorylation quantification with a dynamic range of at least two orders of magnitude, down to 40 fmol of sample on the plate. At levels below this, the spot-to-spot variability in $a$ increases to less reliable levels (not shown).

Figure 2.3. Linearity of double reciprocal equation. The linearity of the double reciprocal procedure is assessed by comparing the experimentally determined and calculated double reciprocal plots using phosphorylated and unphosphorylated analogs of (A) peptide 1 (B) peptide 2 (C) peptide 3A and (D) peptide 3B
To investigate reproducibility, we conducted three experiments in separate days within a two week period under the same instrumental conditions. **Figure 2.5** shows the plots of the concentration curves for analysis of the pep. 2 mixtures with different phosphorylation fractions across the three experiments. Each of the curves was assigned a random range of phosphorylation fractions. Comparison of the curves demonstrates demonstrated good linearity, with $R^2 >0.99$, and good agreement among the three concentration curves, with the mean slope of 0.186 and inter-day CV of 0.65% (n=3).

<table>
<thead>
<tr>
<th>Amount $a$ (pmole)</th>
<th>$S_{un}/S_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.0</td>
<td>6.94 ± 0.35$^b$</td>
</tr>
<tr>
<td>3.5</td>
<td>7.15 ± 0.33</td>
</tr>
<tr>
<td>1.7</td>
<td>7.50 ± 0.90</td>
</tr>
<tr>
<td>0.8</td>
<td>6.93 ± 0.62</td>
</tr>
<tr>
<td>0.4</td>
<td>7.18 ± 0.66</td>
</tr>
<tr>
<td>0.04</td>
<td>7.12 ± 0.55</td>
</tr>
<tr>
<td>Average</td>
<td>7.14 ± 0.21</td>
</tr>
<tr>
<td>RMS error</td>
<td>0.62</td>
</tr>
</tbody>
</table>

*a* the mixture of LRWGFTTPDKKHQKEPPF and LRWGFTpTPDKKHQKEPPF ($x=0.5; 1:1, w/w$) by using 13 μJ laser energy

$b$ standard deviation, n=3
Figure 2.4 Mass spectra of the experiments for dynamic range test in Table 3. Pep 2/p-pep 2 mixtures (LRWGFTTPDKKHQKEPPF/RWGFTpTPDKKHQKEPPF) with phosphorylation fraction (\(x = 0.5\)) in an amount range of 7.0 – 0.04 pmole were analyzed by MALDI-MS.

Figure 2.5 Inter-day reproducibility Double reciprocal plots for peptide 2 obtained from three individual experiments performed on different days in a 2 week period
2.2.2.5 Validation of phosphorylation quantification by isotope dilution MALDI-MS

The previous studies on model peptide systems demonstrate the ability to monitor the amount of phosphorylation in a sample using Eq. 3D. Because consistent \( a \) values can be obtained under a variety of conditions, the DRIP-Q method provides a rapid and low-cost strategy for phosphorylation quantification directly from the observed signal ratios and the \( a \) values without isotopes or chromatography.

In order to validate the proposed DRIP-Q method, it was compared with isotope dilution MS, an established method previously shown to be a robust and precise quantitative MS platform for peptide analysis (2011 Kuklenyik). The isotope dilution strategy utilizes heavy stable isotope-labeled internal standards that are otherwise identical to the analyte; because the analyte and isotopomeric standard have identical chemical structures, they do not suffer from the typical MALDI-related difficulties of heterogeneous crystal distribution and differential ionization. However, its effectiveness as a quantitative method is tempered by the need for cost of producing isotopically labeled standards.

To compare the two methods, mixtures of a peptide \( 1 \), AAAAYRAAR and its phosphorylated form were mixed with 30\% (w/w) and 40\% (w/w) phosphorylation concentrations (\( x = 0.3 \) and 0.4) were subjected to both methods. To perform the isotope dilution analysis, phosphorylated and unphosphorylated versions of peptide \( 1 \) were both synthesized with unlabeled or \( ^{13}\text{C}_3 \)-Alanine at the first position. The labeled and unlabeled peptides were mixed in defined amounts and subjected to MALDI analysis. The results of DRIP-Q method and isotope dilution MALDI-MS method are compared in Table 2.4.
The phosphorylation fraction of each peptide mixture samples was determined at 0.291 and 0.394 in agreement with the results (0.295 and 0.391, p > 0.60) obtained from isotope dilution method. Quantification by isotope dilution yields a smaller standard deviation than the DRIP-Q method (1.7-1.8% RSD vs 4.3-6.3% RSD, p < 0.01), a result that could be expected since the use of isotopic internal standards eliminates some of the intrinsic uncertainty in MALDI quantification. Furthermore, other groups have reported average RSDs of 16.6% and 7.9% for label-free ESI LC-MS method and Isobaric Tags for Relative and Absolute Quantification (iTraq) method, respectively (2012 Wang). While ESI is thought to be more quantitatively reproducible than MALDI, the DRIP-Q method performs comparably to or better than these reported methods, possibly because the MALDI technique allows for longer acquisition time. In this analysis, spectra from approximately 18,000 laser shots were acquired over 5 minutes (laser firing at a repetition rate of 60 Hz), allowing for the averaging of much more data than typically available over the elution time of a peak in LC-MS; indeed MALDI shows substantial improvement in reproducibility as the number of laser shots increases (Figure 2.6). A similar phenomenon is observed by Johnson et al. when evaluating the QconCAT isotope label-based phosphorylation stoichiometry methodology (2009 Johnson). For peptides analyzed by both MALDI and ESI, they observe RSDs of 1.1% - 16.2% for MALDI and 2.0% - 30.5% for ESI, with MALDI ionization giving consistently more reproducible results. Together, these data suggest the variations between DRIP-Q method and

<table>
<thead>
<tr>
<th>fraction (x)</th>
<th>DRIP-Q method</th>
<th>Isotope dilution method</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>0.291 ± 0.018</td>
<td>0.295 ± 0.005</td>
<td>0.58</td>
</tr>
<tr>
<td>0.4</td>
<td>0.394 ± 0.017</td>
<td>0.391 ± 0.005</td>
<td>0.63</td>
</tr>
</tbody>
</table>

\(a\) mixtures of AAAAYRAAR/AAAApYRAAR with phosphopeptide fraction (0.3, 0.4) in an amount range of 0.9-2.0 pmole

\(b\) t-test (two tails, n=5)

Table 2.4 Comparison of the results obtained by the proposed DRIP-Q method and isotope dilution method.
isotope dilution MALDI-MS method are comparable with excellent accuracy, in spite of no isotope labeling or calibration curve needed in the proposed DRIP-Q method. The somewhat better precision of isotope labeling does not lead to an obvious increase in accuracy, and is offset by the lower cost of the isotope-free DRIP-Q method.

2.2.2.6 Applications

This study provides a simple and accurate quantitative method to determine phosphorylation by MS, with potential applications for assessing kinase activity and inhibitor potency. Cyclin-dependent kinases (Cdks) are proline-directed serine/threonine kinases that regulate eukaryotic cell cycle. Inhibitors of Cdks are currently under investigation as potential therapeutics for cancer and viral infection (2014 Holcakova, 2011 Cicenas). Here, the DRIP-Q method is evaluated as a method for assessing the in vitro inhibition of Cdks.

r-Roscovitine and Olomoucine are well known pharmacological ATP-competitive inhibitors of Cdks. In a recent paper, r-Roscovitine has also been reported as a potential pharmacological Cdk inhibitor of HIV-1 replication, showing the higher selectivity in inhibiting HIV-1 infected cells than Olomoucine (2010 Guendel). Peptide 2 (LRWGFTTPDKKHQKEPPF) derived from HIV reverse transcriptase was subjected to in vitro phosphorylation by Cyclin A/Cdk2. After incubation of kinase, peptide and ATP, the reaction mixture was subjected to DRIP-Q analysis to determine the IC$_{50}$ (50% inhibition of the control) values of r-Roscovitine and Olomoucine. The mathematical $a$ value for the system was determined using a single point measurement of phosphorylated and unphosphorylated standards (Figure 2.7A), and found to have a value of $a = 0.184 \pm 0.017$. This $a$ value was then applied to peptide/phosphopeptide signal intensity ratios according to Eq. 3d without the construction of additional calibration curves. Inhibition plots were generated, and IC$_{50}$s
of r-Roscovitine and Olomoucine were found to be 0.92 ± 0.16 μM and 12.28 ± 0.74 μM, respectively (Figure 7B).

**Figure 2.6** Variability in MALDI-TOF MS signal, as measured by the coefficient of variance of the signal intensity ratios for either a non-isotopomeric (blue) or isotopomeric (red) peptide pair decreases as a function of increased laser shots for both isotopomeric and non-isotopomeric pairs. Spectra were acquired using a Waters MaldiMX MALDI-TOF mass spectrometer in positive reflectron mode using unphosphorylated $^{13}$C$_3$ labeled and unlabeled peptide 1 for the isotopomeric pair, and unphosphorylated and phosphorylated unlabeled peptide 1 for the non-isotopomeric pair. MALDI samples were prepared using the dried droplet method as previously described using a 1 pmol 1:1:1 mixture of all three peptides (labeled, unlabeled and phosphorylated) in a saturated solution (solvent 70:29.9:0.1 acetonitrle:water:TFA v/v/v) of alpha-cyano, 4-hydroxycinnamic acid matrix. Spectra were acquired in manual mode using an N$_2$ laser firing at 60 Hz.
This suggests that both r-Roscovitine and Olomoucine have low micromolar inhibitory activity against the substrate peptide2 in the presence of Cdk2/CyclinA. The inhibitory activity of r-Roscovitine is about 10 times higher than Olomoucine for Cdk phosphorylation of peptide 2. Prior published studies report similar IC_{50} range of r-Roscovitine (0.65-1.8 μM) (1997 Meijer, 2005 Bach) and Olomoucine (7- 8.8 μM) (1994 Vesely, 2005 Raynaud) for Cdk2/CycA activity, consistent with the results that r-Roscovitine has also 10 times more inhibitory activity than Olomoucine. The consistency indicated that the proposed DRIP-Q method may have feasibility to provide a reference method for simply, fast and accurately quantify phosphorylation with kinase assays.

Figure 2.7 *In vitro* inhibition of Cdk2 measured by DRIP-Q method. (A) Mass spectra of mixture of phospho – and unphosphopeptide 2 (0.14 pmole) in a Cdk kinase buffer. The response factor ratio $a$ was determined from the observed mass spectrometric signal intensity ratio, $y = 2.93$, and the known phosphorylation fraction, $x = 0.65$ for the measurements of *in vitro* inhibition of Cdk2 phosphorylation of Peptide 2. (B) Inhibitory activities IC_{50} of Roscovitine and Olomoucine against Cdk2/cyclin A. Peptide 2 is used as substrate in *in vitro* assays.
2.2.3 Conclusion

In this chapter 2.2, a novel quantitative strategy combined with isotope-free MALDI-MS to quantify phosphorylation in in-vitro assays is described. No calibration is required in this method. Based on the proposed DRIP-Q method, phosphorylation concentration in an assay sample was simply calculated by using its signal ratios on the mass spectrum with the consistent instrument response factor. Excellent reproducibility, precision, dynamic range and accuracy have been determined by using three different peptides, containing different phosphorylation sites (Serine, Threonine, and Tyrosine respectively). Method validation has been carried out by comparing DRIP-Q method to an isotope-dilution method. In addition, the method has been successfully applied to an in vitro assay to monitor the inhibition activity of Cdk inhibitors by determining IC50 values.
2.3 Label-free MALDI-MS for double-site phosphorylation quantification

Many phosphorylation sites tend to be clustered in disordered or unfolded regions of functionally important protein domains (2007 Chang, 2012 Fisher). Holt et al. found, for example, that multiple Cdk1 phosphorylations at rapidly evolving site clusters appear to regulate the evolution of kinase-signaling circuits (2009, Holt). A gain or a loss of phosphorylation at individual phosphorylation sites may dynamically trigger a chain of events associated different protein functions and biological activities. For example, Aquaporin-2 (AQP-2) shows a significant increase in phosphorylation at Ser-256 in response to short-term vasopressin treatment but, in contrast to Ser-256, dephosphorylation of Ser-261 occurs upon treatment with vasopressin (2010 Xie, 2006 Hoffert).

MALDI-MS analysis may suffer from some problems when applied to identification and quantification of the isomeric AQP-2 phosphopeptides, phosphorylated at Ser-256 (pSer-256) and at Ser-261 (pSer-261). First, the precursor ions of the isomeric phosphopeptides have identical mass in the full scan MALDI-MS spectra. In addition, MS\(^2\) fragmentation of the precursor ions containing labile serine-/threonine phosphorylation sites by collision-induced dissociation (CID) generates primary abundant neutral losses of phosphoric acid (\(m/z\) 98), resulting in indistinguishable MS\(^2\) spectra of the isomeric phosphopeptides.

To solve the MS\(^2\) problem, Electron Capture Dissociation (ECD), developed by McLaffery et al. (1998 Zubarev), and Electron transfer dissociation (ETD) (2004 Syka), induced by Hunt et al., have been recently used in tandem MS to generate c- and z- fragments with cleavage sites at N-C\(\alpha\) bond of the peptide backbone via gas-phase radical process, and leave the side-chains intact (2014 Collins). For phosphorylation analysis, ETD creates MS\(^2\) spectra with no loss of phosphoric acid that show distinguishable fragmentation patterns corresponding to the isomeric phosphopeptide.
backbones, not the abundant neutral loss of phosphoric acid. However, a problem may occur in MALDI coupled to ETD is that singly charged peptide ions, the majority of MALDI generated ions, are not amenable to ETD fragmentation, which requires charge reduction. Moreover, ETD is more expensive than the common CID for the reason that peptides must react with a large number of free electrons for efficient fragmentation, and generally the ECD/ETD technique is coupled to a high-resolution analyzer, like FT-ICR or Orbitrap for top-down analysis.

Instead, economical CID-MS\textsuperscript{3} is proposed in the study as an alternative way to analyze two singly phosphorylated AQP2 peptides at pSer-256 by PKA and pSer-261 by Cdk kinase. Even though the MS\textsuperscript{2} of the phosphoAQP2 peptides contain a dominant peak arising from neutral loss of phosphoric acid, causing identical MS\textsuperscript{2} spectra of the pSer-256 and pSer-261 (shown in Figure 2.8), the following MS\textsuperscript{3} approach provides a platform where the neutral loss peak in MS\textsuperscript{2} is further fragmented resulting in the discrete fragment patterns of remaining peptide backbone of the isomeric phosphopeptides. As shown in Figure 2.9, MS\textsuperscript{3} spectra of pSer-256 and pSer-261 are disguisable with different intensity ratios of the MS\textsuperscript{3} fragment \textit{m/z} 1297.75 to 1495.83. This chapter is to describe a label-free MALDI-CID-MS\textsuperscript{3} method for relative phosphorylation quantification to determine AQP2 phosphopeptides of two discrete phosphorylation sites (Ser-256 and Ser-261) by measuring the change in the intensity ratios of the MS\textsuperscript{3} fragments.

2.3.1 Materials and methods

Materials

HPLC grade water, acetonitrile, and formic acid were purchased as HPLC grade from Fisher Scientific (Pittsburg, PA, USA). Synthetic AQP-2 peptide, RQSVELHSPQSLPR and its two phosphorylated forms at pSer-256 and pSer-261, RQpSVELHSPQSLPR and
RQSVELH$p$SPQSLPR, were purchased from Anaspec (Fremont, CA) for the quantitative method development.

**MALDI-MS preparation**

AQP-2 Peptide standards, RQSVELHSPQSLPR, and its phosphopeptides, RQ$p$VELHSPQSLPR, and RQSVELHpSPQSLPR were selected to establish the label-free MALDI-MS\textsuperscript{2} quantitative method. 1 µL of the sample mixtures of the unphospho- and phosphopeptide standards with various phosphorylation fractions was mixed with 9 µL of matrix solution of either a saturated aqueous solution of a-Cyano-4-hydroxycinnamic acid (CHCA) (Sigma-Aldrich, St. Louis, MO, USA) in 0.1 % trifluoroacetic acid/acetonitril (30/70, v/v) or 2,5-Dihydroxybenzoic acid (DHB) (Acros Organics Morris Plains, NJ, USA) mixed with water / acetonitrile (50/50, v/v) in a 1:4 ratio (v/v). Then, 1 µL of the peptide mixture with the matrix solution was spotted on a MALDI plate prior to MALDI-MS analysis.

**MALDI-MS/MS analysis**

MALDI-MS experiments were carried out on a Thermo LTQ XL linear ion trap mass spectrometer (Thermo Scientific) was equipped with a MALDI ion source. A nitrogen pulsed laser at 337 nm was used to ionize analytes. The intensities of the laser were set at 8 µJ for CHCA matrix samples and at 40 µJ for DHB matrix samples with automated spectrum filter (ASF) off, 10 microscan/step. Automatic gain control (AGC) was set at 30000 for full scans and at 10000 for MS\textsuperscript{n} scans to regulate the ion amount filled in the trap. The acquire time was set to 4 min with 14,400 laser shots. One microscan was acquired per MS/MS spectrum. The protonated AQP-2 ion (parent ion) at \textit{m/z} 1714 subjected to MS\textsuperscript{2} was isolated for CID with 20% normalized collision energy, an isolation window at 3 \textit{m/z}, an activation q at 0.25 and activation time of 30 ms. A subsequent MS\textsuperscript{3}
event was triggered where one of the MS$^2$ product ions at $m/z$ 1616 was selected for CID with 25% or 40% normalized collision energy. The data were processed by Xcalibur$^{TM}$ 2.0.7 software.

**In vitro kinase assays**

To investigate the feasibility of the method, kinase assays were employed by incubating 1 μM AQP-2 peptides with and Cdk1/CyclinB and PKA in different concentration ratios (0.2 to 0.6 μM, and 0.5 to 0.3 μM for Cdk1/PKA) in the presence of 20 mM ATP and 1x NEBuffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl$_2$, and 1 mM DTT, New England BioLabs Inc.) at 30°C for in a time course (10, 30, 90 min). At each time point, 1 μL of each sample was added into 9 μL of the matrix solution for MALDI analysis.

### 2.3.2 Results and discussion

#### 2.3.2.1 Mass spectra of p-AQP-2 isomers (pSer-256 and pSer-261)

Phosphopeptide isomers of AQP-2 pSer-256 and pSer-261 have identical not only mass-to-charge ratio but also CID-MS$^2$ fragmentation. The CID-MS$^2$ spectra (Figure 2.8) of the phosphorylated AQP-2 isomers show a dominant peak corresponding to a neutral loss of phosphoric acid due to the unstable O-P bond of serine/phosphate. This leads to no difference in MS$^2$ fragment pattern that can be identified and used for quantification. To resolve the problem, we proposed to utilize MS$^3$ technique to generate backbone MS$^3$ fragments of pSer-256 and pSer-261, following the strong neutral loss. In Figure 2.9, a different pattern of $m/z$ 1297 to $m/z$ 1460 was observed in the MS$^3$ spectra of the isomers. pSer-256 contained higher abundance of MS$^3$ fragment of $m/z$ 1297, corresponding to $y^*_{12} - 2$NH$_3$; the reason might be that unexpected phosphate rearrangement on residue 12 of pSer-256 in MS$^2$ results in the labile structure of MS$^3$ precursor ion, generating more neutral loss of NH$_3$ around residue 12.
Figure 2.8 MS² fragmentation of pSer-256 and pSer-261.

Figure 2.9 MS³ fragmentation of pSer-256 and pSer-261.
2.3.2.2 Mathematic Approach for double phosphorylation site quantification

The method proposed for double phosphorylation quantification is based on the difference in intensity ratios of MS$^3$ fragments ($m/z$ 1297 ($f_1$) to $m/z$ 1460 ($f_2$)) of AQP-2 phosphopeptides (pSer-256 and pSer-261). It requires the assumption that MS$^3$ fragment ratios of individual phosphopeptide ($\alpha$ for pSer-256 and $\beta$ for pSer-261) remain constant while the amount of the phosphopeptide changes.

The MS$^3$ fragment ratios of pSer-256 and pSer-261 can be represented as:

$$\frac{f_{1p256}}{f_{2p256}} = \alpha \quad \text{(Eq. 1a)}$$

$$\frac{f_{1p261}}{f_{2p261}} = \beta \quad \text{(Eq. 1b)}$$

where $f_{1p256}$ and $f_{2p256}$ represent signal intensity of MS$^3$ fragments $m/z$ 1297 and $m/z$ 1460, respectively for pSer-256; $f_{1p261}$ and $f_{2p261}$ for pSer-261.

Another assumption is that signal responses are directly proportional to sample concentrations. Therefore, taking the amount of phosphopeptides, $A_{p256}$ and $A_{p261}$ into account, $f_{1p256}, f_{2p256}, f_{1p261}$, and $f_{2p261}$ in Eqs. 1 can be represented as:

$$f_{1p256} = \alpha A_{p256} \quad \text{(Eq. 2a)}$$

$$f_{2p256} = A_{p256} \quad \text{(Eq. 2b)}$$

$$f_{1p261} = \beta A_{p261} \quad \text{(Eq. 2c)}$$

$$f_{2p261} = A_{p261} \quad \text{(Eq. 2d)}$$

where if $A_{p256} = A_{p261}$, then $f_{2p256} = f_{2p261}$ (which we has confirmed experimentally in Table 5)
In analyzing a mixture of pSer-256 and pSer-261, Eqs 2a-d can be combined and the MS$^3$ fragments ratio ($f_{1, \text{mix}} / f_{2, \text{mix}}$) becomes:

\[ \frac{f_{1, \text{mix}}}{f_{2, \text{mix}}} = \frac{\alpha A_{p256} + \beta A_{p261}}{A_{p256} + A_{p261}} \quad \text{(Eq. 3)} \]

where the $f_{1, \text{mix}} / f_{2, \text{mix}}$ is proportional to sample amount fraction, either \( \frac{A_{p261}}{A_{p256} + A_{p261}} \) or \( \frac{A_{p256}}{A_{p256} + A_{p261}} \).

Letting the amount ratio \( \frac{A_{256}}{A_{261}} = x \), and the MS$^3$ fragments ratio \( \frac{f_{1, \text{mix}}}{f_{2, \text{mix}}} = R_{\text{mix}} \), the equation becomes:

\[ R_{\text{mix}} = \frac{\alpha x + \beta}{x + 1} \quad \text{(Eq. 4)} \]

According to the proposed Eq. 4, the amount fraction of pSer-256 to pSer-261 can be calculated by using the constant MS$^3$ fragment ratios $\alpha$ and $\beta$, obtained from analyzing pSer-256 and pSer-261 standards, and observed MS$^3$ fragments ratio $R_{\text{mix}}$. In other words, this method only requires determination of $\alpha$ and $\beta$, and therefore avoids time-consuming calibration curve construction.

### Table 2.5 Normalized intensity of MS$^3$ fragments of $f_1$ and $f_2$ for pSer-256 and pSer-261 AQP2 phosphopeptides

<table>
<thead>
<tr>
<th></th>
<th>2 μM</th>
<th></th>
<th>10 μM</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$f_1$</td>
<td>$f_2$</td>
<td>$f_1$</td>
<td>$f_2$</td>
</tr>
<tr>
<td>m/z</td>
<td>1297</td>
<td>1459</td>
<td>1297</td>
<td>1459</td>
</tr>
<tr>
<td>pSer-261</td>
<td>48.93</td>
<td>33.89</td>
<td>47.74</td>
<td>32.54</td>
</tr>
<tr>
<td>pSer-256</td>
<td>29.4</td>
<td>34.03</td>
<td>29.02</td>
<td>33.53</td>
</tr>
</tbody>
</table>

According to the proposed Eq. 4, the amount fraction of pSer-256 to pSer-261 can be calculated by using the constant MS$^3$ fragment ratios $\alpha$ and $\beta$, obtained from analyzing pSer-256 and pSer-261 standards, and observed MS$^3$ fragments ratio $R_{\text{mix}}$. In other words, this method only requires determination of $\alpha$ and $\beta$, and therefore avoids time-consuming calibration curve construction.
2.3.2.3 Assumptions for the proposed Eq. 4

In 2.3.2.1, we describe how the mathematic approach is established for double phosphorylation site quantification, but the proposed Eq. 4 requires two plausible assumptions: 1) It requires the assumption that MS$^3$ fragment ratios of individual phosphopeptide ($\alpha$ for pSer-256 and $\beta$ for pSer-261) remain constant while the amount of the phosphopeptide changes; 2) signal responses are directly proportional to sample concentrations. With the proposed Eq. 4, the amount fraction ($x$) of pSer-256 to pSer-261 can be readily determined by applying the MS$^3$ fragment ratio, $\alpha$ or $\beta$, for pSer-256 or pSer-261 and the MS$^3$ fragment ratio of a mixture shown on a mass spectrum ($R_{\text{mix}}$). To verify the mathematic approach with Eq. 4, the assumptions must be demonstrated.

According to the assumption 1, for the Eq. 4 to hold, $\alpha$ and $\beta$ must be constant. To investigate this, $\beta$ of pSer-261 was evaluated by analyzing pSer-261 samples over a range of amount from 0.20 pmole to 1.00 pmole. According to the data in Table 2.6, the $\beta$ value of pSer-261 is found to be consistent across the amount range with standard deviations varying from 0.012 to 0.025. The overall RMS error for across the experimental $\beta$ values is 0.013 (1.52% of the averaged $\beta$ value). The consistency of the $\beta$ value observed in agreement with our assumption for developing Eq. 4 that $\alpha$ and $\beta$ values remain constant even when the amount of the phosphopeptide changes shows the potential for utilizing Eq. 4 to relative quantification of doublesite phosphorylation by using MALDI-MS$^3$ without isotopes and calibration curves.

After the consistency of $\alpha$ and $\beta$ has been demonstrated, the next step is to investigate if signal responses are directly proportional to sample concentrations (assumption 2). To do this, pSer-256 and pSer-261 are mixed into 11 mixtures with different amount fraction, e.g. 0% to 100% pSer-261, at two concentration ranges, 0.02-0.2 $\mu$M and 0.1- 1 $\mu$M, and subjected to MALDI- MS$^3$ analysis. In the analysis, the changes in $R_{\text{mix}}$ on mass spectra are expected to represent a linear correlation with
amount fraction change in pSer-261 in response to pSer-256 in a mixture. The observed $R_{\text{mix}}$ and the known pSer-261 fraction in the mixtures (% pSer-261) are plotted in Figure 2.10, and good linearity is shown with $R^2$ values above 0.99 across a wide range of concentration between 0.02 $\mu$M and 1 $\mu$M.

Taken together, the value of $\beta$ appears to be constant even when the sample amount changes analyzed, according with assumption 1; the $R_{\text{mix}}$ is response to a changing fraction of pSer-256 and pSer-261 in mixtures that accords with assumption 2. These data allow the availability to further evaluate the proposed Eq. 4 to be applied to MALDI-MS$^3$ analysis for relative quantification of pSer-256 and pSer-261.

<table>
<thead>
<tr>
<th>Amount $^a$ (pmole)</th>
<th>Ratios $^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20</td>
<td>0.885 ± 0.012</td>
</tr>
<tr>
<td>0.40</td>
<td>0.851 ± 0.005</td>
</tr>
<tr>
<td>0.60</td>
<td>0.862 ± 0.002</td>
</tr>
<tr>
<td>0.80</td>
<td>0.860 ± 0.007</td>
</tr>
<tr>
<td>1.00</td>
<td>0.885 ± 0.025</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>0.869</strong></td>
</tr>
<tr>
<td><strong>RMS error</strong></td>
<td><strong>0.013</strong></td>
</tr>
</tbody>
</table>

$^a$ AQP-2 phosphopeptide Ser 261, RQSVELHpSPQSLPR

$^b$ m/z 1297 to m/z 1460

$n=3$
2.3.2.4 Dynamic Range and Accuracy

From the previous results, the β value stays unchanged when the amount of pSer-261 analyzed varies from 0.20 to 1.00 pmole. The remaining question to be resolved in the next piece of the research is the dynamic range of the proposed method by using Eq. 4. To determine the dynamic range of the method, the consistency of α of pSer-256 is evaluated by analyzing the pSer-256 standards with a series dilution. A consistency of α was observed within a sample amount range of 1 fmole to 10 pmole, even though R.S.D. was greater than 25 % as the sample amount is at 1 fmole (Table 2.7). This is because when the concentration is close to detection limit, leading to the spot-to-spot variability in R_{p261} increases to a less reliable level. Overall, the data suggest the method is feasible for double phosphorylation quantification with a dynamic range of at least 3 orders of magnitude, down to 1 fmole of sample on the plate.
To evaluate accuracy of the proposed method, two known mixtures of pSer-256 and pSer-261 (45% and 55% pSer-261) were analyzed with the use of $\alpha$ and $\beta$, determined by using individual phosphopeptide standard under the same instrumental conditions. Also, to validate the proposed single point method, it was compared with a traditional method by generating calibration curve of % pSer-261 across a sample amount fraction of 0% - 100 % pSer-261. In Table 2.8, experimental % pSer-261 were determined at 46.14 and 54.00 % by using the calibration curve method and at 46.89 and 54.49 % by using our proposed single-point method with the use Eq. 4. Both results are in agreement with the true values at 45 and 55 % pSer-261. Together, the current proposed method and the conventional calibration curve method offer similar accuracy but the single-point method would be more convenient and less time-consuming than the calibration curve method.

Table 2.7 Dynamic range of the constant ($\alpha$)

<table>
<thead>
<tr>
<th>Amount of pSer-256 (pmole)</th>
<th>$\alpha$</th>
<th>R.S.D.*</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2.13</td>
<td>1%</td>
</tr>
<tr>
<td>1</td>
<td>2.14</td>
<td>1%</td>
</tr>
<tr>
<td>0.1</td>
<td>2.15</td>
<td>3%</td>
</tr>
<tr>
<td>0.01</td>
<td>2.14</td>
<td>6%</td>
</tr>
<tr>
<td>0.001</td>
<td>2.16</td>
<td>27%</td>
</tr>
</tbody>
</table>

* n= 5
2.3.2.5 Effect of collision energy on double phosphorylation quantification

In tandem MS, different collision energies may vary fragmentation patterns that might further affect the availability of Eq. 4 when α, β, and observed $R_{\text{mix}}$ are used to calculate the phosphopeptide fraction ($x$). Selected ions colliding with insufficient collision energy may lead to low abundance in MS$^3$ fragments and the precursor ions remaining intact, whereas excessively high collision energy may produce a higher MS/MS efficiency but this may cause inconsistent breakdown of the precursor ion. To investigate the effect of collision energy on the $R_{\text{mix}}$ values in response to phosphopeptide fraction ($x$), pSer-256 and pSer-261 were mixed with a range of sample amount fraction (e.g. 0% - 100% pSer-261), and subject to MALDI-MS$^3$ with collision energy of 25 and 40 C.E. (Figure 2.11). Under the two different collision energies, $R_{\text{mix}}$ observed on mass spectra show linearly proportional to the known sample amount fractions ($x$). Experimental sample amount fractions ($x_{exp}$) were further calculated by using Eq. 4, the observed $R_{\text{mix}}$ values, and the previously determined α and β. The values of $x_{exp}$ are in agreement with the known fractions ($x$) with less than 3% relative standard deviation (Table 2.9). An important thing to note is that the change in collision energy can differentially affect the values of α and β, but does not affect the linear correlation between the observed $R_{\text{mix}}$ values and sample fractions – assumption 2 that our single-

<table>
<thead>
<tr>
<th>pSer-261 %</th>
<th>Single-point method</th>
<th>Calibration curve method</th>
<th>p value$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>46.89 ± 1.24$^a$</td>
<td>46.14 ± 1.22</td>
<td>0.25</td>
</tr>
<tr>
<td>55</td>
<td>54.49 ± 2.41</td>
<td>54.00 ± 2.39</td>
<td>0.41</td>
</tr>
</tbody>
</table>

$^a$ n=3
$^b$ t-test (two-tails, n=3)

Table 2.8 Determination of % pSer-261 by using the proposed single-point method and conventional calibration curve method.
point method rests on. As a result, care should be taken to ensure that the same collision energy is applied to all samples in an experiment.

\begin{align*}
\text{y} &= -0.0116x + 2.2071 \\
R^2 &= 0.9964
\end{align*}

\begin{align*}
\text{y} &= -0.0104x + 1.9834 \\
R^2 &= 0.997
\end{align*}

Figure 2.11 The effect of collision energy on \( \alpha, \beta \) and \( R_{\text{mix}} \).

Table 2.9 Effect of collision energies on the amount fractions, % pSer-261. Mixtures of pSer-256 and pSer-261 are analyzed under the collision energy at 25 and 40 C.E.

<table>
<thead>
<tr>
<th>p261 (%)</th>
<th>calculated p261%</th>
<th>calculated p261%</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>12.5 ± 0.2</td>
<td>7.7 ± 0.1</td>
</tr>
<tr>
<td>20</td>
<td>23.7 ± 0.5</td>
<td>22.7 ± 0.4</td>
</tr>
<tr>
<td>30</td>
<td>30.7 ± 0.7</td>
<td>30.4 ± 0.6</td>
</tr>
<tr>
<td>40</td>
<td>43.7 ± 0.6</td>
<td>41.0 ± 0.5</td>
</tr>
<tr>
<td>50</td>
<td>50.9 ± 0.3</td>
<td>48.7 ± 1.1</td>
</tr>
<tr>
<td>60</td>
<td>65.7 ± 1.9</td>
<td>61.4 ± 2.4</td>
</tr>
<tr>
<td>70</td>
<td>72.4 ± 1.4</td>
<td>70.2 ± 0.8</td>
</tr>
<tr>
<td>80</td>
<td>81.6 ± 2.5</td>
<td>80.4 ± 1.2</td>
</tr>
<tr>
<td>90</td>
<td>91.1 ± 0.2</td>
<td>88.9 ± 1.5</td>
</tr>
<tr>
<td>25</td>
<td>26.4 ± 0.6</td>
<td>26.9 ± 0.3</td>
</tr>
<tr>
<td>65</td>
<td>66.7 ± 1.0</td>
<td>65.0 ± 1.1</td>
</tr>
</tbody>
</table>
2.3.2.6 Applications to in vitro kinase assays

This proposed method provides a simple and accurate MS\textsuperscript{3} quantitative method for phosphorylation quantification of aquaporin-2 peptides, which can be phosphorylated at discrete phosphorylation sites, pSer-256 by c\textsuperscript{(PKA)} and pSer-261 by Cdk1/Cyclin B. It is important to distinguish and quantify the pSer-256 and pSer-261 phosphorylation since their phosphorylation regulation leads to different aquaporin-2 protein function. For example, vasopressin mediated PKA monophosphorylation of AQP2 at Ser-256 has shown to be required to increase water permeability and water absorption of renal principal cells, but the adjacent phosphorylation of Ser-261, independent of Ser-256 phosphorylation, does not appear to affect AQP2 trafficking (2010 Moeller).

The proposed single-point method has been demonstrated using Eq. 4 to measure phosphorylation fraction (\(x\)) when AQP-2 pSer-256 and pSer-261 are mixed. To evaluate how Eq. 4 can also facilitate measurements of phosphorylation fraction of pSer-256 to pSer-261 in an in-vitro kinase assay, the AQP-2 peptides are incubated with ATP in a presence of both PKA and Cdk enzymes with different ratios (14:5 and 5:7, PKA:Cdk) in a time course (10 –120 min). Figure 2.12 displays change in pSer-261 at a low or high Cdk level in a time course (10 – 90 min). The results show elevated Ser-261 phosphorylation is observed with a high level of Cdk (5:7 ratio of PKA to Cdk) in comparison to the results with a low level of Cdk (14:5 ratio of PKA to Cdk). This indicates that the proposed method is feasible to provide a method for simply and quickly quantifying relative amounts of phosphorylation at different phosphorylation sites in in vitro kinase assays.
2.3.3 Conclusion

An isotope-free, chromatography-free and calibration curve-free tandem MS³ method has been proposed for relative quantification of AQP-2 double-site phosphorylation. The signal ratios of phosphopeptide MS³ fragments have been shown to be directly proportional to their relative amount fractions in a mixture of AQP2 phosphopeptides (pSer-256 and pSer-261). The values of $\alpha$ and $\beta$ has further demonstrated to be consistent with the change in sample concentration. A wide analytical dynamic range (0.01-10 pmole) was shown by analyzing pSer-256 with a constant $\beta$ value observed with good precision (<6% R.S.D). The accuracy of proposed the single-point method was demonstrated in comparison with calibration construction. Furthermore, the method is shown to be feasible to monitor changes in phosphorylation at differing phosphorylation sites in kinase assays.
Chapter 3: Investigation of Cdk phosphorylation in HIV-1 reverse transcriptase by mass spectrometry

3.1 Introduction

3.2 Materials and methods

3.3 Results and discussion

3.3.1 Host Cdk s support HIV-1 reverse transcription in CD4+ T cells
3.3.2 HIV-1 RT is a substrate for Cdk2-dependent phosphorylation
3.3.3 Substitution of T216 by Alanine reduces HIV-1 RT
3.3.4 Quantification of HIV-1 in vitro Cdk2-dependent phosphorylation
3.3.5 Cdk2-dependent phosphorylation increases function and stability of HIV-1 RT
3.3.6 p21 (waf-1/cip-1) suppresses Cdk2-dependent HIV-1 RT phosphorylation

3.4 Discussions and conclusion

3.1 Introduction

CD4+ T cells are the primary HIV target cells that represent different levels of susceptibility to HIV-1 among a variety of HIV-1 infected persons. The difference in susceptibility appears to be related to host specificity of HIV-1 viruses (2004 Ciuffi; 1995 Spira). Thus, a focus of HIV research has been on identification of relative absence or presence of specific host proteins that modulate the efficiency of HIV-1 replication by inhibiting specific steps of the viral life cycle. Over the recent years, there are several host proteins have been identified, but the virus-host protein interaction network in HIV-1 infected persons still remain unclear (2012 Harris; 2012 Malim).

The host factors may play a dominant role in elite controllers, a small population of HIV-1 infected persons with ability to suppress HIV-1 replication and maintain an undetectable virus level in the absence of antiretroviral therapy. Interesting, in 2011, a significantly reduced susceptibility of host CD4+ T cells to HIV-1 has demonstrated in two geographically distinct cohorts of the elite controllers (2011 Chen; 2011 Saez-Cirion). This indicates that specific host proteins may reduce viral replication steps in the elite controllers and contribute to a CD4+ T cell-intrinsic mechanism of HIV-1 immune defense. Some classical HIV-1 host factors, such as APOBEC3G, TRIM5α and BST2, exert direct inhibitory effects on HIV-1 replication in HIV-1 progressors, but reduced
expression levels of the factors have been observed in CD4 T cells in elite controllers; these known factors are unlikely involved in low susceptibility of host CD4 T cells to HIV-1 infection (2013 Abdel-Mohsen; 2009 Rotger; 2011 Vigneault).

Our collaborators, Lichterfeld et al., have found that a protein called p21 (cip-1/waf-1), a host protein from the cyclin-dependent kinases inhibitor (CDKI) family, to be uniquely unregulated in CD4 T cells from many elite controllers, compared to both HIV-1 negative persons and individuals with progressive infection (2011 Chen). Besides, knockout studies have shown that p21 inhibits HIV-1 replication inhibition in CD4 T cells (2011 Chen), hematopoietic stem cells (2007 Zhang), and macrophages (2013 Allouch), although the underlying mechanisms for HIV-1 restriction seem to vary in each of these cell populations. According to Lichterfeld, after knocking out p21 by siRNA, the susceptibility to HIV-1 significantly increases. More interestingly, the resistance was due to less effective reverse transcription of HIV-1 and HIV mRNA transcription from proviral DNA (2011 Chen).

One proposed hypothesis is that instead of blocking HIV-1 replication by directly interacting with virus proteins, the host p21 protein may indirectly block HIV-1 replication through reducing the functional activity of HIV-1 dependency factors, cyclin-dependent kinases (Cdks). In this project, we show that HIV-1 reverse transcriptase (RT) is a substrate for Cdk2-dependent phosphorylation at a highly conserved Thr 216 residue on the p66 domain (between the finger and palm subdomains) and on the domains p51, the structure shown in Figure 3.1. The Cdk2 phosphorylation of RT is demonstrated to be functionally relevant for maintaining RT activity, stability and viral fitness, and can be effectively blocked by p21. Thus these data suggested a novel, indirect mechanism for inhibition of HIV-1 reverse transcription by the human host protein p21 indirectly blocking Cdk2 phosphorylation of HIV-1 RT in elite controllers.
3.2 Materials and methods

*In vitro* kinase assays. 5μM HIV-1 RT peptides or RT proteins were subjected to a kinase assay in the presence of 1x NEBuffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl₂, and 1 mM DTT, New England BioLabs Inc.) and incubated with 2 mM ATP and 550nM of indicated recombinant Cdk (Life Technologies) at 30°C for 1 or 3 h, as indicated. The kinase reaction was stopped by addition of SDS-PAGE loading buffer. Subsequently, proteins or peptides were analyzed by western blot or mass spectrometry.

**Peptide analysis.** Mass spectrometric analysis of ex-vivo kinase assay products was performed by a Thermo LTQ XL linear ion trap mass spectrometer (Thermo Scientific) equipped with a vacuum MALDI source. Synthetic HIV-1 RT peptide samples were mixed with saturated α-Cyano-4-hydroxycinnamic acid, CHCA (Sigma-Aldrich) in 0.1% trifluoroacetic acid: acetonitrile (70:30, v/v).
A 1.0 µL aliquot of sample/matrix mixture was dispensed onto the MALDI plate, and the solvent evaporated at room temperature.

**In gel digestion.** RT samples separated by NuPAGE and the protein bands were cut out and digested by lysyl endopeptidase (Lys-C, Promega) at 37 °C for 4 h. The enzyme Lys-C was added to a final substrate-to-enzyme ratio of 40:1 with 50 mM ammonium bicarbonate, pH 8.5. After digestion, POROS 20 R2 resin (Applied Biosystems) was added into the digested RT samples for extraction at 4°C for 4h. Prior to MALDI mass spectrometric analysis, the digests were desalted using ZipTip (Millipore) and eluted with CHCA MALDI matrix solution from resins directly onto a MALDI plate, and analyzed in single-stage, MS² and MS³ modes. Protein bands were identified using Mascot (Matrix Science, Boston, MA). Identification of phosphopeptide peaks and interpretation of MS² and MS³ spectra was performed manually, with the assistance of PAWS sequence analysis software (Genomic Solutions, Winnipeg, CA) and ProteinInfo (Rockefeller University, NY; http://prowl.rockefeller.edu) to calculate theoretical molecular weights.

**Western blots.** Protein lysates were prepared from cells infected with indicated viral constructs for 48 hours, using Radio-Immunoprecipitation assay buffer (Sigma) supplemented with the Halt protease and phosphatase inhibitor cocktail (Pierce). Following denaturation at 95°C for 10 minutes, samples were resolved on NuPAGE Bis-Tris 4-12% gels (Invitrogen), electroblotted to PVDF membranes (Life Technologies) and blocked with bovine serum albumin (Sigma). Blots were hybridized with monoclonal antibodies against indicated proteins (p21: clone C-19 (Santa Cruz) 1:1000 dilution; Cdk1: clone 5A6 (Thermo Scientific) 1:1000 dilution, Cdk4: clone PA5-14445 (Thermo Scientific) 1:1000, Cdk9: clone: PA5-19674 (Thermo Scientific) 1:250, Cdk2: clone 78B2 (Cell Signaling) 1:1000 dilution, Cdk7: clone MO1 (Cell Signaling) 1:2000 dilution, Lamin B1: ). For detection of phosphorylated HIV-1 reverse transcriptase, blots were hybridized with antibodies
against phosphorylated HIV-1 reverse transcriptase (1:800 dilution), generated by vaccination of
rabbits with the phosphorylated target peptide (LRWGFTpTPDKKHQKEPPF) according to
standard procedures by a commercial manufacturer (ThermoFisher Scientific). Selective reactivity
against the phosphorylated target peptide, and not against the corresponding unphosphorylated
peptide, was confirmed by Elisa assays. Detection of non-phosphorylated HIV-1 RT was performed
using a commercially available antibody (1:2000 dilution, AB63911 (Abcam)). Control hybridizations
were performed with a β-actin–specific antibody (1:5000 dilution, clone AB8224, Abcam). After
washing, membranes were probed with Cy5-labeled secondary antibody at 1:2500 dilution (ECL
Plex, GE Healthcare) for 1 h for chemifluorescence detection by Typhoon 9410 (Amersham
Biosciences), using excitation/emission wavelengths of 649/670. Alternatively, signals were
visualized using secondary goat-anti-rabbit antibodies labeled with IRDye 800CW, followed by
quantification of signals using Odyssey Image System (Li-Cor) and Image J software (NIH).

**Determination of HIV-1 RT Stability.** Recombinant HIV-1 RT was phosphorylated by a Cdk
kinase assay followed by limited proteolysis with subtilisin (Sigma-Aldrich). The 0.16 mg/ml RT
samples were incubated with subtilisin protease (1 mg/ml) in 100 mM Tris (pH 8.5) buffer at 30 °C
for 2, 30, 60, 120, 180, or 240 min (Serrano et al., 1984). For analysis, the collected samples were
heated with SDS sample buffer at 90 °C, alkylated with iodoacetamide, and separated by SDS gel
electrophoresis. Then, the gels were fixed in 40% methanol and 10% glacial acetic acid for 1 hr and
stained with SYPRO Ruby Protein Gel Stain (Life Technologies) under dark conditions overnight.
After being washed twice in 10% methanol and 7% glacial acetic acid for 1 hr, the stained gels were
scanned on a Typhoon 9410 Imager with a green 532 nm excitation laser and 610BP30 emission
filter. Densitometry quantification was performed with ImageQuant software (Molecular Dynamics).
Functional Assessment of HIV-1 RT. After in vitro kinase assays, 0.25 U of RT and phospho-RT were analyzed with the EnzCheck RT assay (Life Technologies). The HIV-1 RTs were incubated with 0.125 mg/ml Oligo d(T)16 primer, 2.5 mg/ml poly(A) ribonucleotide template, and 5 mM dTTP at room temperature for 10, 20, and 30 minutes. The reaction was quenched by adding EDTA, and the RT activity was measured with a fluorometric assay.

Statistical Analysis. Data are expressed as mean and SD or SEM or as box and whisker plots indicating the median, the 25% and 75% percentile and the minimum and maximum of all data. Differences between different cohorts or different experimental conditions were tested for statistical significance with paired or unpaired Student's t tests as appropriate. A p value of 0.05 was considered significant.

Patients. Peripheral blood mononuclear cells (PBMCs) from HIV-1-infected individuals and HIV-1-negative individuals were used for this study according to protocols approved by the Institutional Review Board of the Massachusetts General Hospital. All study subjects gave written consent to participate.

HIV-1 Viruses and Constructs. The -encoding R5-tropic and VSV-G-pseudotyped HIV-1 plasmids were kindly provided by Dr. Dan R. Littman (New York University). Viral particles were produced by transfecting human embryonic kidney 293T (HEK293T) cells (NIH AIDS Reagent Program) with the respective HIV-1 plasmids, and, if applicable, pCG-VSV-G, using TransIT-293 (Mirus) in OptiMEM. R5-tropic or VSVG- pseudotyped viral variants encoding for a T216A mutation, were generated by site-directed mutagenesis by a commercial supplier (Genewiz); the correct sequence of the variant constructs was confirmed by repeated viral sequencing. Supernatants containing infectious retroviruses were harvested 48 hr after transfection, centrifuged, and filtered.
Viral titers were determined by measuring p24 levels in the supernatant of infected HEK293T cells, and confirmed by measuring TCID50 levels in TZM-bl cells (NIH AIDS Reagent Program).

**In Vitro Infection Assays.** PBMCs were stimulated in RPMI medium containing 10% fetal calf serum (FCS), recombinant IL-2 (50 U/ml), and an anti-CD3/CD8 bispecific antibody (0.5 mg/ml). A homogenous population of activated CD4+ T cells was infected on day 5 with the indicated HIV-1 viruses for 4h (R5) or 2h (VSV-G) at 37_C. After two washes, cells were resuspended and plated at 13106 cells per well in a 24-well plate. The CD4+ T cells were subjected to flow cytometric analysis of GFP+ CD4+ T cells at 96 hr after infection with GFP-encoding R5-tropic HIV-1 or 48 hr after infection with VSV-G-pseudotyped HIV-1 virus. When indicated, proteasome inhibitors (MG-132, Millipore) were added at a concentration of 1mg/ml. For replication capacity assays, CD4+ T cells were infected with either WT or T216A-variant R5-tropic HIV-1 viruses. At indicated time intervals, supernatants were analyzed by p24 ELISA kit (PerkinElmer). Replication capacities were calculated by fitting a linear model to the log-transformed p24 concentration with maximum likelihood models.

**siRNA-Mediated Gene Knockdown.** CD4+ T cells were suspended in 100 ml transfection solution (Lonza), and target-specific or control siRNA (Dharmacon) was added at a concentration of 4 nmol/ml. Then, samples were transferred into nucleofection cuvettes and transfected with T-023 in a Nucleofector device (Lonza). Afterwards, cells were resuspended in medium supplemented with 20% FCS; IL-2 (50 U/ml) was added 2 hr after transfection. Cells were infected with indicated HIV-1 strains at 24 hr after transfection.

**Flow Cytometry.** The proportion of CD3+ CD4+ GFP-positive cells was analyzed on an LSRII flow cytometer (BD Biosciences), and data analysis was performed with FlowJo software (TreeStar). For imaging flow cytometry experiments, PBMCs were stained with antibodies directed against CD3
and CD4, fixed, and permeabilized. Subsequently, intracellular staining was performed with p21-specific antibodies (clone F-5, Santa Cruz Biotechnology). The samples were run on an ImageStream Cytometer (Amnis), and data visualization and colocalization analysis were performed with IDEAS Software (Amnis).

**Detection of HIV-1 Reverse Transcripts.** Cells were harvested 18 hr after infection of activated cells with VSV-G-pseudotyped HIV-1 and 48 hr after infection of activated cells with R5-tropic HIV-1. RT products were amplified from cell lysates with primers and probes hRU5- F2, hRU5-R, hRU5-P (early RT transcripts), FST-F1, SS-R4, P-HUS-SS1 (intermediated RT transcripts) and GagF1, GagR1, and P-HUS-103 (late RT transcripts) (Mbisa et al., 2009). CCR5 was amplified as a housekeeping gene and used for the quantification of input cell numbers (2011 Chen). Serial dilutions of DNA from cell lysates of the HIV-1-infected cell line HEK293T (provided by Dr. Frederic Bushman, University of Pennsylvania) were used for reference purposes.

**Coimmunoprecipitation Assays.** Protein lysates from CD4+ T cells were incubated in spin columns coated with anti-Cdk2 (clone 78B2; Cell Signaling) for 12 hr at 4°C. After washing, protein complexes were eluted, boiled, and subjected to SDS-PAGE followed by western blotting with antibodies against HIV-1 RT (AB63911; Abcam), cyclin A (clone BF683; Cell Signaling), and cyclin T (clone 8744; Cell Signaling).
3.3 Results and discussion

3.3.1 Host Cdk\s support HIV-1 reverse transcription in CD4\(^+\) T cells

In Lichterfeld’s recent study, p21 has been found to be associated with the reduction in HIV-1 susceptibility due to less effective reverse transcription, in which Cdk\s may be important for activating RT, and this process is inhibited by p21. To investigate how Cdk\s play a recognized role in reverse transcription during the HIV-1 life cycle, the influence of pharmacological Cdk inhibition on reverse transcription in CD4\(^+\) T cells was first studied. Primary CD4\(^+\) T cells from HIV-1 negative persons were \textit{ex-vivo} activated with CD3/CD8 bi-specific antibodies and IL-2, followed by infection with a GFP-encoding R5-tropic HIV-1 or VSV-G-pseudotyped HIV-1 virus construct in the presence or absence of Roscovitine or Olomoucine, two chemical inhibitors of Cdk\s. These experiments in \textbf{Figure 3.2A-C} demonstrated that pharmacological inhibition of Cdk\s effectively reduced the proportion of GFP-positive CD4\(^+\) T cells, and the per-cell levels of early HIV-1 reverse transcripts (minus-strand strong-stop DNA), intermediate HIV-1 reverse transcripts (minus strand DNA) and late, double-stranded HIV-1 reverse transcription products; similar observations were made when cells were infected with GFP-encoding VSV-G pseudotyped HIV-1 virus that causes single-round infections of CD4\(^+\) T cells independently of viral coreceptors (\textbf{Figure 3.2D-F}). Altogether, these data strongly indicate blockade of HIV-1 reverse transcription by pharmacological Cdk inhibitors in infected cells with either GFP-encoding R5-tropic HIV-1 or VSV-G-pseudotyped HIV-1 virus construct, therefore, Cdk\s might be required for HIV-1 reverse transcription and it is of interest to better understand the role of the host Cdk\s, the target molecules for Roscovitine and Olomoucine, in supporting HIV-1 reverse transcription.
Figure 3.2 Host CDK supports HIV-1 reverse transcription. Activated CD4+ T cells were infected ex vivo with GFP-encoding, R5-tropic HIV-1, or GFP-encoding VSV-G-pseudotyped HIV-1 in the presence or absence of the pharmaceutical CDK inhibitors Roscovitine and Olomoucine. (A/D) Representative flow cytometry plots showing infected GFP+ CD4+ cells. (B/E) Proportions of GFP+ CD4+ T cells after exposure to indicated CDK inhibitors. (C/F) Quantitative analysis of early, intermediate, and late HIV-1 RT transcripts in CD4+ T cells from the indicated study groups. All statistical comparisons were performed with Student’s t tests. *p<0.05; **p<0.01; ***p<0.001. Data are expressed as box and whisker plots indicating the median, 25% and 75% percentile, and minimum and maximum of all data. These results were generated by Jin Leng at MGH. (2014 Leng)
3.3.2 HIV-1 RT is a substrate for Cdk2-dependent phosphorylation

HIV-1 gene products can represent direct targets for phosphorylation by host kinases (2011 Francis). We hypothesized that Cdk2 can support HIV-1 reverse transcription through phosphorylation of HIV-1 RT, a key enzyme of reverse transcription of viral RNA into dsDNA, which can further be inserted into host DNA genomes and copied by human machine, leading to massive damage to the host cells. Cdk2s are requisite proline-directed serine/threonine kinases with preference for basic amino acids located c-terminal to the phosphate acceptor (1993 Nigg); therefore the canonical Cdk substrate motif is S/T-P-X-R/K, where X can be any amino acid(s). The HIV-1 RT protein contains seven such Cdk motifs (Figure 3.3).

![Figure 3.3 HIV-1 reverse transcriptase sequence](image)

**Figure 3.3 HIV-1 reverse transcriptase sequence.** Seven minimal Cdk phosphorylation consensus motifs (S/T)P are highlighted in red in the HIV-1 reverse transcriptase sequence.

Synthetic peptides were generated with sequences derived from these motifs and their surrounding amino acids for an initial rapid screen of potential Cdk phosphorylation sites. Each of
these peptides was assayed as a Cdk substrate by incubating with ATP and Cdk2/cyclin A, Cdk2/cyclin E, Cdk7/cyclin H or Cdk9/cyclin T1 (Figure 3.4A), followed by mass spectrometric analysis. Robust phosphorylation of the synthetic peptide containing the motif at Thr216 (corresponding to HIV-1 RT amino acid residues [210-227]) was observed using both Cdk2/cyclin A and Cdk2/cyclin E, and trace phosphorylation was observed at two other peptides with Cdk2/cyclin A (Figure 3.4A). Phosphorylation was not observed with any other HIV-1 RT peptides harboring Cdk phosphorylation motifs (data not shown). When the [210-227] peptide was incubated with Cdk2/cyclin A for 0, 1 and 2 hours, a time-dependent increase in mass spectral signal at +80 mass units (corresponding to covalent addition of HPO₃) was observed (Figure 3.4B). When similar assays were performed using Cdk7 or Cdk9, the positive control peptide, but no phosphorylation of the [210-227] peptide or any other HIV-1 RT peptides, were observed (Figure 3.5). These findings suggest that phosphorylation of HIV-1 RT occurs site-specifically at Thr216 by only Cdk2 kinases in vitro. Although Cks1 protein was not used in the in vitro assay, it does not change the importance of Cdk2 phosphorylation site that is necessary to be phosphorylated prior to Cks induced multiple phosphorylation.

Figure 3.4 HIV-1 RT is a substrate for Cdk2-dependent phosphorylation. (A): Level of phosphorylation observed by semi-quantitative MALDI mass spectrometry on synthetic peptides derived from HIV-1 RT after incubation with indicated Cdns and ATP for 3 hours. “+” denotes phosphopeptide levels < 1-25%; “++” indicates phosphopeptide levels of 25%-50%; “+++” reflects phosphopeptide levels >50% of total peptide signal intensity. (B): Time course for in vitro phosphorylation reaction of synthetic peptide [210-227] incubated with ATP and cyclin A/Cdk2, assayed by MALDI-selected ion monitoring mass spectrometry.

83
Figure 3.5 In vitro phosphorylation reaction of synthetic peptide [210-227] by cyclindependent kinases. The peptide [210-227] was incubated with ATP and cyclin H/Cdk7, cyclin T1/Cdk9, or cyclin E/Cdk2 for 3 hours, and assayed by MALDI-selected ion monitoring mass spectrometry. Phosphorylation is not observed for Cdk7 or Cdk9. Dyrktide was used as a positive control.

Lichterfeld’s group further conducted ex-vivo infection experiments of CD4⁺ T cells after siRNA-mediated knockdown of Cdk2 and Cdk7 in the cytoplasm of CD4⁺ T R5-tropic or VSV-G-pseudotyped infected cells where HIV-1 reverse transcription occurs. A significant reduction of proportion of HIV-1 positive cells, and of the frequency of per-cell levels of early, intermediate and late HIV-1 reverse transcripts has only been observed with the down-regulation of Cdk2, but not of Cdk7 (Figure 3.6). Of note, silencing of Cdk9 slightly reduced cellular susceptibility to productive HIV-1 infection, consistent with previous findings (1999 Flores), but had no effect on viral reverse transcription (Figure 3.7). The findings indicate that Cdk2 is critical for supporting HIV-1 reverse transcription in CD4⁺ T cells, corresponding to our prior in vitro observation that HIV-RT has a selective susceptibility to phosphorylation by Cdk2.
**Figure 3.6** Cdk2 supports HIV-1 reverse transcription. (A/D): Representative flow cytometry plots showing infected GFP+ CD4+ cells. (B/E): Proportions of GFP+ CD4+ T cells after exposure to indicated Cdk inhibitors or Cdk-directed siRNAs. (C/F): Quantitative analysis of early, intermediate and late HIV-1 RT transcripts in CD4+ T cells from the indicated study groups. All statistical comparisons were performed using Student’s t tests. *, p<0.05. **, p<0.01. ***, p<0.001. These results were generated by Jin Leng at MGH. (2014 Leng)

**Figure 3.7** Influence of Cdk1, Cdk4 and Cdk9 on HIV-1 replication. (A-B): Infection of CD4+ T cells from HIV-1 negative persons with R5-tropic (A) or VSV-G-pseudotyped (B) HIV-1 in the presence of indicated siRNA. (C-D): Corresponding analysis of HIV-1 reverse transcripts in CD4+ T cells after experimental knock-down of Cdk9. These results were generated by Jin Leng at MGH. (2014 Leng)
To further investigate the Cdk2 phosphorylation site at Thr216, the Cdk2-phosphorylated [210-227] peptide was subjected to tandem mass spectrometry (MS²). Specific fragmentation of the phosphorylated peptide showed a strong signal at -98 mass units from the parent peptide peak, corresponding to the neutral loss of phosphoric acid, a diagnostic characteristic of Ser/Thr phosphorylated peptide ions (Figure 3.8). Furthermore, using backbone fragmentation, we found that the y11 fragment ion, which runs from Pro217 to Phe227 and does not contain Thr216, is observed at a mass indicating no phosphate. In contrast, the y12 fragment ion, which runs from Thr216 to Phe227, is observed at a mass indicating the presence of a phosphate adduct (Figure 3.8). Because the two fragments both arise from the singly phosphorylated [210-227] peptide, and are identical in amino acid sequence except for the Thr216 residue that is selectively contained in the y12 fragment ion, these data indicate that Thr216 is indeed the Cdk2 phosphorylation site (Figure 3.8).

Since the data suggested that Thr216 represents the only site in HIV-1 RT susceptible to Cdk-dependent phosphorylation, we performed phosphorylation analyses on the intact HIV-1 RT protein. Recombinant HIV-1 RT protein was phosphorylated in vitro and subjected to analysis using western blots. For this purpose, we generated rabbit-derived antibodies that specifically recognized the HIV-1 RT [210-227] peptide sequence phosphorylated at position Thr216. In western blot experiments, two RT bands (p51 and p66 subunits) were detected with anti-[phospho-210-227] antibodies after incubation with ATP and Cdk2/Cyclin A for 30 min (Figure 3.9A). An increase in the intensity of the phospho-RT bands was observed after incubation for 120 min. To conclusively determine the phosphorylation site in the RT protein, phosphorylated HIV-1 RT was digested by endoproteinase Lys-C and the digests were analyzed by mass spectrometry. A mass spectral peak at m/z 2413 was observed in the spectrum arising from the Cdk2-phosphorylated but not the unphosphorylated HIV-1 RT digest. This peak corresponds to the phosphorylated Lys-C peptide [202-220], containing the Thr216 Cdk phosphorylation site (Figure 3.9B, bottom panel).
Phosphorylation at Thr216 was confirmed by MS² and MS³ fragmentation (Figure 3.10). In contrast, the spectrum of the unphosphorylated HIV-RT digest contained a significant peak at m/z 2334 corresponding to un-phosphorylated peptide [202-220] (Figure 3.9B, top panel); that peak appears substantially weaker in the spectrum from the Cdk2-phosphorylated RT digest. No other peaks corresponding to phosphorylation at Cdk consensus motifs were observed in either digest (data not shown). Together, these data strongly indicate Cdk2-dependent phosphorylation specifically at Thr216 in HIV-1 RT.

We subsequently analyzed the structural position of the targeted Thr216 residue, using prior crystal structures of HIV-1 RT (1994 Jager; 1992 Kohlstaedt). We observed that Thr216 lies in β strand 11 for both the p51 and p66 subunits, situated near the interface between the fingers and palm of the structure (Figure 3.11A). This solvent-exposed position of Thr216 at the molecular surface is consistent with its role as a target for cytoplasmic host kinases. In addition, we found that the targeted Cdk phosphorylation motif (Thr216 and adjacent Pro217) was highly conserved across more than 5000 viral clade A, B or C sequences derived from the Los Alamos HIV-1 sequence database (www.hiv.lanl.gov), while adjacent amino acid residues were more variable (Figure 3.11B). Together, these observations support the notion that the Thr216 residue in HIV-1 RT is susceptible to Cdk phosphorylation, and functionally and/or structurally important for HIV-1 replication.
Figure 3.8 MS$^2$ spectrum of the phosphorylated synthetic peptide [210-227] analyzed using MALDI-LTQ-MS. MS$^2$ fragmentation reveals a strong neutral loss of 98 m/z units, corresponding to phosphoric acid, characteristic of Ser/Thr phosphorylated peptide ions (McLachlin and Chait, 2001). The signals with m/z of the backbone fragmentation indicated that the underlined Threonine residue is phosphorylated. b ions indicate N-terminal peptide fragments and y'' ions indicate C-terminal peptide fragments, both with cleavage at the peptide bond (Biemann, 1990). Phosphate observed in the y''11 but not the y''12 ion indicates the phosphate group is on the Threonine in the seventh, not the sixth, position in the sequence, which corresponds to Thr216 in the RT sequence.
Figure 3.9 HIV-1 RT is a substrate for CDK2-dependent phosphorylation. (A): Western blot probed by anti-[phospho-210-227] primary antibody showing time-dependence of phosphorylation on recombinant HIV-1 RT protein incubated with cyclin A/CDK2 and ATP. (B): MALDI mass spectra of endoproteinase Lys-C proteolzyed unphosphorylated (top panel) and in vitro cyclin A/CDK2 phosphorylated (bottom panel) HIV-1 RT protein. Signal with m/z corresponding to Thr216 phosphorylated proteolytic peptide is observed in the phosphorylated but not the unphosphorylated digest.
**Figure 3.10** MS$^2$ and MS$^3$ spectra of Thr216 [202-220] phosphorylated Lys-C proteolytic RT peptide. In MS$^2$ spectrum, a strong neutral loss of phosphoric acid indicated the presence of phosphate, and the minor peaks representing backbone fragmentation showed the phosphorylation occurs on the underlined Threonine (in the CDK motif). m/z consistent with phosphorylation on the b15 ion but not he b14 ion indicates that the phosphate group is on the Threonine at the 15th position, which corresponds to Thr216 in the RT sequence. MS$^3$ fragmentation of m/z 2316 (loss of phosphate in MS$^3$) confirmed the sequence of phosphorylated peptide, by presenting backbone fragmentation consistent with the 202-220 peptide, but without phosphate addition.
Substitution of T216 by Alanine reduces HIV-1 RT

To investigate if the Cdk2-dependent phosphorylation of HIV-1 RT has functional implications for HIV-1 reverse transcription, Lichterfeld’s group used site-directed mutagenesis to construct a GFP-encoding R5-tropic or VSV-G-pseudotyped HIV-1 variant in which the identified phosphorylation position Thr 216 is mutated to Ala (T216A), so that the position could no longer be phosphorylated. The replicative activity of the wild-type and the variant virus was assessed in ex vivo activated CD4+ T cells. A direct comparison of the replicative activity of the wild-type and the variant virus over an 8-day period demonstrated a significantly reduced ability of the variant virus to produce HIV-1 p24 antigen, which is consistent with viral fitness (Figure 3.12A). Furthermore, the infection of CD4+ T cells with the T216A variant leads to reduced proportions of GFP-positive CD4+ T cells in comparison to CD4+ T cells with the wild-type (Figure 3.12B-E). The lower replicative activity of the variant virus was associated with reduced levels of early, intermediate and late HIV-1 reverse transcripts, suggesting that the phosphorylation residue Thr216 is needed for effective reverse transcription. Together, these findings suggest that Cdk2-dependent phosphorylation of HIV-1 RT does contribute to viral fitness through modulation of viral reverse transcription.

Figure 3.11 Conservation of Thr 216 in RT. (A): Solvent-exposed position of Thr216 in crystal structures of HIV-1 RT. (B): Entropy of amino acid residues at position 214-219 in indicated clade A, B and C HIV-1 RT sequences obtained from the Los Alamos HIV-1 sequence database. The entropy results are generated by Jin Leng at MGH. (2014 Leng)
HIV-1 RT T216A mutation decreases HIV-1 reverse transcription (A): HIV-1 p24 antigen production in the supernatant of activated CD4+ T cells infected with similar titers of R5-tropic wild-type HIV-1 or the corresponding T216A variant. Data were analyzed by fitting a linear model to the log-transformed p24 concentration using maximum likelihood models. Right panel demonstrates one representative experiment, left panel summarizes the relative replication capacity of wild-type and variant viruses, calculated based on the slope for the p24 production curves in n=4 experiments. (B/D): Proportions of CD4+ T cells infected with R5-tropic or VSV-G-pseudotyped HIV-1 wt and corresponding T216A variants. (C/E): Quantitative analysis of early, intermediate and late HIV-1 RT transcripts in activated CD4+ cells infected with either wild-type R5-tropic or VSV-G-pseudotyped HIV-1 or the corresponding T216A variants. These results are generated by Jin Leng at MGH. (2014 Leng)
To comprehend the amount of HIV-1 RT phosphorylated *in vitro*, relative stable isotope labeling quantification was conducted in conjunction with mass spectrometry. This requires that one sample be chemically modified with a heavy, stable isotope-containing chemical group, while the second sample be modified with the same chemical group that contains only regular isotopes. The two samples are then mixed. Mass spectrometry can distinguish the heavy isotope species from the first sample, and the corresponding regular isotope species from the second sample. The relative amounts of that species are then simply obtained from the heavy-to-light peak intensity ratio. In this experiment, isotope is introduced as deuterated methyl groups after in-gel proteolysis. Free amines from lysine side chains and peptide N-termini are dimethylated either with regular hydrogen-containing methyl groups (2 x CH$_3$, for the unphosphorylated sample) or with doubly-deuterated methyl groups (2 x CD$_2$H, for the phosphorylated sample) (Figure 3.13A). The regular isotope-labeled unphosphorylated RT peptides are then mixed with the heavy isotope-labeled phosphorylated RT peptides, and subjected to MALDI-MS analysis.

Relative quantification of T216 phosphorylation is determined by measuring the decrease in the abundance of unphosphorylated proteolytic peptide fragment ion spanning residues 202-218. Peak intensity of the unphosphorylated T216-containing peptide is decreased for deuterated, Cdk-treated RT (red bar, $m/z = 2091.6$) relative to regular, RT with no kinase treatment (blue bar, $m/z = 2087.6$) (Figure 3.13B). These measurements are obtained from the b$_{17}$ fragment ion spanning RT residues 202-218 arising from the MS$^2$ fragmentation of Lys-C proteolytic peptide spanning residues 202-220. The MS$^2$ fragment ions are used to increase specificity and decrease chemical noise. Quantitative results are derived as the second isotope after deconvolving the overlapping isotope distributions of the H- and D-containing peptides. Similar results are obtained for other, less intense.
fragment ions of the same parent peptide (not shown). Internal standard spectra (inset) are used to correct for potential differences in peptide recovery. By using dimethyl-based stable isotope to determine in vitro phosphorylation stoichiometry for the p51 and p66 subunits of HIV-RT, the results showed that 92.3% Thr 216 on p51 and 62.1% Thr216 on p66 are phosphorylated in *in vitro* kinase assays (Figure 3.13C).

**Figure 3.13** Quantification of *in vitro* phosphorylation of HIV-1 RT by using stable isotope labeling in conjunction with mass spectrometry. (A) quantification workflow. (B) MS² mass spectra of mixtures of Cdk untreated and treated HIV-1 samples, labeled by regular and deuterated methyl groups. p51 and p66 are two subunits of HIV-1 RT. (C) Stoichiometric results for phosphorylation on the p51 and p55 subunits of HIV-1 RT.
We subsequently determined whether Cdk phosphorylation of HIV-1 RT directly modulates the enzymatic activity of reverse transcriptase. For this purpose, HIV-1 RT was exposed to Cdk2/cyclin A, which resulted in 92% and 62% phosphorylation of the p51 and p66 RT subunits, respectively; unphosphorylated HIV-1 RT was used as a control. Either HIV-1 RT was incubated with a poly (A) template, oligo dT primer and dTTP, followed by monitoring of RT activity via the formation of RNA-DNA heteroduplexes. These experiments demonstrated an increased enzymatic activity of phosphorylated HIV-1 RT, suggesting that Cdk2 phosphorylation supports the functional activity of the enzyme (Figure 3.14A).

Since phosphorylation often influences the stability of proteins (2011 Nishi), we considered whether Cdk2 might serve to modulate the resistance of HIV-1 RT to proteolytic degradation. Limited proteolysis was employed as an in vitro tool for probing HIV-1 RT stability. Cdk2/cyclin A-phosphorylated and un-phosphorylated HIV-1 RT was incubated with subtilisin protease - known to be sensitive to the three-dimensional conformation of its substrate - at different reaction times ranging from 2 min to 4 hr. (Figure 3.14 B/C). These experiments demonstrated a markedly more rapid decrease in band intensity for the unphosphorylated HIV-1 RT. Indeed, after 180 minutes of exposure to subtilisin, gel densitometry indicated that 45% of the phosphorylated RT, versus only 20% of the unphosphorylated RT, remained undigested, demonstrating that Cdk2 phosphorylation protects HIV-1 RT from protease cleavage. However, CD4+ T cell infection with the viral wt or the T216A variant in the presence of proteasome inhibitors resulted in proportionally equal increases in HIV-1 replicative activity, suggesting that proteasomal degradation of HIV-1 RT is unaffected by Cdk2-dependent RT phosphorylation (1998 Schwartz). Together, these data indicate that Cdk2 phosphorylation of Thr216 protects HIV-1 RT from protease activity, and suggests that its
mechanism of action may be to stabilize HIV-1 RT against intracellular proteases, perhaps via conformational modulation.

Figure 3.14 Cdk2-dependent phosphorylation enhances activity and stability of HIV-1 RT. (A) Assessment of the functional activity of phosphorylated and unphosphorylated HIV-1 RT in a cell-free system. One representative experiment out of 4 is shown. (B) SDS-PAGE gel of limited proteolysis of unphosphorylated (left) and cyclin A/Cdk2 phosphorylated (right) HIV-1 RT time course. Degradation of the p66 and p51 subunits of HIV-1 RT (major upper and lower bands, respectively) is shown as a function of time and phosphorylation state. (C) Quantitative comparison of the rates of proteolysis for unphosphorylated and phosphorylated HIV-1 RT as measured by SDS-PAGE gel densitometry of Sypro Ruby-stained HIV-1 RT.

3.3.6 p21 (waf-1/cip-1) suppresses Cdk2-dependent HIV-1 RT phosphorylation.

The cell-intrinsic Cdk inhibitor p21 (waf-1/cip1) is upregulated in CD4+ T cells from many persons with natural control of HIV-1 replication (“elite controllers”) (2011 Chen), raising the possibility that p21 has a functional role for inhibiting Cdk2-dependent phosphorylation of reverse transcriptase and reducing the efficacy of viral reverse transcription in CD4+ T cells from these patients. To investigate this, we first analyzed the subcellular localization of p21 in CD4+ T cells from elite controllers, using high-throughput imaging flow cytometry, an approach that combines standard flow cytometry with high-resolution fluorescence microscopy. For this purpose, CD4+ T cells from elite controllers and HIV-1 negative persons were intracellularly stained with antibodies directed against p21, and with the nuclear dye DAPI. This analysis demonstrated that p21 was
located both in the nuclear and in the cytoplasmic subcellular compartment (Figure 3.15A/B), and did not change substantially after in vitro activation. Importantly, we observed that total, nuclear and cytoplasmic cell-associated p21 was expressed at significantly higher levels in CD4+ T cells from elite controllers compared to HIV-1 negative persons, but differences were more obvious for cytoplasmic p21 than for nuclear p21 (Figure 3.15A-B). This suggests that p21 may be functionally relevant for inhibiting Cdk2-dependent phosphorylation of HIV-1 reverse transcriptase in the cytoplasm of CD4+ T cells from elite controllers. To further investigate this, we analyzed HIV-1 reverse transcription in CD4+ T cells from elite controllers after electroporation with p21-specific siRNA, which resulted in effective reduction of p21 protein expression. We found that following infection of CD4+ T cells from elite controllers with either R5-tropic or VSV-G-pseudotyped HIV-1, siRNA-mediated reduction of p21 expression resulted in enhanced levels of HIV-1 positive cells, and in increased levels of viral reverse transcripts. Interestingly, such effects were not visible in CD4+ T cells from HIV-1 negative persons, likely due to lower baseline expression of p21 that minimizes biological effects of p21 silencing (Figure 3.15C-F). Moreover, p21 knockdown did not affect HIV-1 replication in CD4+ T cells from elite controllers infected with the T216A variant, supporting the notion that p21 inhibition of HIV-1 reverse transcription depends on phosphorylation of Thr216 (Figure 3.15C-F). In addition, we observed that in CD4+ T cells from elite controllers, the wild-type virus and the T216A variant showed a similar replicative activity (Figure 3.15G/H), which represented a sharp contrast to the replicative advantage of the wild-type virus in CD4+ T cells from HIV-1 negative individuals (Figure 3.12B-E); this indicates that the increased viral fitness associated with RT phosphorylation at position Thr216 is not visible in cells with high-level p21 expression. Finally, we noted that p21 knockdown led to increased levels of phosphorylated HIV-1 RT in CD4+ T cells from EC infected with R5-tropic or VSV-G pseudotyped HIV-1, providing formal evidence for an inhibitory effect of p21 on Cdk2-dependent
RT phosphorylation (Figure 3.15). p21 silencing also appeared to increase levels of unphosphorylated RT, specifically in cells infected with R5-tropic HIV-1 in which p21-mediated inhibitory effects on viral replication are amplified during multiple rounds of infection. Taken together, these data suggest that high-level p21 expression can indirectly inhibit viral reverse transcription through blockade of Cdk2-dependent phosphorylation of viral RT.

Figure 3.15 Inhibition of p21 (waf-1/cip-1) Increases HIV-1 Reverse Transcription in CD4+ T Cells from Elite Controllers. (A): Analysis of the subcellular localization of p21 in CD4+ T cells. PBMCs from Elite controllers or HIV-1 negative patients were stained with antibodies against CD4 (yellow pseudocolor), p21 (green pseudocolor) and DAPI (red pseudocolor) and subjected to ImageStream analysis. Selected images of single-cell analysis plots from an HIV-1 elite controller and an HIV-1 negative patient are shown. (B): Cumulative average p21 signal intensity in CD4+ T cells from Elite controllers (n=6) and HIV-1 negative persons (n=6) in total, cytoplasmic and nuclear compartments of CD4+ T cells. (C/D): Proportion of HIV-1 positive CD4+ T cells from EC and HIV-1 negative patients after ex-vivo infection with R5-tropic or VSV-G-pseudotyped HIV-1 in the presence of siRNA-mediated knock-down of p21. The T216A variant virus was used where indicated. (E/F): Corresponding analysis of early, intermediate and late RT transcripts in elite controllers. (G-H): Comparison of HIV-1 replication patterns of wild-type HIV-1 and the T216A variant in CD4+ T cells from EC. (G) demonstrates proportion of infected CD4+ T cells, (H) reflects levels of early, intermediate and late reverse transcripts. (I): HIV-1 RT phosphorylation in CD4+ T cells after experimental p21 knock-down. CD4+ T cells were transfected with siRNA directed against p21 or control siRNA, infected with R5-tropic or VSV-G pseudotyped HIV-1 and hybridized with antibody against HIV-1 RT, anti-[phospho-210-227] RT or b-actin. These results are generated by Jin Leng at MGH. (2014 Leng)
3.4 Discussions and Conclusion

Human Cdks have shown its ability to phosphorylate a wide range of host substrates \textit{in vitro} or \textit{in vivo}, and the phosphorylation events significantly contribute to numerous biological functions, including regulation of cell proliferation, survival, differentiation, motility, and gene expression. In addition to cellular regulation, Cdks also play a recognized role in virus-infected cells for regulation, replication, and maintenance of viral functions, such as transcriptional elongation, RNA initiation and processing (2008 Durand, 2010 Zydek). In HIV-1 infection, the host Cdks have been mostly suggested to regulate the post-integration stage of HIV-1. For example, HIV-1 transcription elongation depends on the interaction of HIV-1 Tat with host Cdks. HIV-1 RNA polymerase II is phosphorylated by Cdk9 whose activity is in turn stimulated by Tat phosphorylation catalyzed by Cdk2 (1999 Flores, 2006 Ammosova) However, a possible impact of Cdks on viral pre-integration steps has remained unknown. In this study, we examined whether Cdk2 specifically phosphorylates HIV-1 RT, and whether Cdk2 phosphorylation of RT has regulatory effect on HIV-1 reverse transcription during HIV-1 pre-integration. Our data indicate Cdk2 phosphorylation of RT on Thr 216 facilitates HIV-1 reverse transcription, supporting the emerging recognition of host kinases as modulators of HIV-1 replication in human CD4\(^+\) T cells where HIV-1 virus exploits the virus-host interaction on their own purpose for viral propagation as well as survival in human cells.

Considering that Cdk2 has only \textit{in vitro} phosphorylated the Cdk motif containing Thr 216 among the 7 Cdk consensus motifs of HIV-1 RT in the absence of Cks, therefore we suspect that Cdk2 might phosphorylate other motifs later after its activity is stimulated by Cks binding with the phosphorylated Thr216. Although the role of Cks in the control of cell cycle is still elusive, it has been previously demonstrated that the Cks associate with Cdk-cyclin complexes, modulate their phosphorylation ability, and target the Cdks to phosphorylated substrates in multisite
phosphorylation (1996 Pines, 1999 Patra); for instance, Sic1 multisite phosphorylation is stimulated in the presence of Cks, but the stimulatory effect requires that T5 and T33 of Sic1 are phosphorylated as Cks binding sites (2011 Koivomagi, 2013 Koivomagi). In another study, Cks showed no promotional effect on Wee1 hyperphosphorylation by Cdk1, yet Cks binds to the phosphorylated Cks consensus sites and induce inhibitory phosphorylation in Cdk1 by phosphorylated Wee1, when the Cks consensus sites on Wee1 are phosphorylated in late phosphorylation (2013 McGrath). It remains an open question whether the Cdk motif containing T216 also acts as a Cks-binding consensus site, and it will be interesting to explore conservation of the other Cdk phosphorylation motifs of RT and the role of multisite phosphorylation in HIV-1 infection in which Cks contributes to the phosphorylation process.

How might HIV-1 RT benefit from Cdk phosphorylation? The addition of phosphate might cause conformational and functional changes in RT. The HIV-1 RT is a heterodimer comprised of two separate subunits, referred to as p66 (~ 66kDa) and p51 (~ 51 kDa); the p51 subunit is processed by proteolytic cleavage of the p66 subunit by the viral-encoded protease. Although p66 and p51 have identical amino acid sequences (first 440 amino acid), they are functionally and structurally distinct. The p66 subunit contains active sites as an executor of RT functions with the p51 subunit stabilizing the complex and facilitating RNA/DNA hybrid positioning (2013 Chung), whereas the functional role of p51 subunit remains elusive. According to the quantitative data, 62.1% and 92.3% Thr 261 on p66 and p51 of RT is phosphorylated in vitro, respectively, suggesting the Cdk2 phosphorylation may contribute to modulation of the enzymatic activity and stability of RT and thus results in the enhanced HIV-1 fitness. In agreement with this, the later results showed Cdk2 phosphorylation indeed increases function and stability of HIV-1 RT.
An unresolved issue here is that how the Cdk2 dependent phosphorylation can affect HIV-1 reverse transcription during the viral pre-integration stage on the molecular basis of functional and conformational changes of the complex. With regard to phosphorylation on the catalytic subunit, it seems to interesting that the highly conserved phosphorylation site, Thr 216 residue, is in immediate proximity to the hot spots of RT amino acid residues at Thr 215 and Lys 219 of p66 subunit with known roles for influencing incorporation of nucleotides into viral DNA (2002 Chamberlain, 2010 Tu). Their mutations (T215Y or F, and K219Q) render RT resistant to nucleoside analogue inhibitor, 3'-azido-3'-deoxythymidine (AZT) (Jacobo-Molina, 1991 Biochemistry), suggesting that the two residues are associated with the formation of the triphosphate binding site in HIV-1 transcription in the p66 subunit. In line with this, the triphosphate binding may potentially assist Cdk2 transfers the terminal phosphate from ATP to Thr 216 phosphorylation site during phosphorylation. Moreover, the p51 subunit of RT has been observed being enzymatically inactive because of the alternative configuration and the absence of a C-terminal ribonuclease H (RNase H) domain, but it has a structurally stabilizing contribution to facilitating p66 loading onto nucleic (1998 Harris) and hybrid RNA/DNA binding and orienting the RNA strand for hydrolysis in the p66 RNase H active site (2013 Chung). Prior studies demonstrated that host ERK/MAPK family kinases could enhance chromosomal HIV-1 integration through the induction of phosphorylation-induced conformational changes in HIV-1 integrase (2013 Manganaro). Similarly, phosphorylation of the p51 subunit may play a role in establishing or stabilizing particular conformations of RT by forming more hydrogen bonds, providing structural support.

Because increasing evidence showed that the host Cdk inhibitor p21, a protein best known as cancer suppressor, seems to retain its ability to suppress HIV-1 replication, p21 has received much attention in the context of HIV-1 infection. Intensive studies over the past few years highlighted the p21 is capable of modulating the efficacy of HIV-1 replication in human
macrophages (2009 Bergamaschi), hematopoietic stem cells (2007 Zhang), and CD4+ T cells (2011 Chen), although it seems likely to have different underlying mechanisms associated with HIV-1 restriction in each cell population. In macrophages, p21 was found to inhibit HIV-1 replication by blocking dNTP biosynthesis via down-regulating the expression of RNR2 (ribonucleotide reductase subunit R2), an essential enzyme for the reduction of ribonucleotides to dNTP (2009 Bergamaschi). In hematopoietic stem cells, inhibition of HIV-1 integration by p21 was observed, while HIV-1 reverse transcripts stay unaffected (2007 Zhang). Interestingly, previous findings suggest that selective upregulation of the p21 proteins occurs in activated primary CD4+ T cells of HIV-1 elite controllers, inhibiting transcriptional elongation of HIV-1 mRNA and HIV-1 reverse transcription.

The evidence showed that HIV-1 mRNA transcription affected by p21 was, at least in part, through the indirect blockade of enzymatic activity of a host enzyme Cdk9, known to phosphorylate RNA polymerase II for enhancing HIV-1 mRNA elongation. However, it was not clear how p21 affects HIV-1 reverse transcription during pre-integration stage. Our in vitro and in vivo data in this study conclusively demonstrate a role of p21 for a cell-intrinsic mechanism underlying the inhibition of HIV-1 reverse transcription in CD4+ T cells from elite controllers by blocking phosphorylation of a host enzyme Cdk2, shown its ability to phosphorylate HIV-1 reverse transcription on a highly conserved residue Thr 216. Taken together, the cell-intrinsic inhibitor p21 shows two distinct pathways of indirect suppression of HIV-1 infection at HIV-1 replication steps – mRNA elongation and reverse transcription by indirectly interacting human host cells. Although the indirect inhibition of p21 appears to be less effective than the inhibition of traditional pharmacological RT inhibitors by interacting directly with HIV-1 gene products, but the indirect mechanism can more possibly prevent viral mutational escape, a major cause of drug resistance. This might explain why no selection of viral escape variants of HIV-1 infected lymphocytic and monocytic cells in the presence of pharmaceutical Cdk inhibitors was observed in ex vivo cultures (2001 Wang).
As for the mechanism of host cell control against HIV-1, these data provide very useful insight into an cell-intrinsic immune defense mechanism in CD4+ T cells from elite controllers, where Cdk2 phosphorylation of HIV-1 RT, an essential modification for enhancing efficiency of HIV-1 reverse transcriptase, is inhibited by highly unregulated p21 proteins. Interestingly, a recent report showed that p21 also regulates the HIV-1 restriction factor by inhibiting Cdk1-mediated phosphorylation of SAMHD1 into an active form with its ability to restrict HIV-1 during reverse transcription in macrophages (2014 Pauls). The findings suggest that p21 appears to act as a master of regulator controlling HIV-1 infection in different HIV-1 replication steps by indirect interaction with viral proteins. Of many immune responses to HIV-1 infection, the majority of data suggested the emergence of HIV-specific CD4+ and CD8+ T-cells are correlated to a decline in HIV viral loads as a result of immune protection (2010 Streeck, 1999 Jin, 1999 Schmitz, 1994 Koup, 1994 Borrow). However, further data revealed that effective T cell responses are not always necessary for long-term virus control of HIV-1 infection (2011 Chen, 2011 Jager, 2008 Emu, 1998 Martin). Many HIV-1 elite controllers are, for example, capable of maintaining normal CD4+ T-cell counts and undetectable levels of HIV-1 replication over very long follow-up periods in the absence of strong HIV-1-specific CD8+ T cells (2008 Emu). Overall, although different mechanism may act together or specifically in HIV-1 infected individuals in the synergistic networks among host proteins, viral proteins and the CD8+ T cell-mediated adaptive immunity, it can be speculated at the moment that the cell-intrinsic inhibition of Cdk phosphorylation by p21 may provide a clue to induce a long-term, drug-free remission or a functional cure of HIV-1 infection (2014 Leng).
Chapter 4: Summary and Future Directions

This thesis contributes to developing MALDI mass spectrometric methods without expensive isotope labels and time-consuming calibration curves for single- and double-site phosphopeptide quantification in \textit{in vitro} assays (chapter 2). Through mass spectrometric techniques, Cdk2 phosphorylation of HIV-1 reverse transcriptase at a specific site (Thr 216) has been successfully identified that reveals a cell-intrinsic mechanism of inhibition of HIV-1 reverse transcription with demonstration of \textit{in vivo} experiments (chapter 3).

\textbf{Chapter 2.2} described the results of the DRIP-Q method development and indicated that it may be worthwhile to utilize label-free and calibration-curve-free MALDI-MS as an easy-to-use and robust alternative approach to quantifying phosphopeptides for three different peptide models encompassing phosphoserine, phosphothreonine and phosphotyrosine, and further may be applied in \textit{in vitro} assays, such as for testing drug efficacy. The method is not limited by the difference in MALDI laser energy provided by different MALDI instrument, because the results showed that the change in laser energies of MALDI does not affect the feasibility of the method if the same energy is controlled in one set of experiments. The results obtained demonstrated that the DRIP-Q method provides a wide dynamic range of at least two orders of magnitude and a good reproducibility of inter-day CV of 0.65 \% (n=3). The method was further validated by comparing with an isotope dilution method that most quantification of phosphorylation relies on, showing the two methods are comparable with excellent accuracy, and by assessing the \textit{in vitro} inhibition of Cdk5, showing IC\textsubscript{50} values of pharmaceutical Cdk inhibitors consistent with the literature.

Although the DRIP-Q method demonstrated its feasibility in \textit{in vitro} assay, the analytical performance of the DRIP-Q method will need to be challenged by performing phosphorylation analysis in more complex sample systems, such as peptide digests from purified phosphoproteins.
For example, it would be interesting to synthesize the unphosphorylated and phosphorylated tryptic HIV-1 RT peptides containing the identified phosphorylation site Thr216 for probing phosphorylation quantification of HIV-1 RT protein phosphorylation by using the DRIP-Q method.

In addition, the method can be further refined and expanded into analysis of a larger number of peptides (In this study, there are only three peptides, one each with pSer, pThr, and pTyr). For example, a systematic peptide sequences with different number of basic or acidic residues can be subjected to the proposed DRIP-Q method for testing its general feasibility. The peptide/phosphopeptide samples can also be assigned to different laboratories, and the results from the inter-labs can compared to evaluate the performance of the DRIP-Q method.

Chapter 2.3 proposed another label-free, calibration curve-free method by using tandem mass spectrometric technique for quantification of the double-site phosphorylation of AQP2. This new method overcomes the difficulty in traditional MALDI-MS analysis of two distinct singly phosphorylated peptides with an identical mass. With the use of economical CID-MS$^3$, the method provides an efficient way to quantify singly AQP2 phosphopeptides at each of two discrete phosphorylation sites (Ser-256 and Ser-261) by measuring the change in the intensity ratios of the MS$^3$ fragments in response to the amounts of pSer-256 and pSer-261 in a sample mixture. The results successfully demonstrated two assumptions that support the mathematic approach used in the method: (1) the relative MS$^3$ fragment intensities of individual phosphopeptide ($\alpha$ for pSer-256 and $\beta$ for pSer-261) remain constant while the amount of the phosphopeptide changes; and (2) MS$^3$ signal responses are directly proportional to sample concentrations. It has been observed that the change in collision energy may affect the values of $\alpha$ and $\beta$, but does not affect the feasibility of the method by providing the same collision energy to all samples in an experiment. The data suggest
that the method provides a wide dynamic range of at least 3 orders of magnitude and good accuracy with comparable results to the traditional calibration curve method. The method was further effectively applied to an in vitro kinase assay to measure the change of the phosphorylation at Ser-256 and Ser-261 of AQP2 in the presence of both PKA and Cdk1/Cyclin B.

One of the important future extensions of the work involves application of in vitro kinase assays by assessing AQP2 double-site phosphorylation in the presence of two different kinases. In this study, we observed an increase of pSer-261 at a high level of Cdks. However, insufficient consideration given to this method is that the production of doubly phosphoAQP2 peptides can lead to an unpredictable decrease in the absolute amount of singly phospho- and unphosphopeptides. In order to accurately quantify the total absolute amount of single phosphorylation of AQP2 by PKA or Cdk1 in an in vitro kinase assay, there is a need to obtain standards of doubly phosphoAQP2 peptides on Ser-251 and Ser-261. A peak of doubly phosphoAQP2 will be observed on a MS full scan spectrum with a mass shift of two-phosphate-addition from the unphosphoAQP2 peptide. The amount of occurring double phosphorylation can then readily measured by monitoring the change in signal intensity of the peak of doubly phosphoAQP2 in a sample, compared to the standard. Therefore, the proposed method is able to quantify singly phosphopeptides on the sites Ser-256 or Ser-261 with consideration of the double phosphorylation event. A further extension of this methodology will be to apply it to phosphoprotein digests, to quantify double-site phosphorylation in the context of a whole protein.

Chapter 3 revealed a cell-intrinsic mechanism of blocking HIV-1 reverse transcription by discovery of Cdk2 phosphorylation site at a highly conserved Thr 216 of HIV-1 RT. After assessing phosphorylation of the HIV-1 RT protein using a variety of Cdk kinases in an in vitro kinase assay, the results show that Cdk2 imparts Cdk- and site consensus- specificity to RT phosphorylation. The
host Cdk2 kinase does not only phosphorylate HVI-1 RT protein but also supports HIV-1 reverse transcription. Site-directed mutagenesis experiments was conducted based on substitution of Thr 216 by Ala, demonstrating the vital contribution of the Cdk2-dependent phosphorylation of RT to viral fitness through modulation of viral reverse transcription. In the in vitro phosphorylation assay, the results of dimethyl-based stable isotope quantification showed that 92% Thr261 on p51 subunit of RT and 62.1% Thr216 on p66 subunit of RT are phosphorylated. The in vitro phosphorylated RT was then subjected to an enzymatic activity assay and limited proteolytic experiments, suggesting Cdk2 phosphorylation does increase the functional activity and stability of HIV-1 RT. Further p21 knockdown experiments provide strong evidence that the intrinsic Cdk inhibitor, p21, can indirectly inhibit HIV-1 reverse transcription and reduce the efficacy of viral transcription in CD4+ T cells from elite controllers by blocking Cdk2 phosphorylation of RT.

However, there are several particular questions to the proposed cell-intrinsic mechanism: (1) What does Cdk2 phosphorylation affect a conformational change of the RT protein? (2) Whether Cks1 can be involved the proposed mechanism by which the Cks1 protein may modulate processivity of Cdk2/CyclinA? And (3) How do Cdk kinases associate with other viral HIV-1 Cdk phosphorylation substrates, leading to a change in HIV-1 viral replication?

As mentioned above in Chapter 3, the RT protein was shown to be phosphorylated at the two subunits, p66 and p51, known as to have a functional role and structural role, respectively. The Cdk2 phosphorylation site is situated near the interface between fingers and palm subdomain of the RT structure. It has been reported that, for the p66 subunit, the palm subdomain accommodates the enzymatic active site; the finger subdomain stabilizes the RNA template; the thumb subdomain interacts with the DNA strand (2014 Nowak), and the p51 subunit stabilized RT heterodimer structure that is required for its function (2005 Depollier). In this study, the addition of a phosphate
group during Cdk2 phosphorylation may result in a significant conformational change associated interaction with RNA template as well as the DNA strand, and dimerization of the two subunits, and consequently cause the change in enzymatic activity of RT.

In addition, it is interesting to consider that experiments assessing Cdk2 phosphorylation in the presence of Cks1 to investigate the intriguing role of Cks1 in the RT phosphorylation. The phosphorylated Thr216 site might also be a Cks1 consensus site containing the Cdk-acceptor site (TP) and a bulky hydrophobic residue Tryptophan in the -3 position that facilitates binding of Cks1 to phospho216-Cks1. In this study, the *in vitro* kinase assay performed in the absence of the Cks1 protein might have underestimated the other potential phosphorylation sites that may be induced by Cks1. Considering the western-blotting data cannot give a clue because the data were obtained by probing Cdk2 phosphorylation with only Thr216 specific antibodies; therefore the Cks1 study will be one of the important future directions for better understanding of the cell-intrinsic mechanism.

One more future work will be aimed at characterization of any potential phosphorylation sites on other viral proteins. The pharmaceutical Cdk inhibitors were consistently observed to inhibit HIV-1 replication with no selection of viral escape variants in the infected lymphocytic and monocytic cells (2001 Wang). In addition, another cell-intrinsic mechanism of HIV-1 restriction associated with the Cdk inhibitor p21 was recently discovered that p21 acts as a negative regulatory master to reduce HIV-1 reverse transcription through blocking Cdk1-mediated phosphorylation of a restriction factor SAMHD1. Altogether, Cdk phosphorylation appears to be essential for HIV-1 virus to harnesses host machinery to replicate. In order to comprehend the whole complex Cdk phosphorylation network in HIV-1 replication, it would be necessary to identify other potential phosphorylation sites on the entire HIV viral proteins by using similar strategies to this research in the future studies. The deeper exploration of the role of phosphorylation in HIV-1 replication may
offer long term potential for development of prophylactic or therapeutic strategies for HIV-1 patients.
References


Emu, B.: HLA class I-restricted T-cell responses may contribute to the control of human immunodeficiency virus infection, but such responses are not always necessary for long-term virus control. J. Virol. 82, 5398 –5407 (2008)


Zhang, H., Ge, Y.: Comprehensive Analysis of Protein Modifications by Top-Down Mass Spectrometry, Ge Circ Cardiovasc Genet, DOI: 10.1161/CIRCGENETICS.110.957829 (2011)


