9-2018

Characterization and Quantification of Covalent Modification of Proteins Using Mass Spectrometry

Pratikkumar N. Rathod

The Graduate Center, City University of New York

How does access to this work benefit you? Let us know!
Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

Part of the Analytical Chemistry Commons

Recommended Citation
https://academicworks.cuny.edu/gc_etds/2764

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
Doctoral Dissertation

Characterization and Quantification of Covalent Modification of Proteins Using Mass Spectrometry

by

Pratikkumar N Rathod

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

2018
Characterization and Quantification of Covalent Modification of Proteins Using Mass Spectrometry

by

Pratikkumar N Rathod

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.

Date

Dr. Emmanuel J. Chang
Chair of Examining Committee

Date

Dr. Brian Gibney
Executive Officer

Supervisory Committee:

Dr. Adam Profit
Dr. Mande Holford

THE CITY UNIVERSITY OF NEW YORK
Abstract

Characterization and Quantification of Covalent Modification of Proteins Using Mass Spectrometry

by

Pratikkumar N Rathod

Thesis Advisor: Dr. Emmanuel J. Chang

Identification and characterization of various post-translational modifications of protein is a key to understanding many unknown cellular processes. In the last few decades, mass spectrometry has evolved as an essential and effective analytical tool for qualitative and quantitative analysis of proteins. In this research, we have developed a novel MALDI-MS\(^2\) based quantification method for Desmosine and Isodesmosine, which served as cross-linking amino acids of elastin, in order to measure the elastin degradation in the body. This is the first quantification method that not only illustrates the potential of MALDI-Ion Trap MS\(^2\), but also improvement over the current LC-MS method, in terms of analysis time and solvent consumption, while maintaining similar analytical characteristics. The method is utilized to evaluate the time-dependent degradation of Des upon UV radiation (254nm) and result found to be consistent with quantification by \(^1\)H NMR.

This work also involves the investigation of potential phosphorylation sites and evaluation of its role in various biochemical processes during HIV infection. Based on the results from different phosphorylation prediction algorithms, many in-vitro kinase assays were performed on HIV-derived peptides/proteins in presence of potential kinases. We have successfully identified few novel interactions between host-kinases/HIV phosphorylation substrates. These include the interactions of phosphorylation sites of Vif, Nef and Capsid.
proteins with protein kinase C (PKC), protein kinase A (PKA), and p38 MAPK respectively.

Moreover, this work includes the development of cell-active inhibitors for cysteine cathepsins, a class of enzymes involve in many important cellular processes and in various disorders. In this study, we have synthesized library of two different classes of molecules containing oxirane and vinylsulfonate moieties. Various cell-based experiments were conducted to successfully demonstrate intracellular inhibition of cysteine cathepsin by these developed inhibitory molecules. The result of our study shows 2-(2-ethylphenylsulfonyl) oxirane is cell-permeable and irreversible inhibitor of cathepsin B. On the other hand, peptidyl vinylsulfonate inhibitor (KD-1) is highly potent and selective cathepsin L inhibitor.
Acknowledgement

First of all, I would like to thank Almighty for everything he has offered me and for navigating me on the road of the life.

Most importantly, I would like to offer few words of gratitude to my mentor, Dr. Emmanuel J. Chang for his invaluable guidance and support throughout my PhD life. He remained very patient, approachable and understanding with me. I do not have enough words to thank him and I believe I will be thankful to him for my entire life. I am also very grateful to Dr. Sanjai Kumar for allowing me to become a part of his research team for cathepsin project.

I would really like to acknowledge my parents and my in-laws for their unwavering support and prayers that have helped me to overcome many roadblocks throughout my life and believing in me more than I do in myself.

I am also very grateful to my sister (Falguni), jiju (Amit), brother-in-law (Parth), little nephew (Nivaan), and friends (Boosters group) for boosting me up every time I am felt little down, lightning up my mood and for being my go to people in last few years.

There are number of people who always cheer me up in the lab and supported me in every way they can. I would like to thank Jasy, Dr. Mark and all the undergraduates with whom I have worked with.

Last but not the least; I would like to thank my lovely wife, Dimpy for showering love, care and offering me emotional and mental support throughout my graduation journey.
Abstract

Acknowledgment

Table of Contents

List of Figures

List of Tables

List of Schemes

Chapter 1  Introduction to mass spectrometry

1.1  Background  2

1.2  History  2

1.3  Ionization source  3

1.3.1  Matrix-Assisted Laser Desorption/Ionization  3

1.3.2  Electrospray Ionization  6

1.4  Mass analyzer  8

1.4.1  Ion-trap analyzer  9

1.4.2  Time-of-Flight (ToF) analyzer  11

1.4.3  Hybrid mass analyzer  13

1.5  Tandem mass spectrometry  14

1.6  Application of mass spectrometry  17

1.6.1  Qualitative MS analysis  17

1.6.2  Quantitative MS analysis  18

1.6.2.1  Label-free approach  19

1.6.2.2  Isotopic labeling  20
1.6.2.3 Isobaric labeling

Chapter 2 Development and application of simple and fast quantification method for Desmosine and Isodesmosine using MALDI-Ion Trap tandem mass spectrometry

2.1 Introduction

2.2 Pathological disorders

2.3 Quantification methods

2.4 Development of rapid and sensitive quantification method for desmosine using MALDI-LTQ

2.4.1 Materials and Methods

2.4.1.1 Chemicals

2.4.1.2 Mass spectrometry

2.4.1.3 UV-irradiation of Des

2.4.1.4 Proton NMR

2.4.1.5 Sample preparation for detection limit study in urine and serum

2.4.2 Results and Discussion

2.4.2.1 MALDI-MS² analysis of Des, Isodes and Labeled-Des

2.4.2.2 Linearity and dynamic range

2.4.2.3 Detection limit

2.4.2.4 Reproducibility

2.4.2.5 UV-irradiation of Des

2.4.2.6 Comparison to existing methods
2.4.3 Conclusions

2.5 Application of MALDI-Ion Trap tandem mass spectrometry based quantification method to measure effect of UV-A radiation on the elastin cross links

2.5.1 Material and Methods

2.5.1.1 Sample preparation

2.5.1.2 Histology and microscopy

2.5.1.3 Transmission Electron Microscopy (TEM)

2.5.1.4 $^{13}$C NMR experimental parameters

2.5.1.5 Sample hydrolysis and quantification with Labeled-Des

2.5.2 Results and Discussion

2.5.2.1 Biochemical changes in elastin due to UV-A exposure

2.5.3 Summary and future directions

Chapter 3 Identification and characterization of phosphorylation sites in HIV proteins using mass spectrometry

3.1 Protein phosphorylation

3.2 HIV Introduction

3.2.1 HIV Genome

3.2.2 Role of different HIV proteins

3.2.3 Known phosphorylation events in HIV proteins

3.2.4 Experimental workflow

3.3 Materials and Methods
3.3.1  *In-vitro* kinase assay of HIV-peptides/proteins  61
3.3.2  In-gel digestion  62
3.4  Results and Discussion  62
3.4.1  Investigation of potential CDK-mediated phosphorylation sites of HIV proteins  63
3.4.2  Investigation of potential phosphorylation sites in HIV-1 accessory proteins  67
3.4.3  Investigation of potential phosphorylation sites in HIV-1 structural proteins  71
3.5  *In-vitro* kinase assay on HIV-proteins (Vif, Nef and CA)  75
3.6  Summary and future directions  75

**Chapter 4  Development of cysteine cathepsin inhibitor**

4.A  Introduction to cysteine cathepsins
4.A.1  Background  79
4.A.2  Structure of cysteine cathepsin  80
4.A.3  Substrate binding  82
4.A.4  Localization, Regulation and Physiological role  84
4.A.5  Pathophysiology  86
4.A.6  Known cathepsin inhibitors  87
4.B  Development of cell-active cysteine cathepsin B inhibitor
4.B.1  Introduction  90
4.B.2  Library of cysteine cathepsin inhibitors  91
4.B.3 Result and Discussion
4.B.3.1 Type of interaction and inhibition
4.B.3.2 Computational analysis
4.B.3.3 Selectivity
4.B.3.4 Cell permeability
4.B.4 Conclusion

4.C Development of highly potent, cell-active cathepsin L inhibitor
4.C.1 Introduction
4.C.2 Library of cathepsin L inhibitor
4.C.3 Synthesis
4.C.4 Result and Discussion
4.C.4.1 Selectivity
4.C.4.2 *In-vivo* cathepsin L inhibition
4.C.4.3 Cell migration inhibition
4.C.5 Conclusion

**Contribution to science community**

**References**
# List of Figures

## Chapter 1

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Schematic representation of the components of a mass spectrometer</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Schematic representation of MALDI ionization source</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Schematic representation of ESI source</td>
<td>7</td>
</tr>
<tr>
<td>1.4</td>
<td>Schematic representation of 3D ion trap</td>
<td>10</td>
</tr>
<tr>
<td>1.5</td>
<td>Schematic representation of 2D ion trap</td>
<td>11</td>
</tr>
<tr>
<td>1.6</td>
<td>Schematic representation of linear and reflectron ToF analyzer</td>
<td>12</td>
</tr>
<tr>
<td>1.7</td>
<td>Various tandem mass scanning modes</td>
<td>16</td>
</tr>
<tr>
<td>1.8</td>
<td>Fragmentation of different bonds along peptide sequence</td>
<td>16</td>
</tr>
<tr>
<td>1.9</td>
<td>Workflow of various proteomics quantitative approaches</td>
<td>20</td>
</tr>
<tr>
<td>1.10</td>
<td>Structure of ICAT reagent</td>
<td>21</td>
</tr>
<tr>
<td>1.11</td>
<td>Structure of Isobaric mass tag</td>
<td>22</td>
</tr>
</tbody>
</table>

## Chapter 2

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Structure of desmosine and isodesmosine</td>
<td>25</td>
</tr>
<tr>
<td>2.2</td>
<td>MALDI-MS² spectra of Des, Isodes and Labeled-Des</td>
<td>32</td>
</tr>
<tr>
<td>2.3</td>
<td>Linearity of Des, Isodes and equimolar concentration ratio of Des:Isodes</td>
<td>35</td>
</tr>
<tr>
<td>2.4</td>
<td>Assessment of reproducibility of MALDI-MS² method</td>
<td>37</td>
</tr>
<tr>
<td>2.5</td>
<td>UV-irradiated Des degradation</td>
<td>38</td>
</tr>
<tr>
<td>2.6</td>
<td>Histological image of elastic fibers-with and without UV-irradiation</td>
<td>46</td>
</tr>
<tr>
<td>2.7</td>
<td>TEM image of elastin samples- with and without UV-irradiation</td>
<td>47</td>
</tr>
<tr>
<td>2.8</td>
<td>$^{13}$C MAS NMR spectra of elastin- with and without UV-irradiation</td>
<td>48</td>
</tr>
</tbody>
</table>
Chapter 3

Figure 3.1  Structure of genome organization of HIV-1 particle 56

Figure 3.2  *In-vitro* kinase assay result of VIF-B-21 peptide 65

Figure 3.3  MS$^2$ spectrum of the phosphorylated VIF-B-21 peptide 66

Figure 3.4  Vif multiple sequence alignment and bioinformatics prediction score 68

Figure 3.5  Multiple sequence alignment of HIV-capsid protein 72

Figure 3.6  Mass spectra of GAG-B-22 peptides in presence of p38 MAPK 74

Chapter 4

Figure 4.1  Cysteine cathepsin- structural components of preproenzyme forms 80

Figure 4.2  Crystal structure of cathepsin L 81

Figure 4.3  Pictorial representation of substrate binding sites of cathepsins 83

Figure 4.4  Chemical structure of compound 5-a lead cathepsin B inhibitor 90

Figure 4.5  Origin of arylsulfonyl-substituted oxiranyl moiety 91

Figure 4.6  Inactivation kinetics of cathepsin B in presence of compound 5 95

Figure 4.7  Time and concentration-dependent inhibition of cathepsin B in presence of compound 5 96

Figure 4.8  MS analysis of compound 5 mediated-cathepsin B inhibition 97

Figure 4.9  Computational modeling of compound 5-cathepsin B inhibition complex 98

Figure 4.10  Study of half-life of compound 5 at physiological pH 101

Figure 4.11  Intracellular inhibition of cathepsin B activity by compound 5 102

Figure 4.12  Inhibition of E-cadherin degradation by cathepsin B in presence of compound 5 103
| Figure 4.13 | Effect of compound 5 on cell-migratory behavior of MDA-MB-231 cells | 104 |
| Figure 4.14 | Hybrid design of cathepsin L inhibitory molecule- KD-1 | 110 |
| Figure 4.15 | Concentration-dependent inhibition of cathepsin L activity in presence of KD-1 | 112 |
| Figure 4.16 | Cell-permeability and intra-cellular inhibition of collagen degradation activity of cathepsin L in presence of KD-1 | 114 |
| Figure 4.17 | Inhibition of cell-migratory behavior of MDA-MB-231 cells in presence of KD-1 | 115 |
List of Tables

Chapter 1
Table 1.1 Commonly used MALDI matrices for different analytes 6
Table 1.2 List of other ion sources in MS and its specific characteristics 8
Table 1.3 List of characteristics of mass analyzers 13
Table 1.4 Performances comparison and characteristics of various hybrid analyzers 14

Chapter 2
Table 2.1 Detection limit of MALDI-MS² method 36

Chapter 3
Table 3.1 List of HIV-1 proteins and their primary functions 57
Table 3.2 List of CDK-mediated phosphorylation motif containing HIV-peptides 64
Table 3.3 Result of in-vitro kinase assay of CDK-mediated phosphorylation site containing HIV-peptides 66
Table 3.4 Result of in-vitro kinase assay of 8 HIV-peptide containing CDK-mediated phosphorylation site in presence of other kinases 67
Table 3.5 Result of in-vitro kinase assay of peptides from HIV-auxiliary proteins 69
Table 3.6 Phosphorylation prediction score for the sites of HIV-capsid protein 72
Table 3.7 List of overlapping Capsid-peptides containing Ser33 residue 73

Chapter 4
Table 4.1 Library of synthetic cathepsin B inhibitory molecules 93
Table 4.2 Selectivity of inhibition of compound 5 towards cysteine cathepsins 99
Table 4.3 List of vinylsulfone ester moiety containing inhibitory molecules 109
Table 4.4 2nd order inactivation rate constant of KD-1 for closely-related enzymes.
## List of Schemes

### Chapter 4

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scheme 4.1</td>
<td>Various synthetic approaches in the development of cysteine cathepsin inhibitors</td>
<td>92</td>
</tr>
<tr>
<td>Scheme 4.2</td>
<td>Synthesis steps for the development of Cathepsin L inhibitor (KD-1)</td>
<td>111</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction to Mass Spectrometry

1.1 Background
1.2 History
1.3 Ionization sources
   1.3.1 Matrix-Assisted Laser Desorption/Ionization (MALDI)
   1.3.2 Electrospray Ionization (ESI)
1.4 Mass analyzers
   1.4.1 Ion-trap analyzer
   1.4.2 Time-of-Flight (ToF) analyzer
   1.4.3 Hybrid mass analyzer
1.5 Tandem mass spectrometry
1.6 Application of mass spectrometry
   1.6.1 Qualitative MS analysis
   1.6.2 Quantitative MS analysis
      1.6.2.1 Label-free approach
      1.6.2.2 Isotopic labeling
      1.6.2.3 Isobaric labeling
1.1 Background

Mass spectrometry (MS) is a powerful analytical technique that separates ionized molecules based on their mass-to-charge (m/z) ratio. The overall process of MS involves the conversion of the sample into gaseous ions, separation of gaseous ions based on their m/z followed by measurement of current to determine the abundance of ions produced. Mass spectrometry is utilized for successful qualitative and quantitative analysis of organic and inorganic analytes. Moreover, due to recent advancement, MS has evolved as a method of choice for analysis of complex mixture of large biomolecules such as polypeptides and proteins. A mass spectrometer, the first of which was developed in 1919 by F.W. Aston [1], consists of a sample inlet, an ion source, a mass analyzer, a detector and a data system. The components including an ion source, a mass analyzer and a detector are under the vacuum system.

![Schematic representation of the main components of a mass spectrometer](image)

**Figure 1.1** Schematic representations of the main components of a mass spectrometer

1.2 History

The current state-of-the-art in MS has been achieved as a result of 120 years of incessant developments. A few highlights of this development are listed below:

1897: Discovery of the electrons and determined its m/z - J.J. Thompson [2]

1919: Development of first mass spectrometer - F.W. Aston [1]
1953: Describe quadrupole & ion trap analyzer in a patent- W. Paul & H. Steinwedel [3]

1958: Introduction of first commercial LTOF instrument- Bendix


1999: Development of the orbitrap- A.A. Makarov [8]

1.3 Ionization sources

A prerequisite for successful MS is that the analytes of interest must be in ionized form. Furthermore, the ions must be introduced in the gas phase into vacuum system of mass spectrometer. The vaporization to the gas phase in vacuum system can be easily accomplished for volatile and thermostable analytes while offer difficulties for non-volatile and thermolabile analytes. These analytes require either desorption or desolvation methods to be analyzed by MS, so called soft ionization. In biological mass spectrometry, the polar and non-volatile characteristic of most biomolecules, demand soft ionization methods for MS analysis. Currently two vastly accepted soft ionization methods for MS analysis of the biomolecules are, matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI). In this research, I have employed the MALDI ionization technique for various applications.

1.3.1 Matrix-Assisted Laser Desorption/Ionization (MALDI)

MALDI ionization technique was developed by Karas and Hillenkamp in 1987 [5, 6]. Since then MALDI is widely accepted and emerge as a powerful tool for MS analysis of broad range of large, thermolabile and non-volatile molecules such as proteins, synthetic polymers, and organic and inorganic molecules. MALDI utilizes small organic acid molecules as a matrix to assist desorption and ionization of the analytes. Importantly, a major advantage MALDI offers is
easy sample preparation simultaneously with salt and buffer tolerance [9, 10]. There are mainly two steps in MALDI analysis. In the first step, the analyte is co-crystallized with molar excess of matrix molecules, which usually is small organic acid capable of absorbing laser wavelength. While in the second step upon laser radiation under vacuum, the vaporization of matrix molecules occur which carry the analyte molecules and result in desorption and ionization of the analyte molecules. Thus, matrix play critical role in indirect analyte vaporization and serve as a proton donor to assist ionization of the analyte molecules [11].

The exact mechanism of the MALDI is not completely understood [12, 13]. However, it is believed that irradiation of laser results in accumulation of large amount of energy in matrix molecules due to rapid heating and lead to sequential processes like, sublimation of matrix crystals, ablation of local crystal surface and formation of matrix plume containing intact analyte molecules [14]. Various theories have been introduced to explain the ionization process in MALDI including, ion-molecule reaction, gas-phase photoionization, excited state proton
transfer, and desorption of preformed ions. To date, none of the above processes has been able to explain the exact mechanism of the ionization process in MALDI [15].

The selection of matrix and optimization of sample preparation protocol are two vital criteria in the success of MALDI as ionization source in MS analysis. The important characteristics required for an efficient MALDI matrices are- a) high vacuum stability, b) capable of absorbing laser wavelength, c) lower molecular mass, d) ability of analyte ionization, and e) inert nature. Commonly known MALDI matrix molecules along with the class of analytes for which they used for are listed in Table 1.1. A number of sample preparation methods have been discussed and evaluated for MALDI MS analysis including thin-layer method [16, 17], the sandwich method [18], and dried-droplet method [19]. Among these, the widely accepted sample preparation method is the “dried-droplet method”. In this method 1-2 µL of analyte solution is mixed with 5-10 µL of matrix solution. 0.5-1µL of this mixture is deposited on MALDI-probe; usually consist of a metal plate with regular array for sample application. This solution on a metal plate is set at room temperature to dry before the sample is loaded in mass spectrometer.

Due to the incorporation of pulsed laser in MALDI, it produces ions in bundles during ionization. Hence, the mass analyzers such as time-of-flight (ToF), fourier transform ion cyclotron resonance (FTICR), and ion trap, which are compatible with pulsed ionization, have been successfully coupled with MALDI. Additionally, in the efforts to explore a new method of injecting ions into ToF, Krutchinsky et al., demonstrated successful application of orthogonal injection of MALDI ions to ToF through collision damping interface. This approach makes mass measurement process independent of ionization process and enable MALDI source to couple with nearly any mass analyzer [20]. These instruments have capabilities to perform MS²
analysis, widely utilized in qualitative and quantitative analysis of biomolecules, and offer high selectivity and sensitivity.

**Table 1.1** Commonly used MALDI matrices for different analytes

<table>
<thead>
<tr>
<th>Analyte</th>
<th>MALDI matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peptides/Proteins</td>
<td>α-cyano-4-hydroxyccinnamic acid (CHCA)</td>
</tr>
<tr>
<td></td>
<td>2,5-Dihydroxybenzoic acid (DHB)</td>
</tr>
<tr>
<td></td>
<td>3,5-Dimethoxy-4-hydroccinnamic acid (SA)</td>
</tr>
<tr>
<td>Oligonucleotides</td>
<td>Trihydroxyacetophenone (THAP)</td>
</tr>
<tr>
<td></td>
<td>3-Hydroxyopicolinic acid (HPA)</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>2,5-Dihydroxybenzoic acid (DHB)</td>
</tr>
<tr>
<td></td>
<td>α-cyano-4-hydroxyccinnamic acid (CHCA)</td>
</tr>
<tr>
<td></td>
<td>Trihydroxyacetophenone (THAP)</td>
</tr>
<tr>
<td>Organic molecules</td>
<td>2,5-Dihydroxybenzoic acid (DHB)</td>
</tr>
</tbody>
</table>

1.3.2 *Electrospray Ionization (ESI)*

The 2002 Nobel Prize Laureate in Chemistry, J. Fenn, has first applied ESI to mass spectrometric analysis to measure molecular weight of large molecules in 1988 [7]. As a result of many parallel efforts, ESI extended its application beyond protein analysis and demonstrated success in analyzing other polymers, as well as small polar molecules. The ESI technique made a breakthrough for MS owing to its high sensitivity and ability to readily couple to high-performance liquid chromatography (HPLC).

In ESI, the solution of analyte molecules is allowed to pass through a capillary tube. The strong electric field obtained by applying a potential difference of 3-6kV between capillary and counter-electrode is used to produce electrospray of the analyte solution in ESI [7]. The presence
of an electric field induces a charge accumulation at the liquid surface located at the end of the capillary results in formation of highly charged droplets. A coaxial flow of an inert gas at low flow rate limits the dispersion space of the spray. To evaporate the solvent molecules from the droplet surface, an inert gas flow is directed to the droplets at the spray exit where a “Tylor cone” shape is formed. Evaporation of solvent molecules would shrink the droplets and increase in charge per unit volume. The breakdown of droplets into small highly charged particles occurs when cumblic repulsion between the similarly charged molecules is higher than the force of surface tension of the droplets. This process continue until only one charged analyte molecule is left before entering in mass analyzer.

Beside these two widely accepted ionization techniques, there are also few other techniques have been evolved in early development stage of mass spectrometry with their unique pros and cons. The list of these ionization techniques with their characteristics are stated in Table 1.2.
Table 1.2 List of other ion sources in MS and its specific characteristics

<table>
<thead>
<tr>
<th>Ion source in MS</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Electron Impact (EI)</strong></td>
<td>Use beam of electron to form radical cation</td>
</tr>
<tr>
<td></td>
<td><strong>Advantage</strong></td>
</tr>
<tr>
<td></td>
<td>Well understood</td>
</tr>
<tr>
<td></td>
<td>Applied to all volatile compounds</td>
</tr>
<tr>
<td></td>
<td><strong>Disadvantage</strong></td>
</tr>
<tr>
<td></td>
<td>Requires thermostable and volatile compounds</td>
</tr>
<tr>
<td><strong>Chemical Ionization (CI)</strong></td>
<td>Use ion-molecule reaction</td>
</tr>
<tr>
<td></td>
<td><strong>Advantage</strong></td>
</tr>
<tr>
<td></td>
<td>Simple mass spectra</td>
</tr>
<tr>
<td></td>
<td><strong>Disadvantage</strong></td>
</tr>
<tr>
<td></td>
<td>Thermostable and volatile compounds</td>
</tr>
<tr>
<td></td>
<td>Less fragmentation- not informative</td>
</tr>
<tr>
<td><strong>Fast Atom Bombardment (FAB)</strong></td>
<td>Bombarded with fast atom beam (6keV xenon atoms)</td>
</tr>
<tr>
<td></td>
<td><strong>Advantage</strong></td>
</tr>
<tr>
<td></td>
<td>Simple, rapid</td>
</tr>
<tr>
<td></td>
<td>High tolerance to various samples</td>
</tr>
<tr>
<td></td>
<td><strong>Disadvantage</strong></td>
</tr>
<tr>
<td></td>
<td>High chemical background at low mass region</td>
</tr>
<tr>
<td></td>
<td>Analyte must be soluble in liquid matrix</td>
</tr>
<tr>
<td><strong>Atmospheric Pressure</strong></td>
<td>Use corona discharge</td>
</tr>
<tr>
<td>Chemical Ionization (APCI)</td>
<td><strong>Advantage</strong></td>
</tr>
<tr>
<td></td>
<td>Excellent LC/MS interface</td>
</tr>
<tr>
<td></td>
<td>Compatibility with MS/MS method</td>
</tr>
<tr>
<td></td>
<td><strong>Disadvantage</strong></td>
</tr>
<tr>
<td></td>
<td>Multiply charged species- difficult interpretation</td>
</tr>
<tr>
<td></td>
<td>Very sensitive to contaminant</td>
</tr>
</tbody>
</table>

1.4 Mass Analyzers

The separation of generated gas-phase ions in ionization source takes place in mass analyzer based on their mass-to-charge (m/z). There are mainly 5 different types of mass analyzers: 1) Ion trap, 2) Time-of-flight (ToF), 3) Quadrupole, 4) Fourier Transform-Ion-
Cyclotron Resonance (FTICR), and 5) Sector (magnetic or electric). As a result of continue efforts in mass analyzers development, a new type of mass analyzer, orbitrap, is recently been introduced in 2005. All the mass analyzers use static or dynamic electric and magnetic fields alone or in combination. These mass analyzers can be classified in mainly two categories based on the resolution: a) high resolution analyzer (FT-ICR, ToF and Orbitrap), b) low resolution analyzer (Ion-trap and Quadrupole). The performance of mass analyzer is evaluated based on following characteristics: i) mass range, ii) the scan speed, iii) the transmission, iv) mass accuracy, and v) the resolution. To tackle the current analytical challenges and to improve versatility, many hybrid mass analyzers have been recently developed which offer united advantage of different types of mass analyzers including, IT-LIT, ToF-ToF, Q-Q-ToF, Q-Q-Q and Q-Q-LIT. During my research I have mainly utilized LIT (linear ion trap) and ToF (Time-of-Flight) mass analyzer for MS analysis.

1.4.1 Ion-trap analyzer

The ion trap mass analyzer separates the gas-phase ions based on their $m/z$ as well as store the ions [21]. The ions are trapped in two or three dimensions by oscillating electrical field and RF quadrupole field. Hence, ion trap analyzer can be classified in two different types: 1) 3D ion trap (Quadrupole ion trap or Paul ion trap), 2) 2D ion trap (linear ion trap).

The quadrupole ion trap (QIT), invented by Paul and Steinwedel in early 1950’s [3], and is consists of three electrodes: a donut shaped ring electrode is surrounded by two end cap electrodes on both sides. Upon the application of RF voltage QIT behaves as an ion storage device in which ions are focused in the center of the trap. The motion of ions in the QIT is characterized by one radial and one axial frequency. To circumvent the space charging effect (only limited number of ions can be stored due to the space charge effect), most QIT instruments
have an automatic gain control (AGC). This feature controls the adequate fill time of the trap to enhance sensitivity and to subside resolution losses due to space charge effect.

![Figure 1.4 Schematic representation of 3D ion trap (QIT)](image)

To obtain a mass spectrum, mass-selective ejection of ions is required by continuous increase in RF voltage at the constant rate, in which ions with lower \( m/z \) are ejected first. Because resolution of the spectrum strongly depends on the speed at which the RF voltage is increased, higher resolution is attained with slower speed. Comparatively high sensitivity can be obtained using QIT due to the ability of ion accumulation in the trap. In principle, QIT can perform in time MS\(^2\) and produce fragment ions using neutral gas molecules such as helium, nitrogen or argon.

The 2D ion trap, also known as linear ion trap (LIT), is one of the recently developed mass analyzer, introduced by Schwartz in 2002 [22]. Physically, a linear ion trap is similar to a quadrupole analyzer but in principle it is alike QIT analyzer. LIT is formed by four hyperbolic or circular rods arranged symmetrically that can be divided into three compartments: a) front, b)
central, and e) back compartment. The ions are confined radially by a two-dimensional RF field in LIT.

![Schematic representation of 2D ion trap (LIT)](image)

**Figure 1.5** Schematic representation of 2D ion trap (LIT)

DC voltage is applied to the front and the end compartments to prevent the axial ion escape from the LIT. Application of DC voltage repels the ions from both ends and focus towards the central compartment of the LIT. However, both 3D ion trap and 2D ion trap can be applied to similar type of mass analysis, 2D ion trap offers following advantages over the 3D ion trap: 1) higher ion trapping capacity (due to the ions are focused along the center line rather than to a point), and 2) higher ion trapping efficiency. Because the ions are ejected radially through the slits of the rods from central compartment, placement of two detectors axially on both side of the rods are essential to maximize sensitivity of the LIT.

1.4.2 Time-of-Flight (ToF) mass analyzer

Time-of-flight (ToF) mass analyzer separates the ions based on the time required to cross the field free region, known as flight tube, after initial acceleration by an electrical field [23]. ToF analyzer is well suited for pulsed nature MALDI ionization source. As kinetic energy is
distributed equally among the ions in ToF, the separation of ions is based on the velocities of the ions. Hence, the time required for the ions with lower $m/z$ is shorter than ions with high $m/z$ to travel to the detector. Owing to the good transmission, a spectrum over wide mass range can be generated in short time using ToF analyzer. Moreover, in principle ToF analyzer does not have any upper mass limit; the mass range is mainly limited by decrease in detector response with increase in $m/z$. The mass resolution of simple ToF mass analyzers is relatively poor due to the energy distribution of the similar $m/z$ in the flight time. This issue can be addressed either by reducing kinetic energy distribution of ions before application of acceleration voltage or by increasing the length of the flight tube.

One of the approaches, to reduce the kinetic energy distribution of the ions, is the introduction of time delay between the ionization and acceleration voltage, also known as delayed pulsed extraction. The time delay is generally range from ns to µs before the application of acceleration voltage. An optimum delayed time can compensate for ion energy distribution
and improve mass resolution in ToF analyzer. Another approach to improve the mass resolution of ToF analyzer is by introducing an electrostatic mirror in the drift region to significantly improve the path length for ion travel. In this set-up, the ions with high energy penetrate deeper in the electrostatic mirror region compare to the ions with similar m/z but carry less energy. As a result, all the ions with similar m/z reach at the detector at the same time which reduces the energy distribution. This modification in ToF analyzer not only increases the mass resolution by increasing in the flight path but also maintain the compact size of the instrument. Mostly all the commercial ToF instruments have two detectors (one for linear mode and one for reflectron mode).

1.4.3 Hybrid mass analyzer

The hybrid mass analyzer is a combination of one or more type of mass analyzers. The hybrid analyzer offers the combined pros of different mass analyzers in one instrument at a time. The most common mass analyzers involved in development of the hybrid ones are: quadrupole (Q), ion trap (IT), linear ion trap (LIT), time-of-flight (ToF), and ion cyclotron resonance (ICR). The first hybrid mass analyzer, Q-Q-Q, was developed by Yost and Enke in 1978 led to development of series of hybrid mass analyzers [25]. The hybrid analyzer offers excellent

<table>
<thead>
<tr>
<th>Table 1.3 List of characteristics of various mass analyzers [24]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Mass limit</td>
</tr>
<tr>
<td>Resolution</td>
</tr>
<tr>
<td>FWHM (m/z 1000)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

1.4.3 Hybrid mass analyzer

The hybrid mass analyzer is a combination of one or more type of mass analyzers. The hybrid analyzer offers the combined pros of different mass analyzers in one instrument at a time. The most common mass analyzers involved in development of the hybrid ones are: quadrupole (Q), ion trap (IT), linear ion trap (LIT), time-of-flight (ToF), and ion cyclotron resonance (ICR). The first hybrid mass analyzer, Q-Q-Q, was developed by Yost and Enke in 1978 led to development of series of hybrid mass analyzers [25]. The hybrid analyzer offers excellent
platform for MS\textsuperscript{2} or even MS\textsuperscript{n} analysis, in addition to, excellent dynamic range, high resolution power, and high mass accuracy. The hybrid mass analyzers which have successfully established their importance through their performances are listed in Table 1.4.

**Table 1.4** Performances comparison and characteristics of various hybrid analyzers

<table>
<thead>
<tr>
<th>Dynamic range</th>
<th>IT-LIT</th>
<th>Q-Q-ToF</th>
<th>ToF-ToF</th>
<th>FT-ICR</th>
<th>Q-Q-Q</th>
<th>QQ-LIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution power</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Sensitivity (LOD)</td>
<td>Low</td>
<td>Good</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Mass accuracy</td>
<td>Low</td>
<td>Good</td>
<td>Good</td>
<td>Excellent</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>MALDI</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ESI</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Tandem mass</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Identification</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+++</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Quantification</td>
<td>+</td>
<td>+++</td>
<td>++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td>Modification</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Note: +, ++ and +++ mean moderate, good and high respectively [26]

### 1.5 Tandem mass spectrometry

Tandem mass spectrometry (MS\textsuperscript{2}) is two or multistage mass analysis that involve some type of dissociation process to generate the fragment ions from the mass-selected ions. The tandem mass spectrometry is an excellent tool for determination of peptide/protein sequences, identification of analytes, structure elucidation and quantification through fragment fingerprint. In principle, the tandem mass spectrometry can be conceived in two ways: \textbf{a) in space} by the coupling of two distinct mass spectrometers, \textbf{b) in time} by performing appropriate sequence of events in an ion storage device. The mass spectrometers capable of \textit{in space} tandem mass analysis utilize magnetic sector, quadrupole and ToF/ToF mass analyzers. On the other hand, tandem mass spectrometry can also be achieved through time separation using the mass analyzers such as ion trap, orbitrap and FTICR.
There are main three types of scan modes available using tandem mass spectrometry that utilize inert collision gas for fragmentation: 1) product ion scanning, 2) precursor ion scanning, and 3) neutral loss scanning. 1) Product ion scanning is the most common MS\textsuperscript{2} experiment for identification of the amino acid sequence of specific peptides/proteins. In product ion scanning, the first analyzer (MS\textsubscript{1}) is set to a value that selects one specific precursor ion at a time. The selected ion undergo collision-induce dissociation (CID) in the collision cell, and resulting fragments are analyzed by the second analyzer (MS\textsubscript{2}). 2) Precursor ion scanning sets the second analyzer (MS\textsubscript{2}) to transmit only one specific fragment ion to the detector while MS\textsubscript{1} is scanned to detect all the precursor ions that can generate this specific fragment ion. This mode is used to detect a subset of peptides in a sample that contain a specific functional group such as phosphate ester. 3) Neutral loss scanning scans both analyzers in a synchronized manner, so that the mass difference of the ions passing through MS\textsubscript{1} and MS\textsubscript{2} remains constant. The common application of this scan mode is the detection of peptides phosphorylated at serine or threonine residue via a loss of phosphoric acid.

The most commonly used MS\textsuperscript{2} activation method is collision-induced dissociation (CID), in which collision of selected precursor ion with inert gas, such as helium, results in fragmentation of the mass-selected ion. Beside CID, there are various activation approaches are developed and/or under investigation such as electron capture dissociation (ECD) [27, 28], electron transfer dissociation (ETD) [29], infrared multiphoton dissociation (IRMPD) [30], and surface-induced dissociation (SID) [31]. The development of alternative activation approaches helps to better understand the unimolecular dissociation behavior of a molecule under different excitation methods. The activation methods such as CID and IRMPD are based on vibrational activation, in contrast to ECD and ETD, which are based on radical-induced dissociation.
Peptide sequence identification using MS$^2$ involves fragmentation of a peptide to produce smaller $m/z$ fragment ions, which are assemble together to determine the original peptide sequence. Depending on the cleavage of different bonds along the peptide backbone, different types of fragment ion are produced as shown in Figure 1.8 below. These fragment ions are labeled as $a$, $b$ or $c$ ions if the charge is retained by N-terminal fragments and as $x$, $y$ or $z$ ions if
the charge is retained by C-terminal fragments. Among these fragment ions b and y ions are considered to be the most useful and abundant fragment ions as they correspond to cleavage of amide bond. Interestingly, both fragment ions of a complementary pair are not always detected in equal abundance owing to their difference in stability and instrument discrimination.

1.6 Application of mass spectrometry

The basic principle of mass spectrometry is to ionize the sample and measure m/z of the resulting ions. The qualitative and quantitative analytical capabilities of mass spectrometry efficiently provide information of chemical composition of organic and inorganic compounds. Over the last few decades, mass spectrometry has evolved as a vital tool for analysis of various biomolecules such as peptides/polypeptides, proteins, oligosaccharides, nucleotides etc. The application of mass spectrometry can be classified mainly into two categories: a) qualitative analysis, and b) quantitative analysis.

1.6.1 Qualitative MS analysis

Over the last few decades, the mass spectrometry has evolved as an essential tool for proteomics. The protein analysis can be performed by two approaches using mass spectrometry, “top-down” and “bottom-up”. In top-down approach, the intact proteins are introduced in MS without proteolytic digestion and are fragmented into smaller peptides in mass spectrometer. Due to complex isotopic distribution patterns of high mass proteins in top-down approach, it requires high resolution mass analyzer. Also, employment of ESI in top-down approach can result in even more complex mass spectra due to the presence of multiply charged ions.

In contrast, the bottom-up approach involves the use of proteolytic digestion of intact proteins to produce fragment ions before MS analysis. In bottom-up approach, the identification of protein is performed from measurement of peptides mass generated from protein digestion,
also called peptide mass fingerprint (PMF) followed by comparison with a sequence database. Moreover, MS\textsuperscript{2} analysis can determine the amino acid sequence of each peptide to aid protein identification.

The activation and function of a protein is largely dependent on the post-translational modification. Among many covalent modifications, phosphorylation, glycosylation, and ubiquitination are the most common ones. In order to characterize the protein, it is essential to study these changes in the protein structure. Due to the recent advances in mass spectrometer instruments, MS analysis has replaced many conventional approaches to study protein modifications, for example, MS analysis overshadows the use of radioactive labels and Edman sequencing to study protein phosphorylation over the last two decades.

Beside identification and characterization of proteins, mass spectrometric analysis can also apply to determine chemical structure and chemical formula of small organic compounds. Moreover, the role of mass spectrometry has recently evolved in study of drug metabolism due to the availability of high resolution mass analyzers such as FT-ICR and orbitrap. The MS analysis assists to identify the primary metabolites and to determine pharmacokinetic parameters based on \textit{in vitro} and \textit{in vivo} drug metabolism study [32].

\subsection*{1.6.2 Quantitative analysis}

As a result of many technological advances in mass spectrometer in last few decades, the application of mass spectrometry has not been restricted to successful identification of proteins or protein complexes but has been extended to quantitative proteomics. Quantitative proteomics is critical for understanding protein kinetics and molecular mechanism of biological processes. Although, MS lacks inherent quantitative analysis capabilities due to different ionization efficiency of many peptides and variation in mass spectrometric response of proteolytic peptides
between runs, many approaches have been developed to overcome this limitation such as label-free quantification, metabolic labeling, isotope tags, and isobaric tags. MS-based quantitative proteomics can be divided into two categories: a) relative quantification, and b) absolute quantification. In relative quantification approach, the samples are differentially labeled with stable isotopes. These samples are mixed together followed by quantitative MS analysis. Generally, peak intensity ratios of differentially labeled peptides (light and heavy) are measured to quantitate the relative abundance. Many isotope-based quantification methods have been developed such as metabolic labeling (SILAC) and enzymatic or chemical labeling.

In absolute quantification, the experimental samples are spiked with the known concentration of isotope-labeled target peptides. Hence, the abundance of the target peptides is determined by comparing the abundance of the isotope-labeled target peptides. Although, absolute quantification method seems more obvious, but in routine, relative quantification approach is more often used due to requirement of costly reagent and time-consuming assay development for the absolute quantification of each protein of the interest.

1.6.2.1 Label-free MS quantification

In label-free MS quantification method, samples are separately prepared and analyzed by MS or MS$^2$ because of which this approach is more prone to experimental variations than all other quantification methods. Label-free MS quantification method is cost effective and rapid compared to other quantification methods. Hence, it is an ideal approach when large numbers of samples need to be analyzed such as in clinical screening. However, this method is good for measuring large changes in proteins and less efficient in measuring small changes. Using label-free approaches, the abundance of proteins or peptides can be determined by two ways: a) based on ion peak intensity, and b) based on spectral counting. In the former method, relative
quantification of similar peptides between samples is determined directly from the mass spectrometric ion peak intensities, while in later method; the sum of MS² spectra of LC-MS run of a peptide across different samples is utilized to determine the abundance of a protein.

Figure 1.9 Workflow of various proteomics quantitative approaches [35]

1.6.2.2 Isotopic labeling

The isotopic atom or isotope-coded tags can be introduced to the sample by various means like metabolic labeling, enzymatic labeling and chemical labeling. Metabolic labeling for relative quantification was first introduced by Oda et al., who labeled all amino acids with heavy nitrogen (¹⁵N) [33]. This approach was further developed by Mann et al. and introduced as stable
isotope labeling by amino acids in cell culture (SILAC). Nowadays, SILAC has evolved as the most common approach for in vivo labeling [34]. In SILAC, in contrast to initial approach, cells are now grown in growth medium containing $^{13}$C$_6$-lysine and/or $^{13}$C$_6$-arginine. These residues were chosen so that after trypsin digestion all the peptides contain at least one labeled amino acid and result in constant increase in mass of labeled samples compare to non-labeled sample. Among many advantages of this approach, the key advantage is, as non-labeled (light) and labeled (heavy) samples are combined before sample preparation for MS analysis, the possible experimental bias is minimum compare to all other quantification approach. Additionally, this approach can successfully detect the small changes in a protein between different experimental samples. Other approaches like enzymatic labeling and chemical labeling find their application especially for various biological samples where metabolic labeling approach is not amenable. The isotope labeling can be achieved by different means such as $^{18}$O [36], $^2$H [37], stable isotope dimethylation [38], Isotope-coded affinity tag (ICAT), and isotope-coded protein labeling (ICPL). The ICAT method employs a reagent, which is comprised of three different regions, a sulfhydryl-reactive group, a linker region (heavy or light), and a biotin molecule. Due to the presence of sulfhydryl-reactive group, the ICAT reagents react only with free thiol on cysteine residues.

![Figure 1.10 Structure of ICAT reagent](image)
Owing to the characteristic of only labeling cysteine residues in experimental sample, this approach offers unique capability to minimize sample complexity and assist in identification of low-abundance proteins from the samples [40]. These labeled cysteine containing peptides from an experimental sample are then separated from non-labeled peptides by passing this sample through immobilized avidin, which binds to the biotin tag. Another approach based on the isotope labeling is isotope-coded protein labeling (ICPL), in which lysine residues and available N-termini are isotopically labeled.

1.6.2.3 Isobaric labeling

The isobaric tags are also known as tandem mass tags as isobaric labeling method requires tandem mass spectrometry for quantification. The isobaric tags generally comprise of three components, a) a mass reporter- has variable number of stable isotope substituted atom, b) a mass normalizer- has unique mass to balance the mass of all the isobaric tags to equal mass, and c) a peptide reactive group- react with primary amine or cysteine residues of the experimental samples.

![Figure 1.11 Structure of Isobaric mass tag [41]](image)

These isobaric tags are developed in a way that a mass reporter falls off from each labeled peptide upon collision-induced dissociation (CID) during MS² analysis and yield different size tags. The relative quantification of peptides is performed by quantifying these tags.
The examples of commercially available isobaric tags are iTRAQ and TMT, which can simultaneously analyze 4 to 8 experimental samples. Owing to the multiplex capabilities, high throughput quantitative analysis is possible using the approach.
Chapter-2

Development and application of simple and fast quantification method for Desmosine and Isodesmosine using MALDI-Ion Trap tandem mass spectrometry

2.1 Introduction
2.2 Pathological disorders
2.3 Analytical techniques
2.4 Development of rapid and sensitive quantification method for Desmosine using MALDI-L/TQ
  2.4.1 Materials and Methods
  2.4.1.1 Chemicals
  2.4.1.2 Mass spectrometry
  2.4.1.3 UV-irradiation of Des
  2.4.1.4 Proton NMR
  2.4.1.5 Sample preparation for detection limit study in urine and serum
2.4.2 Results and Discussion
  2.4.2.1 MALDI-MS\textsuperscript{2} analysis of Des, Isodes and Labeled-Des
  2.4.2.2 Linearity and dynamic range
  2.4.2.3 Detection limit
  2.4.2.4 Reproducibility
  2.4.2.5 UV-irradiation of Des
  2.4.2.6 Comparison to existing methods
2.4.3 Conclusions
2.5 Application of MALDI-Ion Trap tandem mass spectrometry based quantification method to measure effect of UV-A radiation on the elastin cross links
  2.5.1 Material and Methods
  2.5.1.1 Sample preparation
  2.5.1.2 Histology and microscopy
  2.5.1.3 Transmission Electron Microscopy (TEM)
  2.5.1.4 \textsuperscript{13}C NMR experimental parameters
  2.5.1.5 Sample hydrolysis and quantification with Labeled-Des
2.5.2 Results and Discussion
  2.5.2.1 Biochemical changes in elastin due to UV-A exposure
2.5.3 Conclusion
2.1 Introduction

Elastic fibers represent the major structural component of the vertebrate extracellular matrix (ECM) assemblies. The prime importance of elastin, a principal component of elastic fibers, is to impart elasticity and resilience to a range of tissues. The elastin contributes to the structural foundation of lung [43], skin [44], aorta [45], bladder [46] and female reproductive tract [47] and play critical role to maintain shape and normal physiological functions of these tissues. Elastin is a highly insoluble polymer protein consists of the cross-linked, soluble monomeric precursor molecule called tropoelastin [48, 49]. Tropoelastin is 60kDa unglycosylated protein encoded by a single human gene and secreted by various cells including fibroblasts [50], smooth muscles cells [51], chondrocytes [52], and endothelial cells [53].

The tropoelastin, a building block of elastin, is mainly cross-linked by two polyfunctional amino acids [54]. The scientists named J. Thomas, E. Eldsen, and M. Patridge have pioneered the study of the cross-linking in elastin. They published the first report in 1963 on the isolation of the cross-linking amino acids from bovine ligamentum nuchae and named them as desmosine and isodesmosine [54, 55]. Subsequently, as a result of their study to understand the biosynthesis

![Figure 2.1 Structure of desmosine and isodesmosine.](image)
of these two amino acids, Anwar and Oda [56] found that these amino acids, desmosine (Des) and its structural isomer isodesmosine (Isodes) are originated from the condensation of four lysine amino acid residues [57-59].

The cross-linked tropoelastin is known as mature elastin, which is mainly produced in early stage of the life. The mature elastin is the longest lasting protein in the body with half-life of about 74 years [60]. However, in the event of damage to the elastin in the skin due to serious injury, sun burn or as a result of aging, the low level of elastin production results in loss of the elasticity or inefficient repair. Moreover, the elastin degradation products activate the elastin receptors, which subsequently stimulate elastase and lead to further degradation of mature elastin. Hence, many serious pathological disorders are associated with degradation of elastin. Since desmosine and isodesmosine are unique to the mature elastin, the presence of these amino acids in various body fluids including urine, plasma or sputum is considered as a potential biomarker for the diseases associated with elastin degradation. Consequently, many reports have been published in last few decades using various analytical techniques to quantify these amino acids to measure elastin degradation.

2.2 Pathological disorders

As elastin is responsible for functionality of many connective tissues, various pathological conditions arise due to the abnormality in elastin (either accumulation or degradation). The evidence suggests that in various destructive lung disorders such as chronic obstructive pulmonary disease (COPD), disseminated bronchiectasis and cystic fibrosis (CF), the degradation of elastin occur due to imbalance between elastin-degrading proteases and protease inhibitors. [61-63]. The pathological conditions associated with elastin can be categories by two
distinct approaches. A) Based on their occurrence on deposition or degradation of elastin. The disorders such as atherosclerosis, abdominal aortic aneurysms, cutis laxa, anetoderma, pulmonary emphysema, chronic pulmonary obstructive disease (COPD), marfan syndrome and menkes’ syndrome are associated with degradation of elastin in the body. In contrast, scleroderma, endocardial elastofibrosis, pseudoxanthoma elasticum (PXE), elastomas, and buschke-ollendorff syndrome are the result of accumulation of elastin. B) Based on the inherited and acquired diseases. For e.g. marfan syndrome and buschke-ollendorff syndrome are inherited elastin associated disease. On the other hand, elastomas and actinic elastosis are acquired elastin associated conditions.

2.3 Quantification methods

Studies have shown a correlation between the amounts of Des/Isodes detected in human body fluids with the degree of elastin degradation [64-66] and therefore, Des/Isodes have been considered as potential biomarkers to determine the extent of tissue damage and elastin breakdown. This is due to Des and Isodes being uniquely associated with mature elastin, not being metabolized after being released from tissue [67] and being independent from the diet [68]. A wide range of analytical techniques have been applied to detect Des/Isodes, which include enzyme-linked immunosorbent assays, [69, 70], amino acid analysis [71], radioimmunoassay, [72, 73], electrokinetic chromatography [74], electrophoresis [75, 76], Nuclear Magnetic Resonance, NMR [77, 78], high performance liquid chromatography, HPLC [79-82], liquid chromatography-mass spectrometry, LC-MS and LC- tandem mass spectrometry, LC-MS/MS [83-85]. Among these methodologies, reported results show that LC-MS/MS gives the greatest selectivity and sensitivity for Des samples where Des detection limits on the order of $10^{-9}$ M
have been obtained [71, 86]. One of the most recent techniques incorporating LC-MS to
determine Des/Isodes amounts involve the use of deuterated Des as an internal standard [86, 87].
Ma et al. results, incorporating the quadruply deuterated (d4)-labeled-desmosine as an internal
standard, demonstrated the concentrations of total Des + Isodes and free Des + Isodes in plasma
were 0.51 and 0.09 ng/mL, respectively [86]. Another study conducted by Albarbarawi et al. to
investigate total Des/Isodes content was performed on COPD urine samples utilized d5-Des
internal standard. The data showed COPD patients on average (20.4 ng Des/mg creatinine) had
three times the total amount of Des/Isodes than healthy patients (6.8 ng Des/mg creatinine) [87].

Beside the advantages, there are some challenges in analysis of Des/Isodes with LC-MS
such as relatively long chromatography runs as well as the labor-intensive process to prepare
samples. One potential analytical technique to improve the Des/Isodes analysis times while
maintaining sensitivity and selectivity is to use Matrix Assisted Laser Desorption/Ionization-
Mass Spectrometry (MALDI-MS). MALDI ionization has the benefit of rapid analysis times
without the requirement for chromatographic separation.

2.4 Development of rapid and sensitive quantification method for desmosine using MALDI-
LTQ

In the present report, we have developed a MALDI-MS\(^2\) based quantification method to
study Des samples mixed in serum or urine. The approach involves synthetic stable-isotope
labeled \(d_4\)-Desmosine (Labeled-Des) as an internal standard to quantify Des and Isodes by
selectively monitoring an ion that arises from the removal of one of the side chains \([M-\]
\(\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}(\text{NH}_2)\text{COOH}]^+\), (130 Da). In the linear ion trap mass spectrometer, this
transition yields a specific and sensitive method for the detection and quantification of Des and
Isodes. In this study, the efficacy of the methodology is determined by evaluating linearity, detection limit, and run-to-run reproducibility of Des, Isodes, and 50:50 Des/Isodes mixtures in water, as well as in backgrounds of serum and urine. Furthermore, we have quantified time-dependent degradation of Des upon UV radiation using MALDI-MS² and discuss the results in comparison to ¹H NMR.

2.4.1 Materials and Methods

2.4.1.1 Chemicals

HPLC grade acetonitrile and water were purchased from Fisher Scientific (Pittsburg, PA, USA). Trifluoroacetic acid was purchased from Thermo Scientific (Rockford, IL, USA). α-cyano-4-hydroxycinnamic acid (CHCA) and deuterium oxide (D₂O) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Des and Isodes standards were obtained from Elastin Products (Owensville, MO, USA). Stable-isotope labeled desmosine standard was obtained from Toronto Research Chemicals (North York, ON, CANADA). Calf serum and human urine was obtained from healthy donors and sterile-filtered before use.

2.4.1.2 Mass spectrometry

Samples were analyzed using Thermo LTQ XL linear ion trap mass spectrometer equipped with a vacuum MALDI source utilizing a nitrogen laser (337 nm) firing at 60 Hz (Thermo Scientific, Waltham, MA, USA). A saturated solution of α-cyano-4-hydroxycinnamic acid (CHCA) was prepared using the dried droplet method as described as previously described [88], and was used as a matrix solution. MALDI-MS² experiments were performed in positive ion mode using 3.0 µJ laser energy. Mass spectra of samples were typically obtained using 400-
600 scans (3-5 minutes) and processed with Xcalibur (2.0.7 SP1) software. Current study involves the use of isotope-labeled internal standard (Labeled-Des) for quantification of Des and Isodes. To study relative abundance of Des or Isodes in samples, the ratio of the generated fragment ion, 397 \textit{m/z} (Des or Isodes) to 400+401 \textit{m/z} (Labeled-Des) were utilized. The precursor ion (\textit{m/z} 528), with isolation width of 6 Da, was chosen to perform MS\textsuperscript{2} experiment of Des or Isodes standard (\textit{m/z} 526.2) mixed with Labeled-Des standard (\textit{m/z} 530.2). Fragmentation of precursor ion was conducted with normalized collision energy 36, activation Q 0.25 and activation time 30 ms.

2.4.1.3 UV-irradiation of Des

The sample, prepared by dissolving Des in D\textsubscript{2}O solvent, was subjected to UV radiation using 12 low pressure mercury lamps of 254 nm in a Rayonet Photoreactor. UV-irradiated Des sample was analyzed using mass spectrometry (MALDI-MS\textsuperscript{2}) and \textsuperscript{1}H NMR at different time intervals (0, 5, 15, 30, and 60 min). The enclosed system was used during UV-irradiation to ensure that Des sample was solely exposed to the specific wavelength of UV light.

2.4.1.4 Proton NMR

Liquid state \textsuperscript{1}H NMR spectra were recorded using Bruker DPX 400MHz FT NMR with automatic sampler. NMR spectra were acquired by accumulating 160 scans and chemical shifts (\textit{\delta}) were reported in parts per million (ppm). The peak at 8.44 ppm was observed in non-irradiated as well as UV-irradiated Des sample, was integrated in reference to the aliphatic region (1.00-5.25 ppm). The relative concentrations of Des were calculated from integration
values of a peak at 8.44 ppm for different time intervals and used to determine the rate constant of UV-induced degradation of Des.

2.4.1.5 Sample preparation for detection limit study in urine and serum

To assess the detection limit of Des, an initial concentration of Des was prepared using urine and serum, using sterilized human urine from healthy donors and commercially prepared newborn calf serum without further dilution or purification. Successive dilutions of Des samples were prepared using HPLC-grade water. 1 µL of analyte solution was mixed directly with 9 µL of CHCA matrix solution without prior enrichment. For mass spectrometric analysis 1 µL of this mixture was deposited onto a stainless-steel plate using dried droplet method.

2.4.2 Results and Discussion

The amounts of Des/Isodes have been shown to be elevated in many pathological disorders associated with elastin degradation [86, 87]. Consequently, there is a need for simple, reliable and sensitive methods to quantify low levels of Des and Isodes. The present study introduces a MALDI-ion trap MS\(^2\) based quantification method for detection of Des and Isodes. Labeled-Des is employed as an internal standard for quantification of Des and Isodes. Due to low abundance and poor signal-to-noise of the precursor ion of Des (\(m/z\) 526.2) and Labeled-Des (\(m/z\) 530.2), single-stage MS analysis is inefficient to measure quantitative changes in the samples (data not shown). Consequently, quantification is performed using the stable fragment ion (\(m/z\) 397) generated by removal of one of the side chains from Des molecule [M-CH\(_2\)CH\(_2\)CH\(_2\)CH\(_2\)CH(NH\(_2\))COOH]\(^+\), 130Da, in tandem-MS mode. Linearity, dynamic range, detection limit and reproducibility of the MALDI-MS\(^2\) method in water and body fluids are
presented in this report. To demonstrate the application of MALDI-MS$^2$ method, quantification of Des degradation upon UV radiation (254nm) at different time points is studied using both MALDI-MS$^2$ method and NMR spectroscopy.

2.4.2.1 MALDI-MS$^2$ analysis of Des, Isodes and Labeled-Des

Mass spectrometry is an analytical technique that separates ions based on their mass-to-charge ($m/z$) ratio. MALDI-mass spectrometry utilizes a small organic molecule (CHCA) as a matrix to assist in the energy transfer and ionization process of the analyte. Due to high background resulting from ions produced from matrix molecule itself, the precursor ions generated from Des samples were difficult to detect in single-stage MS. Therefore, MS$^2$ spectra
Figure 2.2 MALDI-MS² analysis of Des, Isodes and Labeled-Des- (A) MS² spectra of Des, Labeled-Des and Isodes are shown in a top, middle and bottom panel respectively. The highlighted fragment ions resulted from loss of a side chain of Des and Isodes (m/z = 397), and Labeled-Des (m/z = 401) were used for the quantification of Des/Isodes. (B) Isotope distribution of the fragment ion (m/z = 401) of Labeled-Des and the fragment ion (m/z = 397) of Des/Isodes are shown in blue and orange respectively. The result shows two major isotope peaks, 400 m/z and 401 m/z, represent greater than 75 percentage of total peak area of the fragment ion (m/z = 401).

of the two structural isomers were then evaluated. The results showed both structural isomers indeed produce similar fragment ions. However, substantial differences in intensity of the fragment ions were observed. MS² spectra of both structural isomers and Labeled-Des that used as an internal standard to quantify Des, Isodes and equimolar mixture of Des + Isodes are shown in Figure 2.2A.

Since labeled-Des was employed as an internal standard, it was important to evaluate its isotopic purity. To do this, the isotope distribution pattern of one of the most abundant fragment ions generated from Labeled-Des was evaluated. The selected fragment ion, 401 m/z, was generated from loss of one of the side chains, [M- CH₂CH₂CH₂CH₂CH(NH₂)COOH]⁺, of Labeled-Des molecule. (Figure 2.2B) The result of isotopic distribution pattern showed two major isotope peaks, 400 m/z and 401 m/z, representing greater than 75 percentage of total peak
area of the fragment ion. Hence, to compensate for isotopic impurity, instead of incorporating peak area of any single isotope peak, we decided to incorporate the peak area of two most abundant isotope ions \((m/z = 401\) and \(400\)) of Labeled-Des in this investigation. To quantify Des/Isodes, peak area ratio fragment ions generated from Des/Isodes \((397\ m/z)\) to Labeled-Des \((400+401 \ m/z)\) is determined. Linearity, detection limits and reproducibility have been evaluated to assess the applicability, sensitivity and efficacy of the presented MALDI-MS\(^2\) quantification method. As detection of Des and Isodes in different body fluids can be used as biomarkers for elastin degradation in body, we have evaluated these parameters in urine and serum.

2.4.2.2 Linearity and dynamic range

To evaluate the linearity of the MS response, 11 different concentrations of Des range from 125 ng/µL to 3.125 ng/µL were mixed in a 1:1 (v/v) ratio with 25 ng/µL of Labeled-Des. Two sets of each of these mixtures were analyzed in quadruplicate. The linear curve was obtained by plotting the average peak area ratio of Des/Labeled-des vs. concentration of Des. Linear response of Des using Labeled-Des \((r^2 = 0.998)\) over two orders of magnitude was obtained as shown in Figure 2.3A. Inter-assay precision \((\%RSD)\) obtained using MALDI-MS\(^2\) method was from 0.17% to 4.7% for the entire dilution range.

A response curve for Isodes was generated for similar concentration range as of Des. Linear response \((r^2 = 0.998)\) was also obtained from response curve of Isodes (Figure 2.3B). Each of these dilutions was analyzed in quadruplicate. Moreover, because in vivo, Des and Isodes occur in equimolar amounts [89], we mimicked this condition by investigating equimolar concentration ratios of Des and Isodes. These were studied over a concentration range of 125 ng/µL to 3.125 ng/µL mixed with 25 ng/µL of Labeled-Des. A linear response with \(r^2\) value of
0.998 was obtained for equimolar ratio of Des and Isodes as shown in Figure 2.3C. We note that quantitatively reproducible differences in MS\(^2\) fragment intensities could in principle be used to independently quantify Des and Isodes; however prior studies indicate that using the combined concentration of Des and Isodes give a better assessment of elastin breakdown [87].

![Graphs of A, B, C, D, and E](image)

**Figure 2.3** Linearity of Des, Isodes and equimolar concentration ratio of Des: Isodes were evaluated using MALDI-MS\(^2\) quantification method. Linearity of MS response over the dynamic range from 125 ng/µL to 3.125 ng/µL are shown for Des (A), Isodes (B), and equimolar concentration ratio of Des: Isodes (C). Linearity of Des was also assessed in biological fluids- serum (D), and urine (E) over the dynamic range from 43.75 ng/µL to 6.25 ng/µL.

Furthermore, the linearity of Des response in biological fluids was assessed. Both urine and serum were spiked (as described in sample preparation) using wide concentration range from
43.75 ng/µL to 6.25 ng/µL of Des mixed with 25 ng/µL of Labeled-Des standard. The linear curves with $r^2$ value of 0.997 (Figure 2.3D) and 0.998 (Figure 2.3E) were obtained in urine and serum respectively. Results derived from these linearity assays using MALDI-MS$^2$ quantification methods are consistent for Des, Isodes and their equimolar mixtures in water as well as in spiked biological fluids.

2.4.2.3 Detection Limit

To determine detection limit of MALDI-MS$^2$ quantification method, six successive five-fold dilutions of Des were spiked into serum and urine. The peak area ratio of Des/Labeled-Des in both media remains consistent up to the fifth successive dilution of 0.02 ng/µL (Table 2.1), with % RSD less than 5 up to first five dilutions. On the other hand, % RSD observed for the next successive dilution increased to more than 8% in serum. Moreover, due to loss in signal

<table>
<thead>
<tr>
<th>Des Concentration (ng/µL)</th>
<th>Serum</th>
<th></th>
<th>Urine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Peak Area</td>
<td>% RSD</td>
<td>Peak Area</td>
</tr>
<tr>
<td>12.5</td>
<td>0.66</td>
<td>2.53</td>
<td>0.69</td>
</tr>
<tr>
<td>2.5</td>
<td>0.65</td>
<td>1.61</td>
<td>0.68</td>
</tr>
<tr>
<td>0.5</td>
<td>0.66</td>
<td>2.99</td>
<td>0.66</td>
</tr>
<tr>
<td>0.1</td>
<td>0.69</td>
<td>3.76</td>
<td>0.65</td>
</tr>
<tr>
<td>0.02</td>
<td>0.68</td>
<td>1.25</td>
<td>0.64</td>
</tr>
<tr>
<td>0.004</td>
<td>0.74</td>
<td>8.20</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Table 2.1 Detection limit of MALDI-MS$^2$ method- Six successive dilutions of Des in serum and urine were evaluated to assess the detection limit of MALDI-MS$^2$ quantification method. The consistent result in peak area of Des was observed up to fifth successive dilutions with %RSD of less than 5 in both biological fluids, serum and urine.
intensity at lower concentration of 0.004 ng/µL, peak area ratio of Des/Labeled-Des change drastically in urine. This result indicates that MALDI-MS$^2$ method can efficiently determine concentration as low as 0.02 ng/µL of Des in complex biological matrix such as serum and urine. Results from the present study suggest that MALDI-MS$^2$ is a sensitive quantification method.

2.4.2.4 Reproducibility

To evaluate the reproducibility of the present method, we conducted triplicate studies performed by different lab personnel. Response curves for Des in water were generated following a similar approach as described in the linearity and dynamic range sections, over the time span of two weeks. The run-to-run %RSD values for entire concentration range were less than 5% (Figure 2.4) demonstrating the reliability of the current MALDI-MS$^2$ method. During

![Figure 2.4](image)

**Figure 2.4** Assessment of reproducibility of MALDI-MS$^2$ method- Reproducibility of the present method was determined by evaluating the response curve of Des in HPLC water in triplicate by different lab personnel. The run-to-run %RSD values for entire concentration range were < 5.
the two-week test period, Des, Isodes and Labeled-Des underwent more than three freeze/thaw cycles. Thus, these results also demonstrate the stability of Des, Isodes and Labeled-Des standards.

2.4.2.5 UV-irradiation of Des

Detrimental effects of UV radiation on skin such as loss of elasticity, premature aging and wrinkled appearance have been discussed in previous reports [79, 90, 91] which might occur as a result of reduction in Des cross-linking of elastin. Moreover, the results of photolysis study performed by Baurain et al. showed maximum breakdown of Des and Isodes was observed at 274 nm and 285 nm respectively [91]. We therefore utilized the current MALDI-MS² method to evaluate UV-induced breakdown of Des.
Figure 2.5 UV-induced degradation of Des- Degradation of Des was studied by irradiating the sample to UV light at 254nm for different time points (0, 5, 15, 30 and 60 minutes). (A) Represents the MS$^2$ spectra of UV-irradiated Des mixed with 25 ng/µL of Labeled-Des. The fragment ion generated from Des (decrease with time) and Labeled-Des are highlighted in orange and blue box respectively. (B) Represents the UV-induced breakdown by measuring concentration of Des at different time points using MALDI-MS$^2$ and NMR spectroscopy.

To study the effect of UV radiation on Des degradation, samples were aliquoted and analyzed at different time with MALDI-MS$^2$ (Figure 2.5A). Additionally, to validate the change in concentration observed using our method, we have also quantified Des using NMR spectroscopy and compared the results of the two methodologies. The results of MS$^2$ analysis distinctly demonstrate decrease in relative abundance of Des fragment ion (397 m/z) with time. The peak area of 25 ng/µL Labeled-Des fragment ions (m/z= 400 + 401) was used as a reference to determine the concentration of Des standard at zero-time point. In the NMR study, a peak in the aromatic region (8.42ppm) observed in the non-irradiated control spectrum, as well as the spectra of all timepoints, was used to track the UV-induced breakdown Des. The results obtained
from MALDI-MS$^2$ and $^1$H NMR analysis at different time points are shown in Figure 2.5B. The results from both analysis indicate that Des concentration reduces to roughly half after 5 min of UV-radiation. Furthermore, an identical pattern of decrease in concentration of Des with different time intervals was observed from $^1$H NMR and MALDI-MS$^2$ method. The first-order rate constants estimated using MALDI-MS$^2$ method and NMR spectroscopy are $0.110 \pm 0.021$ min$^{-1}$ and $0.124 \pm 0.008$ min$^{-1}$ respectively.

2.4.2.6 Comparison to existing methods

The current state of the art in detection of desmosines utilizes LC-MS/MS methods either with or without stable isotope-labeled standards [86, 92, 93]. The approach presented here shows similar performance characteristics, such as precision, accuracy and sensitivity, while leveraging the strengths unique to MALDI. For example, current LC-MS run times range from 12 minutes [93] and longer and require column equilibration between runs. We typically take a conservative approach and interrogate our MALDI spots for 3-5 minutes, but even with 1 minute scans, we are able to maintain accuracy and reproducibility (data not shown). Additionally, MALDI requires no equilibration between samples; the time to move between MALDI spots is about 5 seconds. In addition, the MALDI approach reduces the expense and environmental impact of solvent usage by multiple orders of magnitude. Each MALDI spot requires 10 microliters or less of solvent to prepare.

The MALDI approach also has the benefit that spectra do not need to be extracted from an LC run, and detection and quantification of peaks can be fully automated via software. Additionally, our MALDI approach consumes only a small fraction of each spot, and therefore unlike destructive LC-MS approaches, our samples can be re-analyzed for Des or for other
purposes. The tolerance of MALDI towards moderate amounts of contaminants allows fewer cleanup steps. For example, no purification is required before the analysis of urine or serum; we have used similar MALDI approaches to quantify desmosines directly from hydrolysates of elastin without purification [78, 79]. Therefore we believe that our MALDI approach represents a robust, novel methodology that can complement existing LC-MS techniques.

2.4.3 Conclusions

Although the application of LC-MS/MS methods for determination of Des/Isodes in various body fluids has met a certain amount of success, there is still a need for a simpler, more rapid, reliable and sensitive quantification method. MALDI-MS$^2$ method offers these traits, including excellent sensitivity and shorter analysis time due to elimination of chromatographic separation, while maintaining similar analytical characteristics. The present study is the first attempt, to our knowledge, to demonstrate and characterize the use of MALDI-ion trap MS$^2$ to quantify Des/Isodes. We have evaluated concentration ranges from >100 ng/µL down to 0.02 ng/µL without prior concentration steps like solid-phase extraction, and found the MALDI-MS$^2$ to perform similarly to existing LC-MS/MS methods with the increase in speed and reduction in per-run expense associated with chromatography-free MALDI ionization. Our results from the evaluation of the linearity and dynamic range, detection limit, and reproducibility successfully demonstrate its efficacy of MALDI-MS$^2$ method for quantification of biomolecules. The case study performed on the quantification of Des breakdown by irradiation of UV light using MALDI-MS$^2$ explicitly demonstrates consistency between results derived from NMR spectra and the current quantification method. Thereby we believe MALDI-MS$^2$ has potential to be applied in clinical lab setting to diagnose conditions which induce elastin degradation.
2.5 Application of MALDI-Ion Trap tandem mass spectrometry based quantification method to measure effect of UV-A radiation on the elastin cross links

While the relationship between UV radiation and photoaging of skin has been established, the exact change that elastin undergoes when exposed to UV radiation is still unclear. One common belief is that elastic fibers are denatured or cleaved by UV radiation, and that most of the elastotic material is produced after exposure to UV radiation or its disorganized structure gives aged skin its characteristic wrinkled appearance. In previously reported studies, electron microscopy, histology, TEM (transmission electron microscopy), and SEM (scanning electron microscopy) have been applied to probe macroscopic changes to the elastic fiber in aged and photodamaged skin [94-96]. For all biological damage associated to sun exposure, UV-B contributes 80% whereas UV-A contributes only 20% even though approximately 95% of terrestrial UV-radiation is UV-A [97]; more work is needed to better understand the microscopic changes that UV-A irradiation causes to the structure of elastin. We conducted this in-vitro study to uncover changes on elastic fibers that undergo when exposed to high intensity UV-A irradiation. Mass spectrometry was also implemented to quantify changes in the relative amount of desmosine cross-links. Elastic fibers were then examined histologically and by TEM to observe macroscopic changes that resulted from irradiation. $^{13}$C solid state NMR spectroscopy was performed to measure possible structural alterations of elastin. These combined methods provide additional information relating to the detrimental effects of UV-A irradiation to the structure of the elastic fiber and of elastin.

2.5.1 Material and Methods
2.5.1.1 Sample preparation

Bovine nuchal ligament elastic fibers purchased from Elastin Products Company, LLC (Owensville, MO) were used for this study. These samples were purified by Elastin products Company using a known protocol [98] and were free of fat, collagen, smooth muscle cells, and other connective tissue. In a prior study, we showed that the protocol used for isolating elastin did not alter the structure of the protein, or the concentration of cross-links [99]. Elastic fibers were completely immersed in distilled water while being irradiated with a 3U40W UV-A lamp (Cnlight Co, China) with a center wavelength of 365 nm which was placed 10 cm above the sample. During irradiation, the system was covered with a shield to ensure that the sample was isolated from other light and were submerged in water during irradiation. The irradiation intensity was 12 mW/cm² and samples were continuously irradiated for 9 days. The intensity of the UV-A lamp is therefore approximately 3 times higher than that of the sun, when directly overhead (located at the zenith).

2.5.1.2 Histology and microscopy

Unirradiated elastic fibers and fibers following 9 days of UV-A irradiation were used for the histological study. For histology, a small amount of sample was placed overnight in 100 ml phosphate buffered saline. The samples were then placed in Lieca Cryo-Gel (SPI supplies Product Ref-02694-AB), sectioned on a Leica CM1850 cryostat at 10 μm, and stained using the Sigma-Aldrich elastic stain kit (REF HT25A-1KT) following a modified version of a previously reported protocol [100]. The slides were gently rinsed with 95% ethanol, and then placed in xylene for a few seconds. Cover slips were then mounted using an Eukitt quick-hardening
mounting medium (Sigma-Aldrich REF 03989) and left to dry overnight. All sections were photographed using a National Optical DC4-156-S digital microscope at a magnification of 10×.

2.5.1.3 Transmission Electron Microscopy (TEM)

Bovine nuchal ligament elastic fibers were immersed in 0.1 M phosphate buffered solution (pH-7.4) for 1 h. All the samples were stained with osmium tetroxide and embedded in epoxy resin. Samples were sliced along the plane that was perpendicular to the fiber axis, with a thickness of 60 nm and examined in a JEM-2000EX transmission electron microscope. The accelerating voltage used was 120 kV.

2.5.1.4 $^{13}$C NMR experimental parameters

Prior to the NMR experiments, unirradiated and 9 days UV-A irradiated samples were immersed in distilled water and solid-state NMR experiments were carried out on hydrated samples. $^{13}$C NMR experiments were performed using a Bruker Avance (Billerica, MA) spectrometer at a magnetic field strength of 21.10 T. All the experiments were carried out using a 4mm center packing rotor with an insert to keep the samples hydrated, as well as to center the samples with respect to the RF coil. $^{13}$C MAS (magic angle spinning) spectra were measured using a DEPTH sequence (to suppress background carbon signals arising from rotor inserts and the probe head) [101, 102] with 80 kHz TPPM decoupling [103] at (300±1) K and direct polarization. The spinning speed was set to 14.5 kHz for all the samples and spectra were acquired by accumulating 18,800 scans. The $^{13}$C π/2 pulse was 54 μs and the recycle delay was 6s. Analysis of data was performed using MATLAB and matNMR with a Gaussian
multiplication broadening factor of 100 Hz. $^{13}$C NMR spectra were referenced to adamantane (TMS=0 ppm).

2.5.1.5 Sample hydrolysis and quantification with Labeled-Des

Elastic fiber samples for this study were lyophilized for 24 h prior to hydrolysis. Approximately, 2.1–2.2 mg of each sample was placed into a solution containing 300 μl of 6 M HCl and 1 μl of 0.5% w/w phenol solution. The sample and solution mixture was placed into a vacuum hydrolysis tube, flushed with nitrogen gas, and then evacuated. The samples were kept at 110 °C for 96 h, afterward the solvent was frozen in liquid nitrogen and lyophilized for 8–10 h. After lyophilization, each sample was suspended in a 50 μl solution of 94.5% 0.14 M sodium acetate, 0.5% triethylamine, and 5% acetonitrile (v/v/v) at a pH of 7.5. Resuspended samples were diluted to 100-fold. Labeled d4-desmosine standard (Toronto Research Chemicals, Toronto, Canada) at final concentration of 10 pmol/μl was mixed with the diluted samples in three different ratios 1:1, 3:1, and 1:3. The relative amount of desmosine in each sample was quantified with respect to mass spectrometric peak area of standard desmosine in MS$^2$ mode. For statistical analysis of the data, a t-test was used assuming our data followed a normal distribution, and using the standard deviations in each sample studied. Null hypothesis probability was measured indicating the level of significance of our data.

2.5.2 Results and Discussion

UV-A induced alterations of the elastic fiber to follow the effect of UV-A exposure on elastic fibers, bovine nuchal ligament fibers with and without UV-A irradiation were investigated. Representative histological images of 9 days UV-A irradiated and unirradiated
elastic fibers are shown in Figure 2.6. From the figure, it is clear that in an unirradiated sample, multiple dense layers of elastic fibers are arranged in a regular fashion; the fibers appear relatively straight and are not fragmented. This regular pattern is dramatically disrupted after UV-A irradiation by 9 days (Figure 2.6, right panel).

**Figure 2.6** Histological images of unirradiated (left) and 9 days irradiated (right) bovine nuchal ligament elastic fibers. Images were photographed at 10× magnification and elastic fibers are stained black. The onset of fragmentation and disruption in ordering of elastic fibers is evident following UV-A exposure. The scale bar in the figures shown is 100 μm.

UV-A exposure results in fragmentation of elastic fibers and alters the regularly ordered arrangement of the fibers as well. Moreover, elastic fibers appear thinner compared to that of unirradiated samples; the observed alterations in the elastic fibers due to UV-A irradiation follows the work previously reported in the literature [95]. It is known that the level of elastase increases due to UV irradiation *in-vivo* which initiates elastin degradation [104, 105]. Imayama *et al.* applied SEM and TEM to study the elastic fibers of dermal connective tissue and showed that UV-irradiation alters the elastic fiber by formation of new elastic fibers which later form an irregular network [91]. In their study, they irradiated the soles of Sprague-Dawely rats with UV-B (130 mJ/cm² for 3 min/day up to 12 weeks) and therefore it was possible to follow the formation of new elastin. In the present *in-vitro* study, cells have been removed by the
purification step, consequently, elastic fibers are not repaired. Thus, this *in-vitro* measurement allows for characterization of alterations that occur as a result of break down in the cross-links of elastin, discussed below, which in turn degrade their structure.

TEM was used to follow the macroscopic alterations of UV-A exposure on elastic fibers. Figure 2.7 shows TEM images of the sample without UV-A exposure and with 9 days of UV-A exposure. In the images, the elastic fibers are in white denoted ‘a’, the interstitial spaced between fibers is denoted ‘c’ and cracks within or on an edge of fibers denoted ‘b’. Due to UV-A irradiation, pronounced cracks are observed in the 9 days irradiated sample, both in the interior of the elastic fiber and on the edges. Similar to the histological measurements, elastic fibers appear sparser and thinner in the 9 days irradiated sample, compared to the unirradiated sample.

![Figure 2.7 TEM images of (left) unirradiated and (right) 9 days irradiated elastin samples taken cross-sectionally (fiber axis points into the page). In the images, the light gray color (a) denotes an elastic fiber, (b) cracks on the edge or within the elastic fiber are as a result of UV-A exposure, and (c) the interstitial space between fibers.](image)

UV-A exposure may initiate microscopic changes of the structure or dynamics of the principal protein, elastin, in the elastic fiber. To follow these changes, $^{13}$C MAS NMR
experiments were carried out. Chemical shift assignments of all major amino acids (e.g. glycine, proline, valine) that comprise the spectra shown in Figure 2.8 were made following previously reported measurements of elastin and its related peptides [99, 106-110]. A detailed discussion of the peak assignments and structures observable in elastin has been reported elsewhere [99, 111].

![Figure 2.8](image-url)  
*Figure 2.8* $^{13}$C MAS NMR spectra of aliphatic region of unirradiated (top) and 9 days UV-A irradiated (bottom) elastic fibers recorded at 300 K. Peak assignments were made using values published in the literature, as discussed in the text. The spectra were acquired by accumulating 18,800 scans at 14.5 kHz magic angle spinning speed at a magnetic field strength of 21.10 T.

Experimental studies of the structure of elastin by modern NMR methodology are difficult due to the large, highly cross-linked nature of the protein. Even with isotopic labeling, degeneracies present in the spectra make structural elucidation challenging—in the present work we apply magnetic resonance spectroscopy to reveal secondary structural changes or dynamics due to UV-A exposure. All $^{13}$C peaks in these samples (unirradiated and 9 days irradiated)
appear in the same position and no significant chemical shift differences were observed. These measurements reveal that any microstructural changes of elastin following 12 mW/cm$^2$ of UV-A irradiation for 9 days do not appear to alter the $^{13}$C NMR chemical shifts of the most abundant amino acids that comprise the protein.

2.5.2.1 Biochemical changes of elastin due to UV-A exposure

The detrimental effects of UV rays on the cell and extracellular components (e.g., elastin, collagen, etc.) of skin are already known to some extent [94, 95, 97, 104, 112, 113]. Sun exposed skin showed abnormalities which include fiber disintegration, thickening, and proliferation of the elastic fiber [94]. A marked decrease in the linearity of the elastic fiber is also observed due to UV exposure [95]. Chatterjee et al. observed skin wrinkling when hairless mice were exposed to UV-B irradiation and followed different biochemical parameters. Biochemical parameters such as water, elastin, and glycosaminoglycan content were observed to increase whereas the collagen content remained the same after UV-B irradiation [104]. However, a different study showed that both UV-A and UV-B irradiation increase the concentration of elastin while the collagen content remains unaffected [113]. To obtain qualitative assessment of the macroscopic damage in elastic fiber that occurs in 9 days of UV-A exposure and to relate these changes to the actinic damage, we compare the UV-A irradiation from the lamp used in this study with that of the sun over the same period of time. Taking the total intensity of light on earth surface (1400 W/m$^2$) one finds that the energy deposition from the UV-A lamp was 2.85 times higher compared to that of continuous exposure to the sun for duration of 9 days. For this calculation we assumed that only 3 % of the incident solar energy is UV radiation, and that the sun is at the zenith [114]. This computation does not account for any attenuation of the UV radiation, e.g. through the dermis,
which may act as a layer of protection to the elastic fibers and reduce the intensity of the incident radiation. Des and Isodes form the cross-linking domains of elastin and their relative concentrations may be used as a measure of biochemical changes. The effects of UV radiation on isolated Des and Isodes have been reported previously by Baurain and coworkers [115]. Their study showed photolytic reaction results free lysine due to breaking down of pyridinium rings of Des/Isodes. However, the production of lysine depends on several factors such as irradiating wavelength, pH, and irradiation time. The formation of lysine was 35% when they irradiated Des at 274 nm for 25 min, whereas, for Isodes the irradiation dose was 285 nm for 50 min and yield was 78%. At wavelengths of 320–400 nm (UV-A), the absorbance in desmosine is ≤0.05 [115].

In the present study, we quantify the reduction of desmosine due to UV-A in bovine nuchal ligament elastic fibers by MS² experiments. In these measurements, the unirradiated sample was used as a control. The relative amount of desmosine in the 9 days UV-A irradiated sample was normalized with the amount of Des in the unirradiated sample. The results indicate the amount of Des in sample without UV radiation was 1.00 ± 0.01 mg, while amount of Des observed after 9-days of UV radiation was 0.89 ± 0.04 mg. The result showed the relative amount of Des in the 9-days irradiated sample was found to decrease by approximately 11%. The decrease in the relative content of Des arises from the breakdown of cross-links as a result of the irradiation, as each measurement was made per unit mass of the lyophilized sample prior to hydrolysis.

A previous in-vivo study on hairless mice skin showed that the desmosine content increases when treated with UV-A or UV-B [104, 113]. Chatterjee et al. observed an increase in elastin content per unit area of mice skin with UV-B exposure (40 mJ/cm²) of 18 weeks
irradiating three times a week [104]. However, Johnston et al. did not observe any statistically significant difference in the relative desmosine content when purified elastin was treated using 1 and 10 J of UV-A [113]. In the current in-vitro study, an 11% decrease in desmosine content is observed in 9 consecutive days of UV-A irradiation using approximately 42 kJ, whereas in the previous work by Johnson et al., 1 J for 6.25 min was employed [113]. Previous studies suggest that UV-B is more detrimental to the asthetic and mechanical functionality of skin than UV-A. However, the results obtained in the present study show that UV-A may play a greater role in photoaging of skin than previously believed. Thus, the findings of this work indicate that cross-linking of tropoelastin is measurably reduced following UV-A irradiation and this breakdown may alter its biomechanical characteristics during prolonged exposure if not repaired. Chung and coworkers [114] have observed a marked increase in MMP-12 activity following UV exposure, which causes degradation of the extracellular matrix, including elastin and collagen. The reduction of Des following UV-A exposure revealed in this in-vitro study of elastic fibers indicates that additional alterations are present before MMP-12 is expressed and other biological processes are initiated.

2.5.3 Summary and future directions

The current study introduces a novel approach to quantify Des and Isodes using Matrix-Assisted Laser Desorption Ionization (MALDI)-tandem mass spectrometry (MS$^2$) in a linear ion trap coupled to a vacuum MALDI source. MALDI-MS$^2$ analysis of Des and Isodes are performed using stable-isotope labeled desmosine (Labeled-Des) as an internal standard in different biological fluids such as urine and calf serum. The method demonstrated good linearity over two orders of magnitude with dynamic range from 3.125 to 125 ng/µL for both Des and
Isodes. Detection limit of current quantification method using Des standard is 0.02 ng/µL in urine and in calf serum. The method reproducibility (%RSD) is < 5% for Des.

The method is used to evaluate the time-dependent degradation of Des upon UV radiation (254nm) and found to be consistent with quantification by NMR. This is the first reported MALDI-MS$^2$ method for quantification of Des and Isodes, and illustrates the potential of MALDI-ion trap MS$^2$ for effective quantification of biomolecules. The reported method represents improvement over current liquid chromatography-based methods with respect to analysis time and solvent consumption, while maintaining similar analytical characteristics.

In future project, we plan to incorporate the solid-phase extraction step prior to MS analysis to be able to analyze the extremely low-level of Des in body fluid such as sputum. We also aim to implement this MALDI-MS$^2$ based quantification method to study the effect of parity on the architecture of vaginal elastic fibers in female Sprague-Dawley rats. The female reproductive tract is rich in elastic fibers. Elastin, a major component of elastic fibers provides elasticity and resiliency to connective tissue of skin, lungs, aorta, and pelvic floor. Many physiological conditions such as aging, normal vaginal birth and multiparity can lead to weakening of the connective tissue of pelvic floor. Consequently, failure to maintain elastic fiber homeostasis after parturition can lead to pelvic floor disorder. Hence, in our collaborative efforts with our collaborator Dr. Boutis, we seek change in Des content of vaginal elastic fibers of virgin and pregnant (1, 3 and more than 5 birth) rat using this developed quantification approach.
Chapter-3

Identification and characterization of phosphorylation sites in HIV proteins using mass spectrometry

3.1 Protein phosphorylation
3.2 HIV Introduction
   3.2.1 HIV Genome
   3.2.2 Role of different HIV proteins
   3.2.3 Known phosphorylation events in HIV proteins
   3.2.4 Experimental workflow
3.3 Materials and Methods
   3.3.1 In-vitro kinase assay of HIV-peptides/proteins
   3.3.2 In-gel digestion
3.4 Results and Discussion
   3.4.1 Investigation of potential CDK-mediated phosphorylation sites of HIV proteins
   3.4.2 Investigation of potential phosphorylation sites in HIV-1 accessory proteins
   3.4.3 Investigation of potential phosphorylation sites in HIV-1 structural proteins
3.5 In-vitro kinase assay on HIV-proteins (Vif, Nef and CA)
3.6 Summary and future directions
3.1 Protein phosphorylation

One of the most intriguing and challenging areas in the current research is to understand cellular mechanism and signaling network integration. An important step to understand these cellular processes is to identify and characterize phosphorylation sites in both, individual proteins as well as human proteome. Protein phosphorylation is one of the most prevalent post-translational modifications which play a pivotal role in many cellular processes such as cell proliferation, differentiation and migration. Protein phosphorylation involves catalytic addition of phosphate group from adenosine triphosphate (ATP) on serine/threonine/tyrosine residues of a protein [116].

Protein phosphorylation is controlled by a wide variety of kinases 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 in the protein to either turn-on or turn-off the protein – an effective way to regulate protein functions [117]. It is estimated that about 30 percent of proteins exist in phosphorylated form at any given time in a cell, which indicates that the phosphoproteome has vast and vital role to play in regulating various cellular processes [118]. However, less than 10 percent of total potential phosphorylation sites have been identified and well-characterized so far [119]. This suggests that even though protein phosphorylation is playing vital role to regulate various cellular processes and has been investigated for long enough time, tremendous amount of work is still needed to be done.

3.2 HIV Introduction:

Human Immunodeficiency Virus (HIV-1) is a causative agent of AIDS. HIV-1 specifically target host defense cells, notably CD4+T cells, and progressively destroys the ability
of the host to fight HIV infection and make the host prone to other diseases. The HIV-1 virus uses various host factors for its replication therefore; the infectivity of HIV-1 virus greatly depends on the interactions of viral components with various host factors [118, 119]. So, in this report we sought to identify and characterize interactions between host-kinases and HIV-protein phosphorylation sites that might provide insight into possible regulatory mechanism of HIV-proteins. Consequently, this investigation might provide a new pathway for development of anti-HIV therapeutic agents.

3.2.1 HIV genome

The HIV-1 genome encodes total of nine open reading frames. Three of these genes encode the major structural proteins for virus particle such as Gag, Pol, and Env polyproteins. These polyproteins further proteolyzed into individual proteins. The Gag polyprotein proteolyzed into four different proteins: Matrix (MA), Capsid (CA), Nucleocapsid (NC) and P6. These proteins from gag gene provide basic infrastructure for the virus particle. The Env polyprotein cleaved into two Env proteins gp120 (SU- surface) and gp41 (TM- transmembrane). Together with Gag proteins, Env proteins are also a part of structural components that make up the core of the virion and outer membrane envelope [119]. The three Pol proteins known as Protease (PR), Reverse Transcriptase (RT), and Integrase (IN), perform essential enzymatic functions, and are involved in viral replication [120]. In addition to these structural proteins, the HIV-1 virus particle also encodes six additional proteins, which are called auxiliary proteins. Three of these proteins, such as Virus infectivity factor (Vif), Viral protein R (Vpr), and Negative regulator factor (Nef) are found in the viral particle known as accessory regulatory proteins. The other two, essential regulatory proteins known as Transactivator of transcription (Tat) and Regulator of
expression of the virion (Rev), provide essential gene regulatory functions. Finally, Vpu indirectly assists in assembly of the virion by enhancing the release of the progeny virion and inducing degradation of the CD4 receptor [121]. These auxiliary HIV proteins play important roles in host-pathogen interactions and have significant impact on course of HIV infection. The retroviral genome also contains two copies of single-stranded RNA molecules. So, in simplistic terms, HIV-1 may be considered as a molecular entity consists of 15 proteins and RNA.

![Figure 3.1](image)

**Figure 3.1** Structure of genome organization of HIV-1 particle [121]

### 3.2.2 Essential roles of HIV proteins

There are 15 different HIV-1 proteins, distinctly categorized into three different classes based on their functions: 1) structural proteins, 2) regulatory proteins, and 3) accessory proteins. Table 3.1 shows the list of HIV-1 proteins and their essential functions.
<table>
<thead>
<tr>
<th>HIV protein</th>
<th>Protein function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matrix, MA (p17&lt;sub&gt;gag&lt;/sub&gt;)</td>
<td>Target Gag to plasma membrane binding for virion assembly</td>
</tr>
<tr>
<td></td>
<td>Facilitate the nuclear transport of viral genome</td>
</tr>
<tr>
<td>Capsid, CA (p24&lt;sub&gt;gag&lt;/sub&gt;)</td>
<td>Formation of virion core structure and assembly</td>
</tr>
<tr>
<td>Nucleocapsid, NC (p7&lt;sub&gt;gag&lt;/sub&gt;)</td>
<td>Virion packaging of genome RNA</td>
</tr>
<tr>
<td></td>
<td>Act as a RNA chaperone</td>
</tr>
<tr>
<td></td>
<td>Facilitate reverse transcription</td>
</tr>
<tr>
<td>p6&lt;sub&gt;gag&lt;/sub&gt;</td>
<td>Promote efficient release of budding virions</td>
</tr>
<tr>
<td>Protease, PR</td>
<td>Proteolytic processing of Gag and Gag-Pol polyproteins</td>
</tr>
<tr>
<td>Reverse Transcriptase, RT</td>
<td>Synthesis of viral DNA from viral RNA</td>
</tr>
<tr>
<td>Integrase, IN</td>
<td>Covalent insertion of viral proviral DNA into cellular DNA</td>
</tr>
<tr>
<td>Surface Glycoprotein, SU (gp120&lt;sub&gt;env&lt;/sub&gt;)</td>
<td>Bind to cell-surface receptors</td>
</tr>
<tr>
<td></td>
<td>Mediate virus attachment and entry</td>
</tr>
<tr>
<td>Transmembrane Glycoprotein</td>
<td>Mediate membrane fusion and virus entry</td>
</tr>
<tr>
<td>TM (gp41&lt;sub&gt;env&lt;/sub&gt;)</td>
<td></td>
</tr>
<tr>
<td>Virion Infectivity Factor, Vif</td>
<td>Suppress APOBEC3G/3F (host factors that inhibit infection)</td>
</tr>
<tr>
<td>Viral Protein R, Vpr</td>
<td>Facilitate nuclear localization of preintegration complex (PIC) G2/M cell-cycle arrest</td>
</tr>
<tr>
<td>trans-Activator of Transcription, Tat</td>
<td>Activate viral transcription through interaction with TAR</td>
</tr>
<tr>
<td>Regulator of Expression of Virion Protein, Rev</td>
<td>Induces nuclear export of intron-containing viral RNAs</td>
</tr>
<tr>
<td></td>
<td>Essential post-transcriptional regulator of virion gene expression</td>
</tr>
<tr>
<td>Viral Protein U, Vpu</td>
<td>Degradation of cellular receptor CD4</td>
</tr>
<tr>
<td></td>
<td>Induces virion release from host cell surface</td>
</tr>
<tr>
<td>Negative Regulator Factor,</td>
<td>Downregulation of cell expression of CD4</td>
</tr>
</tbody>
</table>
3.2.3 Known phosphorylation events of HIV proteins

Many reports have been published to date demonstrating that phosphorylation events regulate the functions of various viral proteins and play a vital role in HIV infection. Hence, strategies that can reverse or inhibit the phosphorylation events have also been proposed as one of the possible mechanisms to halt or slow down the HIV infections in many instances. Some examples of phosphorylation events in HIV proteins are: Vif [122-124], Rev [125-127], Vpr [128, 129], Vpu [130, 131], Nef [132-134], Tat [135-137], P6 [138], Integrase (IN) [139], Capsid (CA) [140], Matrix (MA) [141-144], and Reverse Transcriptase (RT) [145]. These phosphorylation events in different HIV proteins play many vital roles including viral entry to host-immune cell, viral replication, and budding. Some examples are described below:

1. Phosphorylation of Tyrosine residue in Matrix (MA) protein leads to its association with viral Integrase (IN) enzyme in the nucleoprotein complex and aids targeting of Pre-integration complex (PIC) to the nucleus [146, 147].

2. Phosphorylation of Reverse Transcriptase (RT) at Thr216 by host CDK2, increases in-vitro and in-vivo RT activity and viral infectivity [145].

3. Phosphorylation of Tat by RNA dependent cellular protein kinase (PKR) enhances the binding to TAR RNA [136]. Moreover, phosphorylation of Tat by PKR also assists in evasion of host anti-viral mechanism.
4. Vpr is known to be phosphorylated at three sites including Ser79, Ser94 and Ser96. Zhou et al. study shows that phosphorylation of Ser79 is required for cell cycle arrest, while mutation of all three sites attenuated HIV-1 replication in macrophage [128].

5. The 23 kDa Vif protein is phosphorylated at five sites including four sites in the C-terminal region. The substitution of phosphorylatable residues, Thr96 and Ser144 results in loss of 90% of Vif activity and inhibits HIV-1 replication [123, 124].

6. Phosphorylation of Vpr by protein kinase A (PKA) at Ser79 activates Vpr-induced cell cycle arrest during HIV-1 infection [129].

7. Phosphorylation of N-terminal Nef residues by serine kinases influences the viral infectivity [148], while phosphorylation of Nef by protein kinase C (PKC) enhances CD4-down regulation [149]. Furthermore, phosphorylation of Ser6 in Nef modulates HIV transcription, replication rates and subcellular localization [132].

8. Integrase is modified by JNK phosphorylation followed by Pin1-catalyzed prolyl cis/trans isomerization, steps necessary for genomic integration in CD4+ T cells [139].

3.2.4 Experimental Workflow

Preliminary goal of the current investigation is to identify one or more phosphorylation sites across various HIV proteins and study the influence of identified phosphorylation event on HIV infection. This investigation may successfully introduce a definitive database of host-cell kinase/HIV phosphorylation substrate interactions, which may prove as a strong pillar for further investigation of anti-HIV therapeutics. Following are the strategic steps adopted to achieve this aim:
Step-1: Multiple sequence alignments of individual HIV-protein:

We have constructed multiple sequence alignments using 12 different strains of HIV to identify conservation in phosphorylation sites across each HIV protein using ClustalW software. The conservation of phosphorylation sites in a protein reflects their possible biological importance in function of HIV-1 protein. In our study, we focus specifically on the Serine (Ser), Threonine (Thr) and Tyrosine (Tyr) residues that demonstrate low entropy and conservation across different strains.

Step-2: Kinase-specific phosphorylation site prediction algorithms:

HIV protein sequences have been evaluated by web-based kinase-specific phosphorylation site prediction algorithms such as NetPhosK, GPS (Group-based prediction system) and KinasePhos. These algorithms not only provide prediction of potential phosphorylation sites in HIV proteins but also identify the host-kinase with prediction scores for each possible phosphorylation site.

Step-3: In-vitro kinase assay of HIV-peptides and proteins:

Libraries of short synthetic peptides derived from HIV proteins sequence containing candidate phosphorylation sites (identified from step 1 and 2) are subjected to in-vitro kinase assay with respective kinases in presence of ATP. Kinase activity is assayed using MALDI-ion trap mass spectrometer, which has the benefits of less sample requirements, rapid readout, and MS^n capabilities for structural elucidation. Preliminary hits from this initial screening are then be evaluated on protein level. In future, biochemical effects of phosphorylation will be investigated by assessing the differences in activity as well as stability of different HIV proteins as a function
of phosphorylation state. This study will confirm biological relevance of in-vitro phosphorylation sites, and generate hypotheses for future studies focusing on mechanistic characterization of discovered HIV phosphorylation sites.

3.3 Materials and Methods

Various kinases such as Glycogen Synthase Kinase 3 (GSK-3), p42 MAP kinase, cAMP-dependent Protein Kinase (PKA), Casein Kinase I (CKI), Casein Kinase II (CKII), CDK2/CyclinA, CDK2/CyclinE, CDK4/CyclinD, CDK9/CyclinT1, CDK7/CyclinH, CDK1/CyclinB and 1x NEBuffer (50mM NaCl, 10mM Tris-HCl, 10mM MgCl₂, and 1mM DTT) were purchased from New England BioLabs (MA, USA). Other kinases such as CDK5/p25, CDK5/p35, p38 MAPK and Protein Kinase C- [alpha, beta (II), theta, and delta] were purchased from Enzo Life Science (NY, USA). The protease for in-gel protein digestion including trypsin, chymotrypsin, and Arg-C were purchased from Promega (Madison, WI, USA) and Glu-C was purchased from Thermo Fisher Scientific (Waltham, MA, USA). The library of small synthetic HIV-derived peptides together with different full-length HIV-proteins was ordered from National Institutes of Health AIDS Research & Reference Reagent Program.

3.3.1 In-vitro kinase assay of HIV-peptides/proteins

10μM HIV-peptide/protein of interest was incubated with 1μM kinase in presence of 4mM ATP and 1X NEBuffer at 30°C. Sample was aliquoted at different time interval- 30 min, 1 hour, 3 hours and overnight. 1μL of sample was deposited on MALDI sample probe after addition of CHCA (alpha-cyano-4-hydroxycinnamic acid) matrix solution in ratio of 1:14(v/v). Thermo Scientific MALDI LTQ-ion trap mass spectrometer was employed for mass spectrometric analysis.
3.3.2 In-gel protein digestion

After incubation of lead kinase (identified from in-vitro kinase study of HIV peptides) with HIV protein (Vif, Nef or P24) for 4 hours, samples were separated by NuPAGE gel and stained using GelCodeTM Blue Safe Protein Stain (Thermo Scientific). The protein bands were cut out followed by digestion with proteases (trypsin, chymotrypsin, Arg-C and Glu-C) at 37 °C for 4 hr. The protease was added to a final substrate-to-enzyme ratio from 100:1 to 20:1 and optimum pH was maintained (according to standard in-gel digestion protocol provided by vendor). Following the in-gel digestion of protein, peptide extraction was performed using POROS 20 R2 resin (Applied Biosystems) for 4hr at 4°C. The ZipTip (Millipore) was employed to desalt the peptide fragments prior to MALDI mass spectrometric analysis. The sample eluted from ZipTip using CHCA matrix solution was directly delivered onto a MALDI sample probe, and analyzed using MS and MS² modes. Identification of peptide mass, phosphopeptide peaks and interpretation of MS² and MS³ spectra was performed manually using ProteinInfo (Rockefeller University, NY; http://prowl.rockefeller.edu).

3.4 Results and Discussion

The first collaborative project between our lab and Massachusetts General Hospital (Dr. Mathias Lichterfeld) revealed that effective HIV-1 reverse transcription depends on phosphorylation of viral reverse transcriptase (RT) by host cyclin-dependent kinase CDK2 at highly conserved amino acid residue (Thr216) [145]. They investigated functional relevance of CDK2-dependent phosphorylation of HIV-1 RT by substituting Thr216 with Alanine (Ala) and observed that replacement of Thr216 residue results in absence of phosphorylation and reduction in viral infectivity. CDK-2 dependent phosphorylation increased efficacy and stability of viral
reverse transcriptase (RT), an enzyme which is responsible for synthesis of viral DNA from viral RNA, a critical step in HIV infection. Consequently, we decided to further study and characterize potential phosphorylation sites across various HIV-1 proteins to evaluate the effect of HIV-1 protein phosphorylation on the course of HIV infection. This study was conducted mainly in three phases: 1) Investigation of potential CDK-mediated phosphorylation sites of HIV proteins, 2) Investigation of potential phosphorylation sites in HIV auxiliary proteins, and 3) Investigation of potential phosphorylation sites in HIV structural proteins.

3.4.1. Investigation of potential CDK-mediated phosphorylation sites of HIV proteins

The investigation to evaluate potential CDK-mediated phosphorylation sites across various HIV-1 proteins was initiated with a set of 22 potential phosphorylation sites containing HIV-peptides. These HIV-peptides were selected based on the presence of minimal CDK-mediated phosphorylation motif, (S/T)PX, where X represents any amino acid, and (S/T)P means Serine or Threonine residue followed by Proline. Six different cyclin-dependent kinases (CDK)s including CDK2/Cyclin A, CDK2/Cyclin E, CDK4/Cyclin D, CDK9/CyclinT1, CDK7/CyclinH, CDK1/CyclinB were chosen to perform in-vitro kinase assays on these HIV-peptides.
Table 3.2 List of 22 HIV-peptides containing CDK-mediated phosphorylation motif, (S/T)PX are shown below. The last column represents mass of the peptides.

<table>
<thead>
<tr>
<th>Peptide sequence</th>
<th>Name</th>
<th>Peptide mass (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QMVHQAISPRTLNAWVKV</td>
<td>GAG-B-20</td>
<td>2077.25</td>
</tr>
<tr>
<td>WVKVVEEKAFSPEVIPMF</td>
<td>GAG-B-22</td>
<td>2134.11</td>
</tr>
<tr>
<td>PMFSALSEGATPQDLNTM</td>
<td>GAG-B-24</td>
<td>1908.86</td>
</tr>
<tr>
<td>IVRMYSPSILDIRQGPK</td>
<td>GAG-B-39</td>
<td>2073.14</td>
</tr>
<tr>
<td>GEETTPSQQPQEPIDKEL</td>
<td>GAG-B-64</td>
<td>2028.98</td>
</tr>
<tr>
<td>QGYFDPWQNYTPGPGIRY</td>
<td>NET-B-17</td>
<td>2157.99</td>
</tr>
<tr>
<td>FLYQSNPPSPSPEGTRQAR</td>
<td>REV-B-4</td>
<td>2044.01</td>
</tr>
<tr>
<td>TQGVGPSQILVESPAVL</td>
<td>REV-B-14</td>
<td>1693.92</td>
</tr>
<tr>
<td>QILVESPAVLESGTKEE</td>
<td>REV-B-15</td>
<td>1827.95</td>
</tr>
<tr>
<td>REFSSEQTRANSPTRREL</td>
<td>POL-B-3</td>
<td>2163.08</td>
</tr>
<tr>
<td>LGVPTVPNIIQRLNTQI</td>
<td>POL-B-19</td>
<td>1917.14</td>
</tr>
<tr>
<td>GQVDCSPGIWQLDCTHL</td>
<td>POL-B-104</td>
<td>1870.83</td>
</tr>
<tr>
<td>PCVKLTPLCVTNCTDL</td>
<td>ENV-B-17</td>
<td>1831.92</td>
</tr>
<tr>
<td>VRQGYSPLSFQTRLPAPR</td>
<td>ENV-B-95</td>
<td>2072.13</td>
</tr>
<tr>
<td>IRNAILGHIVSPRCEYQA</td>
<td>VIF-B-17</td>
<td>2039.07</td>
</tr>
<tr>
<td>AALITPKKIKPPLPSVTK</td>
<td>VIF-B-21</td>
<td>1901.21</td>
</tr>
<tr>
<td>SDIAGTTSTPQEIQGWM</td>
<td>GAG-A-33</td>
<td>1820.82</td>
</tr>
<tr>
<td>PQRERIRQTAPAEVG</td>
<td>NEF-A-3</td>
<td>1903.04</td>
</tr>
<tr>
<td>PWNHPGSQPTTPCNKCY</td>
<td>TAT-A-2</td>
<td>1928.83</td>
</tr>
<tr>
<td>GRKRRQRQRTPQSSKDH</td>
<td>TAT-A-7</td>
<td>2177.21</td>
</tr>
<tr>
<td>MSPEIWAIVGLIVALIL</td>
<td>VPU-A-1</td>
<td>1950.16</td>
</tr>
<tr>
<td>EAGAEQGTPTFSFPQI</td>
<td>POL-A-7</td>
<td>1834.88</td>
</tr>
<tr>
<td>SFQTLTPNPRGDRLGRI</td>
<td>ENV-C-96</td>
<td>2024.09</td>
</tr>
<tr>
<td>RRRFPASPLRGPPK</td>
<td>Positive Control</td>
<td>1791.15</td>
</tr>
</tbody>
</table>

The result of the *in-vitro* kinase assays indicates that 8 out of 22 potential phosphorylation sites have shown to be phosphorylated in presence of different CDKs. For example, when the peptide derived from VIF protein (VIF-B-21) having sequence of (AALITPKKIKPPLPSVTK) was incubated with CDK2/Cyclin E and ATP, for 1hr, 3hr and overnight, a time-dependent increase in phosphopeptide peak was observe in mass spectrometric analysis (Figure 3.2). The presence of a robust, CDK dependent peak at 1982.82Da with mass difference of +80Da, represent addition of phosphate group (HPO₃) to the control VIF-B-21 peptide peak (1902.73Da). To verify the presence of the mass peak at 1982.82Da was due to
Figure 3.2 In-vitro kinase assay result of VIF-B-21 peptide- Mass spectra of VIF-B-21 peptide (AALITPKIKPPLPSVTK- 1902.73Da) in presence of CDK2/Cyclin E and ATP at different time points (0hr, 1hr, 3hr, and overnight) are shown here. Time-dependent increase in the mass peak at 1982.82Da shows the CDK-dependent phosphorylation of VIF-B-21 peptides.

CDK-mediated phosphorylation of VIF-B-21 peptide, phosphopeptide (1982.82Da) was subjected to MS/MS analysis. In principle, collision induced dissociation (CID) generate a strong signal at -98Da from the phosphorylated mass peak that correspond to the neutral loss of phosphoric acid (H$_3$PO$_4$). The presence of this ion in the MS$^2$ spectra (Figure 3.3) is considered as conclusive evidence of phosphorylated peptide ions. Furthermore, other fragment ions generated from the peptide backbone were also investigated to find out correct phosphorylation site. Fragment ions such as b10, b*13 and b*14 (representing N-terminal peptide fragments) that contain phosphorylated-(Thr5), represent the mass of the fragment ions with attachment of phosphate group, while other fragment ions, y”11 and y”13 (representing C-terminal peptide fragment), which does not contain phosphorylated-(Thr5) residue represent the original mass of fragment ions. Thus, this investigation confirms the phosphorylation of (Thr5) of VIF-B-21 peptide in presence of CDK2/Cyclin E and ATP. The summary of in-vitro kinase assays
performed on the CDK consensus motif containing HIV-peptides are shown in Table 3.3. The list represents only those HIV-peptides that showed the presence of phosphorylation upon incubation with CDKs in presence of ATP.

\[
\text{A A L I pT P K K I K P L P S V T K}
\]

Figure 3.3 MS\(^2\) spectrum of the phosphorylated VIF-B-21 peptide: MS\(^2\) fragmentation indicates the presence of a strong peak with difference of 98Da from the parent mass peak correspond to neutral loss of phosphoric acid. The analysis of various backbone fragmentation of phosphorylate VIF-B-21 peptide suggest the attachment of phosphate group on (Thr5) of the VIF-B-21 peptides.

Table 3.3 represent list of HIV-derived peptides show phosphorylation using different CDKs. (+++, phosphorylated peak is at least equally as intense as unphosphorylated; ++, 25-100%; +, 2-25%; -, <2%)

<table>
<thead>
<tr>
<th>HIV-Peptides</th>
<th>MW (Da)</th>
<th>CDK1/ CyclinB</th>
<th>CDK2/ CyclinA</th>
<th>CDK2/ CyclinE</th>
<th>CDK4/ CyclinD</th>
<th>CDK9/ CyclinT1</th>
<th>CDK7/ CyclinH</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLYQSNPPPSPEGTRQAR</td>
<td>2044.01</td>
<td>+++</td>
<td>+++</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>REFSSEQTRANSQTRREL</td>
<td>2163.08</td>
<td>+++</td>
<td>+++</td>
<td></td>
<td>+++</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>VRQGYSPLSFQTRLPAPR</td>
<td>2072.13</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>AALITPKIKPLPSVTK</td>
<td>1901.21</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>IRNAILGHIVSPRCYQA</td>
<td>2039.07</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>PWNHPGSQQTTIPCNKY</td>
<td>1928.83</td>
<td>+++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SFQTLTPRPGPDRLGRI</td>
<td>2024.09</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>WVKVVEKAFSPEVIPM</td>
<td>2134.11</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Beside the investigation of the aforementioned 8 HIV-peptides in presence of various CDKs, the in-vitro kinase study was also conducted on these HIV-peptides with few other kinases such as, casein kinase I (CK1), Casein kinase II (CKII), mitogen-activated protein kinase
(MAPK), glycogen synthase 3 (GSK-3), protein kinase B (PKB), and different isozymes of protein kinase C (PKC-alpha, beta, delta). These kinases were selected based on their high phosphorylation prediction scores obtained from phosphorylation prediction algorithms.

Table 3.4 Results of in-vitro kinase assay for CDK consensus motif containing 8 HIV-peptides- In-vitro kinase assays performed on 8 HIV-peptides in presence of specific kinases with high prediction score obtained using various phosphorylation prediction algorithms such as NetPhosK, GPS and KinasePhos. (+++, phosphorylated peak is at least equally as intense as unphosphorylated; ++, 25-100%; + 2-25%; ---, no phosphorylation)

<table>
<thead>
<tr>
<th>HIV Peptides</th>
<th>MW</th>
<th>PKA</th>
<th>PKC</th>
<th>CK I</th>
<th>CK II</th>
<th>MAPK</th>
<th>GSK3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Da)</td>
<td>alpha beta</td>
<td>delta</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLQSNPPSPGTRQRAR</td>
<td>REV-B-4</td>
<td>2044.01</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>REFSSEQTRANSFRREL</td>
<td>POL-B-3</td>
<td>2163.08</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>--</td>
<td>+</td>
</tr>
<tr>
<td>VRQGYSLSFQTLAPPR</td>
<td>ENV-B-95</td>
<td>2072.13</td>
<td>--</td>
<td>++</td>
<td>+</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>AALITPJKKPLPSVTK</td>
<td>VIF-B-21</td>
<td>1901.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRNAIHLVISPRCEYQA</td>
<td>VIF-B-17</td>
<td>2039.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWNHPQGQTTPCNKCY</td>
<td>TAT-A-2</td>
<td>1928.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFQILTPPRGBPRLGR</td>
<td>ENV-C-96</td>
<td>2024.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WVKVVEKAFSPEVPM</td>
<td>GAG-B-22</td>
<td>2134.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.4.2 Investigation of potential phosphorylation sites in HIV-1 auxiliary proteins

HIV-1 auxiliary proteins play vital role in HIV replication and infection. Moreover, many publications have been shown the importance of phosphorylation events of these auxiliary proteins in HIV infection [124, 127, 129, 130, and 132]. There are six HIV-1 auxiliary proteins in HIV viral particle such as Nef, Rev, Tat, Vif, Vpr and Vpu. For all these HIV auxiliary proteins multiple sequence alignment using ClustalW was performed to evaluate the conservation of serine/threonine/tyrosine across 14 different strains of HIV-1 (Figure 3.4 top panel- shows ClustalW result of VIF protein sequence). Furthermore, three different phosphorylation prediction algorithms such as NetPhosK, KinasePhos and GPS were employed to identify kinase-specific, potential phosphorylation sites across the HIV-1 protein sequence (Figure 3.4 bottom panel- shows prediction score of VIF obtained using NetPhosK)
Table 3.5: Phosphorylation site conservation and prediction scores for different HIV auxiliary proteins (Vif, Vpr, and Vpu).

<table>
<thead>
<tr>
<th>Subtype</th>
<th>Kinase</th>
<th>SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-20</td>
<td>PKB</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>PKC</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>RSK</td>
<td>0.53</td>
</tr>
<tr>
<td>S-23</td>
<td>DNAPK</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>PKC</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>cdc2</td>
<td>0.51</td>
</tr>
<tr>
<td>Y-30</td>
<td>INSR</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>PKC</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Figure 3.4: Vif multiple sequence alignment and bioinformatics prediction score: Top panel represents Virion infectivity factor (Vif), protein sequence from different strains is aligned together using ClustalW software identifies well-conserved phosphorylation motif in Vif protein across different strains. The residues which can be phosphorylated, such as Serine (S), Threonine (T) and Tyrosine (Y) are highlighted in purple, blue and green color respectively. Bottom panel represents scoring matrix of NetPhosK prediction scores of VIF from each strain. Notation ‘M’ represents mutation at the respective position in particular sequence. High prediction score represents high phosphorylation probability.

Based on conservation of the residues (serine/threonine/tyrosine) across various strains and high prediction scores obtained from different kinase-specific phosphorylation prediction tools, more than ten well-conserved potential phosphorylation sites were identified from three HIV auxiliary proteins (Vif, Vpr, and Vpu). Indeed, seven different peptides containing these ten potential phosphorylation sites were carefully chosen to perform in-vitro kinase study. Simultaneously, more than 8 well-conserved potential phosphorylation sites containing peptides from each HIV auxiliary proteins (Nef, Rev, and Tat) were identified and investigated. The results of the in-vitro kinase study on these peptides are summarized in Table 3.5 below.
Table 3.5 Results of *in-vitro* kinase assay of peptides from HIV auxiliary proteins- *In-vitro* kinase assay performed using potential phosphorylation site containing peptides from HIV auxiliary proteins in presence of potential kinases predicted using bioinformatics tools such as NetPhosK, GPS and KinasePhos. (+++, phosphorylated peak is at least equally as intense as unphosphorylated;++, 25-100%; +, 2-25%; ---, no phosphorylation)

<table>
<thead>
<tr>
<th>Vif, Vpr, Vpu-peptides</th>
<th>MW (Da)</th>
<th>CK I</th>
<th>CK II</th>
<th>PKB</th>
<th>PKC</th>
<th>PKA</th>
<th>GSK3</th>
<th>MAPK</th>
</tr>
</thead>
<tbody>
<tr>
<td>RIRERAEEDSGNESEG</td>
<td>1703.78</td>
<td>++</td>
<td>+</td>
<td>++</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>SLGQHIYETYGDTWA</td>
<td>1739.78</td>
<td>--</td>
<td>+</td>
<td>--</td>
<td>+</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>VDRMRIRTWKLKVHK</td>
<td>1924.09</td>
<td>+</td>
<td>+++</td>
<td>++</td>
<td>+</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>VKKHMYIGAKKGFW</td>
<td>1787.93</td>
<td>++</td>
<td>--</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>RHYESTHRPISSEV</td>
<td>1833.89</td>
<td>+</td>
<td>--</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>VSIEWRKKRYSTQVD</td>
<td>1894.01</td>
<td>--</td>
<td>++</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>DCFSESARINAILGH</td>
<td>1631.77</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>

**Rev peptides**

<p>| MAGRSGDSDEELLKT         | 1607.75 | --   | +++  | --  | --  | --  | --    |      |
| DEELKTVRLIKFLY          | 1879.08 | --   | --   | --  | --  | --  | --    |      |
| RLIKFLYQSNPPSP          | 1755.97 | --   | --   | --  | --  | --  | --    |      |
| WRERQRQIRISGWI          | 1970.07 | +    | --   | +++ | +   | --  | --    |      |
| GWILSTYLGPAEPV          | 1657.88 | --   | --   | --  | --  | --  | --    |      |
| CNEDCGTSGTQGVGS         | 1413.51 | --   | --   | --  | --  | --  | --    |      |
| CGTSGTQGVGSPQIL         | 1403.67 | --   | --   | --  | --  | --  | --    |      |
| GTQGVGSPQILVESP         | 1467.76 | --   | --   | --  | --  | --  | --    |      |
| VGSPQILVESPAVE          | 1536.84 | --   | --   | --  | --  | --  | --    |      |</p>
<table>
<thead>
<tr>
<th>Peptide</th>
<th>Mass</th>
<th>Tat peptides</th>
<th>Nef peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESPAVLESGTKEE</td>
<td>1374.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSPEGTRQARRRREC</td>
<td>1836</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LQLPPLERLTLDCNE</td>
<td>1752.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DPRLEPWKHPGSQPK</td>
<td>1770.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QPKTACTNCYCKKCC</td>
<td>1692.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KGLGISYGRKRRQR</td>
<td>1802.09</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>RQRRRAPQDSQTHQV</td>
<td>1861.97</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>DSQTHQVSLSKQPAS</td>
<td>1611.79</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>HQVLSKQPASQPRG</td>
<td>1618.85</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>LSKQPASQPRGDPTG</td>
<td>1537.78</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>PASQPRGDPTGPKES</td>
<td>1522.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MGGKWSKRSVVGWPT</td>
<td>1674.87</td>
<td>-- ++</td>
<td>+++ +</td>
</tr>
<tr>
<td>SVVGWPTVRERMRA</td>
<td>1798.97</td>
<td>+++</td>
<td>+</td>
</tr>
<tr>
<td>PAADGVGAVSRDLEK</td>
<td>1483.76</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>LEKHGAITSSNTAAN</td>
<td>1512.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAITSSNTAANNADC</td>
<td>1408.59</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>QVPLRPMTYKAADVLD</td>
<td>1700.93</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>YKAADVLSHFLKEKG</td>
<td>1704.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILDLWVVHTQGYFPD</td>
<td>1865.9</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>
The results obtained from kinase-specific phosphorylation prediction algorithms and *in vitro* kinase study shows that our strategy has successfully predicted known host-kinase/HIV-protein interactions. For example, phosphorylation of Nef protein at Ser6 and Ser9 by protein kinase A (PKA) \[133\], phosphorylation Rev protein at Ser5 and Ser8 by casein kinase II (CK II) \[150\], and phosphorylation Vpu protein at Ser52 and Ser56 by casein kinase II (CK II) \[151\] were predicted using our strategy. Additionally, successful assessment of the known phosphorylation interaction between host-kinase and HIV-proteins, the result also demonstrates that our strategy also predicts some novel phosphorylation events in various HIV proteins. For example, our strategy has successfully predicted the phosphorylation of N-terminal of Vif protein with protein kinase C (PKC), and phosphorylation of N-terminal of Nef protein with protein kinase C (PKC) (results are shown in Table 3.5).

### 3.4.3 Investigation of potential phosphorylation sites in HIV-1 structural proteins

Our collaborator at MGH was interested in investigating potential cytotoxic T lymphocytes (CTL) phospho-epitopes in HIV proteins. The N-terminal of capsid (CA) protein (a protein derived from proteolytic cleavage of gag polyproteins) possesses a known CTL epitope, KF-11 (KAFSPEVIPMF) peptide. The *in-silico* prediction and preliminary studies performed at
MGH shown the presence of phosphorylation on Ser33, a residue presence in KF11 peptide. Hence, in our collaborative effort, we decided to investigate the potential host-kinase involved in phosphorylation of Ser33 residue of capsid protein. Similar strategic steps including multiple sequence alignments (results are shown in Figure 3.5 below) and kinase-specific phosphorylation prediction algorithms (Table 3.6) are utilized to identify the possible host-kinases for each potential phosphorylation sites across the capsid protein including the Ser33.

**HIV P24 protein sequence alignment:**

![Multiple sequence alignment of Capsid protein: Capsid (p24) protein sequences from different strains are aligned together using ClustalW software identifies well-conserved phosphorylation motif in Capsid protein across different strains. The residues which can be phosphorylated, such as Serine (S) and Threonine (T) are highlighted in yellow and green color respectively.](image)

**Figure 3.5**

**Table 3.6** Phosphorylation prediction sites with potential kinases: data obtained from various kinase-specific phosphorylation prediction algorithms such as NetPhosK, GPS and KinasePhos are combined and used to predict potential phosphorylation sites and possible kinases.

<table>
<thead>
<tr>
<th>Phosphorylation sites</th>
<th>Predicted by 3 tools</th>
<th>Predicted by 2 tools:</th>
<th>High prediction score only by 1 tool</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NETPHOSK</td>
<td>KINASEPHOS</td>
<td>GPS</td>
</tr>
<tr>
<td>S-16</td>
<td>MAPK(P-38)</td>
<td>---</td>
<td>CK1</td>
</tr>
<tr>
<td>S-33</td>
<td>MAPK(P-38)</td>
<td>CK1</td>
<td>CK1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GSK3</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td>---</td>
<td>CDK</td>
</tr>
<tr>
<td>S-44</td>
<td>CK1</td>
<td>---</td>
<td>CK1</td>
</tr>
<tr>
<td>T-48</td>
<td>CDK5</td>
<td>CDK5</td>
<td>CDK5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T-58</td>
<td>PKC</td>
<td>PKC</td>
<td></td>
</tr>
</tbody>
</table>
The previous publication by Cartier et al. identified three phosphorylation sites across the capsid protein, which includes Ser109, Ser149 and Ser178. The result shows although capsid protein plays a vital role in formation of the viral core structure and assembly, these phosphorylation events occurred in capsid protein do not affect its function in viral structure and assembly formation. In contrast, site-directed mutagenesis study showed that these three mutant viruses were able to assemble and bud but were unable to infect the cell due to incomplete reverse transcription. Hence, these results indicate phosphorylation events are essential for reverse transcription process of virus and viral infectivity [152].

Because all three major phosphorylation sites (Ser109, Ser149, and Ser178) have been identified in previous investigation are present in C-terminal of capsid protein and our specific interest was to investigate phospho-epitope (KF11) presents in N-terminal of the capsid protein, I decided to investigate potential phosphorylation sites in the N-terminal of the capsid protein. As a preliminary goal of investigation was to identify host-kinases responsible for phosphorylation of Ser33 residue from CTL epitope, KF11, all predicted kinases from phosphorylation prediction algorithms (shown in Table 3.7) were evaluated by performing in-vitro kinase assays. Three overlapping peptides containing Ser33 residue were chosen for in-vitro kinase assay.

<table>
<thead>
<tr>
<th>HIV-peptide (Capsid protein)</th>
<th>Identification</th>
<th>Mol. Wt.(Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WVKVVEEKAFSPEVIPMF</td>
<td>Gag-B22</td>
<td>2134.12</td>
</tr>
<tr>
<td>AFSPREVIPMFSALSEGA</td>
<td>Gag-B23</td>
<td>1751.84</td>
</tr>
<tr>
<td>KAFSPE VIPMF</td>
<td>KF11 (p24)</td>
<td>1264.65</td>
</tr>
</tbody>
</table>

The result of the in-vitro kinase assay of these capsid peptides shows phosphorylation of Gag-B22 peptide when incubated with p38 MAPK. Furthermore, none of the other capsid
peptides in presence of any of the above predicted kinases were shown to be phosphorylated (result not shown). The time-dependent increase in phosphorylated peptide mass peak (2215.36Da) with mass difference of +80 \( m/z \) (represent covalent addition of HPO\(_3\)) from unphosphorylated peptide peak was observed. The mass spectra of the time-dependent kinase assay (1/2 hr, 1hr, 3hrs and overnight) are shown in Figure 3.6 below. Moreover, the phosphorylation of Ser33 residue by p38 MAPK indeed support the applied strategy as p38 MAPK was predicted by all three prediction tools. The consistency in the results between \textit{in-vitro} kinase study and predicted host-cell kinase/HIV-phosphorylation site interaction demonstrates the efficiency of our applied strategy.

\textbf{Figure 3.6} Mass spectra of Gag-B22 peptide (WVKVVEEKAFSPEVIPMF- 2134.12Da) in presence of p38 MAPK and ATP at different time points (1/2hr, 1hr, 3hr, and overnight) are shown here. Time-dependent increase in the mass peak at 2214.36Da shows the p38 MAPK-dependent phosphorylation of Gag-B22 peptide.
3.5 *In-vitro* kinase assay on HIV-proteins (Vif, Nef and CA)

The *in-vitro* kinase assays were performed on HIV-derived peptides to find out lead kinase responsible for phosphorylation of individual HIV-protein for the further study. Based on the presence of a significant phosphorylation in *in-vitro* kinase assay performed on HIV-derived peptides, we concluded to investigate following host-kinase/HIV-protein interactions: 1) Vif protein in presence of protein kinase C (PKC), 2) Nef protein in presence of protein kinase A (PKA), and 3) Capsid protein in presence of p38 MAPK. The *in-vitro* kinase assays on the HIV-proteins (Vif, Nef and P24) were performed following similar protocol by incubating them with lead kinase and ATP for 4 hrs. After incubation with kinase, the protein was subjected to SDS-PAGE gel followed by in-gel digestion by specific protease. The extracted peptides were directly subjected to peptide mapping and phosphorylation site analysis using MALDI-LTQ. Importantly, various proteases cleave at different sites, such as Trypsin is known to cleave C-terminal end of Lysine (K) and Arginine (R), chymotrypsin is known to cleave C-termini of phenylalanine (F), tryptophan (W) and tyrosine (Y), and Glu-C is known to cleave C-termini of Glutamic acid (E) residue in the protein sequence. Our efforts to characterize the phosphorylation interactions observed in the host kinase/HIV-peptides at the protein level have not shown any success yet. Various proteases such as Trypsin, Glu-C and chymotrypsin were employed (following in-gel digestion protocol suggested by vendor) to investigate phosphorylation sites in HIV-proteins.

3.6 Summary and future direction

In this project, the preliminary goal was to identify one or more phosphorylation sites across HIV proteins and study the influence of identified phosphorylation event on HIV infection. This study implemented phosphorylation predictions algorithms, not just to sort out the
list of the potential phosphorylation sites across various HIV proteins but also to identify the kinases with high prediction scores. The prediction results of these algorithms authenticate our approach by successfully predicting few known host-kinase/HIV protein interactions. Results of the in-vitro kinase assay performed on potential phosphorylation sites containing HIV-peptides have identified a few novel interactions between host-kinases/HIV phosphorylation substrates. These include the interactions of phosphorylation sites of Vif, Nef and Capsid proteins with protein kinase C (PKC), protein kinase A (PKA), and p38 MAPK respectively.

To achieve our projected aim, there are several steps can be taken in future. 1) Identification and evaluation of potential phosphorylation sites from HIV structural proteins (Gag, Pol, and Env). 2) Study the kinase assay of HIV proteins using specific kinases obtained from initial screening on HIV peptides. Although few attempts have been made so far, various other approaches like a utilization of different proteases for protein digestion, optimization of in-gel digestion protocol and in-solution digestion will be attempted. 3) The western-blot study of kinase treated HIV-protein using site-specific antibody will be performed to evaluate phosphorylation of HIV-protein. 4) Site-directed mutagenesis experiments of phosphorylation sites to Ala (Ser/Thr sites) and Phe (Tyr sites) will be performed to investigate effect of phosphorylation state on function or stability of the HIV-protein. 5) Biochemical effects of phosphorylation will be investigated by assessing the difference in activity of HIV-proteins as a function of phosphorylation state that will confirm biological relevance of identified in-vitro sites. This study will create a definitive database of host-cell kinase/HIV-phosphorylation substrates interaction and might provide better insight into intriguing cellular processes of HIV-1 virus life cycle.
Chapter-4

A) Introduction- Cysteine cathepsins

4.A.1 Introduction
4.A.2 Structure of cysteine cathepsin
4.A.3 Substrate binding
4.A.4 Localization, Regulation and Physiological role
4.A.5 Pathophysiology
4.A.6 Known cathepsin inhibitors

B) Development of cell-active cathepsin B inhibitors

4.B.1 Introduction
4.B.2 Library of cysteine cathepsin inhibitors
4.B.3 Result and Discussion
   4.B.3.1 Type of interaction and inhibition
   4.B.3.2 Computational analysis
   4.B.3.3 Selectivity
   4.B.3.4 Cell permeability
4.B.4 Conclusion

C) Development of highly potent, cell-active cathepsin L inhibitors

4.C.1 Introduction
4.C.2 Library of cathepsin L inhibitor
4.C.3 Synthesis
4.C.4 Result and Discussion
   4.C.4.1 Selectivity
   4.C.4.2 *In-vivo* cathepsin L inhibition
   4.C.4.3 Cell migration inhibition
4.C.5 Conclusion
Chapter-4 A

Introduction- Cysteine cathepsins
4.A.1 Background

Proteolytic enzymes, also called as proteases or peptidases catalyze the hydrolysis of various protein substrates. Proteases are found in almost all living organisms including mammals, viruses, bacteria, plants and fungi [153]. The hydrolysis of protein substrates depends on several factors including, protease specificity, accessibility of susceptible peptide bond in protein substrate, and endogenous protease inhibitors [154]. The human genome expresses more than 670 proteases, categorized in 4 distinct classes based on their catalytic residues: serine proteases (31%), cysteine proteases (25%), metalloproteases (33%), and aspartic proteases (4%) [155]. In addition to these, there are also threonine proteases and few newly evolved proteases with unknown mechanism. More information about various proteases including newly discovered proteases can be found on the online database, MEROPS [156].

In 1937, papain was the first cysteine protease isolated and characterized from *Carica papaya* [157]. The word “cathepsin” was derived from the Greek word kathepsein, which means “to digest” [158]. There are eleven members of cysteine cathepsin expressed in the human genome- cathepsin B, L, K, S, V, F, W, H, Z/X/P, C/J/dipeptidyl peptidase I, and O. The cysteine cathepsins play essential role in many physiological processes associated with the lysosome such as, like cell death, collagen degradation, protein degradation, and immune response [159, 160]. Aberrant expressions of cysteine cathepsins are responsible for many pathological disorders including, bone loss, autoimmune disease, metabolic syndrome and metastatic cancer [161-164]. As cysteine cathepsins can potentially serve as therapeutic targets, development of their inhibitors is crucial.
4.A.2 Structure of cysteine cathepsins

Cysteine cathepsins are generally expressed in preproenzymes form, which have three major compartments- a signal peptide, a propeptide and a catalytic domain [165].

![Figure 4.1- Cysteine cathepsins- Structural compartments of preproenzyme forms](image)

Signal peptides consist of 10-20 amino acid residues and are involved in translocation into endoplasmic reticulum during mRNA translation [166]. Propeptides, generated by removal of signal peptide, are of variable length such as, 36 amino acids in human cathepsin X and 251 amino acids in cathepsin F [167]. Propeptides are known to function- (1) as scaffolds for protein folding of the catalytic domain [166]. (2) as chaperones for the transport of proenzyme to lysosomal compartment following mannose-6-phosphate receptor (M6PR) pathway [166]. (3) as high-affinity reversible inhibitors to prevent premature activation of catalytic domain [168]. The length of catalytic domains of human cathepsins is 214-260 amino acid residues (Figure 4.1).
The first crystal structure among the cathepsins, that of cathepsin B, was determined in 1990 [169]. This was followed by determination of crystal structures of cathepsin L [170], K [171, 172], H [173], x [174], V [175], C [176, 177], S [178] and F [179]. The crystal structures of cathepsin O and W are still not yet determined. Structural elucidation of cathepsins using cathepsin L as a typical representative is shown in Figure 4.2.

**Figure 4.2** Crystal structure of cathepsin L (PDB-2YJC). The active site Cys25 and His163 are shown in ball-and-stick representation (purple color)

It is composed of two main domains, referred to as the left (L) domain and right (R) domain. The L domain consist of three α-helices. The longest central helix is over 30 residues long. The R domain is β-barrel forming a coiled structure. The active-site cleft formed at the interface of the both domains contains Cys25 (from L domain) and His163 (from R domain). These two catalytic residues form the thiolate-imidazolium ion pair that enhances nucleophilicity of the cysteine residue [180]. The propeptide binds to the active site cleft to prevent substrate binding [181, 182]. Evaluation of 3D-based sequence alignment of mature form of cysteine cathepsins exhibit the conservation of active-site residues- Cys25 and His163 (based on cathepsin L numbering) in addition to the residues interacting with main-chain of bound
substrate (Gln19, Gly68, and Trp183), the N-terminus Pro2 and certain Cys residues [183]. Although the three-dimensional structure of two human cysteine cathepsins, O and W, are still unresolved, a 3D sequence alignment performed independently using crystal structure of human cathepsin L as a template shows the same conservation pattern [184].

4.A.3 Substrate binding

Schechter and Berger [185], originally proposed a nomenclature for the positions of the substrate residues (P) and the subsites of cathepsin (S) to understand and elucidate the interactions between cysteine cathepsins and their substrates. The positions (P) and subsites (S) were numbered in both direction of the scissile bond such as position P1 and P1’. The non-primed side refers to the N-terminal part and primed side refers to C-terminal part of the substrate. Berger et al. observed that longer than 7 residues of peptides does not affect the kinetics of their degradation and concluded that there are seven substrate residues binding into seven subsites from S4 to S3’ as shown in Figure 4.3A [185]. Turk et al. revised nomenclature of substrate binding using insight provided by the crystal structures of cathepsin complexes with small-molecule and concluded five substrate residues bind to five subsites from S3 to S2’ as shown in Figure 4.3B [186].
Figure 4.3 A) Pictorial representation of substrate binding sites according to Schechter and Berger. Enzyme pockets have been assigned as “S” and substrate peptide binding to corresponding pockets have been assigned as “P”. B) Revised substrate binding to papain like cysteine proteases according to Turk. This demonstrate the active site cleft binding residues contribution from L-domain as well as R-domain.

The active-site cleft is composed of amino acid residues coming from L- and R-domains of cathepsins [187]. The L-domain provide the surface for the P3, P1 and P2’ residues, whereas the R-domain provide the binding surface for the P2 and P1’ residues [165]. The three-dimensional structures of cathepsin have shown the substrate binding to active-site cleft in extended conformation. The active-site cleft of cathepsins include catalytic residues Cys and His, located onto two different domains. Further detailed study performed by evaluation of superposition of the structures indicates that there are only three well-defined substrate binding sites, S2, S1 and S1’. Among these substrate binding sites only S2 site forms a pocket, on the other hand S1 and S1’ subsites only offer binding surface [189].

Depending upon the cleavage site, cathepsins have been further classified as endopeptidases, exopeptidases or both (Figure 4.1). An endopeptidase cleaves the peptide bonds
of nonterminal amino acids (i.e. within the molecule), in contrast to exopeptidases, which break peptide bonds from end-pieces of terminal amino acids. Among all the cysteine cathepsins, cathepsin K, S, L, V, and F exhibit predominantly endopeptidase activity, in which the active-site cleft extends along the entire length of the two domains [166]. On the other hand, the exopeptidase cathepsins such as H, X and C, [166] are able to dock the substrates through their N or C termini and further classified based on, A) the termini of the substrate, and B) substrate peptide length they cleave. Cathepsin C is an aminodipeptidase [190], cathepsin H is an aminopeptidase [191] and cathepsin X is a carboxypeptidase [192]. Cathepsin B can exert both endo and exopeptidase activity.

4.A.4 Localization, Regulation and Physiological Role

Cysteine cathepsins have long been believed to be found in lysosome and are responsible for normal cellular protein degradation and turnover. Recent studies reveal the presence of cysteine cathepsins also in nucleus, cytoplasm, and in extracellular space [193]. The majority of cathepsins including, cathepsins B, H, L, X, V, F, O, and C are ubiquitously expressed in human tissues. Contrary to, cathepsins K, S and W show a restricted and tissue specific distribution. Cathepsin K is highly expressed in osteoclasts, epithelial cells, and synovial fibroclasts in rheumatoid arthritis joints [194]. Although cathepsin K, L, S, V, and B possess elastolytic and collagenolytic activities, cathepsin K is the only enzyme whose role in bone resorption has been well recognized in humans and mice [195]. Cathepsin S is predominantly expressed in professional antigen-presenting cell (APCs) including B-cells, macrophages and dendritic cells [196]. Cathepsin W is predominantly expressed in CD8+ lymphocytes and natural killer (NK)
cells [197, 198]. Cathepsin V, a homologue of cathepsin L precisely expressed in thymus, testis and cornea [199].

The activity of cysteine cathepsins are mainly regulated by A) zymogen activation, B) pH, and C) through their endogenous protein inhibitors [200]. Cysteine cathepsins show optimal activity at slightly acidic pH while becoming unstable at neutral pH. Hence cathepsins are irreversibly inactivated at neutral pH in the extracellular compartment [201]. Cathepsin S is the only cysteine cathepsin which is stable and active at neutral pH [202], while cathepsin L is the most unstable among cysteine cathepsins at neutral pH [203]. In the presence of high ionic strength and organic solvent content, cathepsin B irreversibly loss the activity in addition to structural changes [204]. Moreover, endogenous cathepsins inhibitors such as stefins control the activity of cathepsins in intracellular compartment; on the other hand, cystatins and kininogens control the activity of cathepsins in extracellular compartment [186]. Hormones and inflammatory cytokines play a key role in regulation of cathepsins B, K, L and S expressed in macrophages, cardiovascular and valve cells [205].

These differential tissue distributions of specific cathepsins refer to their more explicit role in various physiological processes. For example, cathepsin K, B, V, S and L are responsible for degradation of various extracellular proteins such as laminin [206], fibronectin [206, 207], elastin [208, 209] and collagens [206, 210-212]. Cathepsin F, L, S and V are involved in antigen processing and degradation of MHC class II-associated professional antigen-presenting cells such as macrophages and dendritic cells [213-216]. The lysosomal cysteine cathepsins B, K, L, and H target the antiapoptotic molecules such as Bcl-2, Bcl-xL, and Mcl-1 in human cancer cells [217]. Cathepsin B, L, K, and S are involved in degradation of thyroglobulin at pH 7.4 and under
oxidative conditions [218]. Thus, recent studies revealed the role of cysteine cathepsins in various physiological processes beside protein degradation which include keratinocyte differentiation, prohormones activation, angiogenesis, neurobiology, bone remodeling, and immunological responses. [200, 219-222].

4.A.5 Pathophysiology

The over-expression and upregulation of cysteine cathepsins play key roles in many pathophysiological disorders, such as, neurological disorders, cardiovascular diseases, obesity, inflammatory diseases like rheumatoid arthritis, and cancer [223-228]. Under normal physiology, cysteine cathepsins activities are well-regulated in extracellular compartments by neutral pH, endogenous inhibitors including stefins, cystatins, kininogens, and thyropins, and reducing environment. However, acidification of extracellular compartments as well as reduction in concentration of endogenous inhibitors can lead to aberrant expression of cathepsins [166].

Increased expression of cathepsins L, K, S and V lead to degradation of extracellular matrix. Cathepsin K exhibit strong elastolytic activity at neutral pH and collagenolytic activity at slightly acidic pH [229]. While cathepsin V is known for its highest elastolytic activity among cysteine cathepsins [230]. Cathepsin B, S, L, K and H upregulation has been shown in mouse models of IL-13-stimulated chronic obstructive pulmonary disease [COPD] as well as in few other inflammatory conditions such as emphysema, silicosis, and sarcoidosis [231-233]. Several studies demonstrate the role of cathepsin L, S and K in obesity, which is an important risk factor for various cardiovascular and metabolic disorders [234, 235]. The effectiveness of cathepsin K inhibitors in the treatment of rheumatoid and osteoarthritis, atherosclerosis and other bone cancers indicate the critical role of cathepsin K in extracellular matrix degradation [236, 237].
While cathepsin S, B, K and L contribute to cartilage destruction and immune response, aberrant activity of these cathepsins can lead to immune disorder such as rheumatoid arthritis [238-241]. Genetic disorders such as Haim-Munk and Papillon-Lefèvre syndrome are associated with cysteine cathepsin C [242], and pycnodysostosis [243] is associated with cathepsin K.

The association of cathepsin B in tumor progression was first documented 35 years back [244]. Now many studies have shown the elevated cathepsin expression and/or activity associated with cancer progression in number of several types of tumors such as, breast, skin, ovarian, brain, pancreatic, colorectal [245, 246]. Cathepsins B, L and S promote the migration and invasion of tumor cells [247, 248] through ECM degradation and initiation of proteolytic cascade that activate urokinase-type plasminogen activator, matrix metalloproteinases and plasminogens [165]. E-cadherin, a vital adherens junction protein, required to maintain cell to cell adhesion in epithelial cell lines is targeted to proteolysis by cathepsins B, L and S [249]. Hence, loss of E-cadherin leads to cell invasion and metastatic behavior of the cancer cells. Thus, development of an effective inhibitor for the cysteine cathepsins can potentially reverse the functional loss of E-cadherin and can be anticipated to reduce the metastatic behavior of solid tumors [250].

4.A.6 Known cathepsin inhibitors

The discovery of epoxysuccinyl-based inhibitor E-64 was very important for research on cysteine cathepsin inhibitors. Thereafter, due to the association of cysteine cathepsins in various pathological disorders including cancer, several cysteine cathepsin inhibitors have been studied in last few decades. Currently, a few selective inhibitors for cathepsin B, L, K, and S are under drug-discovery process. The molecules containing different electrophilic reactive groups to
interact with active site (nucleophilic cysteine residue) of cysteine cathepsins have been studied so far. The different reactive groups which have shown inhibition of cysteine cathepsin activity are- acyloxymethyl ketone [251], vinyl sulfone [252], nitriles [253], azapenone [254], thiadiazole [255], carbazate [256], epoxide [257], cyclopropenone [258], and furanone [259].
Chapter-4 B

Development of cell-active cysteine cathepsin B inhibitors
4.B.1 Introduction

Cathepsin B is probably one of the most studied cathepsin from the cysteine cathepsin family, due to its role in tumor progression, being highly expressed in a variety of malignant tumors [260]. However, only few of cathepsin B inhibitors have made their way to clinical trials [261]. Hence, the continuation of development and evaluation of cathepsin B inhibitor as an effective anti-cancer agent is essential. A cell-permeable small molecule inhibitory agent which offers covalent and irreversible interaction with Cys25, an active site of cathepsin B, could serve as a lead chemotypes for development of anti-neoplastic agents.

This study was conducted at Dr. Sanjai Kumar’s lab (Queens College-CUNY). In our collaborative effort, we developed a library of compounds containing the sulfonyloxirane moiety and screened them to access their efficacy as a cathepsin B inhibitor in a time-dependent assay. A lead candidate evolved from this study, compound 5 (Figure 4.4), which was further evaluated using enzymology and mass spectrometry-based experiments to study the type of interaction involved between compound 5 and active site, Cys25, of cysteine cathepsin B. *In-vivo* studies including intracellular inhibition of cathepsin B as well as inhibition of cell migratory behavior of highly metastatic breast carcinoma MDA-MB-231 cells demonstrate the effectiveness of compound 5 as a cathepsin B inhibitor.

![Chemical structure of compound 5](image)

**Figure 4.4** Chemical structure of compound 5- a lead cathepsin B inhibitory molecule
4.B.2 Development of library of cathepsin B inhibitory agents

E-64, an epoxysuccinyl-based inhibitory agent, is known as a universal inhibitor of cysteine cathepsins. The structural investigation of the E-64-cathepsins complex assisted the development of selective cysteine cathepsin B inhibitors (e.g., CA030 and CA074) [262, 263]. The evaluation of the structures of CA030 and CA074 in complex with cysteine cathepsins revealed that there exist well-defined binding pockets, S2 and S1’, surrounding the S1 site (Schechter and Berger nomenclature) [183] that prefer aromatic hydrophobic groups [261, 263].

Therefore, we hypothesized that an arylsulfonyl-substituted oxiranyl moiety could serve as a small molecule covalent inactivator of cysteine cathepsins. This approach envisions that (a) an appropriate binding of inhibitory compounds within the active site of cysteine cathepsin will allow the reactive Cys residue to be entrapped by the activated sulfonyloxirane group, (b) the sulfonyl group will likely mimic the potential H-bonding interactions that the carbonyl group α to the epoxy moiety exhibits in the inhibited CA030-cathepsin B complex [262] and (c) the sulfonyl group will further enhance the electrophilicity of the oxiranyl carbon, compared to a carbonyl group, thereby facilitating the entrapment of the active site Cys residue.
Different moieties of sulfonyloxiranes were synthesized following different synthesis protocols-1) Reaction of corresponding aldehyde in presence of potassium hydroxide and chloromethyl phenyl sulfone to yield sulfonyl oxirane (Scheme 4.1A) [264].

**Scheme 4.1** Various synthetic approaches in the development of cysteine cathepsin inhibitors

2) Reaction of substituted thiophenol with 2-bromoethanol in presence of potassium hydroxide under reflux condition yield to arylthioethanol. The arylthioethanol is oxidized by hydrogen peroxide (30%) to provide 2-hydroxy arylsulfonylethane. Treatment of 2-hydroxy arylsulfonylethane with mesyl chloride in presence of triethylamine at 0°C leads to arylsulfonylethene. Further conversion of alkene intermediate to aryl sulfonyloxirane was carried out using n-BuLi in presence of tert-butyl hydroperoxide (Scheme 4.1B) [265]. 3) The aryl sulfonylthiirane molecules were synthesized according to a previously published protocol (Scheme 4.1C) [266].
4.B.3 Results and Discussion

Since overexpression of cathepsin B has been linked to breast, prostate, bladder, stomach, colon, cervix, ovary, bladder, lung, and thyroid cancers, cysteine cathepsin B was chosen as the initial target to evaluate the efficacy of synthesized inhibitory agents [267]. The initial screening was performed by incubating a fixed concentration of compounds 1–13 with cathepsin B under a pseudo-first order condition. A small aliquot of the incubation mixture was withdrawn at a fixed time interval to assess remaining cathepsin B activity in presence of cathepsin B substrate.

Compounds 1 and 2, containing substituent at the second carbon of the oxirane ring were ineffective in cathepsin B inhibition, in addition to, a thiirane analogue, 13, that potentially contains a better leaving group thiolate (compared to alkoxide in 12). This again indicated that proper placing between arylsulfonyl group and the reactive oxirane group within the cysteine cathepsin active site was critically important for its inhibitory effect. Hence, development of
cathepsin B inhibitory agents was then focused on synthesizing analogues of 2-(arylsulfonyl) oxiranes and their functional variants.

Among the synthesized cathepsin inhibitors, compound 5 was found to be most effective inhibitor of cysteine cathepsin B activity. Cathepsin B activity did not improve even after dilution of 5–inactivated cathepsin B complex indicated that the mechanism of inhibition was perhaps covalent and irreversible in nature (data not shown). To assess the inhibitory efficacy, inactivation experiments were performed at appropriate concentrations of compound 5 under pseudo-first order conditions. The experimental data thus obtained were analyzed using a simple two-step inhibition model:

\[
\text{E} + \text{I} \xrightleftharpoons{K_i} \text{E}^*\text{I} \xrightarrow{k_i} \text{E} - \text{I}
\]

where \(K_i\) represents the reversible equilibrium binding constant for the first step, and \(k_i\) is the first order inactivation rate constant for the second irreversible step. The pseudo-first order rate constants of inactivation (\(k_{obs}\)) were determined at appropriate concentration of 5 ([I] >> [cathepsin B]), and a Kitz-Wilson analysis was performed to fit the experimentally obtained data (Figure 4.6A). This procedure yielded the equilibrium binding constant (\(K_i = 86 \pm 3 \, \mu\text{M}\)) for the first equilibrium binding step, and the first order inactivation rate constant (\(k_i = 0.21 \pm 0.01 \, \text{min}^{-1}\)) for the second irreversible step.
Figure 4.6 A) Kitz-Wilson plot of 5-mediated cathepsin B inactivation kinetics. The data points are fitted to a line that yields (a) \( K_i = 86 \pm 3 \) µM for the first equilibrium step of inhibitor binding, and (b) \( k_i = 0.21 \pm 0.01 \) min\(^{-1}\) for the first order inactivation rate constant. B) The rate of 5-mediated cathepsin B inhibition is decreased in the presence of leupeptin (40 nM: blue line), a known competitive inhibitor of cathepsin B, compared to the control (5% DMSO: red line).

4.B.3.1 Type of interaction and inhibition

To determine the type of inhibition and inhibitory concentration of cathepsin B exerted by compound 5, time-dependent as well as concentration-dependent loss of cathepsin B activity in presence of compound 5 was studied (Figure 4.7). In order to be an active-site directed inhibitor, the rate of 5-mediated cathepsin B inactivation reaction postulated to be subsided in presence of a known active site-directed cathepsin B inhibitor, leupeptin (\( K_i = 5 \) nM) [268]. Furthermore, alleviation of inhibition of cathepsin B by compound 5 in presence of leupeptin was observed, indicating that the compound 5 is an active-site directed irreversible inhibitor (Figure 4.6B).
Figure 4.7 A) Progress curves indicating a time-dependent loss of cathepsin B activity. Compound 5 was incubated with cathepsin B under pseudo-first order condition, and the progress curves were obtained at appropriate times of inactivation (red: 0 min, blue: 11 min, green: 17min, black: 25 min and pink: 41 min) using a chromogenic substrate Z-RR-para-nitroanilide. B) Concentration-dependent loss of cathepsin B activity with time. The points are experimental, and the lines joining them are fitted to obtain the pseudo-first order inactivation rate constant ($k_{obs}$).

Furthermore, matrix-assisted laser desorption/ionization (MALDI) - mass spectrometry was also employed to study interaction between compound 5 and cathepsin B. Compound 5-inhibited cathepsin B complex displayed a mass peak increment of 212.1 amu in singly charged state compare to the mass of cathepsin B itself (control). This increment in mass was consistent with the molecular weight of compound 5; indeed, represented the attachment of compound 5 with cathepsin B protein (Figure 4.8). No further mass increase was observed even after prolonged incubation of 5 with cathepsin B. This result demonstrated that the mechanism of compound 5-mediated cathepsin B inhibition is covalent and irreversible in nature.
Figure 4.8 A mass spectrometric analysis of 5- inhibited cathepsin B complex. A mass difference of 212.1 amu, consistent with molecular weight of 5 (212.1), is observed in singly charged 5-inhibited cathepsin B complex (lower panel) compared to uninhibited cathepsin B enzyme (upper panel). This experiment demonstrated that the mechanism of 5-mediated cathepsin B inhibition is covalent and irreversible in nature.

4.B.3.2 Computational Analysis-

The computational modeling studies were conducted to understand molecular interactions of compound 5 with cysteine cathepsin B by making required modification to CA073-cathepsin B complex. (PDB id- 2DC6). A molecular dynamic simulation (100 ps) was performed on the compound 5- cathepsin B complex and the trajectories of the molecular dynamics simulation were analyzed. There were few interesting interactions noted between compound 5 and active site region of cathepsin B (Figure 4.9). The O3 of sulfonyl oxygen formed a H-bond (3.4 Å) with the side chain of Gln23-N_e-a key oxyanion hole residue. Consequently, its O2 atom was engaged in a robust hydrogen bond interaction (2.8 Å) with a tightly-bound active site water molecule (Wtr12).
Figure 4.9 Stereoscopic view of the energy-minimized structure of the covalently modified 5-cathepsin B complex. The low pKa active site Cys29 residue is the site for 5-mediated covalent modification. For clarity, only the inhibitor 5 (yellow) and the key cathepsin B active site residues are shown. The key h-bonding distances of 5 with cathepsin B are shown in green. The intramolecular h-bonding interactions among cathepsin B active site residues near the inhibitor binding are depicted in red (heavy atom h-bond donor and acceptor atoms cut off is 3.6 Å). The position of primed and non-primed binding pockets on the cathepsin B surface is shown in pink text with underline.

The phenyl moiety of the inhibitor was found to be involved in a hydrophobic interaction with a non-polar patch containing Met196, Gly197, and Gly198 residues of S2 pocket. In addition, the ortho-ethyl group interacted with CH–π interactions with the aromatic side chains of Trp221 and His110 from S2’ pocket, albeit weakly [260]. Our modeling study suggests that ortho-substituted aryl groups containing a negatively charged group could prove to be selective inhibitory agents towards cathepsin B. Notably, cathepsin B owned to its unique structural feature that contains a flexible occluding loop, residues 104–126, potential interaction of inhibitory agent with His110 and/or His111 could result in enhanced selectivity and efficacy towards cathepsin B.
4.B.3.3 Selectivity

The inhibitory efficacy of compound 5 towards cathepsin B was tested in presence of an external nucleophile, azide ion (1mM). The absence of substantial changes in inhibitory efficacy indicated that compound 5 is inert to external nucleophiles and covalently interacts with the active site of cathepsin B. To assess the selectivity of compound 5, various enzymes from cysteine cathepsins family were selected (Table-4.2).

Table 4.2 Selectivity of inhibition of compound 5 towards cysteine cathepsins as assessed by the 2\textsuperscript{nd} order enzyme inactivation rate constants.

<table>
<thead>
<tr>
<th>Enzyme (Enzyme Family)</th>
<th>2\textsuperscript{nd} Order Inactivation Rate Constant (M\textsuperscript{-1} Min\textsuperscript{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cathepsin B (Cysteine Cathepsin)</td>
<td>2441 ± 202\textsuperscript{a}</td>
</tr>
<tr>
<td>Cathepsin F (Cysteine Cathepsin)</td>
<td>1300 ± 208\textsuperscript{b}</td>
</tr>
<tr>
<td>Cathepsin S (Cysteine Cathepsin)</td>
<td>1196 ± 96\textsuperscript{b}</td>
</tr>
<tr>
<td>Cathepsin D (Aspartyl Cathepsin)</td>
<td>65 ± 5\textsuperscript{b}</td>
</tr>
<tr>
<td>Cathepsin G (Serine Cathepsin)</td>
<td>NF</td>
</tr>
<tr>
<td>Trypsin (Serine Protease)</td>
<td>NF</td>
</tr>
<tr>
<td>hPTP1B (Protein Tyrosine Phosphatase)</td>
<td>NF</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Calculated from the $k_i$ and $K_i$ parameters obtained using Kitz-Wilson Analysis. \textsuperscript{b}Determined by measuring pseudo-first order rate constant of inactivation $k_{obs}$ first and dividing the $k_{obs}$ by the inhibitor concentration. \textsuperscript{c}NI: No apparent time-dependent inhibition even at concentration as high as 1mM during a 1-hour inactivation reaction.

The compound 5 inhibit cathepsin B 1.9-fold more selectively than the other two closely related cysteine cathepsin F and cathepsin S. Also, cathepsin D (aspartyl cathepsin) was weakly inhibited by compound 5, i.e.,1/38\textsuperscript{th} fold less compare to cathepsin B. The other enzymes that
utilize serine as a catalytic residue including cathepsin G and trypsin protease did not exhibit any signs of inhibition by compound 5. Previous reports suggested that 2-(arylsulfonyl)-oxiranes are modest inhibitors of protein tyrosine phosphatases (PTPs) [265]. Hence, the inhibition of prototypical PTP, hPTP1B, by compound 5 was evaluated but, PTPs was not inhibited by compound 5. Apparently, the presence of an ethyl group at the ortho position of 2-(phenylsulfonyl) oxirane is significantly favored the strong and effective interactions with at the cysteine cathepsin active site. However, it is essential to decipher the three-dimensional structure of compound 5- cathepsin B complex to develop future inhibitory agents based on this moiety.

4.B.3.4 Cell permeability

Due to the presence of electron withdrawing, sulfonyl moiety in adjacent to oxirane ring, it was essential to evaluate the stability of compound 5 at physiological pH before it’s evaluation as a cell-permeable molecule. An NMR- based experiment was performed to evaluate the half-life of compound 5 at physiological pH of 7.4. The result from this study indicated that compound 5 is relatively stable at physiological pH of 7.4, as half-life of compound found to be 3.6 hour (Figure 4.10) and ultimately can be evaluated for in-vivo applications.

Many of the recent reports suggested the presence and vital role of cysteine cathepsins in the compartments beyond lysosomes such as the cytoplasm, nucleus, mitochondrial matrix and plasma membrane [270-272]. For example, overexpression and an enhanced proteolytic activity of cytoplasmic cysteine cathepsins is responsible for tumor progression [224], while successful inhibition of cysteine cathepsins decelerate the tumor growth [273]. These reports highlight the quest for non-peptidyl and cell permeable small molecule inhibitors for cysteine cathepsins.
Figure 4.10 Estimation of the half-life of compound 5 at physiological pH 7.4. The relative peak intensity of 5 was plotted against time and data was fitted to a signal exponential. The $t_{1/2}$ was estimated to be 3.6 hours.

To assess the intracellular efficacy of compound 5 against cathepsin B, the metastatic MDA-MB-231 human breast cell line was chosen since cathepsin B activity was known to be upregulated in these cells [274]. The efficiency in proteolysis of the well-established cathepsin B substrate, Z-RR-AMC ($k_{cat}/K_m = 105 \text{ M}^{-1} \text{ s}^{-1}$) by cathepsin B, was evaluated to confirm the intracellular inhibitory efficacy of compound 5 [275, 276]. The result clearly implies the significant loss in cathepsin B proteolytic activity in the cell treated with compound 5 compare to control MDA-MB-21 cells (Figure 4.11A). Furthermore, dose-dependent inhibition of cathepsin B by compound 5 ratify not only the result from previous experiment but also cell-permeability of compound 5 (Figure 4.11B).
Figure 4.11 Effect of inhibitor 5 on cathepsin B activity in breast cancer MDA-MB-231 cells. A) Cells were incubated with vehicle (DMSO 0.1%) and 20µM of 5 for 20 min, followed by an additional 10 min of incubation with the substrate Z-RR-AMC (25µM net) of 5. Images show a significant loss of signal intensity (green) from the enzymatic turnover of the fluorophore cathepsin B substrate Z-RR-AMC (shown by red arrow in the bottom panel) at 20µM of 5. The control experiment without 5 indicates that a high level of substrate turnover occurs in these cells, thereby indicating a robust cathepsin B activity (white arrow - top panel). Scale bar= 10µm. B) The dose-dependent inhibition of intracellular cysteine cathepsin B activity with inhibitor 5 as measured by average fluorescence intensity over the entire cell. Note: The fluorescent intensity values are mean ± S.D. calculated over a population of at least 5-8 cells for each condition. *Statistical significance was evaluated by the Student’s t-test.

Additionally, the inhibitory efficacy of compound 5 was also investigated using a proteolytic protein target of cathepsin B and S, namely E-cadherin [249]. E-cadherin plays a vital role in cell adhesion and cell morphology through maintaining intercellular contacts [277-279]. Complete or partial loss of E-cadherin occurs as tumor progressively becomes malignant [280-282]. This observation directly associates the loss of E-cadherin to invasive migratory behavior of cancer cells [283, 284].

Due to the presence of low-level of E-cadherin in MDA-MB-231 breast cancer cells [282], this cell line was chosen to investigate the change in the level of E-cadherin using E-cadherin antibody in presence and in absence of compound 5. As anticipated, a dose-dependent increase in the levels of E-cadherin was observed in a population of 5-treated MDA-MB-231 cells, as compared to control cells treated only with DMSO (Figure 4.12A and 4.12B). Although,
the overall cellular morphology of MDA-MB-231 cell line has changed upon treatment of compound 5 (Figure 4.11A and 4.12A), further experiments will be required to understand change in morphology at molecular level. A Western blot analysis was conducted to evaluate the E-cadherin level in compound 5- treated MDA-MB-231 cells (data not shown). Collectively, these studies demonstrated inhibitory intracellular efficacy of compound 5 towards cysteine cathepsin B in the invasive MDA-MB-231 breast cancer cells.

Figure 4.12 A physiological substrate of cysteine cathepsins, E-cadherin, enriched in 5-treated MDA-MB-231 cells. A) E-cadherin expression levels as detected by immunofluorescence in MDA-MB-231 cells. Cells were treated with control (0.1% DMSO) and cysteine cathepsin B inhibitor 5 (20µM) for 20 min, fixed, and then stained for E-cadherin. Bright field and merged fluorescence images of E-cadherin (red) and nuclear stain Hoechst 33342 (blue signal) are shown. Green and yellow arrows show enriched E-cadherin level (red) near the cell periphery (green arrow) and perinuclear region (yellow arrow) respectively in 5-treated cell, compared to control (0.1% DMSO). Scale bar = 10µm. B) A dose-dependent increase of E-cadherin expression levels from immunocytochemistry as measured by average fluorescence intensity over the cell-periphery. The fluorescence intensity values are Mean ± S.D. calculated over a population of 28 cells for each condition. *Statistical significance was evaluated by the Student’s t-test.

Moreover, to demonstrate the functional relevance of cysteine cathepsin B inhibition by compound 5, reduction in cell migratory behavior in metastatic MDA-MB-231 breast cancer cells was evaluated. The cell motility assay was performed in presence of compound 5 (10 µM
and 20 µM) and in absence of compound 5. The result from this assay support that the compound 5 effectively inhibit the cell migratory behavior of MDA-MB-231 cells in dose-dependent manner (Figure 4.13).

**Figure 4.13** The effect of cysteine cathepsin inhibitor 5 on the cell-migratory behavior of metastatic human breast cancer MDA-MB-231 cells. The cell migration phenotype was reduced upon treatment of cells with compound 5 (10 and 20µM), compared to control (0.05% DMSO). Each reported change in area value is the average of triplicate measurements with the corresponding standard deviation value.

### 4.B.4 Conclusion

We have developed a new class of non-peptidyl and cell-permeable small molecule inhibitors of cysteine cathepsin B. The covalent interaction between compound 5 and cysteine cathepsin B was evaluated using mass spectrometric analysis. The inhibitory efficacy of compound 5 was found to be selective towards cysteine cathepsin B. The computational analysis utilized to demonstrate interaction of compound 5 with the active site Cys residue in addition to surrounding S1 and S2 pockets. The inhibitor 5 was shown to be relatively stable to hydrolysis, cell-permeable, and effective in inhibiting intracellular cysteine cathepsin B activity in metastatic
human breast carcinoma MDA-MB-231 cells. Finally, effectiveness of compound 5 in reducing the migratory behavior of the metastatic MDA-MB-231 breast cancer cell lines was also demonstrated. This class of inhibitory agents could provide the platform for development of activity-based probe of cysteine cathepsins in future.
Chapter-4 C

Development of highly potent, cell-active cathepsin L inhibitors
4.C.1 Introduction

Cathepsin L is a ubiquitously expressed lysosomal endopeptidase that is primarily involved in several physiological processes such as cell cycle regulation, immune response, and development of natural killer cells [285-287]. Overexpression and aberrant activity of cathepsin L has been reported in many instances in various human malignancies such as breast, ovarian, prostate, colon, head and neck carcinoma, lung and gastric cancers [288]. Cathepsin L is known to be responsible for degradation of various extracellular matrix (ECM) such as collagen, laminin, and fibronectin and subsequently promote metastatic behavior of cancer. Due to elevated level of cathepsin L along with its vital functionality in cancer progression crafts cathepsin L, as a potential therapeutic target for development of anti-cancer treatment [289].

To date, the role of cathepsin L is remained largely unknown at cellular context. Moreover, the contradiction in functions of cathepsin L has been reported in few instances. For example, Di Piazza et al. found that increased glioma cell death was associated with accumulation of cathepsin L in the cytosol and established the pro-apoptosis role of cathepsin L [290]. By contrast, Gocheva et al. suggested that reduced cathepsin L expression increases apoptosis [249, 288]. In addition, previous investigations also suggested role of cathepsin L in both bone growth [291] as well as bone loss processes [292]. According to recent studies, development of a suitable activity-based cathepsin L probe can be a powerful tool to decipher the role of cathepsin L at cellular level [293]. Hence, we opted to develop a inhibitor that possess specific characteristics such as cell permeability, selectivity towards cathepsin L, covalent irreversible mode of inhibition, and non-basic nature for homogenous distribution in cell and has potential to be developed as a activity-based cathepsin L probe in this report.
4.C.2 Library of Cathepsin L inhibitory agents

Interactions of inhibitory agents with both the active-site as well as with surrounding pockets (i.e. S1, S2, S1’, S2’) of cathepsin L are essential for potency and selectivity. Previous studies including molecular docking of gallinamide A accompanied by X-ray crystallography of cathepsin L revealed several strategically important sites of interactions [294-296]. In another related study, it was noted that an appropriate position of an aromatic moiety could perhaps form more effective inhibitory interactions with the S’ residues of cathepsin L [297]. In addition, a visual analysis of a covalent and reversible inhibitor, Z-Phe-Tyr-(O-tert Bu)-C(O)CHO, bound to cathepsin L indicated that a non-polar cluster from the S’ pocket, involve Trp189, Leu144, Ala138, and Gly139 residues, formed critically important binding interactions with the tert-butyl group [296]. Hence, we sought to first develop a ligand that contain an electrophilic warhead and an aromatic group which is suitable to interact with the prime site residues of cathepsin L active-site.

The vinylsulfonate ester moiety was preferred over vinylsulfone due to the following reasons: 1) The vinylsulfonate ester moiety has never been investigated as electrophilic warhead in contrast to vinylsulfone [252, 298]. 2) The vinylsulfonate ester moiety displays a far superior reactivity as a Michael acceptor than vinyl sulfones [299]. 3) we hypothesized that that an additional S–O bond present in vinylsulfonate ester (but not in vinylsulfone) could rightly place the functionalized aryl group into the S’ pocket of cathepsin L for optimal interaction. Subsequently, the determined inhibitory efficacy of phenyl vinylsulfonate towards cathepsin L was 3-fold higher than that of phenyl vinylsulfone. This result supports the strategy of targeting the S’ pocket and conclusively led to development and screening of various aryl vinylsulfonate
ester molecules. (Table-4.3). The 4-bromophenyl vinylsulfonate ester (Compound 1) was identified as the most effective aryl vinylsulfonate ligand.

**Table-4.3 List of all inhibitory agents containing vinylsulfonate ester moiety**

<table>
<thead>
<tr>
<th>Compound</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cbz-Phe-Leu-CONH₂</td>
<td>[Figure 4.14]</td>
</tr>
</tbody>
</table>

Torkar *et al.* [251] recently reported Cbz-Phe-Leu-CONH₂ (Figure 4.14) as a modest inhibitor (Ki = 14 mM) of cathepsin L. This compound is nonbasic and inhibits cathepsin L activity by presumably occupying the non-prime pockets (i.e. S1, S2, and S3). We postulated that a hybrid design that incorporated the key features of compound 1 and Cbz-Phe-Leu-CONH₂ could render a highly potent and selective inhibitor of cathepsin L. Thus KD-1 was conceived (Figure 4.14).
A hybrid-design approach in the development of highly potent time-dependent inhibitor (KD-1) of human cathepsin L. To generate KD-1, a competitive inhibitor Cbz-Phe-Leu-CONH$_2$ of the modest potency ($K_i = 14\mu$M), known to target S1, S2, and S3 pockets of cathepsin L, is conjugated to a prime site (S') - targeting inhibitory ligand 1 at its C-terminal.

4.C.3 Synthesis

Inhibitor KD-1 was synthesized using the following scheme: N-Boc- L-Leucinal (I) was synthesized by oxidation of N-Boc-L-Leucinol which was utilized for the next step without any further purification [300]. The compound (II) was synthesized from ethyl methanesulfonate by treating with diethyl chlorophosphate in presence of butyllithium and anhydrous [301]. Reaction of N-Boc-L-Leucinal (I) with compound (II) in presence of sodium hydride and anhydrous THF yielded Boc-protected leucine vinylsulfonates (III). In the following step deprotection of Boc group from compound (III) was conducted using 25% TFA and dichloromethane. The free amine synthesized from this step was coupled to commercially available Cbz-protected phenylalanine in presence of coupling agent, HCTU, to generate compound (IV). Treatment of compound (IV) with tetrabutyl ammonium iodide in acetone led to sulfonate salt which was utilized for the next step without any further purification. The sulfonate salt, thus obtained, was converted to sulfonyl...
chloride in presence of triphenylphosphine. The intermediate sulfonyl chloride was treated with 4-bromo phenol in presence of triethyl amine to obtain final compound KD-1.

4.C.4 Result and Discussion

Assessment of potency of KD-1 was performed by measurement of IC$_{50}$ value. An IC$_{50}$ value of 3.6 ± 0.1 nM was observed, indicating excellent binding of KD-1 to cathepsin-L. Additionally, a dose-dependent experiment was conducted at various concentrations of KD-1 in the presence of cathepsin L, and its substrate and the progress curves were generated under pseudo-first order conditions ([KD-1]>>>[Cathepsin L]) (Figure 4.15). Analysis of the progress curve [251] revealed near diffusion-controlled inactivation kinetics (second order inactivation rate constant (k$_{inac}$) = 4.3 x 10$^6$ M$^{-1}$s$^{-1}$). Irreversible inhibition of cathepsin L by KD-1 was concluded due to lack of upturn of enzyme activity even after large dilution of inactivated reaction mixture (data not shown).
Figure 4.15 Concentration-dependent inhibition of cathepsin L by KD-1. Progress curves for cathepsin L-catalyzed hydrolysis of fluorescent substrate, Z-Phe-Arg-(7-amino-4-methylcoumarin, in the absence and presence of appropriate concentration of KD-1.

4.C.4.1 Selectivity

Various cysteine cathepsins such as, L, K, B, S, and H were chosen, in addition to, an aspartyl cathepsin (cathepsin D), a serine cathepsin (cathepsin G), a protein tyrosine phosphatase (hPTP1b), and a serine protease (trypsin) to evaluate the selectivity of KD-1 inhibitory molecule. The second order inactivation rate constants were determined by conducting inactivation kinetics under pseudo-first order condition. The derived values of the second order inactivation rate constants ($k_{intact}$) for different enzymes as mentioned in Table-3.5 clearly indicate the selectivity of KD-1 molecule towards cathepsin L. The measured relative selectivity factor (RSF) indicated 13 and 100-fold selectivity towards cathepsin L compare to the two most closely homologous cysteine cathepsins S and K, while 44,000-fold selectivity compare to cathepsin B. Furthermore,
no reactivity was observed against other enzymes. Owning to this remarkable selectivity, KD-1 was chosen for further evaluation as a cathepsin L inhibitory agent.

4.C.4.2 *In-vivo* cathepsin L inhibition

To evaluate KD-1, as a cathepsin L inhibitory agent *in-vitro* experiment including cathepsin L and collagen type I performed in absence and in presence of KD-1. The results show

Table 4.4 Second order enzyme inactivation rate constant ($k_{\text{intact}}$) of cathepsin L inhibitor KD-1 against a panel of closely-related enzymes.

<table>
<thead>
<tr>
<th>Enzyme Family</th>
<th>Enzyme</th>
<th>$k_{\text{intact}}$ (M$^{-1}$s$^{-1}$)</th>
<th>RSF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cysteine Cathepsins</td>
<td>Cathepsin L</td>
<td>4.3x10$^6$</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Cathepsin K</td>
<td>4.4x10$^4$</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Cathepsin B</td>
<td>96</td>
<td>44,000</td>
</tr>
<tr>
<td></td>
<td>Cathepsin S</td>
<td>3.5x10$^5$</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Cathepsin H</td>
<td>NI*</td>
<td>NA**</td>
</tr>
<tr>
<td>Aspartyl Cathepsin</td>
<td>Cathepsin D</td>
<td>NI*</td>
<td>NA**</td>
</tr>
<tr>
<td>Serine Cathepsin</td>
<td>Cathepsin G</td>
<td>NI*</td>
<td>NA**</td>
</tr>
<tr>
<td>Protein Tyrosine Phosphatase</td>
<td>lhtp1b</td>
<td>NI*</td>
<td>NA**</td>
</tr>
<tr>
<td>Serine Protease</td>
<td>Trypsin</td>
<td>NI*</td>
<td>NA**</td>
</tr>
</tbody>
</table>

Relative Selectivity Factor (RSF) = $k_{\text{intact}}$ (Cathepsin L) / $k_{\text{intact}}$ (other enzyme); NI* = No Inhibition at 0.1mM of KD-1 in 1 hour; NA** = Not Applicable.
the reduction in collagen type I degradation in presence of KD-1 (Figure-4.16A). Furthermore, to evaluate the cell permeability and *in-vivo* cathepsin L inhibitory activity of KD-1, metastatic breast cancer MDA-MB-231 cell line was chosen owing to its known characteristic of overexpression of cathepsin L [302].

A cell permeable fluorescent substrate of cathepsin L, Z-Phe-Arg-(7-amino-4-methylcoumarin), was chosen to evaluate *in-vivo* cathepsin L inhibitory efficacy of KD-1. The loss of the green fluorescent emerging from enzymatic turnover of the substrate by cathepsin L, indicate successful inhibition of intracellular cathepsin L activity by KD-1 (Figure 4.16B). This experiment confirmed the ability of KD-1 to inhibit cathepsin L activity at intracellular environment.

![Figure 4.16 A) Inhibitor KD-1 protects a physiological substrate of cathepsin L, human collagen type I, from proteolytic degradation *in vitro*. The α, β, and γ bands of collagen I are shown by blue arrow (left). B) Effect of KD-1 on cathepsin L activity in a metastatic breast cancer cell line. MDA-MB-231 cells were incubated with (a) vehicle (DMSO 0.05%), and (b) KD-1 (750nM) overnight (22hours). Intracellular inhibition of cathepsin L activity was assessed by exposing the cell to a cell-permeable fluorogenic substrate, Z-Phe-Arg-(7-amino-4-methylcoumarin) (2 minutes). The signal intensity (green) that emerges from the enzymatic turnover of the fluorophore substrate (white arrow) is lost when inhibitor KD-1 is present. This experiment demonstrates that KD-1 is cell permeable and is inhibiting intracellular cathepsin L activity in these cells. The results are representative of 3 independent experiments. Scale bars = 25μm.](image-url)
4.C.4.3 Cell migration inhibition

E-cadherin, an adherens junction protein vital for maintaining intercellular contacts and cellular morphology, is a proteolytic target of cathepsin L [249, 27-279]. The metastatic breast cancer MDA-MB-231 cell line express very low level of E-cadherin due to overexpression cathepsin L [282]. Hence to further evaluate the inhibitory efficacy of KD-1towards cathepsin L, the motility of MDA-MB-231 cells was determined (Figure 4.17). The substantial decrease in

Figure 4.17 A) KD-1 reduces the migratory potential of highly metastatic MDA-MB-231 breast cancer cells, as assessed by the wound healing assays. MDA-MB-231 cells were grown overnight in a 6-well plastic dish before treating with DMSO (control, 0.05%) or KD-1 (750nM). After creating the wounds by drawing a pipet across a confluent field of cells, images were acquired at 0 and 22 hours. Representative images of three independent experiments are shown here. The dotted yellow lines indicate the edge of the wound at time zero, and the heights of the yellow arrows are indicative of wound width. B) The wound closure, a direct measurement of cell migratory potential, was calculated using the following equation: % wound closure = [(A₀ - A₂₂)/A₀] *100%, where A₀ is the area of the wound measured immediately after the scratch, and A₂₂ is the area of wound after 22 hours. The figure represents a statistical analysis from three independent experiments using the unpaired Student’s t test (x, p<0.0001). The data were evaluated as the mean ± 1 standard deviation. The image analysis was performed by ImageJ software.
motility in KD-1 treated cell was observed. This result indicates enhanced integrity of cell-cell junctions due to maintenance of appropriate level of E-cadherin as a result of efficient cathepsin L inhibition by KD-1.

4.C.5 Conclusion

KD-1 is a highly potent, selective and cell-active inhibitory agent of cysteine cathepsin L. The inhibitory efficacy of KD-1 was found to be selective towards cysteine cathepsin L compare to the other closely related cysteine cathepsins such as cathepsin S, and K. The inhibition of collagen type I degradation by cathepsin L in presence of KD-1 in MDA-MB-231 cell indicate the suitability of KD-1 in biological context. Moreover, effectiveness of KD-1 in reducing the migratory behavior of the metastatic MDA-MB-231 breast cancer cell lines was also demonstrated. This inhibitory agent could provide the platform for development of activity-based probe of cysteine cathepsin L and help to understand many undefined roles of cathepsin L.
Contribution to the science community

My dissertation work emphasizes on to study covalent modification of different proteins and development of quantification method for small biological molecules using MALDI-Ion Trap mass spectrometry. Although using advance technologies, expensive softwares and high-end instruments, one can generate high quality data in a short-time, but it is less common to have these benefits at most of the research facilities and labs. The ultimate goal of this dissertation is not to compete against the advance instruments but to demonstrate qualitative and quantitative capability of MALDI- Ion Trap mass spectrometry and to provide viable alternative to the labs and facilities with limited resources.

In this dissertation, we have successfully demonstrated the application of MALDI-MS to identify and characterize various covalent modifications of proteins including phosphorylation and irreversible inhibition of host-enzymes using small molecules. In comparison to widely used LC-MS or LC-MS/MS, MALDI-MS mainly generates singly charged ions and therefore, interpretation of the mass spectra is simple. Furthermore, MALDI-MS exhibits a moderate tolerance to salts and other contaminants. Owing to these advantages, MALDI-MS offer simple and robust platform for the analysis of complex biological sample mixtures. The aim of the research project on HIV protein phosphorylation was to identify and characterize well conserved, potential phosphorylation sites of HIV proteins and to study the effect of phosphorylation in HIV infection and replication. This dissertation has successfully created database demonstrating interactions between HIV-auxiliary proteins/host-kinases. We strongly believe that this database will provide excellent platform to understand HIV cellular mechanism and may assist to identify novel strategic targets to develop anti-HIV therapeutic agents.
The current dissertation also introduces a novel MALDI-MS\(^2\) based quantification methods for small biological molecules. In this work, we demonstrated that even though MALDI-MS has limited applications for quantitative analysis, it certainly offer few advantages over current quantification methods using LC-MS or LC-MS/MS. This novel MALDI-MS\(^2\) based quantification method is not only cost-effective but also completes the analysis in short time with similar sensitivity. Additionally, as MALDI-MS has moderate tolerance towards salt and contaminants, this method can be applied to quantify the biomolecules in complex backgrounds such as serum and urine with prior purification or concentration of the sample.
References


36. Mirgorodskaya OA, Kozmin YP, Titov MI, Körner R, Sönksen CP, Roepstorff P. Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass


51. Prosser IW, Stenmark KR, Suthar MA, Crouch EC, Mecham RP, Parks WC. Regional heterogeneity of elastin and collagen gene expression in intralobar arteries in response to


88. Ho HP, Rathod P, Louis M, Tada CK, Rahaman S, Mark KJ, Leng J, Dana D, Kumar S, Lichterfeld M, Chang EJ. Studies on quantitative phosphopeptide analysis by maldi mass


