Characterization of the MHC Peptidome and Identification of O-GlcNAcylation Sites by Mass Spectrometry

Amanda Sue Wriston

Oak Hill, West Virginia

B.S., Chemistry, West Virginia Wesleyan College, 2010

A Dissertation presented to the Graduate Faculty

of the University of Virginia in Candidacy for the Degree of

Doctor of Philosophy

Department of Chemistry

University of Virginia

May, 2015

_____

_____

_____

_____

_____

Dedicated in memory of my mom and in honor of my dad

I would like to preface this by saying that no words I can come up with will justly express my sincere thanks and gratitude to the people listed below, as I have run out of things to say after writing this dissertation.

There are many people that have made this dissertation possible. Firstly, I would like to thank the good Lord for giving me ability and the wits to make it through graduate school. I would like to thank Dr. Donald F. Hunt for giving me the incredible opportunity to work in his lab. It was his passion for research that brought me here to UVA and I couldn't have asked for a better boss. Secondly I owe a huge amount of gratitude to Dr. Jeffrey Shabanowitz, who I have learned a great deal from and who keeps things running in the lab and makes sure that we are doing quality research. I still have to pinch myself to make sure that working with these two pioneers in mass spectrometry is not a dream. Jeff and Don, thank you both for the opportunity to learn from you and work with you.

I owe a huge debt of gratitude to Dr. Dina Bai for everything that she does in the lab. She goes above and beyond her title of programmer. I'm not sure how I would have gotten through the dissertation writing process without her pep talks and support and her proofreading my dissertation. A simple thank you does not seem like enough, but it'll have to do.

I would also like to express thanks to Dr. Michelle English who took me under her wing and helped me immensely in the lab and collaborated on projects with me. I doubt that I would have gotten to this point without her help, guidance, and

encouragement. Michelle you were right it eventually all worked out. Thank you for being a great mentor and friend.

The foundation for my love of chemistry began in my high school chemistry class taught by Mrs. Vicki Lickliter. Her excitement and passion for chemistry was passed on to me through her teaching. I am most appreciative and grateful for her dedication and giving me a strong base in chemistry. I would also like to think all of my undergraduate professors who strengthened my foundation, especially Dr. Ed Wovchko.

I am indebted to all of my collaborators, without whom, I would have nothing to write about. I have had the privilege of working with some great scientists on exciting projects.

Thank you to my committee members for Dr. Cameron Mura, Dr. David Cafiso, Dr. Charles Grisham, Dr. Ken Hsu, and Dr. Kenneth Tung. I appreciate you reading my dissertation and your guidance through proposal.

To all the members of the Hunt lab during my tenure, thank you. To Dr. Stacy Malaker, ½ of the 2 best friends that anyone could have, you went from mean girl to one of my best friends. I couldn't imagine spending the last four years with anyone but you. Thank you for always making me laugh and for making me do all your chores. Jk. To Paisley Trantham, Pai Pai, I was so glad when another southerner joined the lab. You have become one of my best friends and I'm really going to miss you next year. You always keep me laughing and make things more enjoyable in the lab. Also thank you for reading my dissertation and giving me feedback. Thank you to Ben Barnhill, Dr. Lissa Anderson, Dr. Joe Strukel, Dr. Andrew Dawdy, Dr. Jenn Abelin, Dr. Lichao Zhang, Dr.

Weihan Wang, Dr. Sushmit Maitra, Dr. Aaron Bailey, Dr. Jenny Cottine, Dr. Jessica Chapman, Marshall Chaffee, Stephanie Lehman Miller, Garrett Tanner, and Ellen Speers. If I missed anyone, well they probably won't read this anyway, so I'm safe. Oh and thank you to Scott Ugrin for always keeping me grounded.

I would like to thank my parents, Alden and Marilyn Wriston. Mom, I know that you would be proud of me for becoming Dr. Wriston, and I would not have been able to get to this point without you and Dad showing me where and what hard work can get you. Mom, even though you were able to see me reach this milestone, know that I would not be who I am today without all that you taught me. Mom, thank you for working long, hard hours and setting a good example. Thank you, for always believing in us and for letting us know that nothing was impossible. Thank you, for being my best friend. Dad, you have been there for me and with me every step of the way since day one, and I highly doubt that I would have made it through the last five years without your love and support. Thank you for coming to every school function, every ball game, and every event that I participated in, even if it meant you only got a couple hours of sleep. Thank you for working two jobs. Thank you for being the best dad, and my best friend. I owe a huge thank you to my sister, Tricia (more affectionately known as Mpooka (silent M)), who puts up me on a daily basis, especially these last couple months. I truly hit the sister jackpot with you. You are my best friend and I am so thankful that we have such a close relationship. Thank you for always being up for anything. Thank you for always doing pretty much anything that I ask. Thank you for being such an amazing person to grow up with. I would like to thank my grandparents, Wayne and Patricia Wriston, who have

been nothing but supportive in every endeavor I have undertaken, and have loved me unconditionally. They are truly the epitome of everything I hope to be one day.

I would also like to thank all of my aunts, uncles, and cousins. It is so much easier to go through life with such an extensive and strong support system. To the crazy Keaveny's thank you for including me and my family as your own. Finally, I would like to thank all of my friends, from Oak Hill to Wesleyan to Charlottesville, and everywhere in between who have encouraged me and continue to do so. There are too many to name, but they know who they are.

My freshmen year at Wesleyan I joined an incredible group of women in sisterhood by joining Zeta Tau Alpha, and through that I have met some amazing people that will be lifelong friends. I would like to thank my college roommate who is still one of my best friends and has always been supportive of everything that I have attempted to do. I've said it before and I'll say it again, who would have thought a graphic designer and a chemist would become such good friends. Thank you Jess. Thank you Jocelyn for always being up for an adventure and providing me a getaway from lab. Thank you to Wooten for being such a great friend and believing in me. Thank you to Emily Elborn for always being a source of constant entertainment. Thank you to Kate, Rosgh, Teter, Emmy, Cassie, and the list goes on and on.

A thank you does not seem like enough, but at this point that is all I have. I'll close by saying THANK YOU to everyone who has helped me to get to this point in my life.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Abbreviations

˚C	Degrees Celsius
Å	Angstroms
A	Solvent A, 100 mM acetic acid in water
AcOH	Acetic acid
AD	Alzheimer's disease
ADR	Adverse drug reaction
AIDS	Acquired immunodeficiency syndrome
Angio	Angiotensin 1
Ala, A	Alanine
Arg, R	Arginine
PTM	Post-translational modifications
Asn, N	Asparagine
Asp, D	Aspartic acid
B	Solvent B, 100 mM acetic acid in 70% acetonitrile
c	Centi (1×10^{-2})
C	Carboxyl terminus of a polypeptide
CAD	Collision-activated dissociation
c. eq. cells	Cell equivalents- the MHC peptides from an equivalent number of
CRM	Charged residue model
Cys, C	Cysteine
Da	Dalton
DHS	Drug hypersensitivity syndrome

DILI	Drug induced liver injury
DMF	Dimethylformamide
ECD	Electron-capture dissociation
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
ER	Endoplasmic reticulum
ERAP	Endoplasmic reticulum aminopeptidase
ESI	Electrospray ionization
ETD	Electron transfer dissociation
f	femto (1×10^{-15})
FETD	Front-end electron transfer dissociation
FTICR	Fourier transform ion cyclotron resonance
FT-MS	Fourier transform mass spectrometry
g	Gram
GalNAz	N-azidoacetylgalactosamine
GalT1	Galactosyltransferase 1
GlcNAc	N-acetylglucosamine
Gln, Q	Glutamine
Glu, E	Glutamic acid
Gly, G	Glycine
gS, gT	GlcNAcylated residue
HCl	Hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1piperazineethanesulfonic acid
HGPS	Hutchinson-Gilford progeria syndrome
His, H	Histidine
HIV	Human immunodeficiency virus

HLA	Human leukocyte antigens
HPLC	High performance liquid chromatography
i.d.	Inner diameter
Ile, I	Isoleucine
INM	Inner nuclear membrane
k	kilo
Kasil	Potassium silicate solution
L	Liter
LC	Liquid chromatography
LC-MS	Liquid chromatography-Mass spectrometry
Leu, L	Leucine
LMCO	Low mass cutoff
Lys, K	Lysine
μ	micro (1×10^{-6})
m	milli (1×10^{-3})
m/z	mass-to-charge
M+nH	Molecular ion with n charge(s)
Mane	*macaca neistrina*
Met, M	Methionine
MHC	Major histocompatibility complex
min	minute
$MnCl_2$	Magnesium Chloride
mol	mole (6.022×10^{23})
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry

MS1	Full mass spectrum
MS2	Tandem mass spectrum
n	nano (1×10^{-9})
n-mer	a peptide n residues in length
O-GlcNAc	O-linked β-N-acetylglucosamine
o.d.	Outer diameter
OGA	O-GlcNAcase
OGT	O-GlcNAc transferase
ONM	Outer nuclear membrane
p	pico (1×10^{-12})
PBMCs	Peripheral blood mononuclear cells
PBS	Phosphate buffered saline
PC	Precolumn
pep-V	HSITYLLPV
pep-W	HSITYLLPW
p.i.	Pharmacological interaction
Phe, F	Phenylalanine
ppm	Parts-per-million
Pro, P	Proline
PSCL	Positional scanning combinatorial libraries
psi	Pounds per square inch
s	Second (s)
SAAS	Single amino acid substitution
Ser, S	Serine
SF	Specificity factor

SIV	Simian immunodeficiency virus
SJS	Stevens—Johnson syndrome
SUMO	Small ubiquitin-like modifier
TAP	Transporter associated with antigen processing
TCEP	Tris(2-carboxyethyl) phosphine hydrochloride
TEN	Toxic epidermal necrolysis
THF	Tetrahydrofuran
Thr, T	Threonine
TPP	Triphenylphosphine
Trp, W	Tryptophan
Tyr, Y	Tyrosine
UDP	Uridine diphosphate
UDP-GalNAz	Uridine diphosphate N-azidoacetylgalactosamine
UDP-GlcNAc	Uridine diphosphate N-acetylglucosamine
V	Volt
Val, V	Valine
Vaso	Vasoactive intestinal peptide
WT	Wildtype
X	Any amino acid
XIC	Extracted ion chromatogram

Abtract

Mass spectrometry is an invaluable tool in the study of proteomics. Through the use of mass spectrometry, protein binding partners have been identified, post translational modifications (PTMs) have been site-mapped, and the immunopeptidome has been characterized. The invention of electron transfer dissociation (ETD) greatly facilitated the ability to use mass spectrometry to site-map PTMs. This dissertation discusses four studies that apply and extend the use of mass spectrometry to better understand disease processes, adverse drug reactions, and immunology.

While ETD has enhanced the capability of site-mapping labile PTMs, it does have inherent limitations. ETD requires a charge state of at least +3 or higher to achieve optimal fragmentation, and in areas of charge deficiency sequence ions are not prroduced. This can cause difficulty when attempting to site map O-GlcNAc modifications because they typically occur in poly S/T regions within a peptide, which lack sufficient charge. To overcome this limitation, we have developed a method of adding charge to the O-GlcNAc moiety via the Staudinger reaction to effectively increase charge and enhance fragmentation via ETD.

Lamin A is an intermediate filament nuclear protein, which primarily provides mechanical support and stability to the nucleus. Mutations in the *LMNA* gene, which encodes lamin A, can result in a number of diseases. In the case of Hutchinson-Gilford progeria syndrome (HGPS), a single point mutation causes a deletion of fifty amino acids and causes changes to nuclear shape. Within this 50 amino acid deletion, targets for PTMs are deleted, including both known targets for O-GlcNAcylation. This dissertation

reports the identification of O-GlcNAcylation sites on WT lamin A and the HGPS related Δ35 deletion lamin A to better understand the role O-GlcNAcylation plays in HGPS.

The Major Histocompatibility Complex (MHC) is an essential part of the immune system, which displays peptide antigens to T cells to assess the health of individual cells. If the presented peptide is identified as foreign, or 'non-self' then an immune response is elicited. Understanding the MHC peptidome, peptides presented by MHC molecules, can provide information necessary to understanding mechanisms of disease as well as provide insight into the adverse effects of pharmaceutical drugs. This dissertation reports the characterization of the Chinese rhesus macaque's MHC peptidome by mass spectrometry to provide information relative to the development of an HIV/SIV vaccine.

Finally, adverse drug reactions (ADR), which cause immune-mediated responses can be linked to a specific MHC allele. For example, abacavir, a nucleoside analog used in the treatment of HIV, exclusively causes an ADR in individuals expressing the HLA-B*57:01 allele. MHC peptides from untreated cells and from abacavir treated cells were analyzed by mass spectrometry to identify potential differences between them. An altered peptide repertoire is presented in the presence of abacavir, stimulating a T cell response. This is a novel mechanism of T cell stimulation, which can be applied to other drugs causing an ADR.

Chapter 1: Introduction to the Dissertation

1.1 Mass Spectrometry and Proteomics

Mass spectrometry has been an indispensable tool in proteomic research for the past 30 years. In the Hunt laboratory mass spectrometry is used to identify and sequence proteins and peptides. The complexity of the proteome is amplified by the presence of post translational modifications (PTMs). Proteins are adorned with a variety of different modifications including phosphorylation, glycosylation, ubiquitination, acetylation, and methylation among others. These modifications can be first site-localized by mass spectrometry and then their function can be elucidated through biological assays. Mass spectrometry is also utilized to gain a better understanding of the immune system by sequencing peptides displayed by proteins of the major histocompatibility complex (MHC). This dissertation will investigate changes to the MHC peptidome when influenced by pharmaceutical drugs, identify the MHC peptidome of a species used as an animal model, will propose and test a method to enhance the site-mapping of O-GlcNAcylated peptides, and will identify O-GlcNAc sites on the nuclear protein lamin A. At the core of all these projects is mass spectrometry.

Mass spectrometry is a comprehensive and versatile technique that has been at the forefront of proteomic research for the past 30 years. During the 1990's mass spectrometry replaced Edman degradation as the primary method to sequence peptides and proteins.(1) Mass spectrometry offers high resolution, high mass accuracy (<5ppm) and a wide dynamic range for complex analyses.(2) High performance liquid chromatography (HPLC) is usually directly coupled to a mass spectrometer (LC-MS) to

continuously introduce sample as it is separated. Typically in proteomic LC-MS experiments, peptides are gradient eluted using reverse phase HPLC, which separates compounds based on their hydrophobicity.(2)

A typical mass spectrometry experiment begins with on-line separation by reverse phase HPLC. The HPLC column is equipped with an electrospray emitter tip which facilitates ionizing the sample as it is introduced to the mass spectrometer. Upon introduction to the mass spectrometer ions are analyzed by either a low resolution analyzer (linear ion trap) or a high resolution analyzer (Fourier transform ion cyclotron resonance (FTICR) or an orbitrap). Typically a full mass spectrum (MS) is analyzed at high resolution to obtain accurate mass measurements and an MS2 is analyzed in the ion trap. Fragmentation of peptides is performed by both collisionally activated dissociation (CAD) and electron transfer dissociation (ETD).

Electrospray ionization was a major breakthrough for analyzing large biomolecules like proteins and peptides. Electrospray ionization is a soft ionization technique that is essential when working with peptides and proteins because of their polar, nonvolatile, and thermally unstable nature which requires an ionization technique that will not lead to degradation.(2) Analytical HPLC columns are constructed with an electrospray emitter tip that is approximately 2 μm in diameter. As electrospray ionization requires a high electric field, a voltage of 2 kV is applied to the column with the counter electrode being the capillary inlet to the mass spectrometer.(3) In positive mode, positive ions move toward the counter electrode and accumulate at the surface of the tip.(4) The meniscus of the tip deforms when the positive ions overcome the surface

tension of the liquid and expand the liquid into a Taylor cone, which continuously produces droplets enriched in positive ions.(1, 4, 5) Solvent evaporation occurs on the droplets formed by the Taylor cone and the radius of the droplet decreases at constant charge until the Coulombic repulsion between the charges overcomes the surface tension resulting in Coulombic fissions, shown in **Figure 1.1**.(1, 4) There are two theories explaining the desolvation of the droplets, but only one is applicable to peptides and proteins—the charged residue model (CRM).(4) The CRM assumes that a series of droplet fission events leads to a final droplet containing a single analyte molecule.(4)

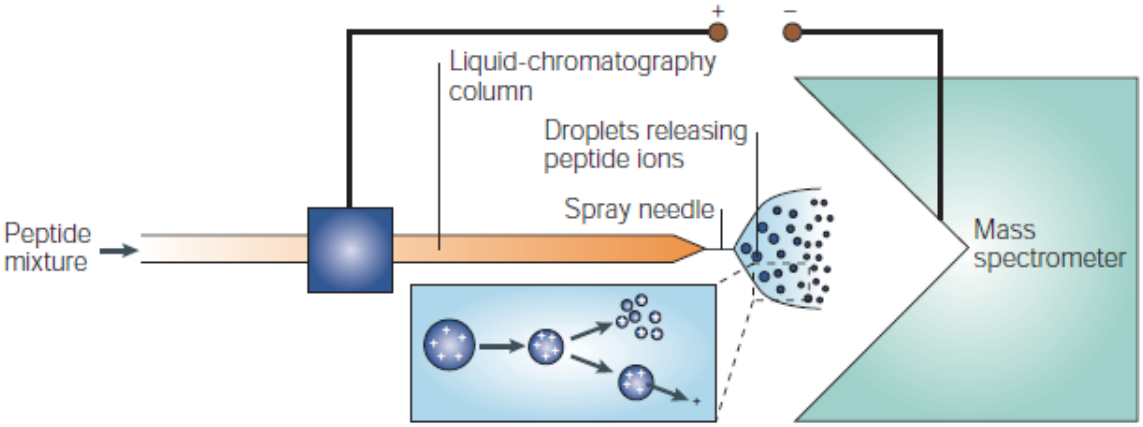

Figure 1.1. Schematic of electrospray ionization directly coupled to a mass spectrometer.{{82 Steen, H. 2004}} A positive voltage is applied to the column and the positive peptide ions move toward the capillary inlet of the mass spectrometer. While the droplets are airborne they decrease in size and the charge density increases as the solvent evaporates. Repetitive droplet fission occurs until each droplet contains only one analyte molecule.{{82 Steen, H. 2004}}

After the ions are electrosprayed into the mass spectrometer they undergo mass spectral analysis. The instrument is typically operated in a data-dependent mode, meaning ions are selected for fragmentation from the full MS based on a set of parameters the use selects centered on the characteristics of the precursor ion. Usually, a

top 6 data-dependent, toggle method is used. This type of method selects the top 6 most abundant, non-excluded ions for CAD and ETD fragmentation. The full MS scan is acquired in the high resolution mass analyzer and the MS2 scans are typically acquired in the low resolution mass analyzer, though they can also be acquired in the high resolution mass analyzer.

In CAD, protonated peptides are accelerated by an electric potential and undergo multiple collisions with helium gas in the ion trap. During these low-energy collisions, the kinetic energy of the peptide is converted to vibrational energy, which is distributed throughout the molecule.(6, 7) Once the peptide has undergone a sufficient number of collisions for the vibrational energy of the peptide to exceed the bond energy of the weakest bond, the peptide will dissociate, usually at the amide bonds along the peptide backbone. (6) CAD fragmentation, shown in **Figure 1.2**, produces b-ions, containing the N-terminus, and y-ions, containing the C-terminus.(6, 7) Sequence information can be

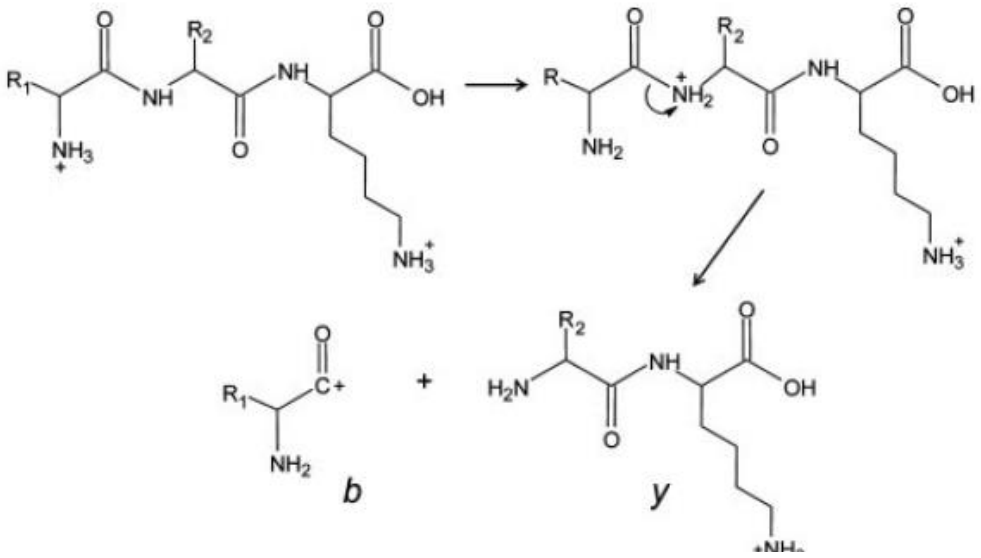

Figure 1.2. CAD reaction scheme.(10)

obtained from these ions by determining the Δm between ions of the same series. This Δm will be equal to the mass of an amino acid.

Collisionally activated dissociation is the most popular fragmentation technique for peptide analysis, and is particularly amenable to short peptides with low charge states. However, CAD can be problematic when trying to sequence post translational modifications, especially labile modifications like phosphorylation and O-GlcNAcylation. During CAD fragmentation, the bond with the weakest energy is preferentially cleaved. Typically the weakest bond of a modified peptide is the bond between the amino acid and the modification. When an O-GlcNAcylated peptide undergoes CAD fragmentation, the resulting spectrum is dominated by an ion at m/z 204,

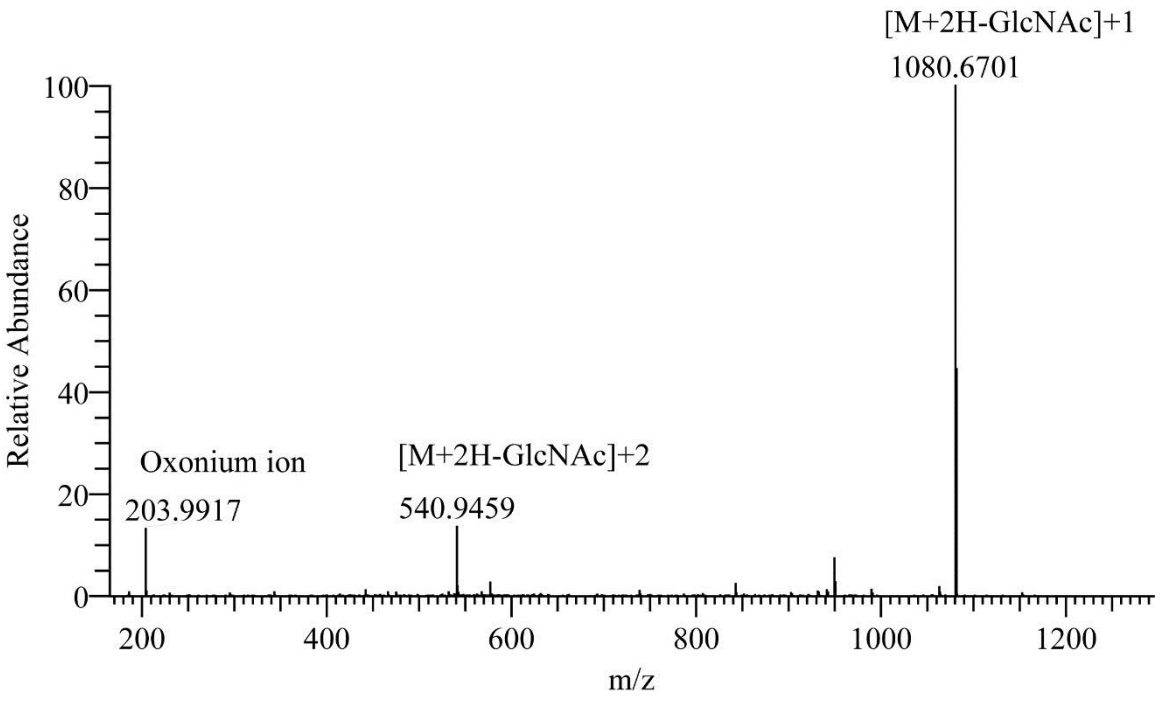

Figure 1.3. CAD spectrum of an O-GlcNAcylated peptide. The spectrum is dominated by the O-GlcNAc oxonium ion at m/z 204 and the charge reduced product ion with the loss of 203 amu.

which corresponds to the O-GlcNAc oxonium ion, and the corresponding charge reduced product ion with a loss of 203 amu, shown in **Figure 1.3**.(8) The spectrum will have very few sequence ions, and they will be present at low abundance. In most cases, these spectra will not result in sequence identification. Additionally, CAD is ill-suited for long peptides and peptides with a charge state greater than +3. Due to CAD's incompatibility with certain peptides, it is best used in conjunction with electron transfer dissociation (ETD), which complements CAD.

Electron transfer dissociation (ETD), developed in the Hunt lab, is a fragmentation method suitable for highly charged peptides, and it preserves labile PTMs during fragmentation.(9) In this fragmentation process, multiply protonated peptides undergo ion-ion reactions with a radical anion, such as azulene or fluoranthene.(9) During the reaction the radical anion transfers an electron to the backbone of the multiply protonated peptide triggering highly selective fragmentation of the N-C_α bonds along the peptide backbone, shown in **Figure 1.4**.(10) This fragmentation process results in sequence informative c-ions, containing the N-terminus, and $z^\bullet$-ions, containing the C-terminus.(10) During the ETD reaction, a charge is consumed; therefore, peptides with a charge state of +3 or higher are more suitable for ETD than those with a charge state of +2.

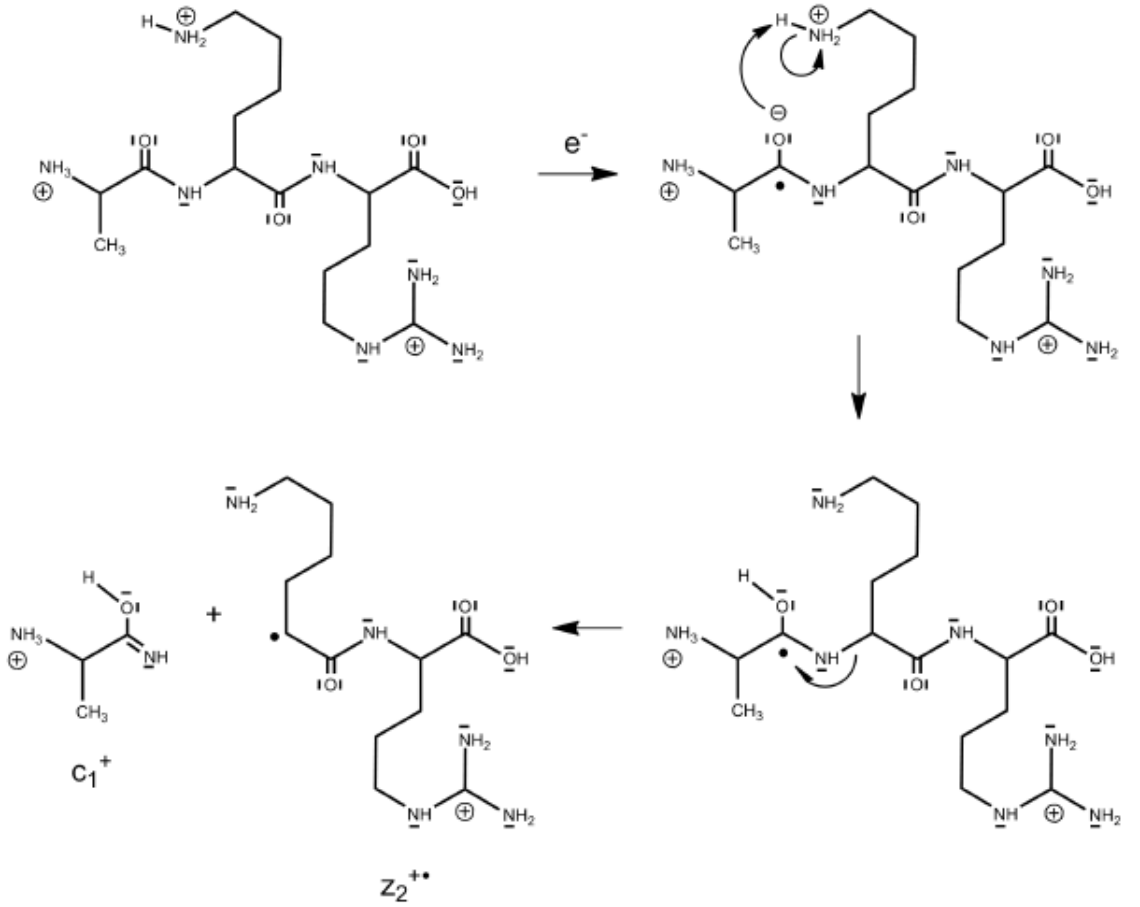

Figure 1.4. ETD reaction scheme.(10)

Electron transfer dissociation has been instrumental in the identification and site-localization of many phosphorylation and O-GlcNAcylation modifications in our lab. Chapter 5 will highlight the ability of ETD to assist in the localization of highly modified peptides. However, ETD does have some inherent limitations. ETD of proline residues does not result in fragmentation because even though there is cleavage of the N-C_α bond in the ring system, the fragments remain attached through other atoms in the ring.(9) Additionally, ETD fragmentation also tends to be poor in sections of peptides that lack charge. This aspect of ETD can be problematic when site-mapping O-GlcNAc

modifications, because they tend to occur in poly S/T regions that are deficient of charge.(11) Chapter 4 of this dissertation will address adding charge to the GlcNAc moiety to enhance ETD of O-GlcNAc modified peptides.

After mass spectral analysis is completed, the data is searched against a database to identify the contents of a sample. In our lab data is searched, against either the RefSeq(12) or SwissProt human databases(13), using the open mass spectrometry search algorithm (OMSSA)(14). The peptide matches from the search are given a score that corresponds to the statistical likelihood that the spectrum matched the sequence given.(14) The search results above a certain threshold are manually interpreted and validated by annotating the spectrum and by assessing the mass accuracy.

Mass spectrometry can be applied to study a wide range of proteomic applications from identifying protein binding partners to site-mapping PTMs to characterizing the MHC peptidome. Characterizing the MHC peptidome is important to reveal how a disease progresses, and potentially how to treat a disease. Identifying the MHC peptides provides a great deal of insight into how the adaptive immune system works and what triggers an immune response. In the chapters that follow mass spectrometry will be applied to elucidate information about the immune system to gain better insight into the causes of adverse drug reactions and how the MHC immunopeptidome of macaques are important to the study of HIV/SIV.

1.2 References

1. Steen H, Mann M. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Bio.* 2004; 55: 699-711.

2. Yates JR, Ruse CI, Nakorchevsky A. Proteomics by mass spectrometry: Approaches, advances, and applications. *Annu Rev Biomed Eng.* 2009; 11: 49-79.

3. Hofstadler SA, Bakhtiar R, Smith RD. *J Chem Edu.* 1996; 73(4): 82-88.

4. Gabelica V, De Pauw E. Internal energy and fragmentation of ions produced in electrospray sources. *Mass Spectrom Reviews.* 2005; 24: 566-587.

5. Kebarle P. A brief overview of the present status of the mechanisms involoved in electrospray mass spectrometry. *J Mass Spectrom.* 2000; 35: 804-817.

6. Hunt DF, Yates JR, Shabanowitz J, Winston S, Hauer CR. Protein sequencing by tandem mass spectrometry. *Proc Nat Acad Sci.* 1986; 83: 6233-6237.

7. Boersma PJ, Mohammed S, Heck AJR. Phosphopeptide fragmentation and analysis by mass spectrometry. *J Mass Spectrom.* 2009; 44: 861-878.

8. Mikesh LM, Ueberheide B, Chi A, Coon JJ, Syka JEP, Shabanowitz J, Hunt DF. The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta.* 2006; 1764: 1811-1822.

9. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Nat Acad Sci.* 2004; 101(26): 9528-9533.

10. Udeshi ND, Compton PD, Shabanowitz J, Hunt DF, Rose KL. Methods for analyzing peptides and proteins on a chromatographic timescale by electron-transfer dissociation mass spectrometry. *Nat Protocols.* 2008; 3(11): 1709-1717.

11. Chalkey RJ, Thalhammer A, Schoepfer R, Burlingame AL. Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry of native peptides. *Proc Nat Acad Sci.* 2009; 106(22): 8894-8899.

12. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts, and proteins. *Nucl Acids Res.* 2007; 35: D61-D65.

13. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucl Acids Res.* 2000; 28(1): 45-48.

14. Greer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, Bryant SH. Open mass spectrometry search algorithm. *J Prot Res.* 2004; 3: 958-964.

Chapter 2: Drug Hypersensitivity caused by the Alteration of the HLA-presented Self-peptide Repertoire

2.1 Introduction

2.1.1 Immunology

The pharmaceutical industry is one of the largest worldwide, worth an estimated $300 billion. The primary goal of pharmaceutical companies is to develop and manufacture drugs that help the body combat infection and disease. However, there are instances where a drug produces harmful side effects as opposed to benefiting the patient. This can be caused by an allergic reaction to the drug, or in more severe cases a drug can cause an immune mediated reaction called an adverse drug reaction (ADR) or drug hypersensitivity syndrome (DHS). An adverse drug reaction can manifest into a variety of illnesses and disorders which in extreme cases can cause death. To comprehend what causes these adverse reactions, one first must have an understanding of how the immune system works.

The immune system is a complex entity that constantly surveys the health of all different types of cells in the body. The immune system is responsible for identifying and eliminating foreign substances from our bodies. There are two types of immunity that our body uses in its defense: humoral immunity, which provides protection in the extracellular fluids by employing antibodies to recognize and neutralize foreign invaders, and cell-mediated immunity, which provides surveillance intracellularly through the presentation of peptide antigens to T cell receptors.(1) Presentation of peptide antigens is a critical component of the immune system, as it enables the immune system to determine

if a foreign substance has invaded the cell and eliminates the threat. Major histocompatibility complex (MHC) molecules present peptide antigens to T cell receptors (TCR) located on T cells through a process called antigen presentation, a process which is the basis of cell-mediated immunity.

The major histocompatibility complex was originally discovered as transplantation antigens, both donor and acceptor were required to have the same haplotype to avoid graft rejection. The MHC locus is comprised of over 100 separate genes and is one of the most polymorphic genes known.(2) In humans the MHC locus is known as the human leukocyte antigen (HLA). There are two classes of MHC molecules: class I molecules, which present intrinsic antigens, and class II molecules, which present extrinsic antigens. Even though these two classes have slightly different attributes, their principle objective is the same: to present peptide antigens to TCRs. Since they possess different types of peptide antigens, they present them to different types of T cells. Class I MHC molecules present antigens to $CD8^+$ T cells while class II MHC molecules present antigens to $CD4^+$ T cells. Both types of T cells recognize an antigen as either self or non-self, foreign, and will elicit an immune response if the antigen is found to be non-self. This process will be discussed later in further detail.

There are three genes that comprise the class I MHC in humans, HLA-A, HLA-B, and HLA-C; and every human inherits one of each of these three genes from each parent which can result in a maximum of six different alleles or as few as three different alleles. Interestingly, there are over 4,000 different HLA allele variants. (1) A cell typically expresses several hundred thousand copies of MHC molecules and each molecule

displays a single peptide, thus a single cell is capable of presenting thousands of distinct peptides at any one time.(2) It is advantageous to have a more diverse repertoire of HLA haplotypes, as this allows for the potential to present a larger variety of peptides to T cells, and this a greater chance of being able to detect a foreign invader and elicit an immune response.(3)

The MHC class I molecule, shown in **Figure 2.1**, is comprised of three extracellular polymorphic chains, α_1, α_2, α_3, a transmembrane domain, and a cytoplasmic tail.(4) The alpha domain forms a complex with β_2-microglobulin, which is required

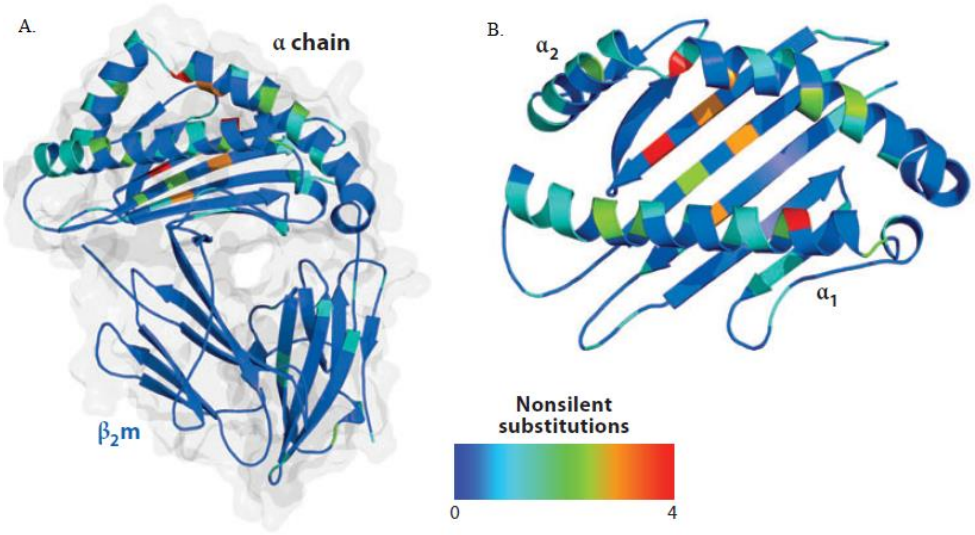

Figure 2.1. Structure of MHC class I molecule. **A.** Structural representation of MHC class I molecule with β_2-microglobulin. **B.** Top view of the MHC class I binding groove highlighting nonsilent substitutions, which are substitutions that give rise to the high number of variants, that occur in HLA-A alleles.(4)

before a peptide can be loaded onto the MHC molecule for antigen presentation, shown in **Figure 2.1 A**. The α_1 and α_2 domains form the peptide binding cleft, shown in **Figure 2.1 B**. The peptide-binding cleft has two anchor positions where the 2nd and 9th residues of a peptide dock in the groove, which typically accommodates peptides between 8-10 amino acid residues in length. In the binding cleft the middle of the peptide is solvent exposed and displays different structures to TCRs.(3) The polymorphism of the MHC proteins results in amino acid variations in the peptide-binding groove, the α_1 and α_2 domains, depicted by the color changes in **Figure 2.1**, which produce a wide diversity of peptide motifs, examples of which can be seen in **Figure 2.2**.(5) The α_3, transmembrane region, and cytoplasmic tail domains are much more conserved. The peptide motif provides different HLA alleles with an important and distinct characteristic, as it determines which peptides are presented to the T cell.

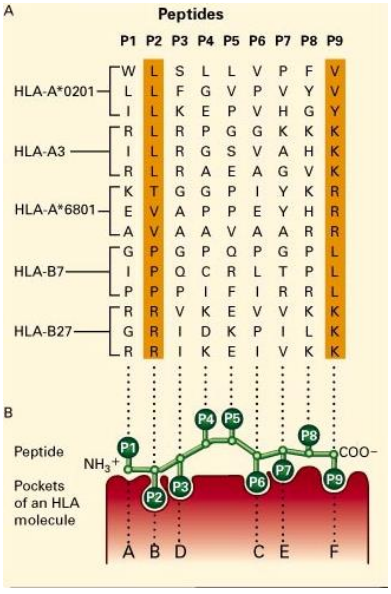

Figure 2.2. A. Examples of motifs for common HLA alleles. **B.** Interaction between MHC molecule and peptide, and the orientation of the peptide in the groove. The B-pocket anchors the 2nd residue and the F-pocket anchors the 9th residue of the peptide antigen.(5)

T cells are only capable of recognizing peptides that are bound to MHC molecules, so how is a peptide loaded onto an MHC molecule? The initial step in the MHC class I processing pathway is the degradation of cytosolic proteins that have been marked for degradation by the addition of ubiquitin. Proteins are degraded by the 26s proteasome, which is a multi-protein, barrel-like complex that defines the carboxyl terminus of the resulting peptides.(6) Peptides generated from the proteasome are transported into the endoplasmic reticulum (ER) by the peptide transporter associated with antigen presentation (TAP).(6) TAP preferentially transports proteins of 8-16 amino acids in length, and longer peptides are trimmed by ER aminopeptidase 1 (ERAP1).(7) This aminopeptidase specifically trims the N-termini of longer peptides resulting in peptides that are 8-9 residues in length, at which point ERAP1 loses its catalytic properties.(7)

Meanwhile in the ER, newly synthesized MHC molecules are being primed to be loaded with a peptide antigen. The MHC forms a peptide-loading complex with chaperone proteins to stabilize the molecule prior to peptide loading. This complex consists of TAP, which acts as a platform for the folding of MHC molecules; and three chaperone proteins—tapasin, calreticulin, and ERp57. This complex oversees efficient peptide loading onto the MHC molecule. After a peptide is loaded the MHC molecule is released from the chaperone complex. The MHC-peptide complex then exits the ER and travels through the Golgi apparatus to the cell surface for antigen presentation to T cells. **Figure 2.3** summarizes the MHC class I antigen processing pathway.(8)

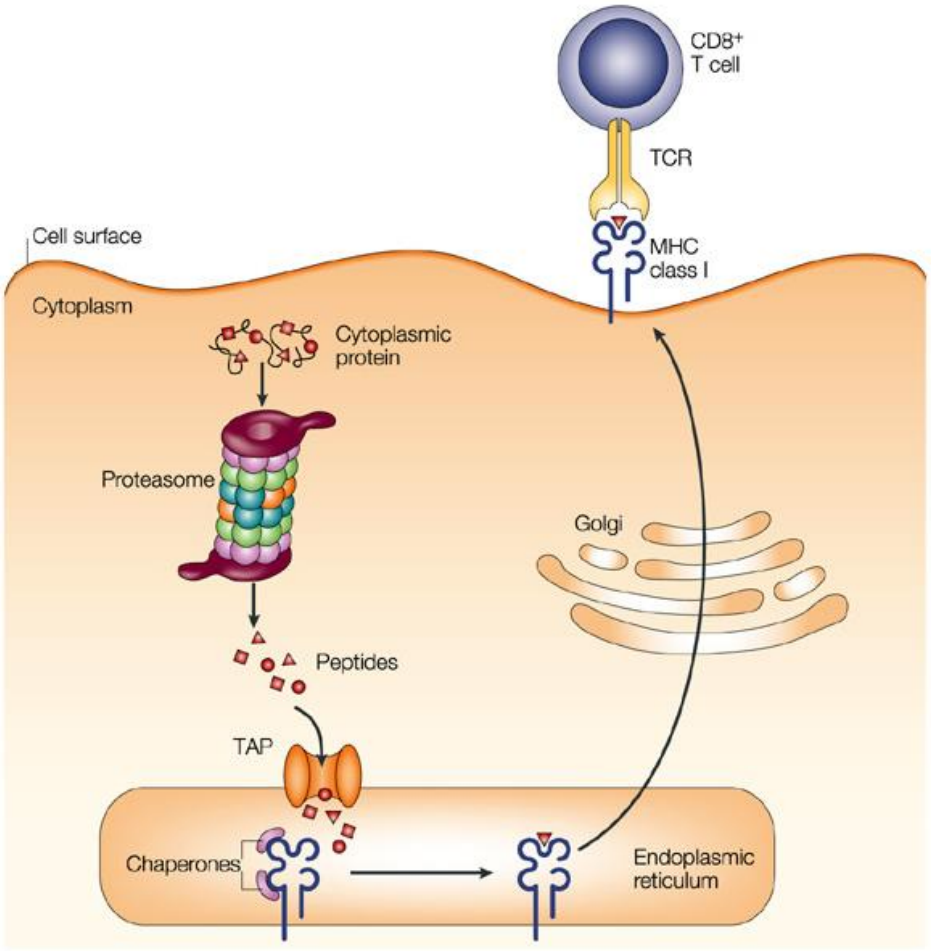

Figure 2.3. MHC class I antigen processing pathway. Proteins are degraded by the proteasome, the peptides are then transported into the ER by TAP and then loaded onto a MHC molecule. The peptide-MHC complex exits the ER and passes through the Golgi apparatus to the cell surface for antigen presentation to T cells.(8)

There are three types of T cells: cytotoxic T cells, regulatory T cells, and helper T cells; cytotoxic T cells are part of MHC class I antigen presentation. Only a specific subset of T cells, which have the CD8 glycoprotein attached to them ($CD8^+$ T cells) are able to interact with MHC class I molecules. The CD8 glycoprotein serves as a co-receptor for the T cell receptors when interacting with a peptide.(3) T cells and TCRs are

the surveillance system of the immune system. They determine if a presented peptide antigen is self or non-self, and depending on their conclusion will either elicit an immune response or continue their surveillance.

The main purpose of T cell receptors, located on the surface of T cells, is to recognize the antigen and then signal to the interior of the T cell if an immune response is necessary. The interaction between the TCR and MHC-peptide complex is depicted in **Figure 2.4**. TCRs consist of two pairs of subunits that are very similar to

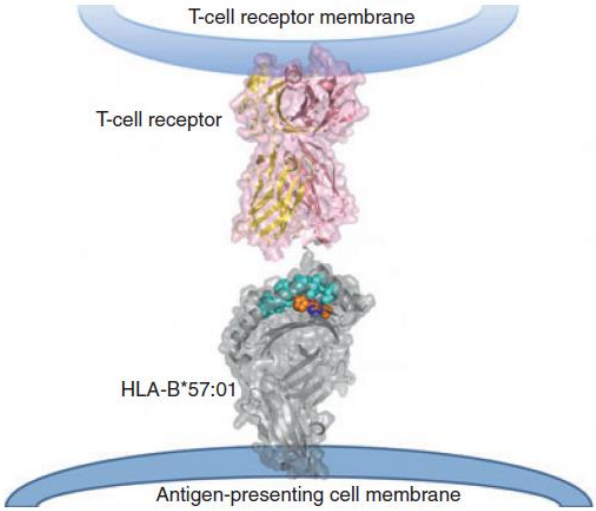

Figure 2.4. Model of a TCR interacting with a peptide presented by an MHC molecule.(11)

immunoglobulins. TCRs are similar to MHC molecules in that they will only recognize specific antigens. How does a T cell know if an antigen is self or foreign? Naïve T cells (thymocytes) are educated in the thymus where they go through a negative and positive selection process. Positive selection is the process by which immature double positive ($CD8^+/CD4^+$) thymocytes expressing T cell receptors with an intermediate affinity and/or avidity for self-peptide-MHC complexes are induced to differentiate into mature single-

positive thymocytes, either $CD8^+$ or $CD4^+$.(9) Negative selection is the intrathymic elimination of double-positive or single-positive thymocytes that express TCRs with a high affinity to self-antigens, shown in **Figure 2.5**.(10)(9) Double-positive thymocytes will also be eliminated if they fail to interact with self-peptide-MHC complexes during positive selection in a process called death by neglect.(9) This education is necessary to ensure that self-peptides do not induce an immune response and that non-self-peptides do induce an immune response.

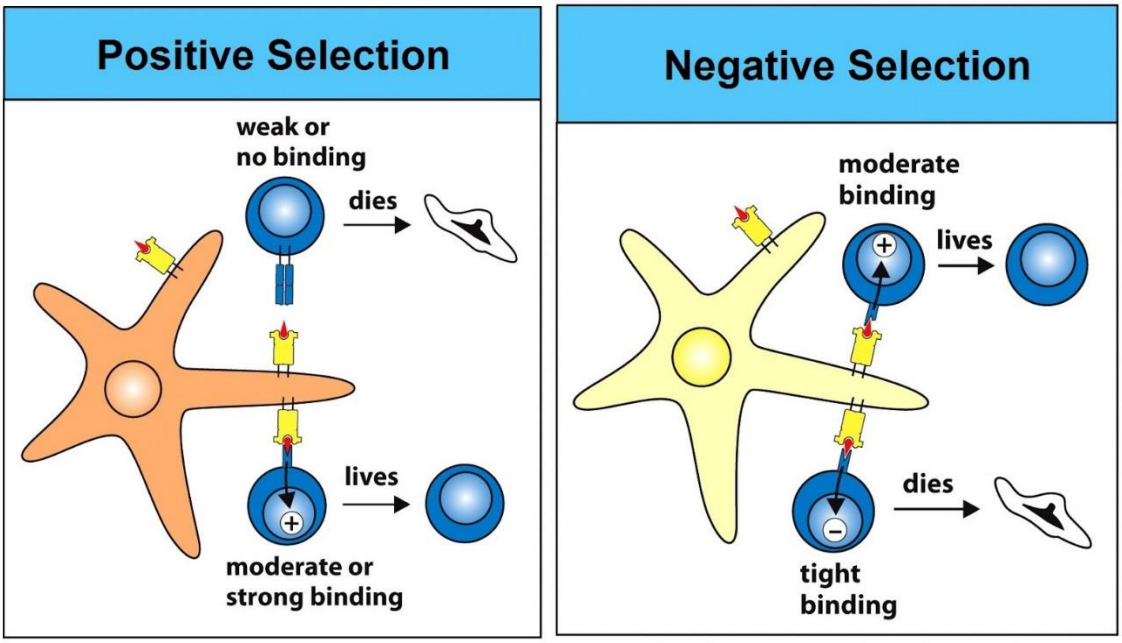

Figure 2.5. Positive and negative selection of T cells.(10)

With this understanding of T cells, it becomes clear how important their role is in the immune system. The MHC complex and T cells are crucial to maintaining health. When an antigen is recognized as foreign the TCR sends signals to the interior of the T cell and mounts an immune response, but what happens if a self-peptide is recognized foreign? This phenomenon is the basis of autoimmunity and can wreak havoc on an

individual. One instance of self-peptides being recognized as non-self is if there is a foreign entity in the MHC peptide-binding pocket, for example a drug or a small molecule can be bound in the binding pocket. What effect does this altered recognition have and what are the repercussions for the affected individual? That question is the basis of this chapter and will be discussed in detail below.

2.1.2 Adverse Drug Reactions

What do you do when you get an infection, a headache, or have high blood pressure? Chances are you take an aspirin or some sort of pharmaceutical drug to help alleviate the symptoms. What if taking medicine does not alleviate your symptoms and instead has the opposite effect, causing you to get even sicker? When a drug causes an individual more harm than good it is called an adverse drug reaction (ADR). Adverse drug reactions are one of the leading causes of morbidity and mortality affecting more than 7% of the general population(4, 11), and costs an estimated $136 billion annually in the United States (more than the cost for cardiovascular or diabetic care).(12) An ADR is a harmful or significantly unpleasant effect caused by a drug at doses intended for therapeutic effect which warrants reduction of dose or complete withdrawal of the drug.(13) There are two major types of ADRs: Type A reactions are predictable, common, and related to the pharmacological activity of the drug while Type B reactions are uncommon, unpredictable, and normally not related to the pharmacological activity of the drug.(14) Type B reactions are commonly called idiosyncratic ADRs caused by an immune-mediated response to the drug and are often more serious than Type A reactions.

Idiosyncratic ADRs, also known as drug hypersensitivity syndrome (DHS), are more severe reactions characterized by fever, rash, and the failure of multiple organs.(4) DHS can also lead to more serious conditions including Stevens-Johnson syndrome (SJS), drug induced liver injury (DILI), toxic epidermal necrosis (TEN), and in some instances death. The immune-mediated response in DHS is caused by the stimulation of T cells by a drug, but how does a particular drug stimulate these T cells? There are currently three models to explain the mechanism of T cell stimulation by a drug: the hapten/prohapten model, pharmacological interaction (p.i.) with immune receptors concept, and the danger/superantigen model.

The hapten/prohapten model, depicted in **Figure 2.6**, involves the covalent binding of a drug to a peptide or protein.(15) This 'new' modified peptide is then loaded onto the MHC molecule and presented to T cells where it is recognized as foreign and the T cell elicits an immune response as a result. The prohapten part of this model refers to a

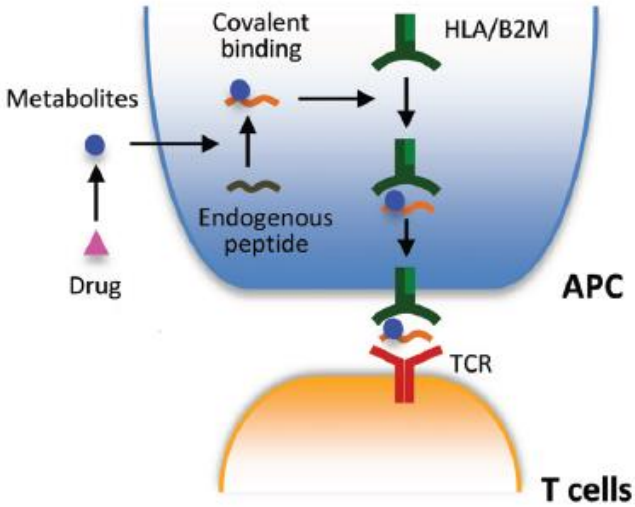

Figure 2.6. Model of the hapten/prohapten mechanism of T cell stimulation. A drug or metabolite of the drug covalently binds to a peptide and is presented to the T cell inducing an immune response.(15)

metabolite of the drug covalently binding to a protein. However, this model is contingent on the haptenated peptide's ability to undergo antigen processing and presentation.(4, 16)

Chemically inert drugs, unable to covalently bind to proteins can also stimulate a T cell response. Alternatively to the hapten/prohapten model, the p. i. concept involves the noncovalent, labile interaction between a drug and MHC molecule or peptide antigen at the cell surface, which would then be seen by the T cell as foreign and elicit an immune response.(16) It has not yet been explained how or where the drug-MHC-peptide interaction occurs. The p.i. concept is depicted in **Figure 2.7**.

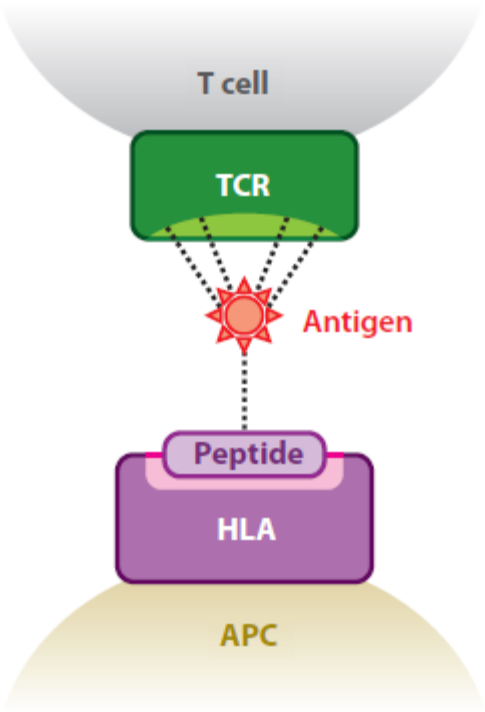

Figure 2.7. Model of the mechanism of T cell stimulation by the p. i. concept. A drug or small molecule non-covalently (dashed lines) interacts with the peptide and TCR.(5)

The third model is the danger/superantigen model which states that danger signals other than the drug itself are required to overcome immune tolerance barriers that otherwise suppress DHS reactions.(17) A superantigen directly links the T cell to the MHC molecule in a peptide independent fashion, as shown in **Figure 2.8**.(4) This type of interaction is less likely to have a direct and specific relationship to a particular HLA haplotype. These three models explain the basis for most ADRs; however, some ADRs do not fit any of these models.

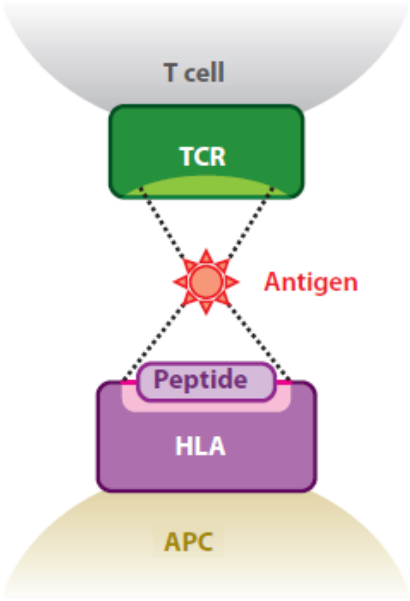

Figure 2.8. Model of the mechanism of T cell stimulation by the superantigen/danger model. A drug or small molecule non-covalently (dashed lines) interacts with the MHC molecule and TCR independent of the peptide.(5)

Early studies based on familial occurrences of ADRs and on occurrences in identical twins indicated that there may be a genetic link to idiosyncratic ADRs.(11) Since then several drugs that cause an ADR have been linked to specific HLA alleles, or

a small subset of HLA alleles, as outlined in **Table 2.1**. One drug of particular importance is the nucleoside analog abacavir, shown in **Figure 2.9**, which is used in the

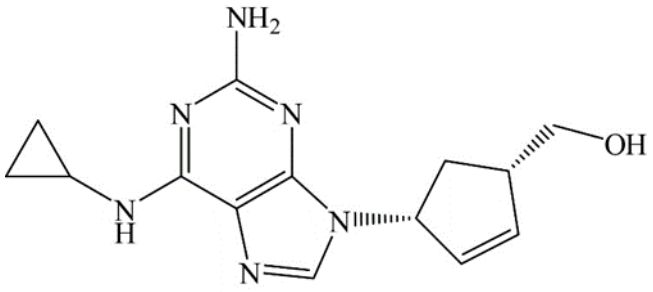

Figure 2.9. Structure of abacavir.

treatment of AIDS. Abacavir inhibits HIV-1 reverse transcriptase suppressing HIV replication, and in approximately 8% of recipients causes an idiosyncratic ADR.(17, 18) Abacavir hypersensitivity is characterized by fever, malaise, and internal organ failure.(19) A 2002 study linked abacavir DHS with the HLA-B*57:01 allele.(20) Abacavir hypersensitivity is a delayed hypersensitivity reaction that occurs several days after the drug has been administered, and upon cessation of abacavir treatment symptoms are resolved.

Table 2.1. HLA Associations with Drug Hypersensitivities.(5, 11, 18)

Drug	Adverse Reaction	HLA association	Class
Abacavir	AHS	B*57:01	I
Allopurinol	SJS/TEN	B*58:01	I
Carbamazepine	SJS/TEN	B*15:02	I
Dapsome	DRESS/DIHS	B*13:01	I
Flucloxacillin	Hepatitis	B*57:01	I
Levamisole	Agranulocytosis	B27	I
Oxicam	SJS/TEN	A2, B12	I
Phenytoin	SJS/TEN	B*15:02	I
Aspirin	Asthma	DPB1*03:01	II
Hydralazine	SLE	DR4	II
Penicillamine	Penicillamine toxicity	DR3	II

Abbreviations: AHS—abacavir hypersensitivity syndrome; SJS/TEN—Stevens-Johnson syndrome/toxic epidermal necrolysis; DRESS—drug reaction with eosiniophilia and systemic symptoms; DIHS—drug-induced hypersensitivity syndrome; SLE---systemic lupus erythematosus.

Abacavir hypersensitivity has been shown to be exclusively linked with the HLA-B*57:01 allele. Interestingly, abacavir does not have an effect on any other alleles, even alleles that differ from HLA-B*57:01 by only a few amino acid residues. Illing and colleagues investigated this phenomenon by testing alleles closely related to HLA-B*57:01 with abacavir-specific T cells.(20) They stimulated antigen presenting cells displaying HLA-B*57:02, HLA-B*57:03, HLA-B*58:01, or HLA-B*57:11 with abacavir specific T cells and measured their responses using flow cytometry in both abacavir treated and untreated cells. They found that only the HLA-B*57:01 allele gave a response, as shown in **Figure 2.10**.(21) They concluded that abacavir must associate with the F-pocket of HLA-B*57:01, as that is where the sequences of these alleles are different.(21)

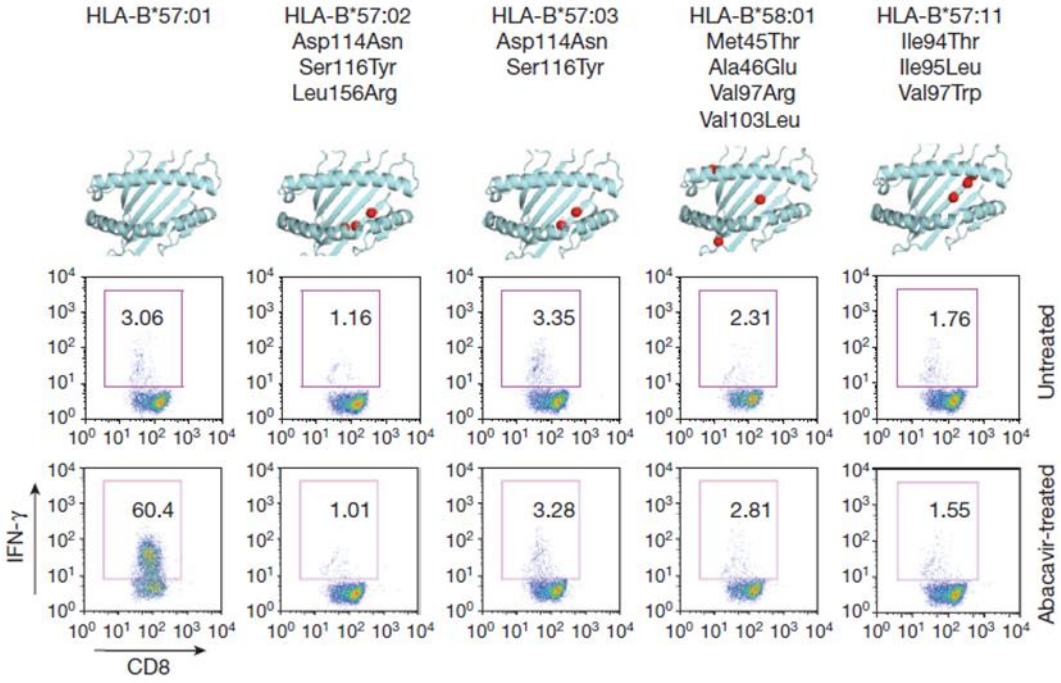

Figure 2.10. Specificity of abacavir T cell responses and binding to HLA-B*57:01. Results of four closely related alleles to HLA-B*57:01 and their T cell response when treated with abacavir and untreated. Only HLA-B*57:01 shows a response.(21)

How does the interaction between HLA-B*57:01 and abacavir stimulate an immune-mediated T cell response? The current models for T cell stimulation do not hold credence for abacavir. It is improbable that abacavir covalently modifies antigens that are presented by only HLA-B*57:01, eliminating the hapten/prohapten model. Similarly, for the p.i. concept to apply abacavir would have to uniquely bind with a surface patch of HLA-B*57:01 capable of inducing TCR recognition.(4) However, this is unlikely since abacavir-insensitive HLA-B*57:03 differs from HLA-B*57:01 by two residues located in the F-pocket of the binding groove, which would be inaccessible by the TCR. Additionally, the danger model could apply, but does not take into account the HLA

restriction. Therefore, we propose a new mechanism of T cell stimulation responsible for the abacavir immune-mediated response. In this mechanism, abacavir binds to the F pocket, which is typically in contact with the side chain of the C-terminal residue of the bound peptide, in the binding groove of HLA-B*57:01 thereby altering the motif presented to the T cells.

2.2 Materials

2.2.1 Reagents

Fisher Scientific, Waltham, MA

Methanol

Honeywell, Morristown, NJ:

Acetonitrile, HPLC grade, ≥99.8% purity

Pierce, Rockford, IL:

LC-MS grade water

Sigma-Aldrich, St. Louis, MO:

Acetic Acid (AcOH, glacial, 99.99%)

Angiotensin I human acetate hydrate, ≥90% purity (Angio)

Azulene

Formamide

1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC)

Histamine·HCl

Pyridine

Vasoactive intestinal peptide fragment 1-12, human ≥97% (Vaso)

Thermo-Fisher Scientific, Waltham, MA

Hydrochloric acid (HCl)

2.2.2 Column Making Materials

Polymicro Technologies, Inc., Phoenix, AZ:

Polyimide coated fused silica capillary

360 μm o.d. × 50 μm i.d.

360 μm o.d. × 75 μm i.d.

360 μm o.d. × 150 μm i.d.

Zeus Industrial Products, Inc., Orangeburg, SC:

Teflon tubing, 0.012” i.d. × 0.060” o.d.

PQ Corporation, Valley Forge, PA:

Kasil-Potassium silicate solution

YMC Corp., Morris Plains, NJ (available through Waters Corp.):

ODS-AQ, C-18, 5 μm spherical silica particles, 120 Å pore size

ODS-AQ, C-18, 5-20 μm spherical silica particles, 300 Å pore size

2.2.3 Equipment

Agilent Technologies, Palo Alto, CA:

1100 Agilent HPLC

Branson, Danbury, CT

Branson 1200 Ultrasonic Bath

Eppendorf, Hamburg, Germany:

Centrifuge 5424

Labconco, Kansas City, MO:

CentriVap Vacuum Concentrator System

Sutter Instrument Co., Novato, CA:

P-2000 laser puller with fused silica adapter

Thermo Fisher Scientific, Waltham, MA:

LTQ mass spectrometer (custom modified with backend ETD)

LTQ FT-ICR hybrid mass spectrometer (custom modified with front-end ETD)

LTQ-Orbitrap mass spectrometer (custom modified with front-end ETD)

LTQ-Orbitrap Velos mass spectrometer (custom modified with front end ETD)

VWR, Radnor, PA:

Mini vortexer

2.3 Methods

John Sidney and Imir Metushi, of the Sette Laboratory at the La Jolla Institute for Allergy and Immunology, isolated MHC peptide samples and conducted peptide binding assays.

2.3.1 Isolation of Class I MHC Peptides from Cultured Cells Treated with Abacavir

Samples were prepared from the HLA-B*57:01 single allele transfected 721.221 cell line. Treated cells were incubated with either abacavir or acyclovir at a concentration of 500 µg/mL for 14 hr at 37°C. Cells were harvested and stored at -80°C. Upon thawing, cells were lysed using cell lysis buffer in the presence of protease inhibitors. The lysate was centrifuged at 100,000 × g for 1 hr and the supernatant was collected and passed through a 0.8/0.2 µm filter. The filtrate was collected and passed through a sepharose CL-4B column to capture any remaining materials in the lysate that may clog the antibody column. The filtrate was then passed through a column packed with protein A sepharose beads coated with MK-D6 antibody which served as an irrelevant antibody used to derive a negative control peptide extract. Next, the filtrate was passed through a second protein A sepharose column coated with W6/32 antibody which captures all HLA class I molecules. The columns were washed sequentially with: 1) 2 column volumes of lysis buffer, 2) 20 column volumes of 20 mM Tris-HCl (pH 8.0, 1150 mM NaCl), 3) 20 column volumes of 20 mN Tris-HCl (pH 8.0, 1 M NaCl), and 4) 20 column volumes of 20 mM Tris-HCl (pH 8.0), and then eluted with 4 column volumes of 0.2 M glacial acetic acid. The eluted peptides were then collected and spun at 3,500 × g at 4°C until 98% of the solution had passed through Millipore ultrafiltration units with a

3,500 Da cut-off to exclude the β-2 microglobin subunit. The filtrate was then collected and concentrated in vacuum for LC-MS analysis.(22)

2.3.2 Cleanup of Samples to Remove Non-Peptidic Impurities

An aliquot of peptide sample was brought up to a total volume of 50 μL by adding 0.1% AcOH. The sample was pressure loaded (200 psi, 1 μL/min) onto a fused silica clean-up column (360 o.d. ×150 i.d. μm) with a Kasil frit and packed 10 cm with C18 resin. The sample was then washed with 0.1% AcOH (1 μL/min, 50 μL) by pressure loading. The column was then rinsed on an HPLC for 10 minutes with solvent A (0.1M AcOH). The sample was eluted into an Eppendorf tube over a 70 minute gradient (0-80% Solvent B (70% acetonitrile/ 30% 0.1M AcOH)). At the end of the gradient the column was rinsed with a liquid composition of 80% B and 20% A for an additional 30 minutes, followed by a 15 minute rinse of 100% A. Sample was taken to dryness on a vacuum concentrator and reconstituted with 0.1% AcOH to a concentration of 1e7 cell equivalents/ μL (c. eq.).

2.3.3 Peptide Derivatization with Histamine

An aliquot of cleaned up peptide sample was taken to dryness on a vacuum concentrator. Peptides were amidated in a solution that contained 1 M histamine·HCl (20 μL) in 1 M pyridine·HCl buffer (pH 5-5.5) and 0.1 M EDC (5 μL) in pyridine·HCl buffer. The solution was sonicated for 2 hours before being dried in a vacuum

concentrator and then reconstituted in 0.1% AcOH to a concentration of 1e7 cell equivalents/ μL.

2.3.4 Assembly of Capillary HPLC Columns

Precolumns (PC) were constructed by creating a Kasil frit in the end of a fused-silica capillary column (360 μm o.d. × 75 μm i.d.). Kasil frits were formed by dipping the end of a capillary column into a mixture of Kasil:formamide (3:1) and heating in an oven overnight at 70˚C. The fritted column was packed with 5-8 cm of irregular, 5-20 μm diameter, 300 Å pore size, C18 reverse-phase resin using a pressure bomb (500 psi). Analytical columns (AC) were constructed with integrated electrospray emitter tips.(23) A Kasil frit was created in a short section of the fused silica (360 μm o.d. × 50 μm i.d.). Columns were then packed with 6-9 cm of 5 μm diameter, 120 Å pore size, C18 reverse-phase resin using a pressure bomb (500 psi). A micropipette laser puller was used to generate a tip (2 μm i.d.) approximately 5 mm from the Kasil frit. Analytical columns were butt-connected to the precolumns using Teflon tubing (0.012" i.d. × 0.060" o.d.).

2.3.5 Loading and Rinsing of Samples

Samples were pressure loaded (~200 psi, ~1 μL/min.) on a PC for each mass spectrometric analysis. Approximately 100 fmol each of two standard peptides (angiotensin I and vasoactive intestinal peptide), were also loaded onto the column for

quantitation purposes. The precolumn was then rinsed on an HPLC with solvent A to remove salts and other small molecules that may interfere with mass spectrometric analysis (30 bar, 30-40 min.). The analytical column was then connected to the precolumn using Teflon tubing (0.012" i.d. × 0.060" o.d.) and rehydrated with solvent A for ~15 min.

2.3.6 Mass Spectrometric Analysis

Samples were analyzed by nanoflow-HPLC/microelectrospray ionization coupled directly to a Thermo Orbitrap Classic, Orbitrap Velos, or FT-ICR mass spectrometer with a home built front-end electron transfer dissociation (FETD) source.(24) After loading the sample on the PC and reattaching the PC to the AC the sample was gradient eluted through a laser-pulled electrospray tip directly into the mass spectrometer with an Agilent 1100 series binary LC pump at a flow rate of ~60 nL/min. The elution gradient was as follows: 0-60% B in 40 minutes, 60-100% B in 3 minutes, hold at 100% B for 3 minutes, 100-0% B in 2 minutes, and hold at 0% B for 20 minutes.

For standard peptide analysis the instrument was operated in a data-dependent mode, performing a high resolution MS1 scan followed by acquisition of MS2 spectra in the linear ion trap on the most abundant peak, first for CAD then ETD. The ETD reagent used was azulene for Thermo Orbitrap and FT-ICR or fluoranthene for Orbitrap Velos, and ETD reaction times were either 40 or 50 ms. For peptide analysis the instrument was operated in a data-dependent mode similar to the standard peptide analysis selecting the

top five most abundant non-excluded ions in the MS1 for fragmentation by both CAD and ETD, instead of just the most abundant ion. For ETD only analyses the instrument was operated in a data-dependent mode, performing a high resolution MS1 followed by acquisition of MS2 spectra in the ion trap on the top 10 most abundant non-excluded ions observed in the MS1.

2.3.7 Data Analysis

MS/MS spectral data were searched against the RefSeq human protein database(25) or the SwissProt human protein database(26) using the Open Mass Spectrometry Search Algorithm (OMSSA) software. (27) OMSSA search tolerances were ± 0.01 Da and ± 0.35 Da for precursor and product ion masses, respectively. For ETD spectra, search parameters were set to exclude reduced charge species from peak lists prior to searching. Data were searched using variable modifications of phosphorylation of serine and threonine residues, and oxidation of methionine, with 1024 variable modifications per peptide being allowed. For derivatized samples, data were searched using fixed modifications of histamine on the peptide C-terminus and aspartic and glutamic acid residues, in addition to the variable modifications. Other search parameters utilized were: +2 charge state products allowed for peptides of charge +3 and above, peptide size range of 7-15 a. a. with no enzyme restrictions. Peptide sequences were manually validated.

Histamine derivatized peptides were also analyzed by in house software that searches raw data files for user specified m/z values. For this experiment the program

was set to search for m/z values of 111 m/z and 195 m/z, the masses for a histamine signature peak and a derivatized C-terminal valine residue, respectively.

2.4 Results and Discussion

2.4.1 Abacavir

In a multi-collaborative effort abacavir's effects on HLA-B*57:01 were investigated. First, to determine if abacavir has an impact on the peptide-binding specificity, our collaborators tested the binding affinity using positional scanning combinatorial peptide libraries (PSCL) in the presence and absence of abacavir. Each library used random peptide sequences of nine residues in length. Normally, the HLA-B*57:01 allele favors a small polar amino acid such as an alanine, threonine, or serine residue in position 2 and a large hydrophobic residue such as a phenylalanine or tryptophan residue in position 9.(28) However, in the presence of abacavir there is a shift in preference for the residue in the ninth position from a large hydrophobic to a small hydrophobic residue. The PSCL studies showed an increased affinity for valine, alanine, and isoleucine in the ninth position, shown in **Figure 2.11A**. This suggests that abacavir occupies the F-pocket of the binding-groove. The binding-groove of HLA-B*57:01 is in fact large enough to accommodate both abacavir and a 9-mer peptide. The effect of abacavir was also tested on the HLA-B*58:01 allele which differs from the HLA-

B*57:01 allele by only two residues. Abacavir did not have a significant effect on the binding affinity of the HLA-B*58:01 allele, shown in **Figure 2.11B**.

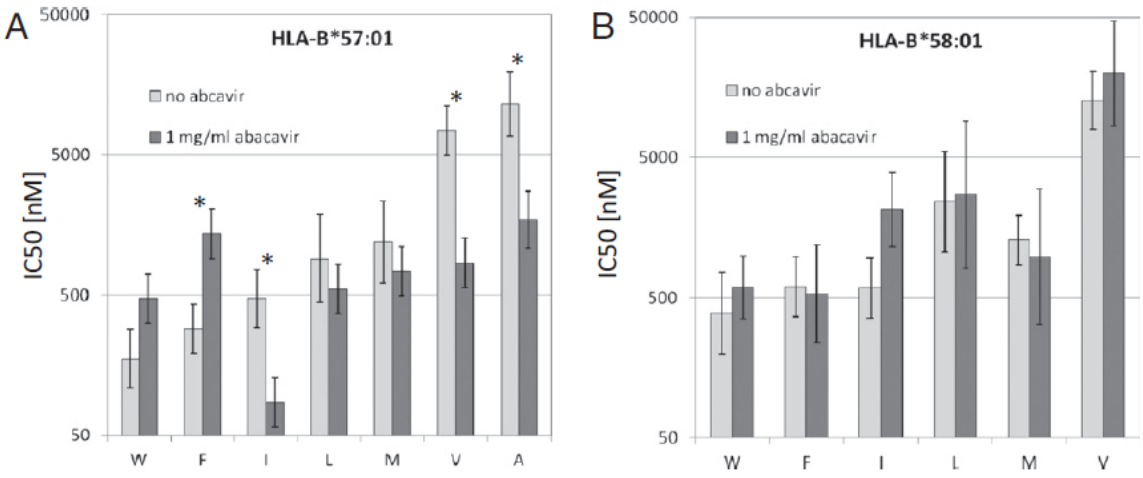

Figure 2.11. A. Abacavir alters the binding specificity of HLA-B*57:01. **B.** Abacavir does not alter the binding specificity of a closely related allele, HLA-B*58:01. Error bars indicate 95% confidence intervals for the mean, and residues marked with an asterisk had significantly different IC_{50} values in the presence vs. absence of abacavir. (17)

Based on the results from the PSCL assay, our collaborators had two peptides synthesized, HSITYLLPV (pep-V) and HSITYLLPW (pep-W), to test their binding efficiency in the presence (in varying concentrations) and absence of abacavir. **Figure 2.12A** shows a linear increase in binding affinity to dose of abacavir for pep-V. **Figure 2.12B** shows that abacavir has very little effect on the affinity of pep-W with the most notable effect observed at the highest dose of abacavir where the affinity for pep-W was decreased by approximately 15%. This result supports the hypothesis that abacavir has an effect on the F pocket of the binding groove. When abacavir is present it not only

increased the affinity for smaller hydrophobic peptides, but also in higher doses had a small effect on the binding of larger hydrophobic residues.

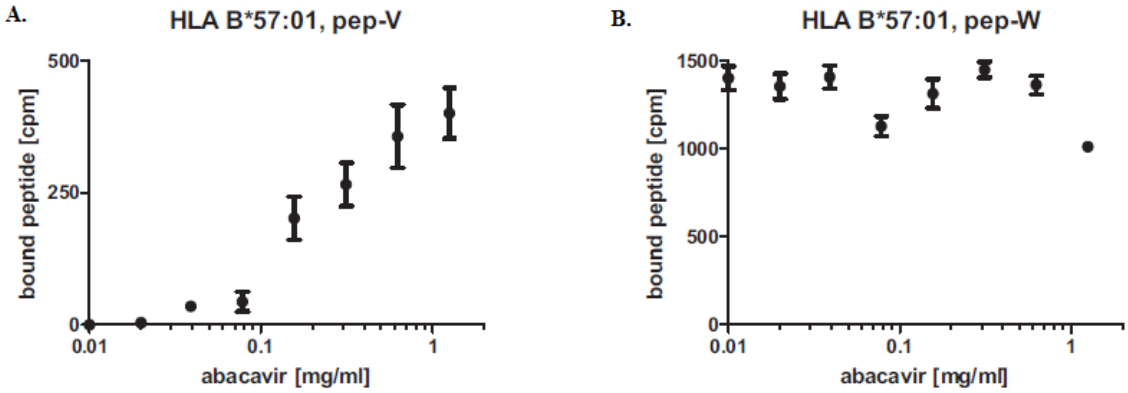

Figure 2.12. A. Binding efficiency of HSITYLLPV (pep-V) to HLA-B*57:01 in increasing doses of abacavir. **B.** Binding efficiency of HSITYKKPW (pep-W) to HLA-B*57:01 in increasing doses of abacavir.(17)

To investigate this further, our collaborators solved the crystal structure of abacavir-HLA-B*57:01-pep-V complex, and found that abacavir does in fact occupy the F pocket of the peptide-binding groove, as can be seen in **Figure 2.13**. Abacavir is bound to a largely hydrophobic pocket in the binding groove forming van der Waals contacts with both HLA-B*57:01 and pep-V. Abacavir directly interacts with the two residues, Asp114 and Ser116, that distinguish HLA-B*57:01 from abacavir-insensitive HLA-B*57:03. Abacavir forms hydrogen bonds with both Asp114 and Ser116, shown in **Figure 2.14**. The HLA-B*58:01 allele has the same residues at positions 114 and 116 as

HLA-B*57:01, but differs at five other residues. HLA-B*58:01 is likely insensitive to the effects of abacavir due to the replacement of Val97 to Arg97 in HLA-B*58:01 because Val97 is part of a hydrophobic pocket in HLA-B*57:01 that forms van der Waals contacts with abacavir, shown in **Figure 2.13** represented by rays. The minor differences between HLA-B*57:01 and these very similar allelic variants prove to be significant in determining abacavir hypersensitivity.

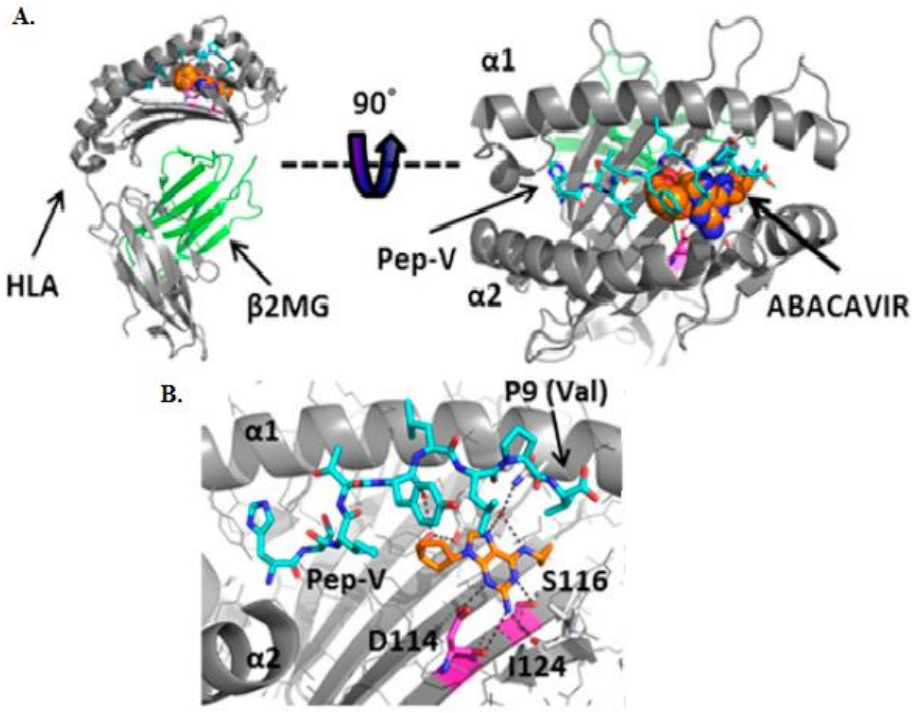

Figure 2.13. A. Crystal structure of the abacavir-peptide-MHC complex solved to a resolution limit of 2.0 Å. The peptide HSITYLLPV is shown in cyan carbons and abacavir is shown as spheres, orange for carbon, blue for nitrogen, and red for oxygen. Abacavir occupies the F-pocket of the binding-groove. **B.** Zoomed in view of abacavir in the F-pocket of the binding-groove. Abacavir forms hydrogen bond interactions (dashed lines) with both peptide and HLA-B*57:01.(17)

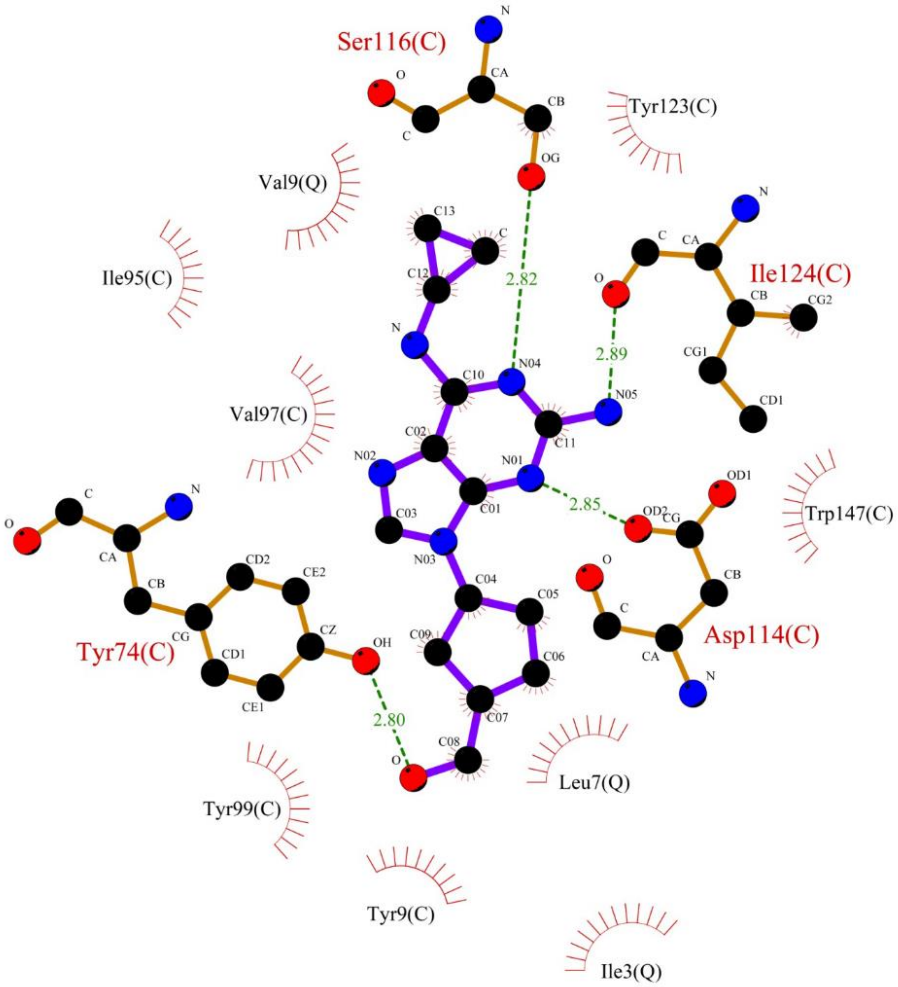

Figure 2.14. Interactions of abacavir, shown in purple, with the surrounding MHC and peptide. H bonds are shown as dashed lines and van der Waals contacts are shown as rays.(17)

To explore the biological effects of these findings, live cells were cultured by our collaborators in the presence and absence of abacavir. MHC peptides were eluted from an HLA-B*57:01 single allele-transfected 721.221 cell line treated with and without abacavir. Peptides from both samples were analyzed separately by HPLC-MS/MS to look for differences in the presented peptide repertoire. Mass spectrometric analysis of the untreated samples resulted in 682 peptides identified by search results. Peptides from

the untreated cells revealed a motif consistent with what would be expected, a small polar residue in position 2 and a large hydrophobic residue in position 9, as seen in **Figure 2. 15**.(29) An initial screen of the abacavir treated cells did not reveal any significant

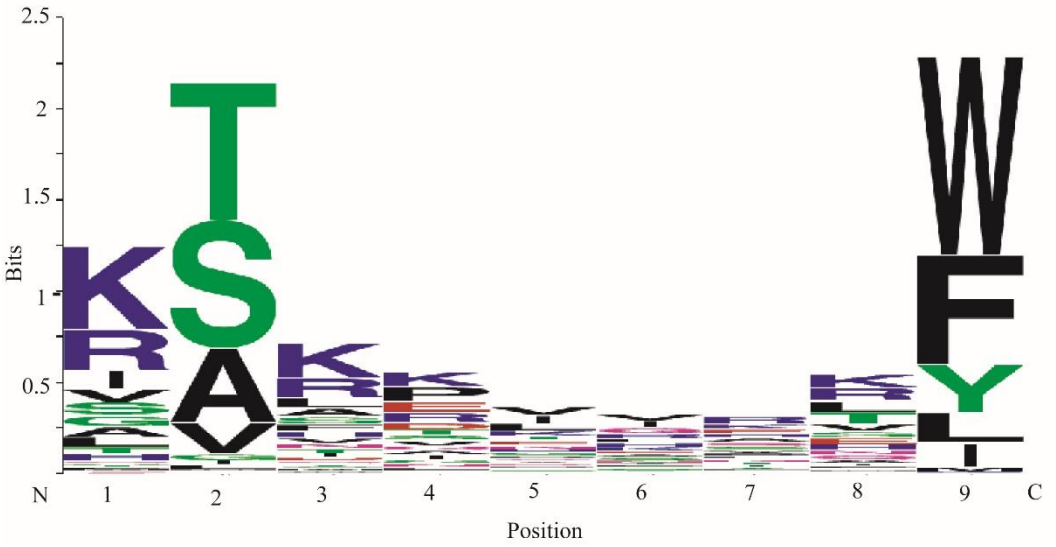

Figure 2.15. Sequence motif of eluted HLA-B*57:01 peptides.(29)

differences in the motif. This could be due to the fact that small hydrophobic residues on the C- or N-terminus have a mass that falls below the low mass cutoff (LMCO) of the ion trap making that ion unobservable. The LMCO typically follows the "1/3 rule", which sets the LMCO at 1/3 the value of the parent m/z. For example, a species selected for an MS2 with a parent m/z of 444 would have a LMCO of 144 m/z, and a C-terminal valine has a z-ion mass of 102 m/z and a y-ion mass of 118 m/z. This is problematic for this experiment, because most of the peptides from this sample have a charge of +2 and would have a parent m/z and LMCO similar to the example, thus preventing the $z\bullet_1$ or y_1 peak of a C-terminal valine residue to be present in an MS2 spectrum.

The problem presented by the LMCO can be overcome by manually setting the low mass cutoff, but this could cause suboptimal conditions for the MS2 and may

sacrifice more sequence information than what would be gained. Another way to overcome this obstacle is to derivatize the C-terminus of the peptide. Previous work in our lab has shown that derivatization of the C-terminus with histamine chemistry adds 94 Da to the residue mass of the C-terminal residue; in addition, it also adds charge to the peptide making it more suitable for ETD fragmentation.

The histamine derivatization reaction, shown in **Figure 2.16** is a simple, one-pot

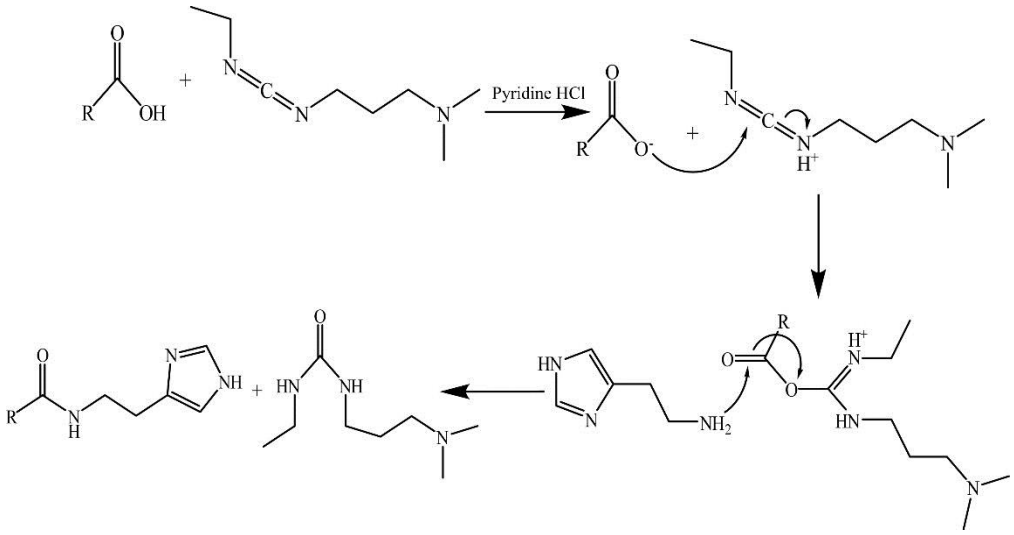

Figure 2.16. Reaction scheme for derivatization of carboxylic acids with histamine.

reaction between histamine and EDC that modifies the C-terminus and aspartic and glutamic acid residues. Derivatization with histamine increases the mass of the C-terminal ion above the LMCO. Histamine derivatization provides multiple advantages including increasing the mass of the C-terminal residue and increasing the charge of the peptide making it more compatible for ETD. An ETD spectrum of a histamine derivatized peptide produces a diagnostic peak at 112 m/z, the mass of a protonated histamine molecule, indicating that the peptide has, in fact, been derivatized. This gives

added confidence that when a peptide with a C-terminal residue of valine or isoleucine is observed, it is that residue and not a dipeptide with the same mass. **Figure 2.17** shows the advantages of performing histamine derivatization by comparing a non-derivatized peptide to its derivatized counterpart.

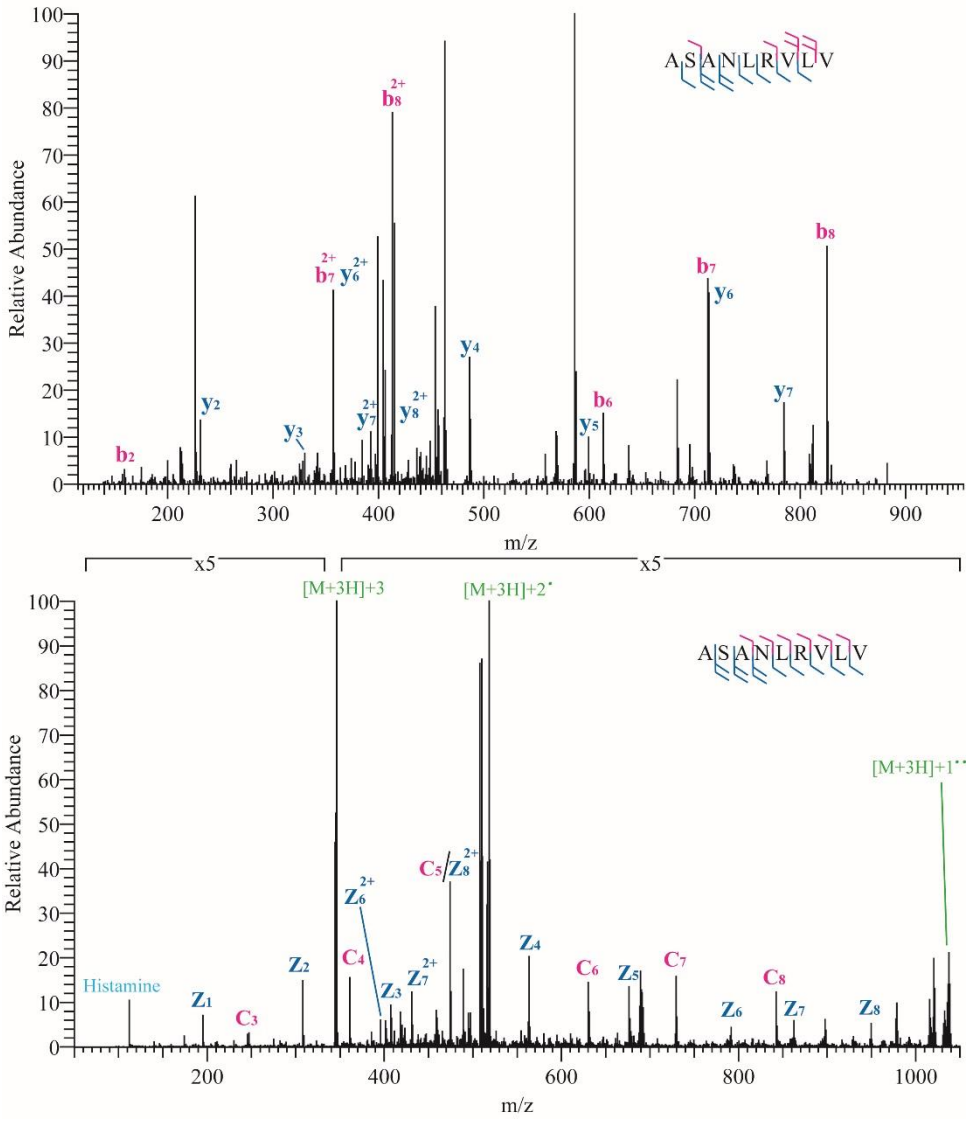

Figure 2.17. Top—2+ CAD spectrum of ASANLRVLV. Bottom—3+ ETD spectrum of histamine derivatized ASANLRVLV.

After histamine derivatization, the sample was analyzed by HPLC-MS/MS, and the instrument was operated in a data dependent mode fragmenting the top 10 most abundant, non-excluded peaks for ETD fragmentation only. This run showed promising results and confirmed the PSCL binding studies. In addition to performing OMSSA searches, in-house software was used to identify spectra with the mass of a C-terminal V, 195 m/z, and the mass of the signature protonated histamine ion, 112 m/z. This program searches all masses present in each spectrum from the run and reports the spectrum number, along with the +1 precursor masses, masses found, and intensity of the masses found. Once these results are obtained the spectra can be evaluated and de novo sequenced to identify peptides with a C-terminal V. Search results identified 539 peptides in the abacavir treated sample, and 15 of these peptides were manually confirmed to have a valine residue at the C-terminus, shown in **Table 2.2**. There was also an increase in the number of peptides with an isoleucine at the C-terminus and a decrease in the number of peptides with either a tryptophan or phenylalanine residue at the C-terminus compared to the untreated sample, as shown in **Figure 2.18**. The combinatorial assay also predicted an increase in C-terminal alanine residues which was not observed in the eluted peptides. This is can be explained by the antigen processing machinery, including proteasomal cleavage and TAP transport, which restricts the peptide repertoire available for binding to MHC and disfavors peptides with a C-terminal alanine.(17) Even though only 15 peptides with a C-terminus valine were identified, 3 of these peptides were present at levels (>100 copies/cell) placing them in the top 5% of all

Table 2.2. Summary of peptides with a C-terminal valine in the abacavir treated sample.

Sequence	Length	Accession	Source Protein	Start	Stop
ASANLRVLV	9	Q8WW24	Tetkin-4	262	270
ATIKLQSTV	9	Q8IZT6	Abnormal spindle-like microcephaly-associated	1681	1689
HSIPVTVEV	9	Q9UQL6	Histone deacetylase 5	26	34
KAAKIRVSV	9	Q6P1K8	General transcription factor IIH subunit 2-like protein	190	198
KIYEGQVEV	9	P46777	60S ribosomal protein L5	117	125
KSNGTIIHV	9	P10747	T-cell-specific surface glycoprotein CD28	127	135
KTIHLTLKV	9	O15205	Ubiquitin D	72	80
KVAKVEPAV	9	Q9P2E9	Ribosome-binding protein 1	145	153
RSARVTVAV	9	Q9UDX5	Mitochondrial fission protein MTP18	73	81
RTFHHGVRV	9	Q8NFH4	Nucleoporin Nup37	67	75
RTLAEIAKV	9	Q15233	Non-POU domain-containing octamer-binding protein	119	127
RTTETQVLV	9	P12081	Histidyl-tRNA synthetase, cytoplasmic	405	413
RVAGIHKKV	9	P26373	60S ribosomal protein L13	82	90
TVAPFNPTV	9	Q96BU1	S100P-binding protein	147	155
VTTDIQVKV	9	O00267	Transcription elongation Factor SPT5	980	988

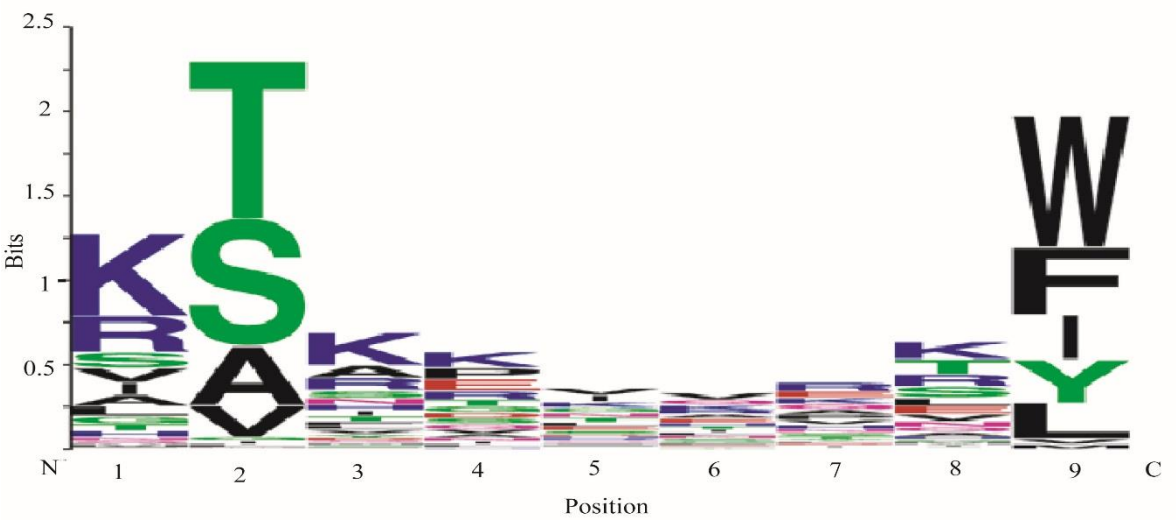

Figure 2.18. Sequence motif of eluted HLA-B*57:01 peptides treated with abacavir. There is a shift in preference in the C-terminal position.(29)

peptides presented. This is significant because increased abundance of 'non-self' peptides increases the likelihood of eliciting a T cell response. Multiple copies on the cell surface are required to trigger an immune-mediated response. **Table 2.3** summarizes the results of the C-terminal residues observed in the abacavir treated and untreated cells. In summary, 539 peptides were identified in the drug-treated sample and 682 peptides in the untreated sample, 287 of which were found in both samples. These findings confirm that there is a significant alteration in the peptide repertoire presented in the presence of abacavir.

Table 2.3. Distribution of C-terminal residues in peptides uniquely presented by abacavir treated and untreated cells.

C-terminal residue	Untreated	Treated
W	218	↓95
F	89	↓31
Y	42	↓33
L	25	25
I	14	↑45
M	6	7
V	0	↑15

The final test performed by our collaborators was to determine whether T cells could be detected in hypersensitive patients who recognize HLA-B*57:01-restricted peptides that are presented only in the presence of abacavir. Peripheral blood mononuclear cells (PBMCs) from HLA-B*57:01 positive donors with a clinical history of abacavir hypersensitivity were incubated with four pools, each of which contained three or four peptide antigens found in the elution study ending in V. This test was performed in the presence and absence of abacavir. The peptides from the pool with the

highest response were tested individually. Peptide VTTDIQVKV showed the greatest response. A significantly greater response was detected when cells were pulsed with both peptide VTTDIQVKV and abacavir compared with the response against cells pulsed with either abacavir alone or peptide alone, as shown in **Figure 2.19**. This demonstrates that memory T cell responses in abacavir-hypersensitive donors are directed against a self-peptide that requires abacavir to efficiently bind HLA-B*57:01.

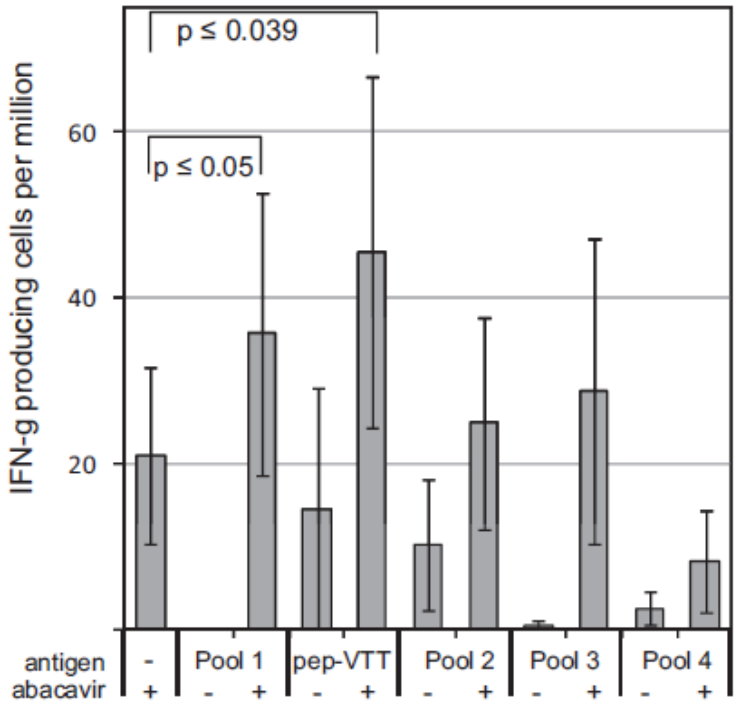

Figure 2.19. T cells from hypersensitive donors respond to specific self-peptides in an abacavir-dependent fashion. PBMCs from five HLA-B*57:01 donors with a clinical history of abacavir hypersensitivity were pulsed with peptide antigens in the presence or absence of abacavir. The figure shows the calculated mean (± SEM) IFN-γ pots per million input PBMCs. Statistically significant responses compared with the response induced by the abacavir pulse alone were obtained for peptide pool 1 and the individual peptide VTTDIQVKV.(17)

2.4.2 Acyclovir

Due to the fact that ADR's usually go undetected during the drug discovery process, our collaborators decided to explore the utility of the tests described above to drugs similar to abacavir to see if they also elicit an immune response. Ideally, there are two *in vitro* assays that could be implemented in the drug discovery process to determine if there is the potential of an ADR. The first assay is a purified HLA binding assay, which can be used together with a positional scanning combinatorial peptide libraries (PSCPL) approach to measure changes in the peptide-binding specificity in the presence or absence of the drug in question. The second is a peptide elution assay performed on single HLA allele transfected cell lines performed in the presence and absence of the drug directly identifying the effects on the self-peptide repertoire by mass spectrometry. The first step in these assays becoming a part of the drug discovery process is to determine the threshold between a drug that is safe and one that causes an immune mediated response.

Above, it was shown that the nucleoside analog abacavir induces an immune mediated response in individuals with the HLA-B*57:01 allele. Our collaborators identified drugs that were structurally similar to abacavir through a computational screening process. This screening yielded a total of seven drugs roscovitine, cladribine, acyclovir, arranon, minoxidil, sangivamycin, and bohemine; and each of these were tested for binding specificity of HLA-B*57:01 against three peptides with a C-terminal V, KVAKVEPAV, RVAGIHKKV, and HSITYLLPV. Of these seven drugs, only one, acyclovir, had an increased affinity to the three peptides with a C-terminal valine and the HLA-B*57:01 allele, shown in **Figure 2.20**. Acyclovir, structure shown in **Figure 2.21**,

has been on the market for over 30 years and is an acyclic purine nucleoside analog that is a highly potent inhibitor of herpes simplex virus and chicken pox.(30)

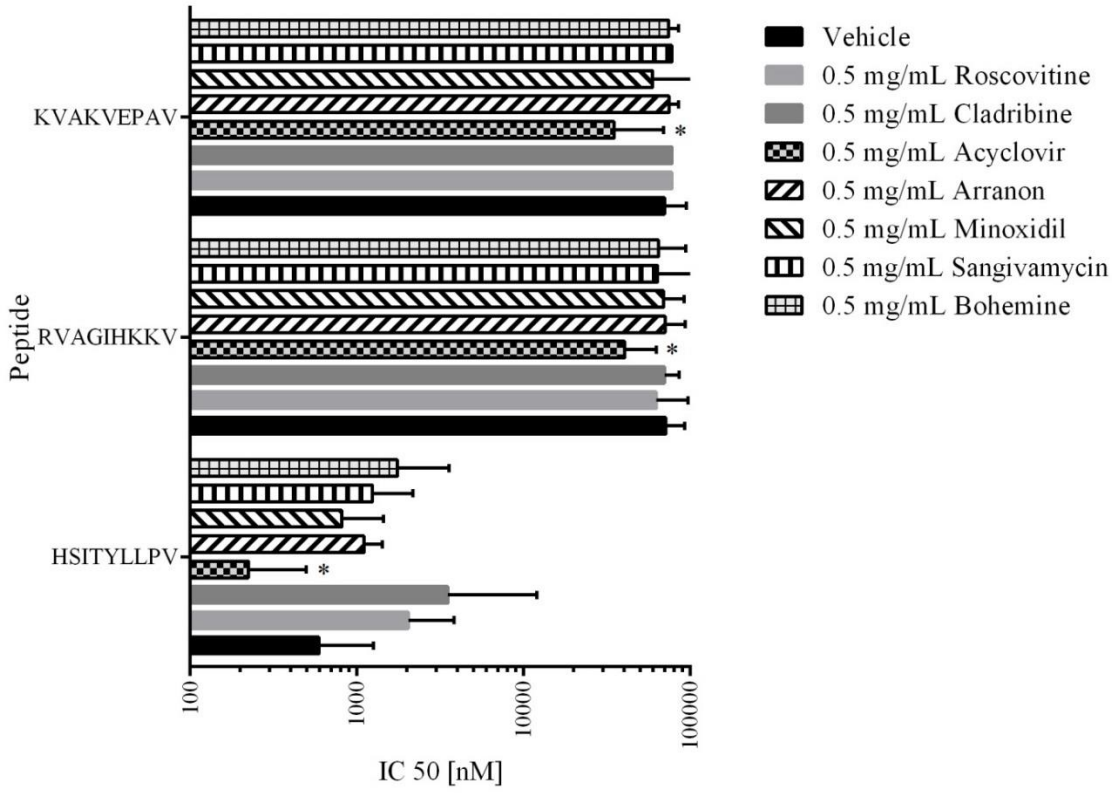

Figure 2.20. Effect of seven abacavir similar drugs on binding specificity of HLA-B*57:01. Three different peptides with a C-terminal valine were tested for binding to HLA-B*57:01 in competitive binding assays. * indicates statistical significance only in the presence of acyclovir. (22)

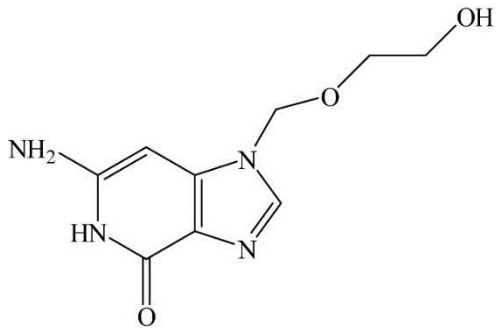

Figure 2.21. Structure of acyclovir.

Binding studies performed *in vitro*, by our collaborators, showed that acyclovir had a dose-dependent increase in the HLA-B*57:01 affinity for peptides ending in V. Abacavir was used as a positive control for these binding experiments to compare the effect acyclovir has on peptide binding. As expected, in the presence of abacavir HLA-B*57:01 showed a significant increase for peptides with a C-terminal V. Similarly, in the presence of acyclovir this increase in affinity was also observed; however, the effect was much smaller as shown in **Figure 2.22**. This study demonstrated that acyclovir has a similar effect to abacavir.

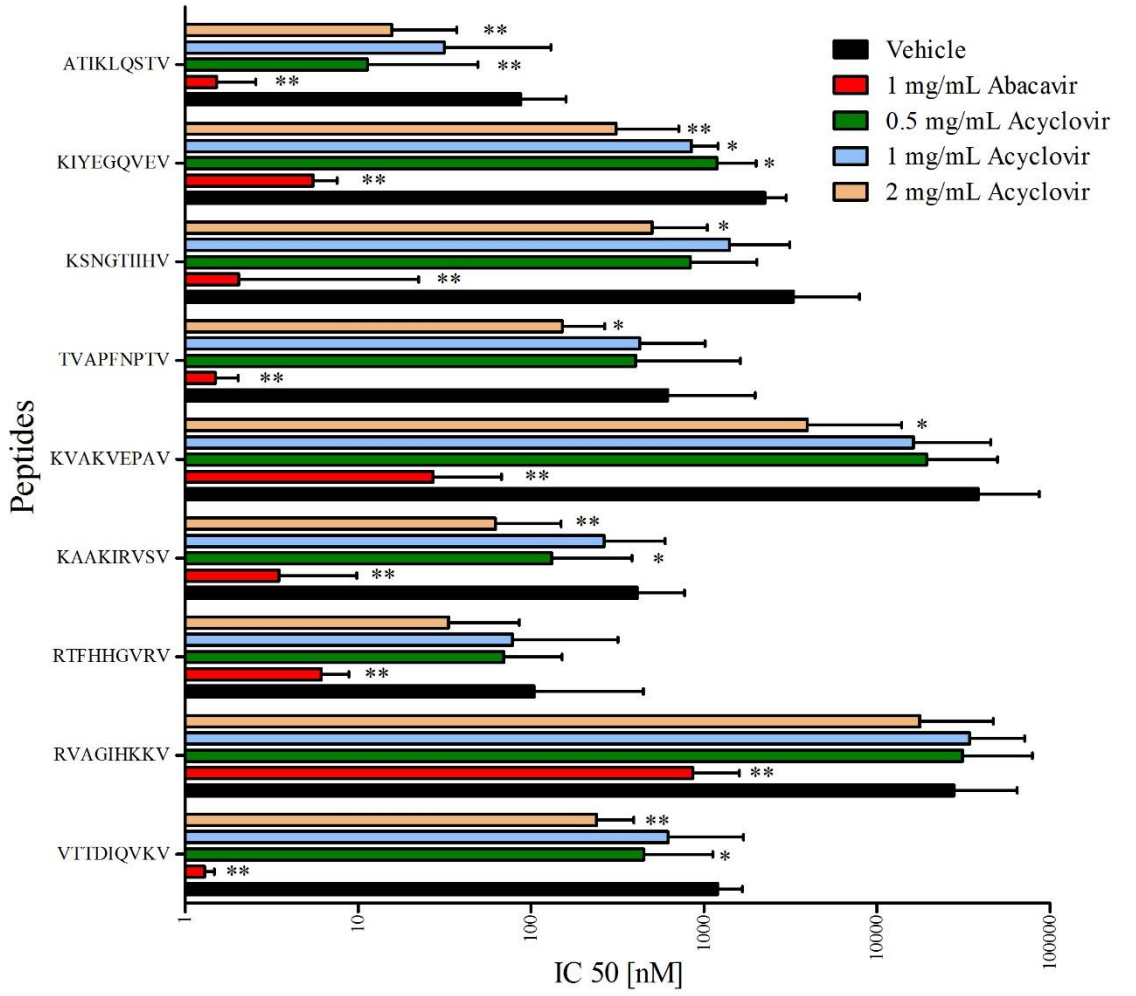

Figure 2.22. Effects of abacavir and acyclovir on HLA-B*57:01 binding specificity. Specific peptides with a C-terminal valine that showed an increased affinity for HLA-B*57:01 in the presence of abacavir were tested. Values are represented as geometric mean with 95% CI of two independent runs in triplicates, analyzed for statistical significance by Mann-Whitney U test comparing log IC_{50} values vs. vehicle, $p<0.05$ was considered significant (*$p<0.05$, **$p<0.001$).(22)

Next, PSCPL studies were performed by our collaborators to show the effect acyclovir has on HLA-B*57:01 peptide-binding. This assay systematically scans the effect of each naturally occurring amino acid in the ninth position. It was found that the greatest increases observed were for the C-terminal residues of C, I, and V, each of these residues showed a 2-fold increase in HLA-B*57:01 affinity. **Figure 2.23** shows the binding affinity to HLA-B*57:01 of all 20 amino acids in the C-terminal position in the presence of acyclovir. Abacavir, on the other hand, induces a 10-fold increase in affinity for I and V. Both acyclovir and abacavir induce a change in affinity for the C-terminal residue for HLA-B*57:01, but do so to a different degree.

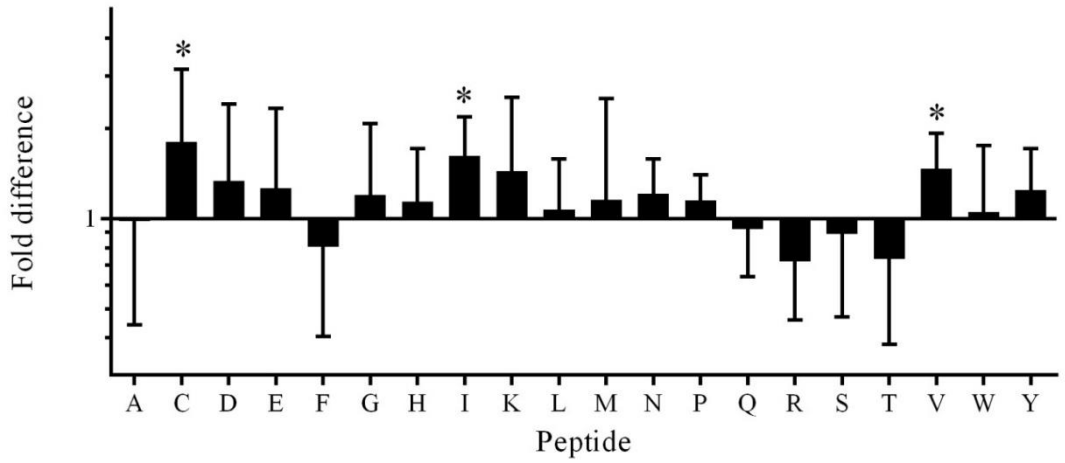

Figure 2.23. Effects of acyclovir (2 mg/mL) on the affinity of C-terminal residues for HLA-B*57:01. The highest affinity increases in the presence of acyclovir were found for peptides with a C, I, and V at the C-terminus. Values are represented as geometric mean with 95% CI of the fold difference between vehicle/acyclovir treatment. *indicates statistical significance.(22)

To test these results, seven peptides were selected with random sequences, with the only requirement of having an I in the ninth position, and seven other peptides were selected in the same way but with a C in the ninth position. These peptides were tested by our collaborators in the presence of acyclovir at varying concentrations, and also in the presence of abacavir, as a positive control. **Figure 2.24** shows that of the 7 peptides

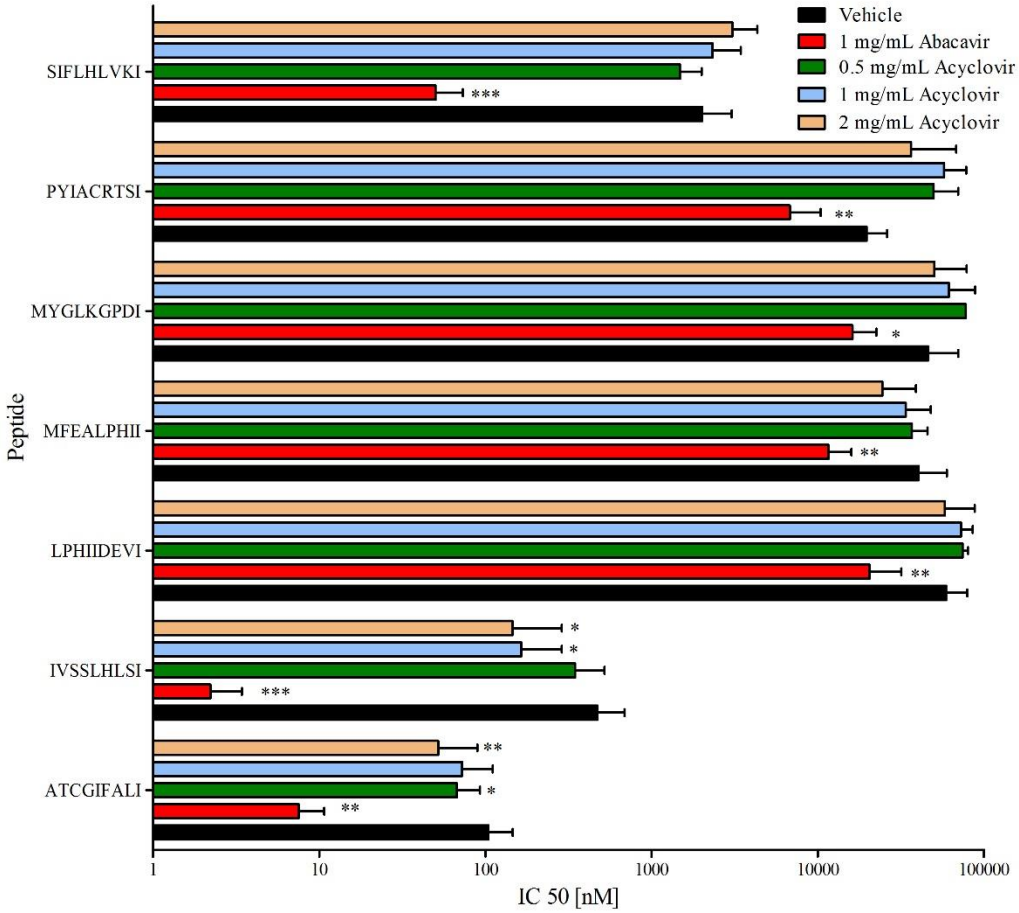

Figure 2.24. Effects of abacavir and acyclovir on HLA-B*57:01 binding specificity. Specific peptides with a terminal isoleucine that showed an increased affinity for HLA-B*57:01 in the presence of abacavir were tested. Values are represented as geometric mean with 95% CI of two independent runs in triplicates, analyzed for statistical significance by Mann-Whitney U test comparing log IC_{50} values vs. vehicle, $p<0.05$ was considered significant (*$p<0.05$, **$p<0.01$, ***$p<0.001$).(22)

with a C-terminal I, only two showed a dose-dependent increase in affinity for HLA-B*57:01 in the presence of acyclovir. In contrast, all seven peptides with a C-terminal I showed increased affinities in the presence of abacavir. None of the peptides with a C in the ninth position showed an increased affinity in the presence of either acyclovir or abacavir. These results indicate that acyclovir does have an effect on the peptide-binding affinity for HLA-B*57:01, but does so to a smaller degree than abacavir.

The ability of a peptide to cause an immune response hinges on its ability to bind to an MHC molecule. To measure this, our collaborators had a peptide with the sequence KAAKYRVSV was synthesized. The direct peptide binding assay demonstrated that in the presence of acyclovir the peptide KAAKYRVSV binds with increasing affinity in a dose-dependent manner. Similarly, this is also observed in the presence of abacavir, but with a greater effect, up to five-fold compared to two-fold in the presence of acyclovir, as shown in **Figure 2.25A**. As a control, binding to the structurally similar HLA-B*58:01 was tested, and was found to be unaffected by the presence of either acyclovir or abacavir, as shown in **Figure 2.25B**. This assay demonstrates that in the presence of acyclovir a peptide with a C-terminal V is able to bind to the MHC molecule, an integral component in being able to induce an immune response.

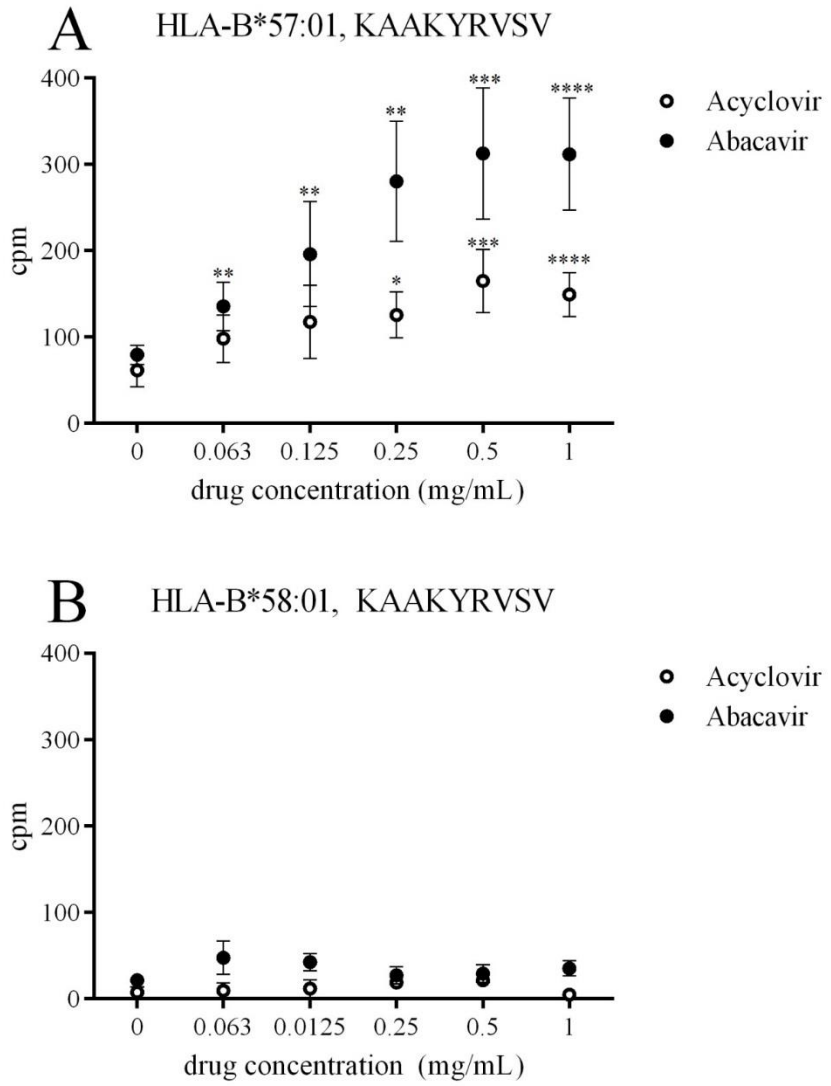

Figure 2.25. Acyclovir and abacavir alter the binding specificity of **A.** HLA-B*57:01 and **B.** HLA-B*58:01. The peptide KAAKYRVSV was radiolabeled and tested for binding to HLA-B*57:01 in increasing doses of acyclovir and abacavir. Values are represented as geometric mean with 95% CI of four experimental runs in triplicates, analyzed for statistical significance by one-sided Mann-Whitney test (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).(22)

Predictions made by the binding studies with the computational library assays were tested by analyzing peptides eluted from an HLA-B*57:01 single allele transfected cell line in the presence or absence of acyclovir. Peptide samples were analyzed by LC-MS/MS. An initial screen of both samples did not yield any peptides that were specific to the sample treated with acyclovir. These results did not provide support for the results from the combinatorial assays. Since the drug treated sample displayed favor for smaller residues at the C-terminal position similar to abacavir, we decided to perform histamine chemistry on the sample. Histamine chemistry increases the mass of the C-terminal residue above the LMCO making the C-terminal z•- and y-ion observable in the ETD and CAD MS2 spectra, respectively. Additionally, histamine derivatization allows us to use an in-house software alongside OMMSA searches to identify as many peptides as possible. While there were a higher number of peptides detected and identified with a C-terminal isoleucine in the acyclovir treated sample, there was only one peptide exclusively detected in the drug sample after searching for the mass of the other identified peptides in the MS1 of the untreated sample. After derivatization with histamine, we were able to identify one peptide, RARQLNYTI, detected exclusively in the acyclovir treated sample. This peptide, RARQLNYTI, was synthesized to undergo further tests by our collaborators to investigate its effect on binding to HLA-B*57:01 in the presence of acyclovir. In the presence of acyclovir the binding of the peptide was increased by approximately two-fold, as shown in **Figure 2.26**. These results do not show a noticeable difference between the acyclovir treated sample and the untreated sample.

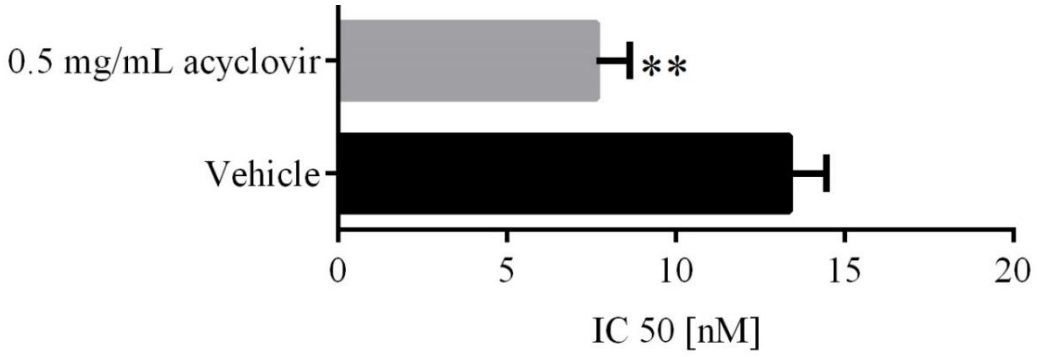

Figure 2.26. Effect of acyclovir on HLA-B*57:01 binding specificity for RARQLNYTI. Values are represented as Mean ± SEM and analyzed for statistical significance by Mann-Whitney U test comparing IC_{50} values vs. vehicle; *$p<0.05$ was considered significant (**$p<0.01$).(22)

The final test examined the ability of acyclovir to induce an HLA-B*57:01 T cell response. The effect of acyclovir was tested using PBMC from two healthy HLA-B*57:01 positive, HIV-negative, and abacavir unexposed donors, and was compared to abacavir, positive control, and no drug, negative control. As expected, in the abacavir treated sample an expanded population of CD8 positive T cells were observed. By contrast, no expansion of T cells were observed in the acyclovir treated cells and the untreated sample. This suggests that under *in vitro* conditions acyclovir does not have the capacity to stimulate T cells and elicit an immune mediated response. This conclusion is further validated by the fact that acyclovir has been on the market for over 30 years with no evidence of an ADR.

2.5 Conclusion

These results indicate that an alternative mechanism is responsible for the activation of T cells in the case of abacavir. The above data provide a mechanistic explanation for abacavir-induced drug hypersensitivity. The altered peptide binding repertoire induced in the presence of abacavir provides evidence for a different mechanism of T cell stimulation, the altered peptide repertoire model, shown in **Figure 2.27**. This model states that the drug covalently binds to the binding-groove of the HLA molecule preventing the normal peptide repertoire from binding to the HLA molecule. Instead novel peptides are able to bind to the HLA molecule. This discovery was made simultaneously with two other labs all of whom published similar findings within weeks of each other.(21, 31) Our results are unique compared to these other studies, as we were the only one able to identify peptides in the treated sample with a C-terminal V residue. The incorporation of this altered repertoire is a unique scenario since it is technically 'self' peptides being presented as 'non-self' causing an autoimmune response.

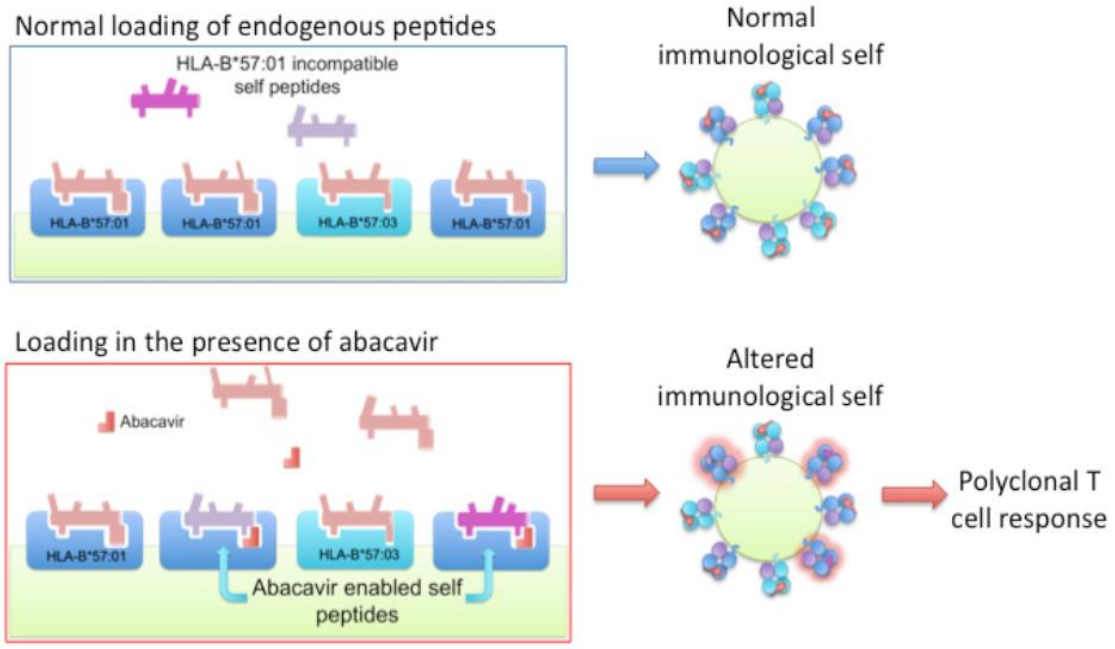

Figure 2.27. Mechanism of T cell stimulation by the altered peptide repertoire caused by the presence of abacavir.(20)

The altered-peptide repertoire model for T cell stimulation is a new concept that potentially explains the mode of T cell stimulation for other drugs. Currently this model explains abacavir induced hypersensitivity and carbamazepine induced hypersensitivity. Carbamazepine hypersensitivity has a strong association with the HLA-B*15:02 allele in Asian populations. Illing and colleagues investigated the association between HLA-B*15:02 and carbamazepine by identifying MHC eluted peptides in the presence and absence of the drug similar to the assay for abacavir. They found a shift in the preferred amino acid side-chain at positions 4 and 6.(21) They found a preference for smaller residues at these positions, but did not observe a shift in anchor residue preference. The magnitude of this shift in preference was about 15% smaller than that observed for abacavir.(21) This illustrates two examples that are consistent with the mechanism for T

cell stimulation in adverse drug reactions, and the plethora of drug hypersensitivities known make it likely that this is a general mechanism.

Since idiosyncratic adverse drug reactions are unpredictable, they usually go undetected during the drug discovery process. However, the assays performed to study the effect of abacavir could be utilized in the drug discovery process. They could be used as a screening strategy to identify drugs that have the highest risk of causing an ADR. Then subsequent *in vitro* testing could be performed on the identified drugs to determine the likelihood that they would in fact cause an ADR. Furthermore, patients could then be tested for the specific allele that might cause an ADR prior to administration of the medication. In the case of abacavir, patients are now HLA typed prior to taking the medication to prevent negative side effects from the prescription. Screening for drugs that have similar structures to those that cause an ADR, and performing similar tests, both computational and biological assays on those drugs as was done for abacavir and acyclovir can potentially identify drugs that could cause an allele specific ADR and require HLA typing before administering the drug. Implementing such assays as those performed in the study of abacavir and acyclovir will not only decrease the number of ADR's in patients, it will also save a significant amount of money.

2.6 References

1. Reinherz EL. A false sense of non-self. *Nature.* 2012; 486: 479-481.

2. Shastri N, Schawb S, Serwold T. Producing nature's gene-chips: The generation of peptides for display by MHC class I molecules. *Annu Rev immunol.* 2002; 20: 463-493.

3. Male D, Brostoff J, Roth DB, Roitt IM. Immunology. 8th ed. United States: Elsevier/ Saunders; 2013.

4. Bharadwaj M, Illing P, Theodossis A, Purcell AW, Rossjohn J, McCluskey J. Drug hypersensitivity and human leukocyte antigens of the major histocompatibility complex. *Annu Rev Pharmacol Toxicol.* 2012; 52: 401-431.

5. Klein J, Sato A. The HLA system. *N Engl J Med.* 2000; 343: 702-709.

6. Neefies J, Jongsma MLM, Paul P, Bakke O. Towards a systmes understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immuno.* 2011; 11: 823-836.

7. Chang S, Momburg F, Bhutani N, GOldberg AL. The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a "molecular ruler" mechanism. *Proc Nat Acad Sci.* 2005; 102(47): 17107-17112.

8. Yewdell JW, Reits E, Neefjies J. Making sense of mass destruction: Quantitating MHC class I antigen presentation. *Nat Rev Immuno.* 2003; 3: 952-961.

9. Klein L, Kyewski B, Allen PM, Hogquist KA. Positive and negative selection of the T cell repertoire: What thymocytes see (and don't see). *Nat Rev Immuno.* 2014; 14: 377-391.

10. Parham P. The immune system. 3rd ed. London; New York: Garland Science; 2009.

11. Chung W, Hung S, Chen Y. Human leukocyte antigens and drug hypersensitivity. *Curr Opin Allergy Clin Immunol.* 2007; 7: 317-323.

12. Pompeu YA, Stewart JD, Mallal S, Phillips E, Peters B, Ostrov DA. The structural basis of HLA-associated drug hypersensitivity syndromes. *Immunological Reviews.* 2012; 250: 158-166.

13. Edwards IR, Aronson JK. Adverse drug reactions: Definitions, diagnosis, and management. *Lancet.* 2000; 356: 1255-1259.

14. Adam J, Pichler WJ, Yerly D. Delayed drug hypersensitivity: Models of T cell stimulation. *Br J Clin Pharmacol.* 2010; 71(5): 701-707.

15. Pan R, Wu Y, Chung W, Hung S. HLA and TCR recognition of medications in severe cutaneous adverse reactions. *Curr Immuno Rev.* 2014; 10: 51-61.

16. Pichler WJ. Pharmocological INteraction of drugs with antigen-specific immune receptors: The p-i concept. *Curr Opin Allergy Clin Immunol.* 2002; 2: 301-305.

17. Ostrov DA, Grant BJ, Pompeu YA, Sidney J, Harndahl M, Southwood S, Oseroff C, Lu S, Jakoncic J, de Oliveira, C. A. F., Yang L, Mei H, Shi L, Shabanowitz J, English AM, Wriston A, Lucas A, Phillips E, Mallal S, Grey HM, Sette A, Hunt DF, Buus S, Peters B. Drug hypersesitivity caused by alteration of the MHC-presented self-peptide repertoire. *Proc Nat Acad Sci.* 2012; 109(25): 9959-9964.

18. Illing P, Vivian JP, Purcell AW, Rossjohn J, McCluskey J. Human leukocyte antigen-associated drug hypersensitivity. *Curr Opin Immunol.* 2013; 25: 81-89.

19. Pavlos R, Mallal S, Phillips E. HLA and pharmacogenetics of drug hypersensitivity. *Pharmacogenomics.* 2012; 13(11): 1285-1306.

20. Mallal S, Nolan D, Witt C, Masel G, Martin AM, Moore C, Sayer D, Castley A, Mamotte C, Maxwell D, James I, Christiansen FT. Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV1 reverse-transcriptase inhibitor abacavir. *Lancet.* 2002; 359: 727-732.

21. Illing P, Vivian JP, Dudek NL, Kostenko L, Chen Z, Bharadwaj M, Miles JJ, Kjer-Nielsen L, Gras S, Williamson NA, Burrows SR, Purcell AW, Rossjohn J, McCluskey J. Immune self-reactivity triggered by drug-modified HLA-peptide repertoire. *Nature.* 2012; 486: 554-560.

22. Metushi IG, Wriston A, Banerjee PS, Gohlke BO, English AM, Lucas A, Moore C, Sidney J, Buus S, Ostrov DA, Mallal S, Phillips E, Shabanowitz J, Hunt DF, Preissner R, Peters B. Acyclovir has low but detectable influence on HLA-B*57:01 specificity without inducing hypersensitivity. *PLOSONE.* 2015.

23. Martin SE, Shabanowitz J, Hunt DF, Marto JA. Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem.* 2000; 72: 4266-4274.

24. Earley L, Anderson LC, Bai DL, Mullen C, Syka JEP, English AM, Dunyach J, Stafford GC, Shabanowitz J, Hunt DF, Compton PD. Fron-end electron transfer dissociation: A new ionization technique. *Anal Chem.* 2013; 85(17): 8385-8390.

25. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts, and proteins. *Nucl Acids Res.* 2007; 35: D61-D65.

26. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucl Acids Res.* 2000; 28(1): 45-48.

27. Greer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, Bryant SH. Open mass spectrometry search algorithm. *J Prot Res.* 2004; 3: 958-964.

28. Marsh SG, Parham P, Barber LD. The HLA FactsBook. 1st ed. San Diego: Academic Press; 2000.

29. Crooks GE, Hon G, Chandonia J, Brenner SE. WebLogo: A sequence logo generator. *Genome Res.* 2004; 14: 1188-1190.

30. Elion GB. Mechanism of action and selectivity for acyclovir. *Am J Med.* 1982; 73: 7-13.

31. Norcross MA, Luo S, Lu L, Boyne MT, Gomartelli M, Rennels AD, Woodcock J, Margulies DH, Mcmurtrey C, Vernon S, Hildebrand WH, Buchli R. Abacavir INduces loading of novel self-peptides into HLA-B*57:01: An autoimmune model for HLA-associated drug hypersensitivity. *Aids.* 2012; 26(11): F21-F26.

Chapter 3: Elucidation of Macaque MHC Peptide Motifs

3.1 Introduction

Animal models have been used in scientific research studies for over a hundred years, and have been instrumental across many areas of research including diabetes, epilepsy, and HIV/AIDS to name a few. Throughout the years, numerous different species have been utilized for research purposes from rabbits to mice to worms to monkeys. The animal model used is dependent upon which disease or stressor is under investigation. Animal models have become integral in the development of drugs, insight into disease progression, and understanding mechanisms by which a disease affects the body. Animal models are advantageous because they allow researchers to better understand human illnesses and diseases without doing harm to individuals. The translational significance that animal models provide is invaluable to the understanding of disease mechanisms and therapeutics in human beings.

There have been numerous scientific breakthroughs and advancements using animal models. However, animal models do not always provide positive results and can hinder scientific advancement, as there is no perfect complete model system for the human body. One example of interference came at the turn of the 20th century when Dr. Simon Flexner was attempting to develop a polio vaccine. Dr. Flexner used a rhesus macaque of Indian origin as the animal model. The misstep occurred because rhesus macaques are a rare breed of monkey in which the poliovirus does not replicate in the digestive tract. This led to the hypothesis that a vaccine candidate would need to be designed to grow only in neural cells, which is not the case.(1) The use of this

incompatible animal model delayed the discovery of a polio vaccine for over 49 years, but fortunately did not prevent one from being developed with time. The lesson from this unfortunate event is that there is a responsibility of researchers to fully understand the differences between animal models and the human body.

Fortunately, the use of rhesus macaques in scientific research was not tainted by its interference in the discovery of a polio vaccine. In fact, rhesus macaques have been an integral part of scientific research into human illnesses and diseases for several decades. Rhesus macaques and humans shared a common ancestor 25 million years ago, and this is reflected by a highly similar immune system making the rhesus macaque an attractive animal model in biomedical research.(2) Rhesus macaques serve as non-human primate models and have been used extensively in the study of over 70 human infectious diseases including bacteria, viruses, fungi, parasites, and prions.(3) Macaques have advanced the understanding of influenza, smallpox, rabies, and rotaviruses among others.(3) Most notably, macaques have provided worthwhile insight into the pathogenic mechanisms of acquired immunodeficiency syndrome (AIDS), caused by the infection of human immunodeficiency virus (HIV). In some cases macaques are inefficient at modeling human diseases due to incompatibility between the two species; however, parallels can still be drawn between the species which can aid in the design of future experiments and clinical trials.(4)

Currently there are three species of macaques typically used in biomedical research: rhesus macaque, pig-tailed macaque, and cynomolgus macaque; shown in **Figure 3.1**.(5) The most popular macaque animal model has been the rhesus macaque of

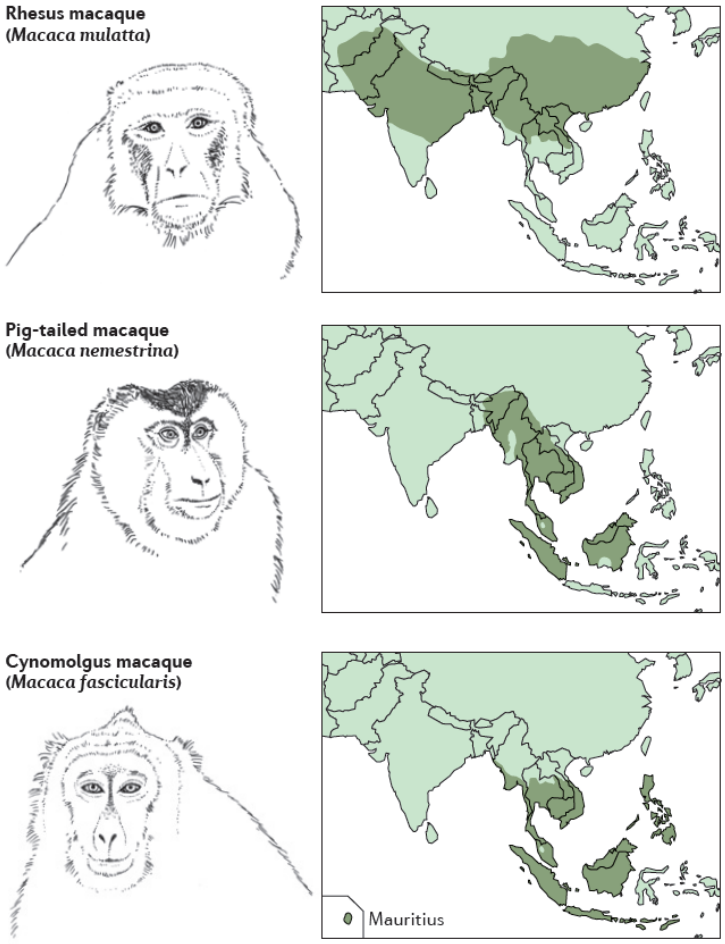

Figure 3.1. Species of macaques commonly used in scientific research and their geographical ranges.(5)

Indian descent. The Indian rhesus macaque has been well characterized and provides researchers with a useful animal model. However, an export ban on Indian descent rhesus macaques placed in 1978 has made this species increasingly harder to obtain for research.(6) Currently, Indian rhesus macaques have a high degree of inbreeding, and are only available from U.S. breeders causing a discrepancy between supply and demand.

This gap has been decreased through an increase in the use of rhesus macaques of Chinese descent. The pig-tailed macaque also has the potential to assume a larger role and shoulder some of the burden caused by the lack of availability of the Indian rhesus macaques.(5) However, before these species can be used as a substitute filling the void made by the shortage of Indian rhesus macaques for research, a more in-depth understanding of their immune system is required.

When investigating a human disease in an animal model, it is important understand both species' immune systems. This a necessity to apply what is learned in the animal system to the human system. The immune system is one of the most intricate systems in the human body and is instrumental in the progression or abatement of illnesses and diseases. The immune system is comprised of two different types of immune responses: the innate immune response which is the first line of defense against foreign invaders and results in the same response every time; and the adaptive immune response which becomes more powerful following repeated encounters with the same antigen.(7) Characteristics of each arm of the immune system are outlined in **Table 3.1**.(7, 8) As previously mentioned, the adaptive arm of the immune system employs T cells to survey cell health by interacting with MHC-peptide complexes. This is accomplished through antigen processing and presentation of peptide antigens to T cell receptors. Briefly, proteins are degraded by the proteasome, the resulting peptides are transported to the ER where they are loaded onto MHC class I molecules, and the stable MHC-peptide complexes then exits the ER and travels through the Golgi apparatus to the cell surface where peptide antigens are presented to $CD8^+$ T cells, as shown in **Figure 3.2**.(9) The T cell distinguishes between self and foreign peptide antigens and depending

on the outcome will either elicit an immune response or continue its surveillance. As demonstrated in the previous chapter this arm of the immune system can provide valuable insight as to how the immune system functions upon on an attack. This chapter will focus on MHC molecules and their peptide antigens in Chinese rhesus macaques and pig-tailed macaques.

Table 3.1. Characteristics of innate and adaptive immunity.(7,8)

Innate Immunity	Adaptive Immunity
• General protection	• Highly specific for a particular antigen
• Early phase of host response to pathogens without requiring prior exposure	• Late phase response of antigen-specific lymphocytes to antigens
• Immediate maximal response	• Lag time between exposure and maximal response
• Is not altered on repeated exposure	• Improves with each successive exposure (immunological memory)
• Rapid response, non-specific	• Slower, exhibits specificity
• Limited diversity	• Wide diversity

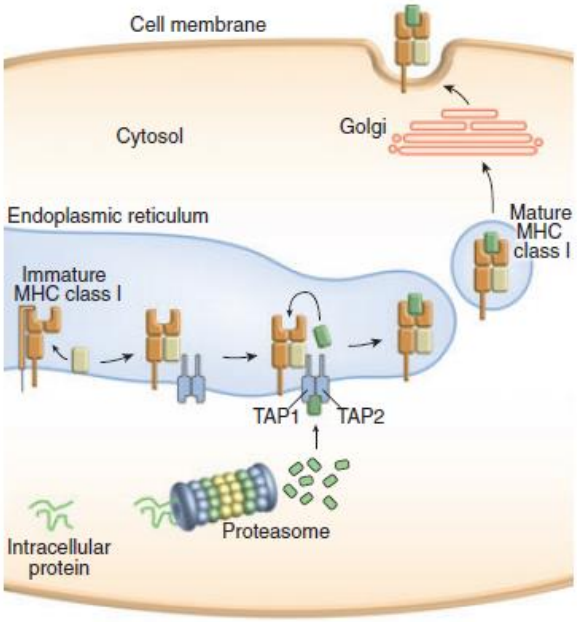

Figure 3.2. MHC class I antigen processing and presentation pathway.(9)

Rhesus macaques have been an invaluable resource in the study of HIV/AIDS, even though they are not capable of contracting HIV. Macaques acquire a lentivirus, simian immunodeficiency virus (SIV), which is very similar to HIV. HIV/AIDS was first discovered in 1981 and has since claimed the lives of millions. It is a global pandemic that affects over 35 million people worldwide.(1) HIV has very few genes, less than 10, and produces more virus particles by invading a cell and converting viral RNA into DNA and then reverse transcriptase converts it back to RNA.(10) This trait of HIV makes it very difficult to treat because it increases the likelihood of mutations due to reverse transcriptase's lack of 'proofreading'.(10) These mutations give rise to different sub-types, clades, which can be identified in patients so an appropriate treatment plan can be designed. Interestingly, it is possible to be infected with multiple different clades of HIV at the same time.(10)

Normally when a virus is detected in our body, our immune system attacks the invader and destroys it; however, HIV is not normal, it attacks the $CD4^+$ T cells and macrophages of the immune system. HIV typically manifests into AIDS 5-10 years after exposure.(11) Currently, there are drugs on the market that can stave off the transition to AIDS for a couple of years, but these drugs do not keep the disease at bay forever. Captivatingly, it is thought that HIV was transfected to humans through chimpanzee and monkey blood during the butchering of bush meats in the 1930's. A cure or vaccine for HIV has remained elusive for the past thirty years, despite initial thoughts that it would only take two years to discover a cure.(12)

SIV, the lentivirus found in rhesus macaques, is asymptomatic and can be transmitted through wounds, bites, and sexual contact.(3) While SIV has been beneficial to the study of HIV in humans, it does have shortcomings. HIV and SIV are very similar in their respective hosts in pathogenesis, pathology, and molecular makeup; however, small differences have major consequences when switching species.(3) This led researchers to generate a better mimetic for HIV, which they did by incorporating genes and aspects of HIV in the SIV, thus creating a chimeric HIV/SIV, SHIV.(14) This strain of the virus has produced promising results, and even led to clinical trials for some vaccines and treatments.

Macaque animal models are an integral component of HIV/AIDS research due, in part, to the species being well characterized, including its immune system. The challenge that has been presented by the ban on Indian rhesus macaques has brought about the need to find not only an adequate replacement, the Chinese rhesus macaque and pig-tailed macaque, but also to characterize and understand the immune systems of these species. MHC alleles have a critical role in adaptive immunity, which is utilized in the fight against HIV. In fact, certain HLA alleles have shown to be elite controllers of HIV, meaning they are capable of reducing the viral load of HIV, delaying the onset of AIDS. Specifically, HLA-B*27 and some HLA-B*57 alleles possess this trait. Elite controllers provide the potential to harness their ability in the form of a vaccine or drug to treat HIV. Additionally, $CD8^+$ T cells play a central role in viral containment of SIV infection illustrating how important adaptive immunity is to the study of HIV.(15) A deeper understanding of how SIV affects adaptive immunity is dependent on our knowledge of

the characteristics of macaques' MHC alleles and their similarities to human MHC alleles.

Human classical MHC class I alleles are composed of a single polymorphic HLA-A, HLA-B, and HLA-C locus; alternatively rhesus macaque (*macaca mulatta*) MHC class I alleles have variable numbers of Mamu-A and Mamu-B loci, shown in **Figure 3.3**.(16) Human HLA genotypes affect disease progression in HIV infection, and knowing which alleles display delayed progression could be the difference in designing

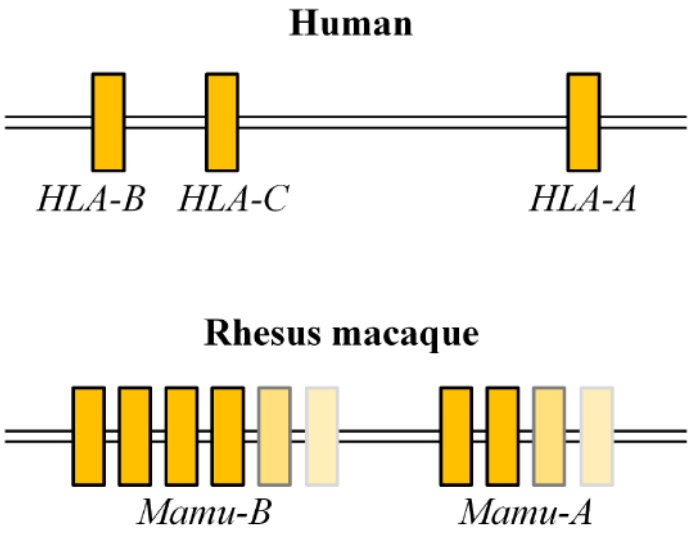

Figure 3.3. Comparison of genome structure of MHC class I alleles in human and rhesus macaque.(16)

an effective vaccine or treatment.(16) To do this, animal models are required to gain deeper insight into the interaction between host and virus. The key to relaying lessons learned in animals to humans is to determine the similarities between the two. In this study, the peptide binding motifs of MHC class I alleles from Chinese rhesus macaque and pig-tailed macaques are characterized.

3.2 Materials

3.2.1 Reagents

Fisher Scientific, Waltham, MA

Methanol

Honeywell, Morristown, NJ:

Acetonitrile, HPLC grade, ≥99.8% purity

Pierce, Rockford, IL:

LC-MS grade water

Sigma-Aldrich, St. Louis, MO:

Acetic Acid (AcOH, glacial, 99.99%)

Angiotensin I human acetate hydrate, ≥90% purity (Angio)

Azulene

Fluoranthene

Formamide

Vasoactive intestinal peptide fragment 1-12, human ≥97% (Vaso)

3.2.2 Column Making Materials

Polymicro Technologies, Inc., Phoenix, AZ:

Polyimide coated fused silica capillary

360 μm o.d. × 50 μm i.d.

360 μm o.d. × 75 μm i.d.

360 μm o.d. × 150 μm i.d.

Zeus Industrial Products, Inc., Orangeburg, SC:

Teflon tubing, 0.012” i.d. × 0.060” o.d.

PQ Corporation, Valley Forge, PA:

Kasil-Potassium silicate solution

YMC Corp., Morris Plains, NJ (available through Waters Corp.):

ODS-AQ, C-18, 5 μm spherical silica particles, 120 Å pore size

ODS-AQ, C-18, 5-20 μm spherical silica particles, 300 Å pore size

3.2.3 Equipment

Agilent Technologies, Palo Alto, CA:

1100 Agilent HPLC

Branson, Danbury, CT

Branson 1200 Ultrasonic Bath

Eppendorf, Hamburg, Germany:

Centrifuge 5424

Labconco, Kansas City, MO:

CentriVap Vacuum Concentrator System

Sutter Instrument Co., Novato, CA:

P-2000 laser puller with fused silica adapter

Thermo Fisher Scientific, Waltham, MA:

LTQ mass spectrometer (custom modified with backend ETD)

LTQ FT-ICR hybrid mass spectrometer (custom modified with front-end ETD)

LTQ-Orbitrap mass spectrometer (custom modified with front-end ETD)

LTQ-Orbitrap Velos mass spectrometer (custom modified with front end ETD)

VWR, Radnor, PA:

Mini vortexer

3.3 Methods

John Sidney and Carrie Moore, of the Sette Laboratory at the La Jolla Institute for Allergy and Immunology, isolated MHC peptide samples and conducted peptide binding assays.

3.3.1 Isolation of Class I MHC Peptides from Cultured Cells Transfected with Macaque Alleles

Stable MHC class I transfectants were produced in the MHC class I-deficient EBV-transformed B-lymphoblastoid cell line 721.221. The cell lines transfected with a single macaque MHC class I allele were expanded and harvested to generate a pellet equivalent to approximately 5×10^9 cells. Cells were lysed using cell lysis buffer in the presence of protease inhibitors. The lysate was centrifuged at 100,000 × g for 1 hr and the supernatant was collected and passed through a 0.8/0.2 μm filter. The filtrate was collected and passed through a sepharose CL-4B precolumn, an irrelevant MHC class I antibody affinity column to assess nonspecific binding (antimouse MHC class I MKD6), and then W6/32 antibody affinity column to bind MHC class I molecules. The columns were washed sequentially with: 1) 2 column volumes of lysis buffer, 2) 20 column

volumes of 20 mM Tris-HCl (pH 8.0, 1150 mM NaCl), 3) 20 column volumes of 20 mN Tris-HCl (pH 8.0, 1 M NaCl), and 4) 20 column volumes of 20 mM Tris-HCl (pH 8.0). Separately, each affinity column was eluted with 4 column volumes of 0.2 N glacial acetic acid. The eluted peptides were then collected and spun at 3,500 × g at 4°C until 98% of the solution had passed through Millipore ultrafiltration units with a 3,500 Da cut-off to exclude β-2 microglobin. The filtrate was then collected and concentrated in vacuum for LC-MS analysis.(17)

3.3.2 Assembly of Capillary HPLC Columns

Precolumns (PC) were constructed by creating a Kasil frit in the end of a fused-silica capillary column (360 μm o.d. × 75 μm i.d.). Kasil frits were formed by dipping the end of a capillary column into a mixture of Kasil:formamide (3:1) and heating in an oven overnight at 70˚C. The fritted column was packed with 5-8 cm of irregular, 5-20 μm diameter, 300 Å pore size, C18 reverse-phase resin using a pressure bomb (500 psi). Analytical columns (AC) were constructed with integrated electrospray emitter tips.(18) A Kasil frit was created in a short section of the fused silica (360 μm o.d. × 50 μm i.d.). Columns were then packed with 6-9 cm of 5 μm diameter, 120 Å pore size, C18 reverse-phase resin using a pressure bomb (500 psi). A micropipette laser puller was used to generate a tip (2 μm i.d.) approximately 5 mm from the Kasil frit. Analytical columns were butt-connected to the precolumns using Teflon tubing (0.012” i.d. × 0.060” o.d.).

3.3.3 Loading and Rinsing of Samples

Samples were pressure loaded (~200 psi, ~1 μL/min.) on a PC for each mass analysis. Approximately 100 fmol each of two standard peptides (angiotensin I and vasoactive intestinal peptide), were also loaded onto the column for quantitation purposes. The precolumn was then rinsed on the HPLC with solvent A to remove salts and other small molecules that may interfere with mass spectrometric analysis (30 bar, 30-40 min.). The analytical column was then connected to the precolumn using Teflon tubing (0.012" i.d. × 0.060" o.d.) and rehydrated with solvent A for ~15 min.

3.3.4 Mass Spectrometric Analysis

Samples were analyzed by nanoflow-HPLC/microelectrospray ionization coupled directly to a Thermo Orbitrap Classic, Orbitrap Velos, or FT-ICR mass spectrometer with a home built front-end electron transfer dissociation (FETD) source.(19) After loading the sample on the PC and reattaching the PC to the AC the sample was gradient eluted through a laser-pulled electrospray tip directly into the mass spectrometer with an Agilent 1100 series binary LC pump at a flow rate of ~60 nL/min. The elution gradient was as follows: 0-60% B in 40 minutes, 60-100% B in 3 minutes, hold at 100% B for 3 minutes, 100-0% B in 2 minutes, and hold at 0% B for 20 minutes.

For standard peptide analysis the instrument was operated in a data-dependent mode, performing a high resolution MS1 scan followed by acquisition of MS2 spectra in the linear ion trap on the most abundant peak, first for CAD then ETD. The ETD reagent used was azulene for Thermo Orbitrap Classic and FT-ICR or fluoranthene for Orbitrap

Velos, and ETD reaction times were either 30 or 50 ms. For peptide analysis the instrument was operated in a data-dependent mode similar to the standard peptide analysis selecting the top five most abundant non-excluded ions in the MS1 for fragmentation by both CAD and ETD, instead of just the most abundant ion.

3.3.5 Data Analysis

MS/MS spectral data were searched against the refseq human protein database(20) or the SwissProt human protein database(21) using the Open Mass Spectrometry Search Algorithm (OMSSA) software.(22) OMSSA search tolerances were ± 0.01 Da and ± 0.35 Da for precursor and product ion masses, respectively. For ETD spectra, search parameters were set to exclude reduced charge species from peak lists prior to searching. Data were searched using variable modifications of phosphorylation of serine and threonine residues, and oxidation of methionine, with 1024 different potential combinations of variable modifications per peptide being allowed. For derivatized samples, data were searched using fixed modifications of histamine of the peptide C-terminus and aspartic and glutamic acid residues, in addition to the variable modifications. Other search parameters utilized were: +2 charge state products allowed for peptides of charge +3 and above, peptide size range of 7-15 a. a. with no enzyme restrictions. Peptide sequences were manually validated.

3.4 Results and Discussion

3.4.1 Pig-tailed Macaque

The pig-tailed macaque (*Macaca nemistrina*) has become an attractive alternate for HIV/SIV research with the demands overtaking the supply of Indian rhesus macaque (*macaca mulatta*). The identification of the peptide-binding motif is of great value because control of HIV and SIV infection has been linked to $CD8^+$ T cell responses that are MHC class I restricted.(16) In this work the pig-tail macaque MHC class I Mane-A1*082:01 allele was characterized. This was accomplished through combinatorial assays and identification of eluted peptides by mass spectrometry.

Initial characterization of the peptide binding motif of Mane-A1*082:01 was determined by our collaborators by testing the binding specificity of purified MHC class I molecules to various peptides using a radiolabeled probe. The radiolabeled probe used was an analog (peptide 3265.0005; sequence DHQAAFQYI) of the SHIV derived Mane-A1*082:01-restricted T cell epitope. A panel of individual single amino acid substitution analogs, corresponding to substitutions with each of the 20 naturally occurring amino acids at each position of peptide 3265.0005, was tested for its capacity to bind Mane-A1*082:01.(6) In total, a panel of 172 analog peptides were tested. A detailed binding motif was elucidated from the IC_{50} (in nanomoles) values for each single substituted peptide, which were standardized as a ratio to the geometric IC_{50} value of the entire panel, and then normalized at each position so that the value associated with optimal binding at each position corresponds to 1.(6) **Table 3.2** summarizes the binding data from the SAAS assay. Histidine was the most dominant amino acid in position 2 with the next

closest amino acid having a 30-fold reduction in binding. The most preferred residues at the C-terminus were the aromatic residues F and Y, with F being more strongly preferred. Methionine also displayed a favorable binding specificity. Also shown in **Table 3.2** is the specificity factor (SF) which identifies positions where the majority of residues are associated with significant decreases in binding capacity, signifying position specific stringency.(6) Positions 5 and 7 have the next highest specificity factors after position 2 and the C-terminus, making them secondary anchor positions.

Next, to verify the findings of the SAAS binding assay, MHC class I eluted peptides from Mane-A1*082:01 were sequenced by mass spectrometry. From this analysis, 46 peptides were identified and sequenced, listed in **Table 3.3**, ranging in length from 8-15 residues in length. Peptides nine residues in length were most common,.

Table 3.2. SAAS-derived matrix describing 9-mer binding to Mane-A1*082:01.(6)

	Position								
Residue	1	2	3	4	5	6	7	8	9
A	0.034	0.033	0.371	0.348	0.014	0.368	0.017	0.273	0.006
C	0.169	0.005	0.110	0.448	0.001	0.295	0.143	0.160	0.047
D	0.144	0.005	0.025	0.344	0.005	0.037	0.018	0.161	0.004
E	0.038	0.010	0.013	1.000	0.019	0.131	0.004	0.253	0.001
F	0.206	0.002	0.703	0.586	1.000	0.271	1.000	1.000	**1.000**
G	0.064	0.002	0.094	0.979	0.031	0.362	0.009	0.132	0.000
H	0.177	**1.000**	0.330	0.631	0.041	0.110	0.055	0.151	0.006
I	0.207	0.003	0.993	0.041	0.051	0.266	0.040	0.288	0.022
K	0.012	0.001	0.046	0.790	0.018	0.427	0.005	0.149	0.005
L	0.036	0.000	1.000	0.851	0.046	0.024	0.081	0.203	0.089
M	0.264	0.005	0.844	0.729	0.039	0.562	0.060	0.194	**0.158**
N	0.093	0.000	0.546	0.109	0.033	0.091	0.016	0.025	0.010
P	0.038	0.001	0.020	0.321	0.022	0.590	0.021	0.217	0.001
Q	0.120	0.020	0.318	0.181	0.024	1.000	0.028	0.311	0.001
R	0.070	0.006	0.034	0.253	0.013	0.291	0.016	0.126	0.011
S	0.276	0.000	0.255	0.359	0.008	0.286	0.011	0.171	0.004
T	0.160	0.000	0.233	0.182	0.022	0.238	0.009	0.268	0.006
V	0.324	0.000	0.705	0.089	0.044	0.249	0.013	0.139	0.048
W	1.000	0.000	0.496	0.117	0.061	0.438	0.175	0.799	0.013
Y	0.226	0.000	0.796	0.136	0.105	0.334	0.287	0.388	**0.232**
SF	0.51	24.05	0.28	0.19	2.27	0.25	1.87	0.28	5.26

Table 3.3. List of peptides identified from Mane-A1*082:01.

Peptide Sequences		
8-mers	FQYESKVFY	NHIHTGHHSSM
HHSFPYDY	LHSDFIVKY	HNNAEILKTKF
HLAKIYAM	NQVIFPVSY	SHFDLNKKTLY
	NQYPDFNNY	VHREIPQAERY
9-mers	NRQPEWVVY	HHNEAVLHLRF
GHLAKIYAM	NRVGLNLLY	NHLAEWKSLEY
HHIEAQQRL	VQKVDQYLY	NKMPLSVFPYY
HHIPDVITY		NQFVNKFNVLY
HHVIARHEY	**10-mers**	VHFTGEKELEF
KHQEHILRF	DHIKAQIHQF	
NHHLQETSF	DHIQSRIVYY	**12-mers**
NHKKWKVKY	DHRREDRVHY	NHEGEVNRARYM
NHQEQREHY	HREIPQAERY	SHQSQDKKIHVY
NQHGIILKY	NHKDFKELRY	MHLDPKKAGTYF
QHAKNRYYY	NHQALEQLHY	
RHIYNREEY	NQQLPIESQM	**≥13-mers**
YHHSFPYDY		NHMNLAEKSTVHM
DHSFFKISY	**11-mers**	NHNHQLKGDAQRY
FQYESVFY	NHEGEVNRARY	TGSWIGLRNLDLKGF
LHSDFIVKY	NHESRQITQVY	

followed by 11-mers and 10-mers, collectively accounting for 82.6% of the peptides identified. Longer peptides and 8-mers were less abundant with 8-mers being the least tolerated. The sequences of the eluted peptides were analyzed to determine the frequency of each amino acid occurring in positions 1 through 8 and the C-terminus. The results are displayed in **Table 3.4**. The results from the eluted peptides verified the results from the SAAS assay, with a histidine at position 2 and an aromatic residue at the C-terminus. However, the eluted peptides displayed a higher preference for Y than for F. This result differed from the SAAS assay which displayed a four-fold higher preference for F than

Y. This discrepancy could be contributed to the cellular environment and/or limitations of the protein processing machinery.(6)

Table 3.4. Frequency of amino acid residues in eluted peptide ligands from Mane-A1*082:01 and respective positions.(6)

	Peptide position								
Residue	1	2	3	4	5	6	7	8	C-term
A	–	–	0.043	0.087	0.087	0.022	0.043	0.065	–
C	–	–	–	–	–	–	–	–	–
D	0.087	–	–	0.087	0.065	0.043	0.022	0.022	–
E	–	–	0.087	0.130	0.130	0.065	0.109	0.087	0.022
F	0.022	–	0.065	0.043	0.109	0.022	0.022	0.043	0.152
G	0.022	0.022	–	0.087	0.022	0.043	–	0.022	–
H	0.152	0.717	0.065	0.043	0.022	–	0.043	0.087	–
I	–	–	0.130	0.065	0.087	0.130	0.109	–	–
K	0.022	0.022	0.065	0.087	0.022	0.130	0.130	0.152	–
L	0.022	0.022	0.065	0.043	0.109	0.022	0.130	0.087	0.022
M	0.022	–	0.043	–	–	–	–	0.022	0.130
N	0.457	–	0.065	0.022	0.065	0.043	0.065	0.043	–
P	–	–	–	0.087	0.087	0.065	–	–	–
Q	0.022	0.152	0.130	0.022	0.087	0.109	0.065	–	–
R	0.022	0.065	0.043	0.022	0.022	0.109	0.022	0.109	–
S	0.043	–	0.087	0.065	0.043	0.022	–	0.109	–
T	0.022	–	–	0.022	0.022	–	0.022	0.043	–
V	0.065	–	0.065	0.043	–	0.087	0.130	0.065	–
W	–	–	–	0.022	0.022	0.043	–	–	–
Y	0.022	–	0.043	0.022	–	0.043	0.087	0.043	0.674
Median	0.022	–	0.054	0.043	0.033	0.043	0.033	0.043	–
Deleterious	6	14	6	2	4	4	6	5	15

This study utilized MHC elution analyses and panels of single amino acid substitution analog peptides to probe the binding specificity and determine the peptide-binding motif of Mane-A1*082:01. This particular allele is the second most common pig-tail macaque MHC allele with a frequency of 23%.(6) The peptide motif of Mane-A1*082:01 was found to have a H in position 2 and either an F or Y at the C-terminus. This motif is very similar to that of HLA-B*27:05, which has been associated with long-term disease non-progression in HIV-infected individuals.(6) The findings in this study

expand the usefulness of these animals in scientific research, most notably in HIV/SIV studies.

3.4.2 Chinese Origin Rhesus Macaque

The MHC peptidome of Indian rhesus macaques (*macaca mulatta)* has been well characterized, and these animals have been shown to be highly valuable for HIV/SIV research. However, the ban placed on exporting this species has developed a need to find a replacement. The Chinese rhesus macaque is one species capable of filling this void, and has already been used in biomedical research. Nevertheless, a greater understanding of the MHC peptidome of this species is required. This project aims to do just that. In the rhesus macaque of Chinese origin, no specific MHC class I allele is present at a frequency of >9%, and because of this, several alleles need to be characterized to understand the complexity of immune responses in these animals. The present study identifies the peptide binding motif of 5 alleles to bring the total number characterized to 12. These 12 alleles account for approximately 60% of Chinese rhesus macaque MHC class I alleles, shown in **Table 3.5**.(23) The alleles characterized in this study include Mamu-A1*026:01, Mamu-A2*01:02, Mamu-B*066:01, Mamu-B*010:01, and Mamu-B*087:01.

Table 3.5. MHC class I alleles of Chinese rhesus macaques studied to date.(23)

Mamu allele	Phenotypic frequency (%)[a]	HLA homology
A1*022:01	8.9	B07-like
A1*026:01	6.7	A02-like
A7*01:03	6.1	A01-like
B*083:01	5.8	A03-like
B*066:01	5.8	A03-like
B*039:01	5.8	–
B*010:01	5.8	B27-like
B*087:01	5.8	B44-like
B*003:01	5.8	B27-like
B*001:01:02	5.8	B44-like
A2*01:02	5.6	B27-like
B*090:01	2.2	A03-like

Peptide motifs were determined by sequencing eluted MHC peptides from a single allele transfected cell line for each of the 5 alleles by MS/MS. Peptides were manually validated from search results to form a consensus motif. In addition, our collaborators performed positional scanning combinatorial library assays on each allele.For Mamu-A2*01:02 18 peptide ligands, listed in **Table 3.6**, were identified ranging in length from 8-11 residues with the majority being 9-mers. The peptides displayed a preference for an H in position 2 and primarily a Y at the C-terminus. Four of the eighteen ligands had an L at the C-terminus. Positional scanning combinatorial libraries (PSCL) of randomized 9-mer peptides tested the binding capacity of the purified MHC molecules. The PSCL assay confirmed the results of the eluted peptide assay

revealing a preference for H at position 2 and a preference for Y at the C-terminus. Additionally, W was also preferred at the C-terminus with M, L, and F being tolerated.(23) **Figure 3.4** summarizes the motif for Mamu-A2*01:02 from the PSCL assay and from the analysis of the eluted peptides. The peptide binding motif for Mamu-A2*01:02 is associated with an HLA-B27-like motif, which has been shown to be an elite controller.(23)

Table 3.6. Sequences of eluted peptides from Mamu-A2*01:02.

Peptide Sequences	
8-mers	SHSPLRTEY
HSKVLYVY	SHSQITRLY
	SHSSHFSEF
9-mers	THAPTHSDY
AHNAVVHEL	YHSPYPEEY
AHQFLQNEY	
IHSPVVNEL	**10-mers**
NHFEGHYQY	SHDHPTDVDY
QHAKDFKEY	SHNPNTTNNF
SHIPQSSSY	
SHMEFYNQY	**11-mers**
SHSHVGYTL	SHITPSTQQHL

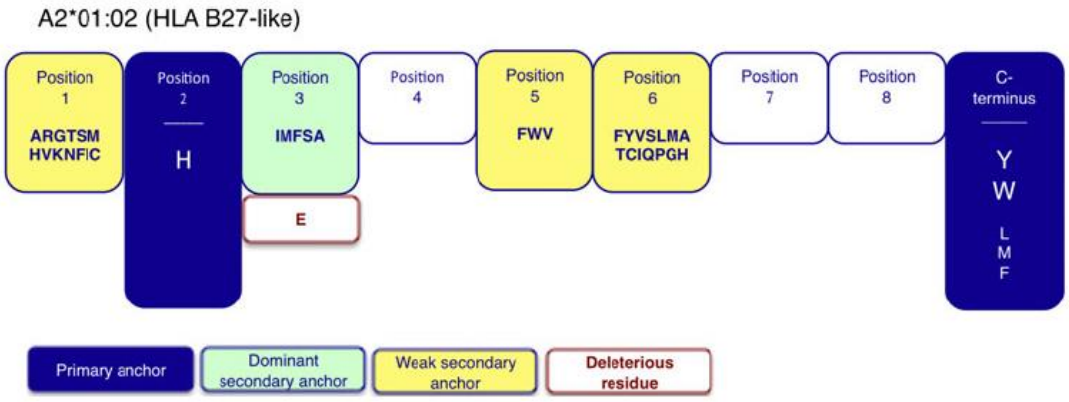

Figure 3.4. Peptide motif summary for Manu-A2*01:02.(23)

Six peptides, listed in **Table 3.7**, were sequenced from the Mamu-A7*03:01 allele with five of the six being 9-mers and one 10-mer. Half of the peptides favored an A residue at position 2 and five peptides favored an L at the C-terminus. PSCL analysis found promiscuity at position 2 where residues S, V, I, T, A, and L were preferred. This assay also identified position 3 as a primary anchor favoring M and Q residues. The C-terminus, identified as a secondary anchor, showed a preference for L, W, I, M, V, and F.(23) These analyses revealed a motif that resembles that observed for the HLA-A*01:01 allele, shown in **Figure 3.5**.(23)

Table 3.7. Sequences of eluted peptides from Mamu-A7*01:03.

Peptide Sequences	
9-mers	STEEKAYEI
ATESVQDVL	
RAEDNADTL	**10-mers**
RAEDNADYL	SVQTTEGEQI
RAENTQEEL	

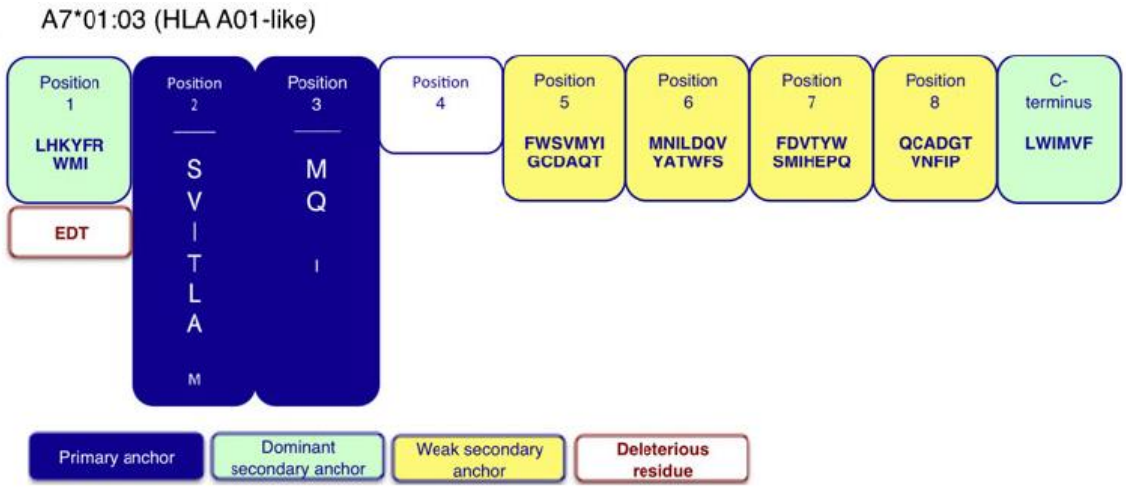

Figure 3.5. Peptide motif summary for Manu-A7*01:03.(23)

For the Mamu-B*010:01 allele, 30 peptides, listed in **Table 3.8**, were identified and sequenced with peptides ranging from 8-12 residues in length. These peptides showed a preference for an H at position 2 and a preference for L at the C-terminus. The PSCL assay identified position 2 and the C-terminus as primary anchor sites. Consistent with the eluted peptides, an H was preferred at position 2 with an A being tolerated. However, the C-terminus was found to prefer F primarily along with M, L, and I, shown in **Figure 3.6**, whereas the eluted peptides showed a strong preference for an L at the C-terminus, and only 3 peptides had an F at the C-terminus.(23) This allele displays a motif similar to that of HLA-B27 and Mamu-A2*02:01.(23)

Table 3.8. Sequences of eluted peptides from Mamu-B*010:01.

Peptide Sequences	
8-mers	THVMDYRAL
ANLKVSQI	VHSDAAKQI
SHAQTVVL	VHFEESSKL
VRVPARVL	
	10-mers
9-mers	HHVENKGATL
DHARLQLQL	VHAPPMQLGF
DHIRFISEL	
HHADKALTM	**11-mers**
IHQDGIHIL	IHEDSTNRRRL
IRSSYIRVL	NHEDGPTAKKL
NHHLQETSF	NHHFFDGKTAL
NQDRYIKKL	NHREAGKEAAL
NQQEFVRAL	WHADGRLTKKL
NQSKFFRVL	VHAPSSRRTTL
RHISIQRQL	
RHKsDSISL	**12-mers**
RHVMLPREL	NHEGEVNRARYM
SHIDRVYTL	NHQQPPQQNTGF
SQSPQGRVM	

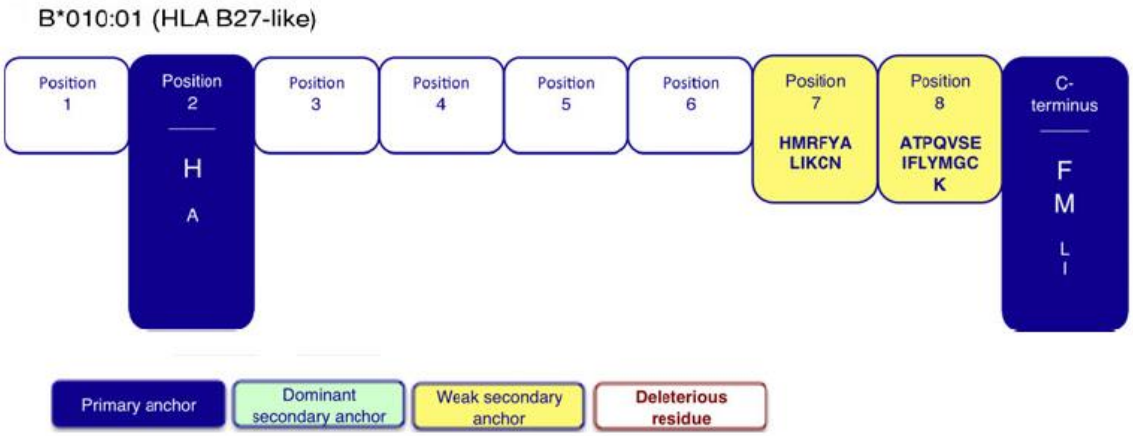

Figure 3.6. Peptide motif summary for Manu-B*010:01.(23)

For the Mamu-B*066:01 allele 14 peptides were identified and sequenced, listed in **Table 3.9**, ranging from 9-11 residues in length. There was not a dominant preference for position 2, G was the most common being present in 5 of the 14 peptides. The lack of a preferred residue is consistent with the PSCL studies that identified position 2 as a secondary anchor and not a primary anchor, as is normally the case. Both the peptides and the PSCL assay displayed a preference for the positively charged residues R and K at the C-terminus, shown in **Figure 3.7**. Mamu-B*066:01 has a motif similar to HLA-A3.

Table 3.9. Sequences of eluted peptides from Mamu-B*066:01.

Peptide Sequences	
9-mers	HGAAPDISVK
AQIDEKTEK	LTAAKKARAG
AVAAVAARR	YFAIAENESK
FAIAENESK	YGYEGIDAVK
SIPEKNRPL	
SLADIMAKR	**11-mers**
YGINTDPPK	AGASGGAYEHR
	KTYIPPKGETK
10-mers	TSSEITTKDLK
AGFAGDDAPR	

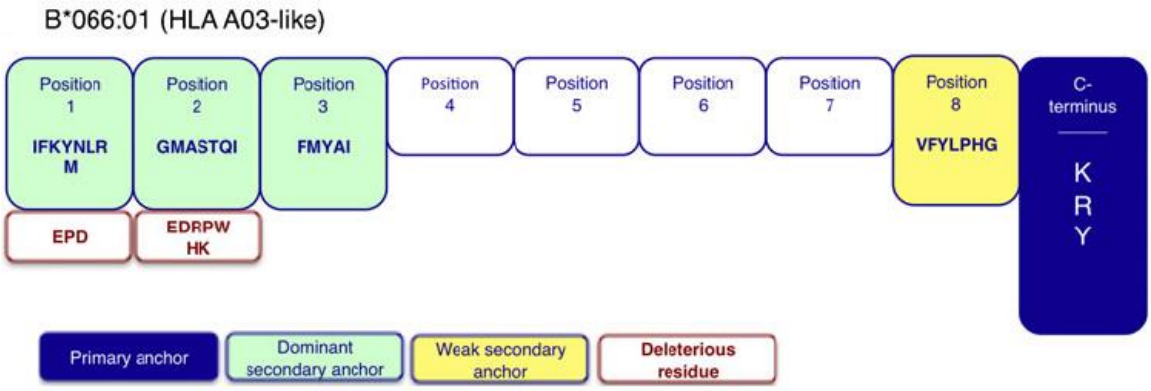

Figure 3.7. Peptide motif summary for Manu-B*066:01.(21)

There were 27 peptides identified and sequenced, listed in **Table 3.10**, in the Mamu-B*087:01 allele ranging from 9-14 residues in length. The majority of the peptides were 10 residues in length. Consistent with the other B-alleles, Mamu-B*087:01 showed a preference for longer peptides. These peptides showed a strong preference for D at position 2, L was also preferred at this position, and a strong preference for L was observed at the C-terminus. PSCL studies identified position 3 and the C-terminus as primary anchors. (23) Aspartic acid was found to be dominantly preferred at position 3, with F and S being tolerated, and the C-terminus showed a strong preference for the hydrophobic residues L, F, M, and I, shown in **Figure 3.8**.(23) This motif closely resembles the motif associated with the HLA-B44 supertype.

Table 3.10. Sequences of eluted peptides from Mamu-B*087:01.

Peptide Sequences	
9-mers	TEDDPQSQHM
FDDEANHLL	
GDVEKLSKF	**11-mers**
KDVEKFIRL	FIDKEDLHDML
SDIDGDYRV	VDIsPTRLHSL
SDVGYERF	
VLDEADEML	**12-mers**
	HDISPQAPTHFL
10-mers	HIDEVQNEIDRL
ADDSSSRDSF	KLDDNEALIEKL
ADENPDKSTL	RIEEELGSKAKF
ADIEPKFDRL	
ALDTSEQERM	**13-mers**
FLDEKTHELL	KDIDPEGKKFDRV
GLDDPRLEKL	NDIsPESSPGVGR
KLDRPRFERV	RDIDAATEAKHRL
RDIDPIITRL	
RDLAPEVEKL	**14-mer**
SLDPGPLEQF	ADIDGDGQVNYEEF

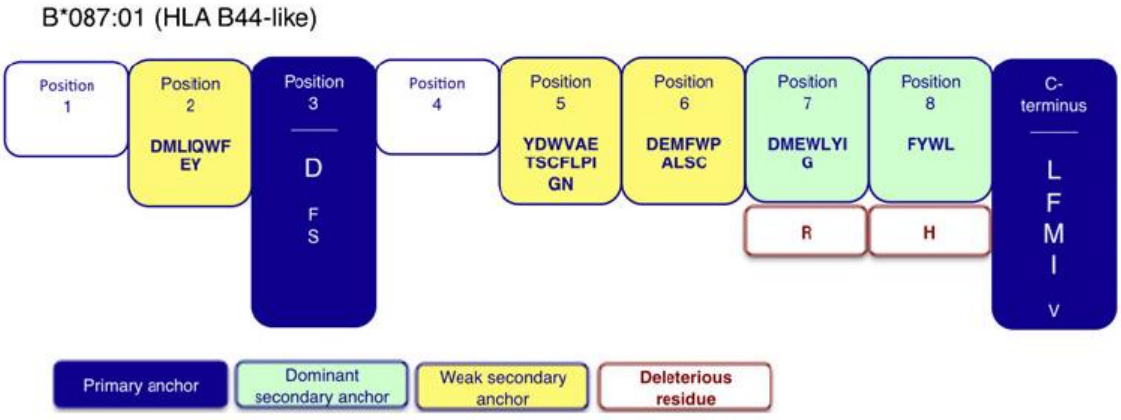

Figure 3.8. Peptide motif summary for Manu-B*087:01.(21)

3.5 Conclusion

The ban placed on Indian rhesus macaques has distressed the current population in the U.S. for use in biomedical research. Due to this, it has become necessary to study a new species of animal to provide new insights into viruses and diseases. Currently, the Chinese rhesus macaque and pig-tailed macaque have been used in biomedical research in addition to Indian rhesus macaques. However, this transition is not just a simple replacement. The Indian rhesus macaque has been well characterized, whereas these new species have not been exhaustively characterized. To understand the progression of viruses and diseases, it is essential to understand the mechanisms by which the immune system fights them off. Of particular importance is the role of MHC molecules in disease progression. When working with animal models it is vital to know how their immune system compares to the human immune system to be able to relate findings from the animal to the human. The aim of this study addresses this issue. This study characterized six total macaque MHC alleles, five Chinese rhesus macaque alleles and one pig-tailed macaque allele.

Prior to this study, only seven Chinese rhesus macaque alleles had been characterized. In the case of Chinese rhesus macaques, no MHC allele has a frequency >9%, unlike Indian rhesus macaques which have MHC allelic frequencies of up to 27%.(23) The alleles characterized in this study all showed similarities to HLA supertypes; on the other hand, approximately half of the Indian rhesus macaque alleles have HLA homology, as seen in **Table 3.11**.(23) HLA homology is an important factor when designing human vaccines in animal models.

Table 3.11. HLA homology of characterized rhesus macaque MHC alleles.(23)

Chinese origin		Indian origin	
Allele	HLA homology[a]	Allele	HLA homology
A1*022:01	B07	A1*001:01	B58
A1*026:01	A02	A1*002:01	A01
A2*01:02	B27	A1*007:01	B27
A7*01:03	A01	A1*011:01	B44
B*001:01:02c	B44	B*001:01:02c	B44
B*003:01c	B27	B*003:01c	B27
B*010:01	B27	B*004:01	–
B*039:01	–	B*008:01	B27
B*066:01	A03	B*017:01	–
B*083:01	A03	B*022:01	–
B*087:01	B44	B*029:01	–
B*090:01	A03	B*048:01	–
		B*052:01	–

Of particular interest are the alleles with HLA homology to B27 as these have been shown to be elite controllers and progress slowly to AIDS. Determining how these alleles are capable of this has the potential to provide valuable information in the design of a vaccine for HIV. Similarly, the pig-tailed MHC allele characterized has similarities to alleles that have been associated with elite controllers.(6) These HLA similarities provide the scientific community valuable tools to evaluate the disease pathogenesis and vaccine concepts.

3.6 References

1. Shedlock DJ, Silvestri G, Weiner DB. Monkeying around with HIV vaccines: Using rhesus macaques to define 'gatekeepers' for clinical trials. *Nat Rev Immuno.* 2009; 9: 717-728.

2. de Goot NG, Heijmans CMC, de Ru AH, Hassan C, Otting N, Doxiadis GGM, Koning F, van Veelen PA, Bontrop RE. Unique peptide-binding motif for mamu-B*037:01: An MHC class I allele common to indian and chinese rhesus macaques. *Immunogenetics.* 2013; 65: 897-900.

3. Gardner MB, Luciw PA. Macaque models of human infectious disease. *ILAR J.* 2008; 49(2): 220-255.

4. Pereira LE, Srinivasan P, Smith JM. Simian-human immunodeficiency viruses and their impact on non-human primate models for AIDS. In: Metodiev K, editor. Immunodeficiency. INTECH; 2012. p. 311-356.

5. Hatziioannou T, Evans DT. Animal models for HIV/AIDS research. *Nat Rev Micro.* 2012; 10: 852-867.

6. Moore C, Sidney J, English AM, Wriston A, Hunt DF, Shabanowitz J, Southwood S, Bradley K, Lafont BAP, Mothe BR, Sette A. Identification of the peptide-binding motif recognized by the pigtail macaque class I MHC molecule mane-A1*082:01 (mane A*0301). *Immunogenetics.* 2012; 64(6): 461-468.

7. Male D, Brostoff J, Roth DB, Roitt IM. Immunology. 8th ed. United States: Elsevier/ Saunders; 2013.

8. The innate (or non-specific) immune response [homepage on the Internet]. The Hong Kong Polytechnic University [cited 1/31/15]. Available from: http://www.acad.polyu.edu.hk/~bcbio/images/LearnImmunology2.jpg.

9. Anderson MH, Schrama D, Straten P, Becker JC. Cytotoxic T cells. *J Invest Dermatol.* 2006; 126: 32-41.

10. Whiteside A. HIV/AIDS: A very short introduction. New York, NY: Oxford University Press; 2008.

11. Simon V, Ho DD, Karim QA. HIV/AIDS epideiology, pathogenesis, prevention, and treatment. *Lancet.* 2006; 368: 489-504.

12. Barry SM, Mena AJ, Novak L, Novak RM. Trial, error, and breakthrough: A review of HIV vaccine development. *J AIDS Clin Res.* 2014; 5(11): 359.

13. Ambrose Z, KewalRamani VN, Bieniassz PD, Hatziioannou T. HIV/AIDS: In search of an animal model. *Trends in Biotech.* 2007; 25(8): 333-337.

14. Shibata V, Kawamura M, Sakai H, Hayami M, Ishimoto A, Adachi A. Generation of a chimeric human and simian immunodeficiency virus infectious to monkey peripheral blood mononuclear cells. *J Virology.* 1991; 65: 3514-3520.

15. Schimitz JE, Korioth-Schimtz B. Immunopathogenesis of simian immunodeficiency virus INfection in nonhuman primates. *Curr Opin HIV AIDs.* 2013; 8: 273-279.

16. Nomura T, Matano T. Association of MHC-I genotypes with disease progression in HIV/SIV infections. *Front Microbio.* 2012; 3: 1-6.

17. Metushi IG, Wriston A, Banerjee PS, Gohlke BO, English AM, Lucas A, Moore C, Sidney J, Buus S, Ostrov DA, Mallal S, Phillips E, Shabanowitz J, Hunt DF, Preissner R, Peters B. Acyclovir has low but detectable influence on HLA-B*57:01 specificity without inducing hypersensitivity. *PLOSONE.* 2015.

18. Martin SE, Shabanowitz J, Hunt DF, Marto JA. Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem.* 2000; 72: 4266-4274.

19. Earley L, Anderson LC, Bai DL, Mullen C, Syka JEP, English AM, Dunyach J, Stafford GC, Shabanowitz J, Hunt DF, Compton PD. Fron-end electron transfer dissociation: A new ionization technique. *Anal Chem.* 2013; 85(17): 8385-8390.

20. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts, and proteins. *Nucl Acids Res.* 2007; 35: D61-D65.

21. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucl Acids Res.* 2000; 28(1): 45-48.

22. Greer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, Bryant SH. Open mass spectrometry search algorithm. *J Prot Res.* 2004; 3: 958-964.

23. Mothe BR, Southwood S, Sidney J, English AM, Wriston A, Hoof I, Shabanowitz J, Hunt DF, Sette A. Peptide-binding motifs associated with MHC molecules common in chinese rhesus macaques are analogous to those of human HLA supertypes and include HLA-B27-like alleles. *Immunogenetics.* 2013; 65(5): 371-386.

Chapter 4: Addition of Charge to O-GlcNAc via Staudinger Reduction

4.1 Introduction

4.1.1 O-GlcNAcylation

Post translational modifications (PTMs) play a key role in the life of a cell. There are numerous types of PTMs including phosphorylation, O-linked β-N-acetylglucosamine (O-GlcNAc), acetylation, methylation, and many others. These PTMs have a wide range of functions which include the regulation of cellular processes and participation in signaling pathways. Additionally, it has been shown that PTMs are widely associated with many diseases, including cancer.(1-4) In order to determine the implications and roles that these PTMs play in the life of a cell, it is important to identify which proteins are modified and where the modification occurs. This can be very challenging due to the sub-stoichiometric abundances and labile nature of the modifications, specifically phosphorylation and O-GlcNAcylation. In this chapter, the O-GlcNAc modification will be examined in greater detail.

In 1984 Hart and coworkers discovered the presence of O-GlcNAcylation on cell surface proteins when probing for N-acetylglucosamine in cells.(5) Their observations showed that O-GlcNAcylation occurred in the cell, most abundantly in the nucleus.(5) This was a surprising finding given that at the time protein glycosylation was thought to occur primarily on the cell surface and within the secretory pathway.(6) This discovery jump-started a resurgence in the field of glycobiology to determine which proteins have this modification, the proteins responsible for modifying O-GlcNAcylated proteins, and the role that O-GlcNAcylation plays.

O-GlcNAcylation is similar to phosphorylation in multiple ways including its abundance in the cell, its cyclic nature, and its role in cell signaling.(7, 8) Interestingly, there are over 500 kinases and over 200 phosphatases(9) which catalyze the addition and removal of phosphate from serine, threonine, and tyrosine residues, respectively; on the other hand, there is only one enzyme that catalyzes the addition of O-GlcNAc and one enzyme that facilitates the removal O-GlcNAc from serine and threonine residues. These two regulatory enzymes for O-GlcNAcylation are O-GlcNAc transferase (OGT) discovered in 1990, which catalyzes the addition of an O-GlcNAc moiety to serine and threonine residues; and β-N-acetylglucosaminidase (OGA) discovered in 2002, which catalyzes the removal of O-GlcNAc.(6, 7) The genes that encode for these enzymes are highly conserved from rat to worm to human. The crystal structure of OGT in **Figure 4.1** displays two domains: an 11.5 tetratricopeptide repeats (TPRs) domain at the N-terminus

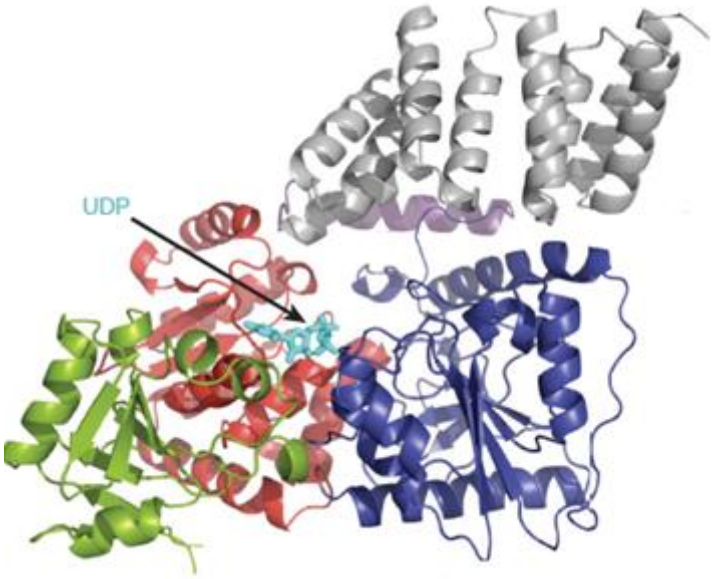

Figure 4.1. Crystal structure of OGT complexed with UDP. The TPR domain is shown in grey and the catalytic domain is shown in green, red, and blue. UDP is shown as a ribbon structure.(7)

and a putative catalytic domain at the C-terminus.(7) The TPRs serve as a protein-protein interaction domain where regulatory and targeted proteins bind. Although the crystal structure of OGA has not yet been solved, its sequence suggests that it also has two domains. The N-terminus of the enzyme is responsible for the removal of O-GlcNAc and the C-terminus contains a putative acetyl transferase domain.(6, 10) While OGA and OGT are important for the regulation of O-GlcNAcylation, they are not the most important regulators. The most important factor in the regulation of O-GlcNAcylation is the amount of available uridine 5`-diphosphate (UDP)-GlcNAc, the donor substrate of OGT.(6, 7, 10) UDP-GlcNAc is synthesized through the conversion of glucose/glucosamine via the hexosamine biosynthetic pathway, shown in **Figure 4.2**.(1, 4, 7) When the GlcNAc moiety is transferred to a protein UDP is released, and UDP is a potent feedback inhibitor of OGT.(10) UDP-GlcNAc is one of the most highly concentrated small molecules in the cell, resulting in the role of UDP-GlcNAc to serve as a nutrient sensor since the levels of UDP-GlcNAc directly affect the extent of O-GlcNAcylation in the cell.(10)

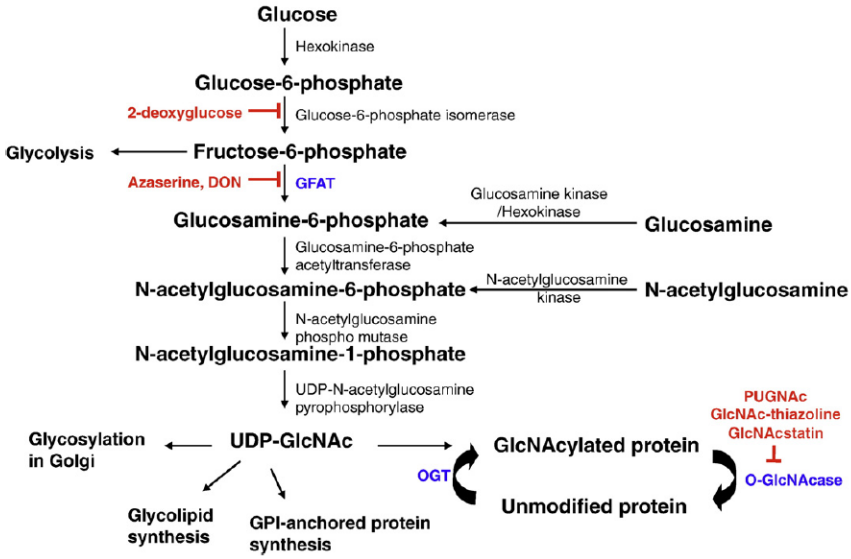

Figure 4.2. Hexosamine biosynthetic pathway (HBP). The synthesis of UDP-GlcNAc from glucose and enzymes involved in the process are shown.(7)

O-GlcNAcylation is a dynamic PTM analogous to phosphorylation, and is one of the most abundant PTMs within the nucleocytoplasm. OGT modifies the hydroxyl groups on serine and threonine residues of target proteins, as shown in **Figure 4.3**. O-linked glycosylation is unique in that it is a single sugar attached to a serine/threonine residue unlike N-linked glycosylation where more complex, branched sugars are common. (11) Studies have shown that O-GlcNAcylation is involved in transcription/translation, nutrient sensing, neural development, cell cycling, and stress response.(8, 12) O-GlcNAcylation is essential to cell viability; for example, when OGT is knocked out in mice it is embryonic lethal.(13) Additionally, it has been shown that OGT is required for embryonic cell development.(13)

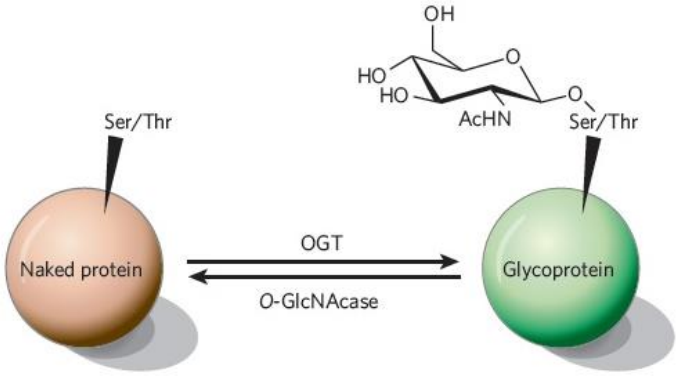

Figure 4.3. O-GlcNAcylation cycles similar to phosphorylation. There is one enzyme that catalyzes the addition of GlcNAc, OGT, and one enzyme that catalyzes the removal of GlcNAc, OGA.(6)

A dynamic interplay exists between O-GlcNAc and phosphate modifications. Every O-GlcNAc modified protein identified to date can also be modified by phosphorylation. Several studies have shown that there is crosstalk between these two modifications, as shown in **Figure 4.4**. For example, disruption of phosphorylation events can affect the level of O-GlcNAcylation and vice versa. This interplay between

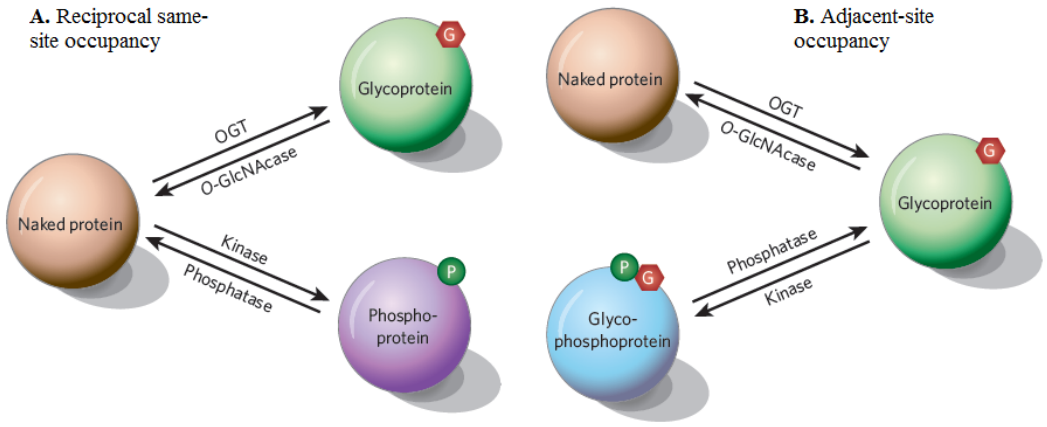

Figure 4.4. There is a complex and dynamic interplay between O-GlcNAc and phosphorylation. **A.** Phosphorylation and O-GlcNAcylation can compete for occupancy at the same site. **B.** O-GlcNAcylation and phosphorylation can occupy adjacent sites on a protein where they act in conjunction with each other.(6)

these two modifications is not a simple on/off switch, but is in fact more complicated.

Figure 4.5 shows that the diversity of the proteome is further complicated by the relationship between O-GlcNAc and phosphate. Therefore, it is important to identify proteins that are O-GlcNAcylated and localize the modifications to gain a better understanding of the role O-GlcNAc plays in the general biology of the cell. Obtaining this information will help to elucidate the relationship between O-GlcNAc and phosphate.

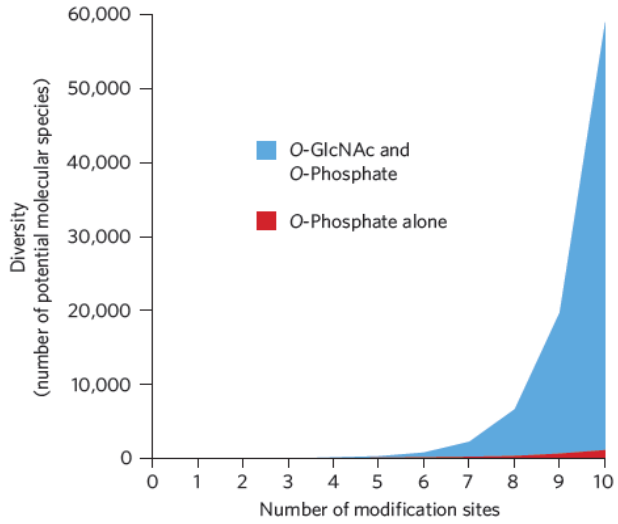

Figure 4.5. The interplay between O-GlcNAc and phosphate lends itself to enormous molecular diversity. (6)

One example of the complex O-GlcNAc/phosphate relationship is the role of these modifications in Alzheimer's disease (AD). A hallmark of AD is the formation of neurofibrillary tangles which are comprised of hyperphosphorylated tau proteins.(3) In a healthy brain, tau is extensively O-GlcNAcylated; this O-GlcNAcylation acts as a negative regulator for phosphorylation.(14) In AD, a patient's neuron glucose

metabolism is markedly decreased, resulting in fewer O-GlcNAcylation sites and, in turn, an increase in phosphorylation leading to the formation of neurofibrillary tangles.(14)

A 2007 review in *Nature* reported that only fifty O-GlcNAcylation sites had been mapped since the discovery of the modification in 1984(6); however, since then this number has skyrocketed to over a thousand sites. Even though over a thousand sites have been identified, there is still no well-defined consensus sequence. This observation lends itself to the possibility that protein-protein interactions govern the specificity of this enzyme as opposed to a certain motif.(15) Why then has it taken so long to identify O-GlcNAc sites? One of the main reasons is a lack of sensitive detection methods. The technology to detect these modified proteins has been deficient in sensitivity and/or specificity with the main methods being lectins, antibodies, and galactosyltransferase1 (Gal-T1).(10) The most sensitive tool for identifying these sites is mass spectrometry.

Mass spectrometry has the sensitivity and the capability to detect O-GlcNAc modifications, and can localize and site map these modifications as well. However, mass spectrometry does come with its own set of challenges. First, O-GlcNAcylated peptides are often found in sub-stoichiometric amounts and are regularly suppressed by their unmodified counterparts; however, despite this, modified peptides can be detected with mass spectrometry. Second, O-GlcNAcylation is a labile modification, which is not preserved in collisionally activated dissociation (CAD). The O-GlcNAc bond is preferentially cleaved in CAD resulting in a signature spectrum like that shown in

Figure 4.6 which is dominated by the loss of the O-GlcNAc moiety from the precursor and a peak corresponding to the GlcNAc oxonium ion. This issue is overcome with the

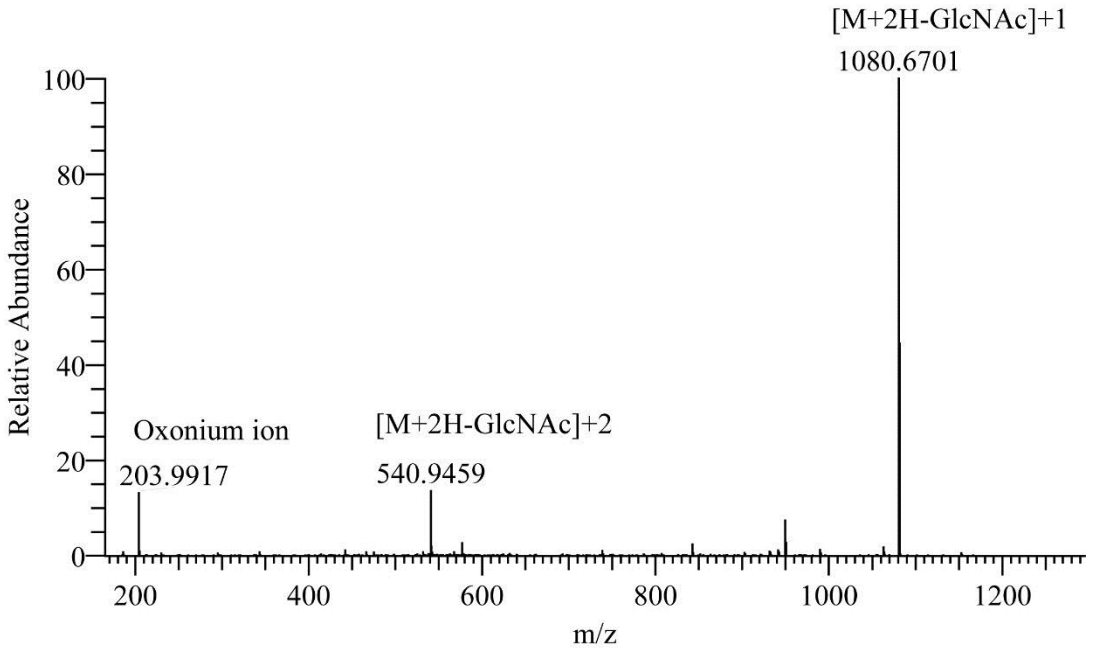

Figure 4.6. Signature spectrum resulting from CAD fragmentation of an O-GlcNAcylated peptide. The spectrum is dominated by the loss of GlcNAc from the precursor and the oxonium ion.

use of electron transfer dissociation (ETD) which conserves labile modifications upon dissociation. ETD occurs when multiply charged peptide cations in the gas phase react with the radical anion, such as azulene.(16, 17) During the reaction, the radical anion transfers an electron to the multiply charged peptide cation triggering the highly selective fragmentation of N-C_α bonds along the peptide backbone.(16) However, ETD is not a cure all, as it is best suited for peptides with a charge state of +3 or higher.

Finally, O-GlcNAcylated proteins are often modified in a poly S/T region which lacks charge. A 2009 study which site-mapped 58 O-GlcNAcylation sites found this to be the case, and also found that a proline was commonly found three residues prior to the modification site, as shown in **Figure 4.7**.(18) These findings present a challenge for ETD analysis. Proline residues do not result in fragmentation by ETD. Also, ETD ions are often not observed in a region that is deficient of charge. This can be problematic when site mapping O-GlcNAc sites, and if O-GlcNAc sites are not mapped then the biological consequences of the modification will be difficult to elucidate. The aim of this work is to develop a means of overcoming this obstacle by adding charge to the O-GlcNAc moiety on a modified peptide.

Figure 4.7. Sequence of 58 O-GlcNAcylated peptides with the modified residue at position 0. There is no strict consensus sequence. The region of modification is rich in S/T residues.(18)

4.1.2 Staudinger Reaction

In 1919, German chemist Hermann Staudinger discovered that a reaction between an azide and a phosphine yields an amine and a phosphine oxide, as shown in **Figure 4.8**. Mechanistically, the Staudinger reaction proceeds through a nucleophilic attack of the

phosphine on the azide followed by loss of nitrogen resulting in the aza-ylide intermediate.(19, 20) In aqueous environments the aza-ylide is rapidly hydrolyzed to

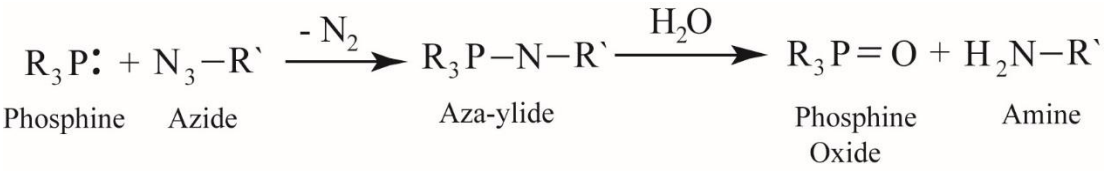

Figure 4.8. Schematic of the Staudinger Reaction between a phosphine and an azide forming a phosphine oxide and an amine.

produce a phosphine oxide and an amine.(21-23) The amine produced by the reaction is attractive for ETD because of its ability to carry charge. However, for several decades after its introduction the Staudinger reaction did not receive much attention or development, until 2000 when Carolyn Bertozzi and colleagues introduced the Staudinger ligation method.(21)

Bertozzi harnessed the selectivity and reliability of the Staudinger reaction to use as a chemically orthogonal tool. However, the traditional Staudinger reaction did not give the resultant that would truly unleash the highest potential of this reaction. The real promise lay in the ability to attach different entities to the azide, such as biotin, prompting Bertozzi to develop the Staudinger ligation reaction. In the ligation technique the aza-ylide intermediate does not undergo hydrolysis. The Staudinger ligation uses specifically engineered phosphines that contain an electrophilic trap, for example a methyl ester, which captures the aza-ylide by intramolecular cyclization resulting in an amide bond, as shown in **Figure 4.9**.(21, 22) The Staudinger ligation is a highly selective and reliable method that has been used in several applications, including modification of cell surfaces, protein engineering, specific labeling of nucleic acids, organic synthesis, and as a general tool for bioconjugation.(20, 23)(24)(24)

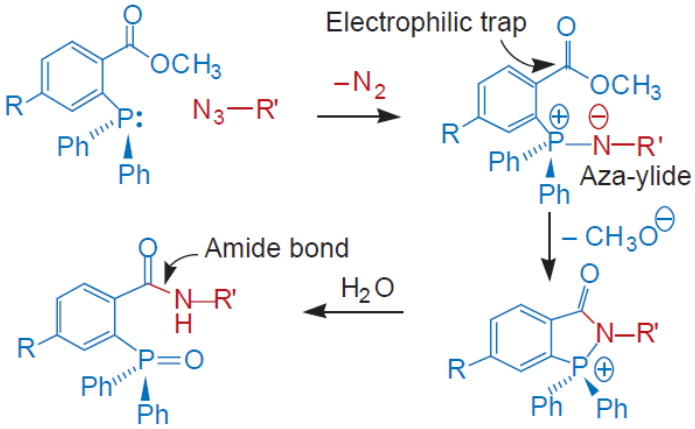

Figure 4.9. Modified Staudinger reaction developed by Bertozzi, known as the Staudinger ligation. The Staudinger ligation results in an amide bond.(21)

In Bertozzi's landmark paper, which launched the field of bioorthogonal chemistry, she reported the Staudinger ligation as a method to engineer cell surface glycoconjugates.(24) This was accomplished through the metabolism of synthetic azidosugars. The azide component on the sugar was then reacted with a specifically designed biotinylated phosphine. The biotin's fluorescence was then measured to calculate the amount of sugar on the cell surface. The critical component for a successful ligation is a phosphine that prevents hydrolysis of the intermediate resulting in a primary amine, and instead forms an amide bond.

The modified Staudinger reaction has shown to be an invaluable tool in the field of glycobiology. Staudinger ligation has enabled selective biomolecule labeling in a variety of environments including the complex environment of live mice.(20) This has allowed researchers to investigate glycan-dependent cell-cell and virus-cell interactions.(25) Most interesting for our purposes is the modification of azidosugars. This is particularly attractive because our group has previously derivatized O-

GlcNAcylated peptides with an azidosugar, GalNAz, as part of an enrichment technique.(26) With this proven technology we propose that adding a GalNAz sugar to an O-GlcNAc moiety followed by the Staudinger reaction will add charge to the GlcNAc and, in turn, enhance ETD fragmentation for improved site mapping of O-GlcNAcylated peptides and proteins. Here we show the application of the Staudinger reaction chemistry to the enhancement of charge on GlcNAc moieties to enhance ETD fragmentation using conditions compatible with HPLC/MS. Specifically we have developed a protocol that is highly efficient, has no competitive side reactions, is applicable to small sample sizes (pmols), allows direct infusion into RP-LC electrospray ionization MS, runs at room temperature, and has a short reaction time.

4.2 Materials

4.2.1 Reagents

Fisher Scientific, Waltham, MA

Methanol

Honeywell, Morristown, NJ:

Acetonitrile, HPLC grade, ≥99.8% purity

Invitrogen, Carlsbad, CA:

Uridine diphosphate N-azidoacetylgalactosamine (UDP-GalNAz)

Galactosyltransferase (GalT1)

Magnesium chloride

Pierce, Rockford, IL:

LC-MS grade water

Sigma-Aldrich, St. Louis, MO:

Acetic Acid (HOAc, glacial, 99.99%)

Angiotensin I human acetate hydrate, ≥90% purity (Angio)

Azulene

Formamide

HEPES

Phosphate buffered saline (PBS)

Tris(2-carboxyethyl) phosphine hydrochloride (TCEP)

Vasoactive intestinal peptide fragment 1-12, human ≥97% (Vaso)

Thermo-Fisher Scientific, Waltham, MA

Dimethylformamide

4.2.2 Column Making Materials

Polymicro Technologies, Inc., Phoenix, AZ:

Polyimide coated fused silica capillary

360 μm o.d. × 50 μm i.d.

360 μm o.d. × 75 μm i.d.

360 μm o.d. × 150 μm i.d.

Zeus Industrial Products, Inc., Orangeburg, SC:

Teflon tubing, 0.012” i.d. × 0.060” o.d.

PQ Corporation, Valley Forge, PA:

Kasil-Potassium silicate solution

YMC Corp., Morris Plains, NJ (available through Waters Corp.):

ODS-AQ, C-18, 5 μm spherical silica particles, 120 Å pore size

ODS-AQ, C-18, 5-20 μm spherical silica particles, 300 Å pore size

4.2.3 Equipment

Agilent Technologies, Palo Alto, CA:

1100 Agilent HPLC

Eppendorf, Hamburg, Germany:

Centrifuge 5424

Labconco, Kansas City, MO:

CentriVap Vacuum Concentrator System

Sutter Instrument Co., Novato, CA:

P-2000 laser puller with fused silica adapter

Thermo Fisher Scientific, Waltham, MA:

LTQ mass spectrometer (custom modified with backend ETD)

LTQ FT-ICR hybrid mass spectrometer (custom modified with front-end ETD)

LTQ-Orbitrap mass spectrometer (custom modified with front-end ETD)

LTQ-Orbitrap Velos mass spectrometer (custom modified with front end ETD)

VWR, Radnor, PA:

Mini vortexer

4.3 Methods

Reactions were performed on a mixture of three synthetic peptides:

TVA(GlcNAc-T)QASGLLSNK, APAVQHIVV(GlcNAc-T)AAK,

APA(GlcNAc-T)QAYGHQIPLR.

4.3.1 Chemical Derivatization of O-GlcNAc Modified Peptides

A mixture of 25 pmols of each of the above mentioned peptides were added to a 0.5 mL Eppendorf tube and taken to dryness on a vacuum concentrator. The peptides were resuspended in 1 μL of $MnCl_2$, 50 μM of UDP-GalNAz, 5 μL of GalT in HEPES. The sample was vortexed and spun down and the reaction was allowed to shake for 5 hours. After 5 hours the reaction was quenched with 0.2 μL of glacial acetic acid (AcOH).

4.3.2 Cleanup of Derivatized Peptides to Remove Excess UDP-GalNAz

The sample was brought up to a total volume of 50 μL by adding 45 μL of 0.1% AcOH. The sample was pressure loaded (200 psi, 1 μL/min) onto a fused silica clean-up

column (360 o.d. ×150 i.d. μm) with a Kasil frit in the end and packed to 10 cm with C18 resin. The sample was then washed with 0.1% AcOH (1 μL/min, 50 μL) by pressure loading. The column was then rinsed on HPLC for 10 minutes with solvent A (0.1M AcOH). The sample was eluted into an Eppendorf tube over 70 minute gradient (0-80% Solvent B (Solvent B was 70% acetonitrile/ 30% 0.1M AcOH)). At the end of the gradient the column was rinsed with a liquid composition of 80% B and 20% A for an additional 30 minutes, followed by a 15 minute rinse of 100% A. The sample was taken to dryness on a vacuum concentrator and reconstituted with 0.1% AcOH to a concentration of 1.25 pmol/μL.

4.3.3 Staudinger Reaction of GalNAz Derivatized Peptides with Triphenylphosphine

The Staudinger reaction was carried out in a total volume of 10 μL. To an empty 1.5 mL Eppendorf tube 6 μL of tetrahydrofuran (THF), 2 μL of AcOH, 1 ng of triphenylphosphine (TPP), and 3.75 pmols of peptide were added. The reaction was immediately vortexed, spun down, and put on the shaker for 2 hours. Then the reaction was taken to dryness in a vacuum concentrator and resuspended in 10 μL of 0.1% AcOH for further analysis.

4.3.4 Staudinger Reaction of GalNAz Derivatized Peptides with Tris(2-carboxyethyl) Phosphine Hydrochloride in PBS

A solution of tris(2-carboxyethyl) phosphine hydrochloride (TCEP) was prepared at a concentration of 3 μg/μL in phosphate buffered saline (PBS) pH 7.5. An aliquot of 2 pmols of peptide was taken to dryness. The following were added to the dried peptides: 16 μL of PBS pH 7.5, 3 μL of bottled water, and 3 μg of TCEP. After either 5 minutes or 2 hours of incubation at room temperature the reaction was taken to dryness in a vacuum concentrator and brought up in 10 μL of 0.1% AcOH for further analysis.

4.3.5 Staudinger Reaction of GalNAz Derivatized Peptides with Tris(2-carboxyethyl) Phosphine Hydrochloride in H_2O

A solution of TCEP was prepared at a concentration of 3 μg/μL in bottled water. An aliquot of either 2 or 10 pmols of peptide was taken to dryness. Then, 49 μL of bottled water was added to the dried down peptides along with either 3 or 15 μg of TCEP. The reaction was put on the shaker for either 4 hours or 5 minutes. Afterward the reaction was taken to dryness in a vacuum concentrator and resuspended in 10 μL of 0.1% AcOH for further analysis.

4.3.6 Assembly of Capillary HPLC Columns

Precolumns (PC) were constructed by creating a Kasil frit in the end of a fused-silica capillary column (360 μm o.d. × 75 μm i.d.). Kasil frits were formed by dipping the end of a capillary column into a mixture of Kasil:formamide (3:1) and heating in an oven overnight at 70˚C. The fritted column was packed with 5-8 cm of irregular, 5-20 μm diameter, 300 Å pore size, C18 reverse-phase resin using a pressure bomb (500 psi). Analytical columns (AC) were constructed with integrated electrospray emitter tips.(27) A Kasil frit was created in a short section of the fused silica (360 μm o.d. × 50 μm i.d.). Columns were then packed with 6-9 cm of 5 μm diameter, 120 Å pore size, C18 reverse-phase resin using a pressure bomb (500 psi). A micropipette laser puller was used to generate a tip (2 μm i.d.) approximately 5 mm from the Kasil frit. Analytical columns were butt-connected to the precolumns using Teflon tubing (0.012" i.d. × 0.060" o.d.).

4.3.7 Loading and Rinsing of Samples

Samples were pressure loaded (~200 psi, ~1 μL/min.) on a PC for each mass analysis. Approximately 100 fmol each of two standard peptides (angiotensin I and vasoactive intestinal peptide), were also loaded onto the column for quantitation purposes. The precolumn was then rinsed on the HPLC with solvent A to remove salts and other small molecules that may interfere with mass spectrometric analysis (30 bar, 30-40 min.). The analytical column was then connected to the precolumn using Teflon tubing (0.012" i.d. × 0.060" o.d.) and rehydrated with solvent A for ~15 min.

4.3.8 Mass Spectrometric Analysis

Samples were analyzed by nanoflow-HPLC/microelectrospray ionization coupled directly to a Thermo Orbitrap Classic or FT-ICR mass spectrometer with a home built front-end electron transfer dissociation (FETD) source.(28) After loading the sample on the PC and reattaching the PC to the AC, the sample was gradient eluted through a laser-pulled electrospray tip directly into the mass spectrometer with an Agilent 1100 series binary LC pump at a flow rate of ~60 nL/min. The elution gradient was as follows: 0-60% B in 40 minutes, 60-100% B in 3 minutes, hold at 100% B for 3 minutes, 100-0% B in 2 minutes, and hold at 0% B for 20 minutes.

The instrument method used for analysis of standard peptides was data dependent CAD and ETD fragmentation of the most abundant peak in the MS1. The ETD reagent used was azulene, and ETD reaction times were either 40 or 45 ms. The instrument method used for the analysis of the derivatized samples was a data dependent method that selected the top three non-redundant, most abundant peaks in the MS1 for fragmentation by both CAD and ETD. Spectra were manually interpreted to assess the practicality of the derivatization.

4.4 Results and Discussion

4.4.1 Method Development

Prior to developing the methodology of the Staudinger reaction, the GlcNAc moieties first needed to be tagged with a GalNAz sugar by GalT1.(26) This reaction was

previously optimized in our lab as part of an O-GlcNAc enrichment method. A mixture of 25 pmols of each of the three GlcNAcylated standard peptides was derivatized with GalT1 to add a GalNAz moiety to the GlcNAc moiety. This reaction yielded an average of 88% conversion, as shown in **Table 4.1**. Conversion rates were calculated by dividing the area of the derivatized peptide by the sum of the areas of the derivatized and underivatized peptide.

Table 4.1. Conversion rate of the GalT1 reaction.

Peptide	% Conversion
TVAgTQASGLLSNK	94%
APAVQHIVVgTAAK	75%
APAgTQAYGHQIPLR	93%

Methodological development of the Staudinger reaction began by adapting previously published protocols.(29)(21) Triphenylphosphine (TPP), shown in **Figure 4.10A**, which is soluble in organic solvents, was selected as the phosphine. The first set of reaction conditions tested consisted of a 7:3 ratio of tetrahydrofuran (THF):0.1% AcOH with 1ng of TPP. This reaction did not yield any detectable trace of the Staudinger product. Observations at the end of the reaction revealed that the THF had evaporated during the course of the two hour reaction. To combat this, the amount of THF was increased for the next experiment to a 13:3 ratio of THF:0.1% AcOH. Once again the reaction yielded only a slight trace of the Staudinger product from two of the

peptides, and the THF had evaporated by the end of the reaction. Since the THF was evaporating by the end of the reaction, the amount used was decreased to 1%. THF could not be eliminated altogether because it was needed to dissolve the TPP. Less than 1% of product was observed with the decreased amount of THF.

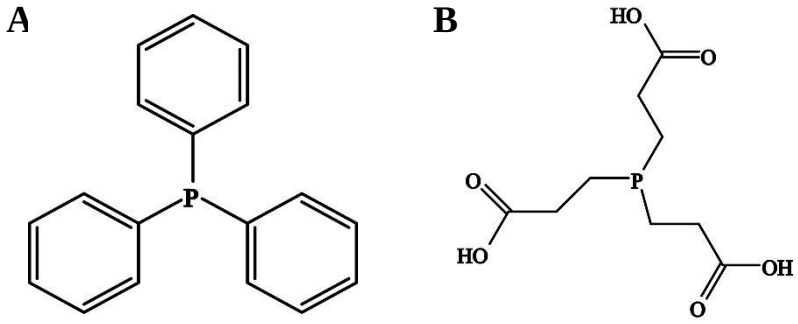

Figure 4.10. Structure of phosphines used in the Staudinger reaction. A. Triphenylphosphine. B. tris(2-carboxyethyl) phosphine.

With reactions carried out in THF not yielding any product we turned our attention to a different solvent, dimethylformamide (DMF). For this reaction the amount of TPP was increased 10-fold to 10 ng and DMF was substituted for THF. Under these conditions an average of 4% conversion to the Staudinger product was observed. The main difference between the reactions was that the DMF did not evaporate throughout the course of the reaction. It is likely that the increase in TPP and the change to DMF accounted for the increased conversion to the Staudinger product, the GalNAm.

To further develop the reaction we decided to increase the volume of the reaction to a total of 50 μL, instead of 10 μL, and double the amount of peptide. The reaction was carried out in DMF using 1 ng of TPP with a 2 hour reaction time on the shaker at RT. The conversion to the Staudinger product fell to <1%. This observation supports the idea that the increase in product previously seen can be attributed to the increase in TPP.

With little success we delved back into the literature for a new direction. Brik and colleagues reported that the water-soluble phosphine tris(2-carboxyethyl) phosphine (TCEP), structure shown in Figure 4.9B, efficiently reduced an azido-peptide in phosphate buffer, pH 7.5.(30) They reported that the reaction was pH dependent, with higher yields at higher pH values up to 7.5. They reported complete reduction of the azido-peptide after only 6 minutes.(30)

Following in their footsteps, we designed a very similar experiment. The subsequent reactions were carried out on a single peptide instead of a mixture of three peptides as in the above experiments. Experiments designed after Brik's yielded an 11% conversion after a five minute reaction time. An 11% conversion in only five minutes was a substantial improvement over only 4% after 2 hours as seen in previous experiments. Next, the reaction time was extended to investigate its effect on conversion rates. Also, we adjusted the amount of PBS to a 1:6 PBS:H_2O ratio. This was done because there was a considerable amount of salt in the dried down reaction. An increased reaction time of 1 hour increased the amount of conversion to the Staudinger product to 36%.

Even though promising results were seen with these conditions, a high amount of salt was present in the dried down product derived from PBS. When a sample is very salty, it requires extended rinsing of the PC to try and eliminate the salt from the column. However, even with extensive rinsing, salt can remain on the column. When salt is present on the column it can affect the chromatography, and, consequentially, unreliable results are obtained. This led to the decision to carry out the reaction in water rather than

PBS, which in turn lowered the pH of the reaction. Brik reports that this is a suboptimal condition for the reaction. However, performing this reaction in water with 3 μg of TCEP resulted in 69% conversion, the highest observed up to this point. The pH of this reaction was 3.5. These results suggest that using the water soluble phosphine is a step in the right direction.

Next, with the reaction environment showing promise, we evaluated the amount of reagent needed by increasing the amount of TCEP five-fold. This reaction, which reacted for 4 hours on the shaker, resulted in 94% conversion. The same reaction conditions were carried out on a three peptide mixture, increasing the amount of TCEP for the increased amount of peptide. This resulted in an average of 88.7% conversion. From these results we conclude that this is a fairly robust method that would be applicable to a real sample.

The final variable assessed was the amount of time it takes for the reaction to go to completion. We found a conversion of 94% after only five minutes, which is the same rate that was observed after 4 hours. We concluded that 5 minutes is a sufficient reaction time for the Staudinger reaction. We also showed that a lower pH does not affect the efficiency of the reaction under our conditions. **Table 4.2** summarizes key experiments in the development of the Staudinger reaction applied to O-GlcNAcylated derivatized peptides. This is shown to be a highly selective technique with no observed side reactions.

Table 4.2. Summary of key experiments in the development of the Staudinger reaction. * denotes experiments carried out using three peptide mixture.

Phosphine	Amount of phosphine	Amount of peptide (pmol)	Solvent	Total Volume (µL)	Time	pH	% Conversion
TPP	1 ng	3.75*	70% THF	10	2 hrs.	ND	0
TPP	1 ng	3.75*	80% THF	16	2 hrs.	ND	<1
TPP	1 ng	3.75*	1% THF	10	2 hrs.	ND	<1
TPP	10 ng	3.75*	70% DMF	10	2 hrs.	ND	4
TCEP	3 µg	2	PBS	20	5 min.	7.56	11
TCEP	0.3 µg	2	5% PBS	35	2 hrs.	7.67	36
TCEP	3 µg	10	H_2O	50	4 hrs.	3.56	69
TCEP	15 µg	2	H_2O	50	4 hrs.	3.56	94
TCEP	15 µg	2	H_2O	50	5 min.	3.56	94
TCEP	37.5 µg	3.75*	H_2O	50	5 min.	3.56	88.7

4.4.2 Spectral Evaluation of Staudinger Product

The main goal of this project is to facilitate spectral interpretation of O-GlcNAcylated peptides by derivatizing the GlcNAc to contain a primary amine that will allow for charge enhancement of the modified residue to aid in site mapping. So far we have shown that we can, in fact, derivatize the GlcNAc to contain a primary amine with high selectivity and high conversion rates. Now we will discuss the effect this chemistry has on the chromatography and fragmentation of these derivatized, modified peptides.

In this protocol the modified peptide undergoes two chemical reactions, each of which changes the hydrophobicity of the peptide. This change in hydrophobicity has an effect on the retention time of the peptide. As can be seen in the extracted ion chromatogram (XIC) in **Figure 4.11B** the addition of the GalNAz moiety slightly increased the hydrophobicity of the peptide, shifting the retention time to slightly later in the gradient. Conversely, the reduction of the azide group on the GalNAz decreased the hydrophobicity which resulted in the peptide eluting two minutes earlier in the gradient

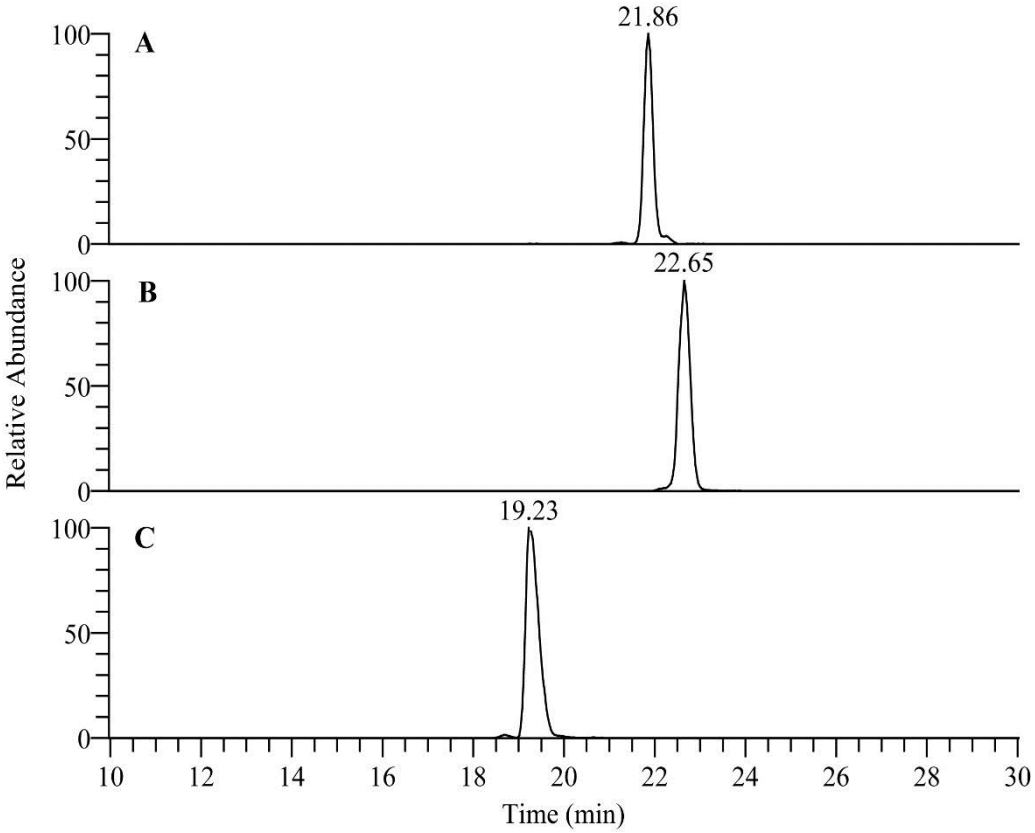

Figure 4.11. Extracted ion chromatogram showing the retention times of the modified peptide TVAgTQASGLLSNK in three different forms (A) GlcNAc, (B) GlcNAc+GalNAz, and (C) GlcNAc+GalNAm.

compared to the GlcNAc, shown in **Figure 4.11A&C**. This provides an added benefit, by eluting earlier in the gradient, the modified peptide would be further separated from its unmodified counterpart reducing the effect of ion suppression.

In addition to the changes in hydrophobicity, there is also a change in the mass of the modification. As shown in Table 4.3, the addition of the GalNAz adds a mass of 244.0808 Da increasing the mass of the modification to 447.1601 Da. Upon reduction of the azide, the resulting GalNAm has a mass of 218.0903 Da decreasing the total mass to 421.1696 Da.

Table 4.3. Masses of sugar residues.

Sugar	**Mass**
GlcNAc	203.0793 Da
GalNAz	244.0808 Da
GalNAm	218.0903 Da
GlcNAc+GalNAm	421.1696 Da

If the Staudinger reaction is successful, then there should be an increase in the charge state of the peptide, or at the very least a shift in the most abundant ion present from a lower charge state to a higher charge state. This is in fact the case for the peptide, TVAgTQASGLLSNK, shown in **Figure 4.12A**, where the most abundant species is the +2 and the +3 is at approximately 20% relative abundance. In a complex sample the likelihood of the +3 species being selected for an MS2 is low; however, the +2 species

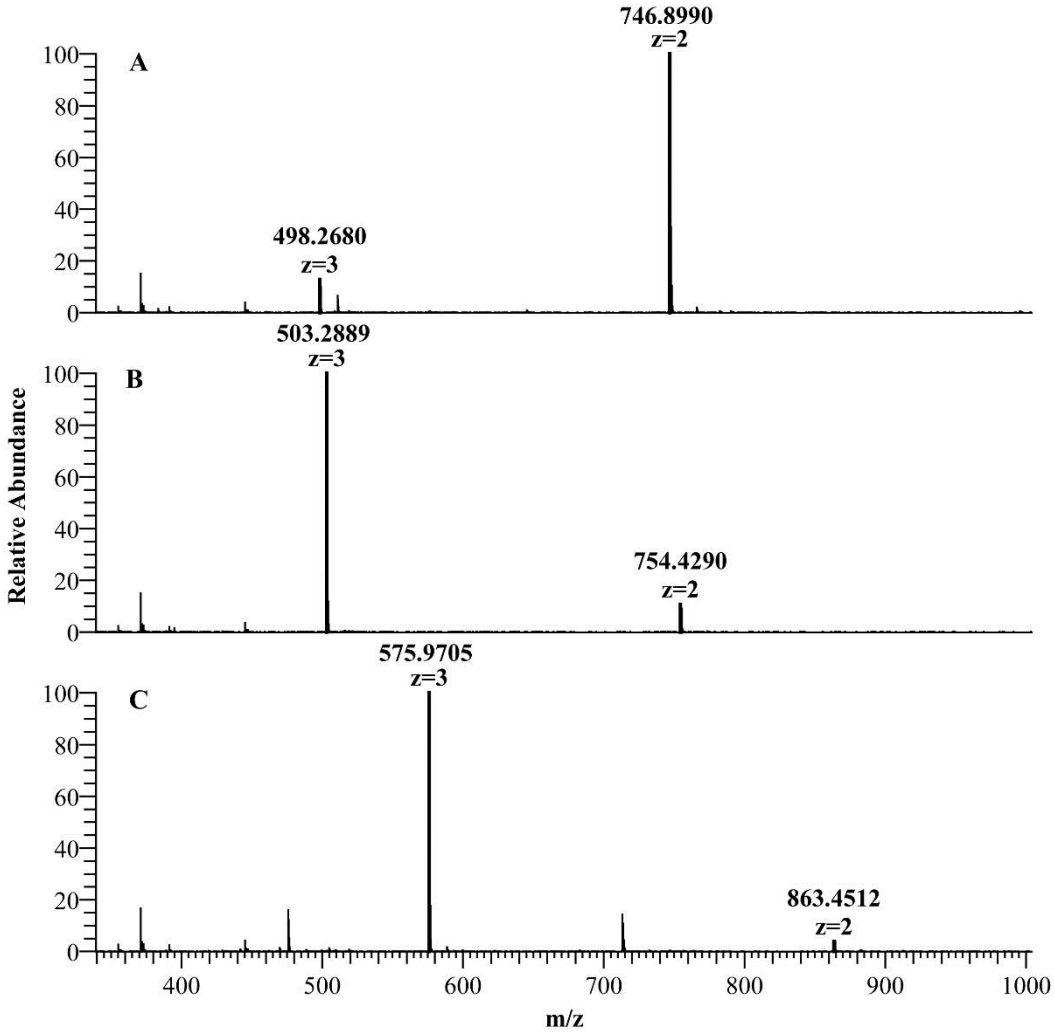

Figure 4.12. Charge state distribution of GlcNAcylated peptides prior to derivatization. (A) TVAgTQASGLLSNK, (B) APAVQHIVVgTAAK, and (C) APAgTQAYGHQIPLR.

would likely be selected and would result in a lack of sequence information necessary to site map the modification. On the other hand, post Staudinger reaction the most abundant peak is the +3 species with the +2 species at approximately 20% relative abundance, as shown in **Figure 4.13A**. Even though the charge state did not increase for this peptide, there was a shift in the charge state distribution which would greatly enhance the probability that the +3 species would be selected, providing a greater chance of being able to site map the modification. For the peptide APAVQHIVVgTAAK there is not only a change in charge state distribution, but the +4 species is present post Staudinger derivatization. The most abundant form of the underivatized form is the +3 species and there is no trace of the +4 species, as shown in **Figure 4.12B**. After conversion to the Staudinger product the +4 species is not only present, but is the most abundant species, as shown in **Figure 4.13B**. Lastly, a charge state distribution shift is also seen for APAgTQAYGHQIPLR. Similarly to APAVQHIVVgTAAK, the most abundant form of APAgTQAYGHQIPLR is the +3 with no detection of a +4 species shown in **Figure 4.12C**. After derivatization the +4 species is present and is the most abundant species, shown in **Figure 4.13C**.

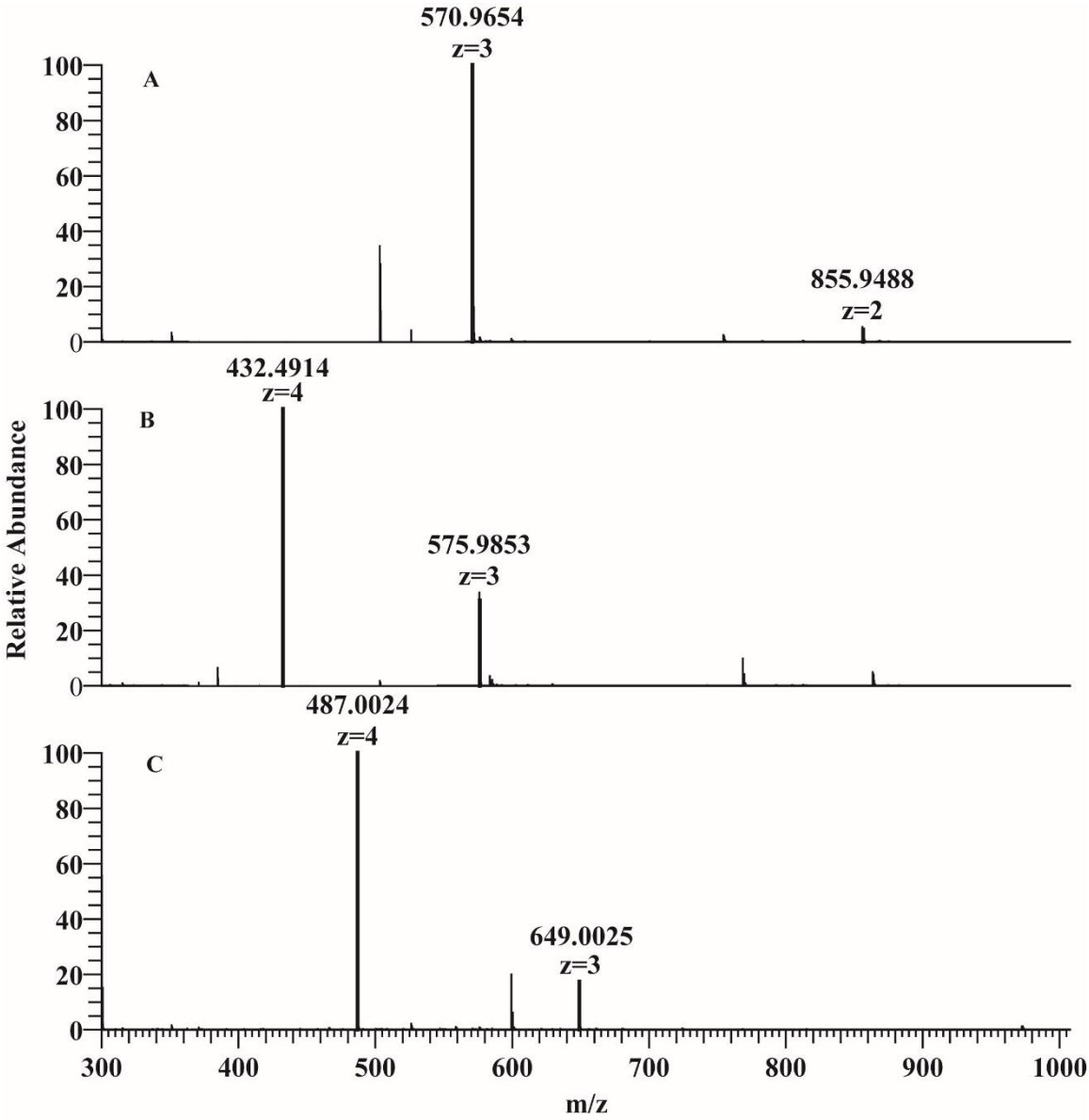

Figure 4.13. Charge state distribution post Staudinger reaction. (A) TVAgTQASGLLSNK, (B) APAVQHIVVgTAAK, and (C) APAgTQAYGHQIPLR.

Above we showed that when an O-GlcNAcylated peptide is fragmented by CAD it produces a signature spectrum with the reduced charge precursor minus the loss of the GlcNAc as the most abundant peak. The loss of the GlcNAc residue yields a peak at 204 m/z. The derivatized peptide still produces a signature spectrum; however, in addition to

the peak observed at 204 m/z, there are also peaks at 219 m/z, the mass of GalNAm, and at 422 m/z, the mass of GlcNAc+GalNAm, as shown in **Figure 4.14**. The loss of these moieties from the precursor are also observed.

We have shown that the Staudinger product shows a shift in charge state

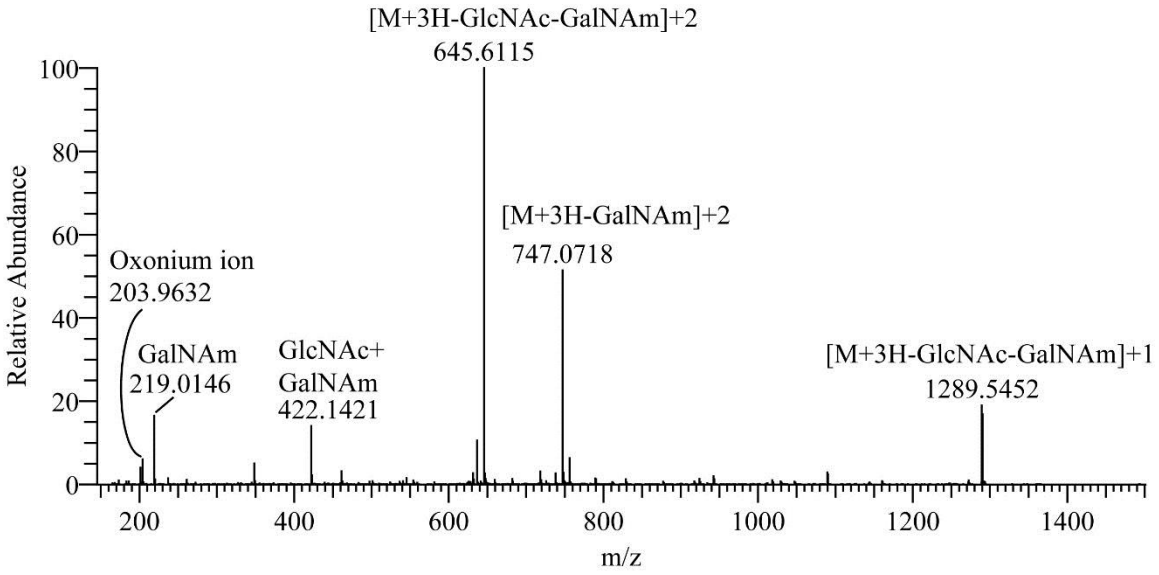

Figure 4.14. CAD fragmentation of the +3 species of the TVAgTQASGLLSNK peptide post Staudinger highlighting the signature ions.

distribution from a lower charge state to a higher charge state and that signature ions are present in CAD spectra. Now we will analyze the effect that Staudinger reduction has on the ETD MS2 of the peptides to assess whether or not this process actually increases the likelihood of site mapping. In the case of the TVAgTQASGLLSNK peptide, the highest charge state present is the +3 for both the modified peptide and the derivatized peptide. In this case we see very similar sequence coverage between the two forms. **Figure 4.15** shows the spectra of the post Staudinger product. For the peptide APAVQHIVVgTAAK,

we see a major increase in the amount of the doubly charged fragment ions present between the +3 species, shown in **Figure 4.16A** and the +4 species, shown in **Figure 4.16B**, the product of the Staudinger reaction. We observe a similar pattern for the third peptide, APAgTQAYGHQIPLR, with an increase of 8 doubly charged ions between the +3, shown in **Figure 4.17A**, and the +4 species, shown in **Figure 4.17B**. In the case of the latter two peptides, the additional ions increase confidence in the assignment of the modification. This would be beneficial in identifying O-GlcNAcylated peptides in a more complex sample, and would be especially beneficial for tryptic digests where short,

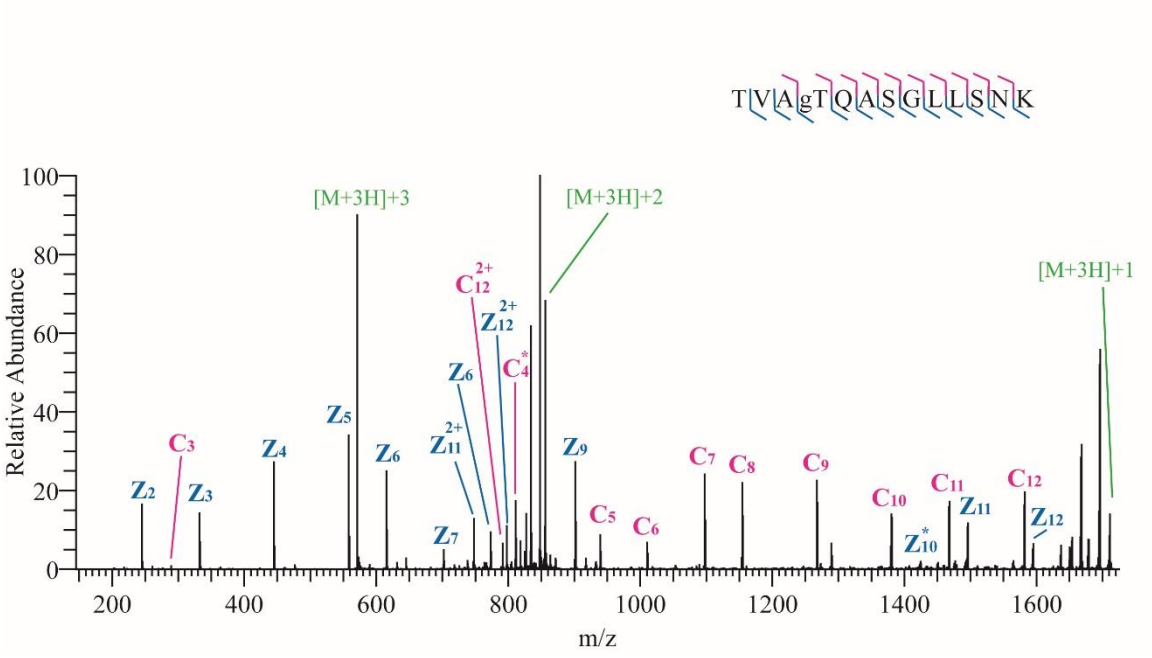

Figure 4.15. ETD fragmentation of the +3 species of the TVAgTQASGLLSNK peptide post Staudinger.

+2 peptides are common.

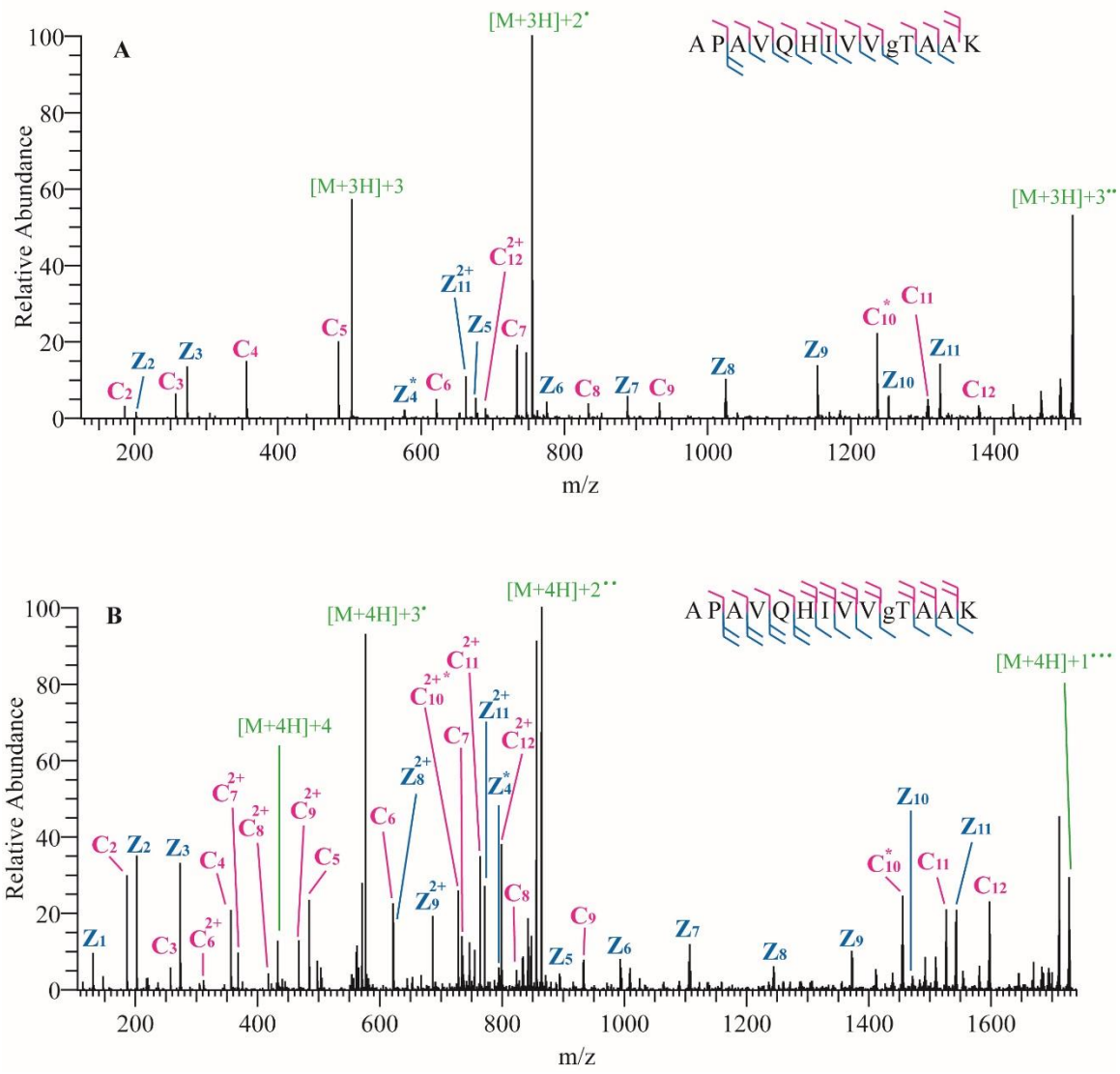

Figure 4.16. (A) ETD fragmentation of the +3 species of the APAVQHIVVgTAAK peptide. (B) ETD fragmentation of the +4 species of the APAVQHIVVgTAAK peptide post Staudinger.

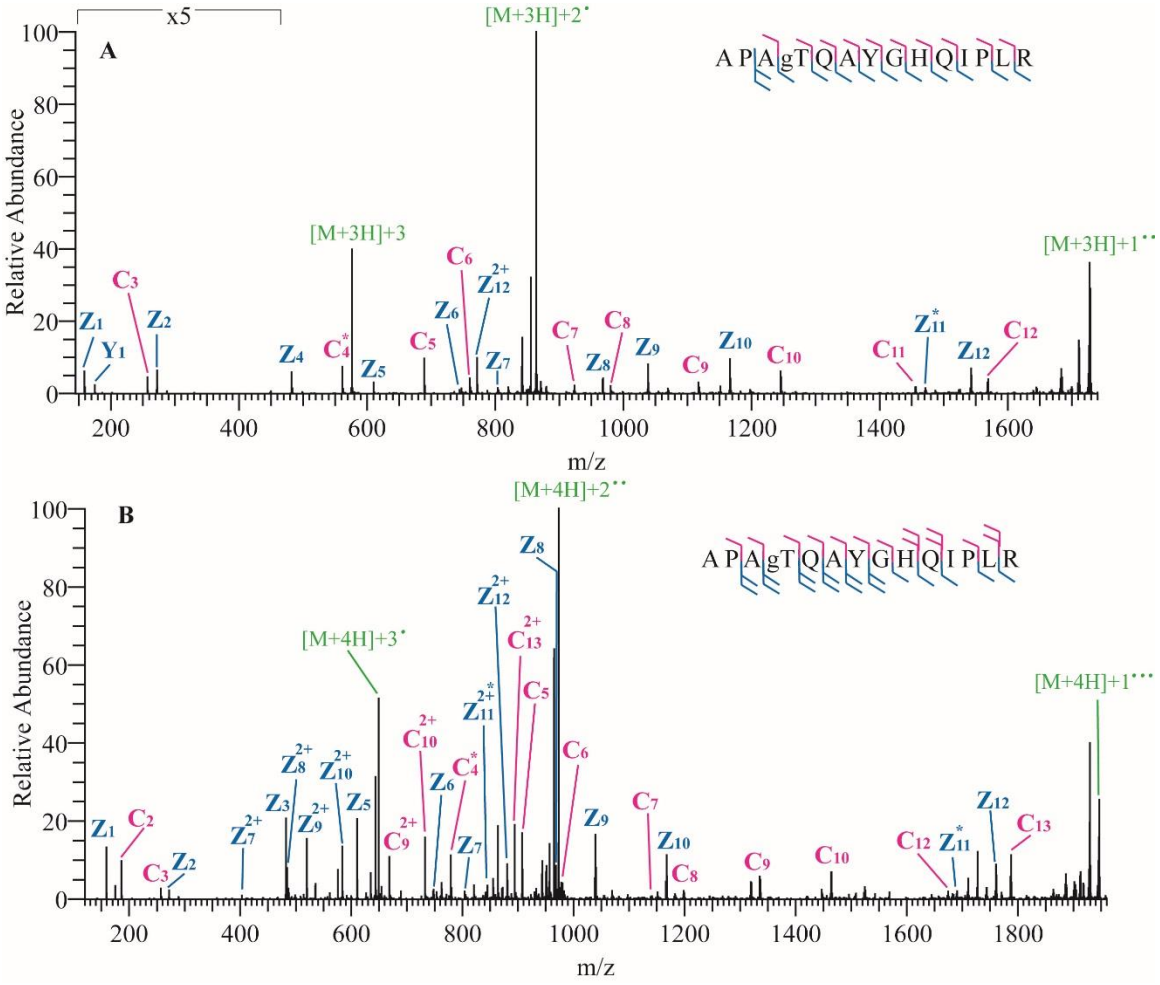

Figure 4.17. (A) ETD fragmentation of the +3 species of the APAgTQAYGHQIPLR peptide. (B) ETD fragmentation of the +4 species of the APAgTQAYGHQIPLR peptide post Staudinger.

4.5 Conclusion

Demonstrated herein is the successful derivatization of O-GlcNAcylated peptides, which increased their charge and enhanced ETD fragmentation. This was accomplished

by tagging the GlcNAc with an azidosugar which is then reduced to a primary amine via the Staudinger reaction. This is a simple two-pot experiment that is highly selective and highly efficient with high conversion rates. This methodology would be particularly amenable to trypsin digests of O-GlcNAcylated proteins, as the majority of tryptic peptides are +2 species. Derivatization of the sample would ideally increase the charge of the peptide making it more suitable for ETD fragmentation, and, therefore, increase the ability to site localize the modification. This would however involve an added step of removing the trypsin prior to performing the derivatization. This methodology also has the added benefit of the shifting the retention time of the modified peptide earlier in the gradient which separates it from its unmodified counterpart reducing, if not eliminating, ion suppression.

Another potential application for this methodology would be to introduce GalNAz to the cell. GalNAz has been shown to be metabolized by numerous cell types.(31) Introducing the azidosugar directly to the cell eliminates the need for the GalT reaction resulting in a one-pot reaction for the Staudinger reaction. This holds great promise for identifying and site-mapping O-GlcNAcylated peptides in the future.

4.6 References

1. Zachara NE, Hart GW. Cell signaling, the essential role of O-GlcNAc! *Biochem Biophys Acta.* 2006; 1761: 599-617.

2. Zachara NE, Hart GW. The emerging significance of O-GlcNAc in cellular regulation. *Che Rev.* 2002; 102: 431-438.

3. Wells L, Whelan SA, Hart GW. O-GlcNAc: A regulatory post-translational modification. *Biochem Biophys Res Commun.* 2003; 302: 435-441.

4. Zachara NE, Hart GW. Cell signaling, the essential role of O-GlcNAc! *Biochem Biophys Acta.* 2006; 1761: 599-617.

5. Torres CR, Hart GW. Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. evidence for O-linked GlcNAc. *J Biol Chem.* 1984; 259: 3308-3317.

6. Hart GW, Housley MP, Slawson C. Cycling of O-linked β-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature.* 2007; 446: 1017-1022.

7. Butkinaree C, Park K, Hart GW. O-linked β-N-acetylglucosamine (O-GlcNAc): Extensive crosstalk with phosphorylation to regulate signaling and transcription in response to nutrients and stress. *Biochem Biophys Acta.* 2010; 1800: 96-106.

8. Wang Z, Udeshi ND, Slawson C, Compton PD, Kaoru S, Cheung WD, Shabanowitz J, Hunt DF, Hart GW. Extensive crosstalk between O-GlcNAcylation and phosphorylation regulated cytokinesis. *Sci Signal.* 2010; 3(104): 1-12.

9. Sacco F, Perfetto L, Castagnoli L, Cesareni G. The human phosphatase interactome: An intricate family portrait. *FEBS Letters.* 2012(17): 2732-2739.

10. Hart GW, Akimoto Y. The O-GlcNAc modification. In: Varki A, Cummings RD, Esko JD, et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2009. p. 236-279.

11. Moremen KW, Tiemeyer M, Nairn AV. Vertebrate protein glycosylation: Diversity, synthesis and function. *Nat Rev Mol Cell Bio,.* 2012; 13: 448-462.

12. Wells L, Vosseller K, Hart GW. Glycosylation of nucleocytoplasmic proteins: Signal transduction and O-GlcNAc. *Science.* 2001; 291: 2376-2378.

13. Slawson C, Hart GW. O-GlcNAc signalling: Implication for cancer biology. *Nat Rev Cancer.* 2011; 11: 678-684.

14. Liu F, Grundke-Iqbal I, Hart GW, Gong CX. O-GlcNAcylation regulates phosphorylation of tau; A mechanism involved in alzheimer's disease. *Proc Nat Acad Sci.* 2004; 101(29): 10804-10809.

15. Banerjee PS, Hart GW, Cho JW. Chemical approaches to study O-GlcNAcylation. *Chem Soc Rev.* 2013; 42: 4345-4357.

16. Udeshi ND, Compton PD, Shabanowitz J, Hunt DF, Rose KL. Methods for analyzing peptides and proteins on a chromatographic timescale by electron-transfer dissociation mass spectrometry. *Nat Prot.* 2008; 3(11): 1709-1717.

17. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Nat Acad Sci.* 2004; 101(26): 9528-9533.

18. Chalkey RJ, Thalhammer A, Schoepfer R, Burlingame AL. Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry of native peptides. *Proc Nat Acad Sci.* 2009; 106(22): 8894-8899.

19. Hang HC, Yu C, Kato DL, Bertozzi CR. A metobolic labeling approach toward proteomic analysis of mucin-type O-linked glycosylation. *Proc Nat Acad Sci.* 2003; 100(25): 14846-14851.

20. Sletten EM, Bertozzi CR. From mechanism to mouse: A tale of two bioorthognonal reactions. *Acc Chem Res.* 2011; 44(9): 666-676.

21. Saxon E, Bertozzi CR. Cell surface engineering by a modified staudinger reaction. *Science.* 2000; 287: 2007-2010.

22. Hang H, Bertozzi CR. Chemoselective approaches to glycoprotein assembly. *Acc Chem Res.* 2001; 34(9): 727-736.

23. Lin FL, Hoyt HM, van Halbeek H, Bergman RG, Bertozzi CR. Mechanistic investigation of the staudinger ligation. *J Am Chem Soc.* 2005; 127: 2686-2695.

24. Saxon E, Armstrong JI, Bertozzi CR. A "traceless" staudinger ligation for the chemoselective synthesis of amide bonds. *Org Lett.* 2000; 2(14): 2141-2143.

25. Prescher JA, Dube DH, Bertozzi CR. Chemical remodelling of cell surfaces in living animals. *Nature.* 2004; 430: 873-877.

26. Wang Z, Udeshi ND, O'Malley M, Shabanowitz J, Hunt DF, Hart GW. Enrichment and site mapping of O-linked N-acetylglucosamine by a combination of Chemical/Enzymatic tagging, photochemical cleavage, and electron transfer dissociation mass spectrometry. *Molecular & Cellular Proteomics.* 2010 January 01; 9(1): 153-160.

27. Martin SE, Shabanowitz J, Hunt DF, Marto JA. Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem.* 2000; 72: 4266-4274.

28. Earley L, Anderson LC, Bai DL, Mullen C, Syka JEP, English AM, Dunyach J, Stafford GC, Shabanowitz J, Hunt DF, Compton PD. Fron-end electron transfer dissociation: A new ionization technique. *Anal Chem.* 2013; 85(17): 8385-8390.

29. Saxon E, Luchansky SJ, Hang HC, Yu C, Lee SC, Bertozzi CR. Investigating cellular metabolism of synthetic azidosugars with the staudinger ligation. *J Am Chem Soc.* 2002; 124: 14893-14902.

30. Nepomniaschiy N, Grimminger V, Cohen A, DiGiovanni S, Lashuel HA, Brik A. Switch peptide via staudinger reaction. *Org Lett.* 2008; 10(22): 5243-5246.

31. Vocadlo DJ, Hang HC, Kim E, Hanover JA, Bertozzi CR. A chemical approach for identifying O-GlcNAc-MOdified proteins in cells. *Proc Nat Acad Sci.* 2003; 100(16): 9116-9121.

Chapter 5: *In vitro* Mapping of O-GlcNAc Sites on the C-Terminal Tail of Lamin A

5.1 Introduction

The nuclear envelope contains DNA and is the most highly organized organelle in the body. It is essential in forming the boundaries between the nucleoplasm and the cytoplasm. The nuclear envelope is comprised of the inner nuclear membrane (INM) and the outer nuclear membrane (ONM). The nuclear face of the INM consists of the nuclear lamina, which primarily provides structural support for the nucleus. The main constituents of the nuclear lamina are lamin proteins, lamin A/C and Lamin B1 and B2. The lamin proteins not only provide structural integrity to the nuclear envelope, but are also important in a variety of nuclear functions.(1) However, the lamin proteins are not indestructible. Changes in their structure are the causes of numerous devastating diseases, termed laminopathies.(2) Of particular interest is the laminopathy—Hutchinson-Gilford progeria syndrome (HGPS), which is an accelerated aging disease.(3)

The nuclear lamina is a protein meshwork that forms a thin barrier, approximately 20-50 nm, around the edge of the INM.(1) This is fashioned primarily by the lamin proteins and the proteins with which they interact.(1) The nuclear lamins are type V intermediate filaments found exclusively in the nucleus, and closely resemble cytosolic intermediate filaments. There are two main types of lamins, A-type and B-type lamins. A-type lamins are encoded by the *LMNA* gene, which encodes lamin A and lamin C through alternative splicing(2); and the B-type lamins are encoded by two separate genes-*LMNB1* and *LMNB2*, which encodes for lamin B1 and lamin B2, respectively.(4)

A-type lamins are synthesized as precursors that undergo a series of post translational modifications essential to form mature lamin A, which then gets incorporated into the nuclear lamina.(5, 6) Lamins contain a C-terminal CaaX motif, where C is cysteine, a is any amino acid with an aliphatic side chain, and X is any amino acid. This motif undergoes a series of post translation modifications, beginning with the addition of a farnesyl group to the cysteine.(5) After farnesylation, the aaX amino acids are cleaved by the ER-associated endoproteinase, Zmpste24.(5, 7) Following cleavage by Zmpste24, the C-terminus is methylated by an ER-associated methyltransferase rendering the protein more hydrophobic and making it more amenable to associate with the INM.(5) Approximately 30-60 minutes post-synthesis and after assembly into the nuclear lamina, lamin A undergoes another cleavage by Zmpste24 removing an additional 15 amino acids from the C-terminus including the modified cysteine residue.(5) This sequence of events results in the formation of mature lamin A, shown in **Figure 5.1A**.(8)

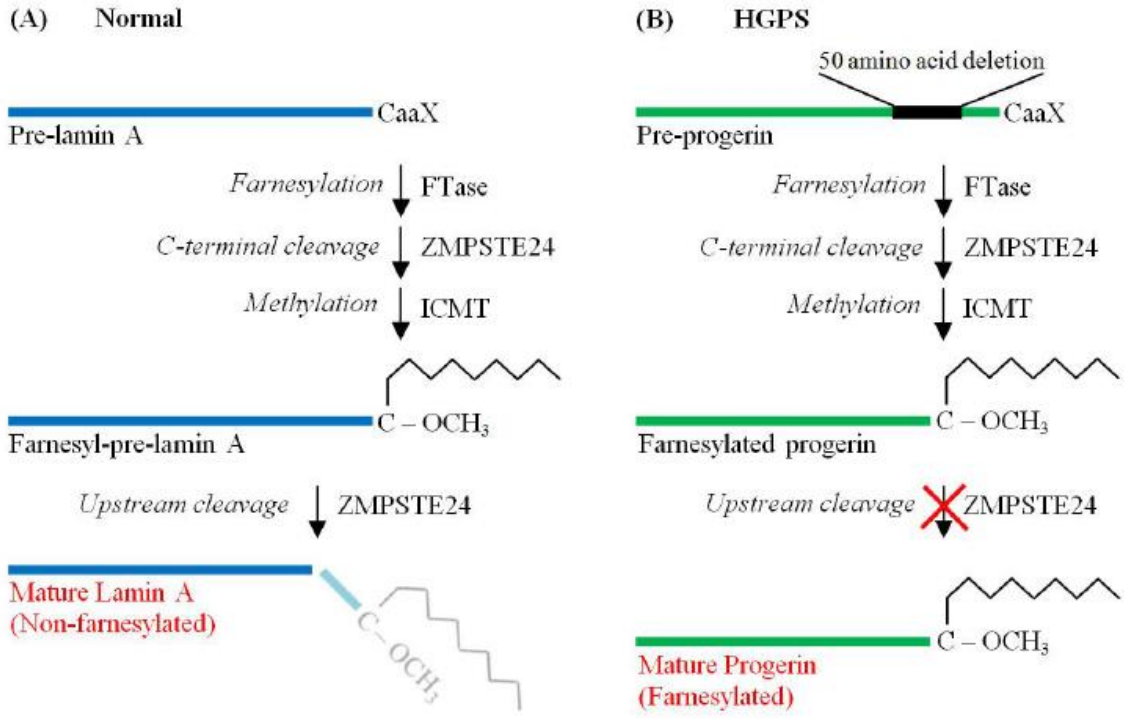

Figure 5.1. Summary of the post translational processing prelamin A. **A.** Prelamin A undergoes a series of post translational processes to form mature lamin A, which is incorporated into the nuclear lamina. **B.** Post translational processing of Pre-progerin results in mature progerin, which retains the farnesyl group due to the deletion of 50 amino acids.(8)

Lamins and cytoplasmic intermediate filaments share a very similar structure, which consists of a central α-helical coiled coil domain flanked by a small nonhelical N-terminal head and a larger non-helical C-terminal tail, outlined in **Figure 5.2A**.(9, 10)(5) The C-terminus also contains a nuclear localization sequence and a highly conserved immunoglobulin fold.(11) The initial step in the polymerization of lamins is the formation of lamin dimers. These lamin dimers assemble into head-to-tail polar polymers, which then laterally aggregate in an anti-parallel fashion forming non-polar

filaments, as shown in **Figure 5.3**.(2, 11) The formation of the lamins determines the shape of the nuclear envelope.

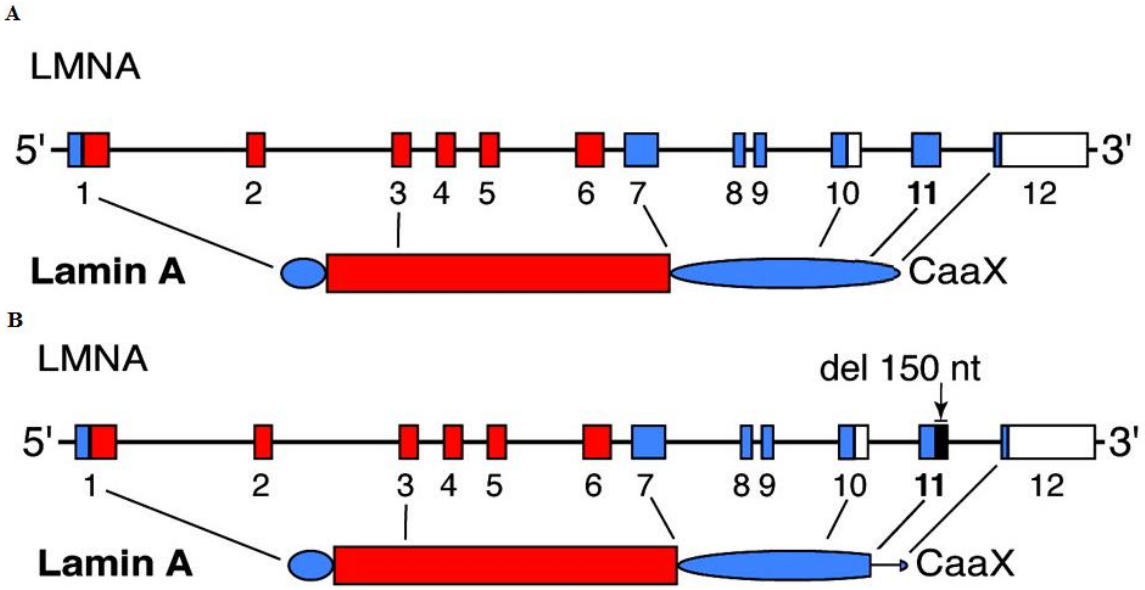

Figure 5.2. Representation of the *LMNA* gene and lamin A protein. The small blue circle represents the N-terminal tail, the red rectangle represents the α-helical coiled coil domain, and the large blue oval represents the C-terminal tail, which includes a nuclear localization sequence and an immunoglobulin fold. **A.** Normal expression and **B.** G608G mutation creating a cryptic splice site at exon 11.(10)

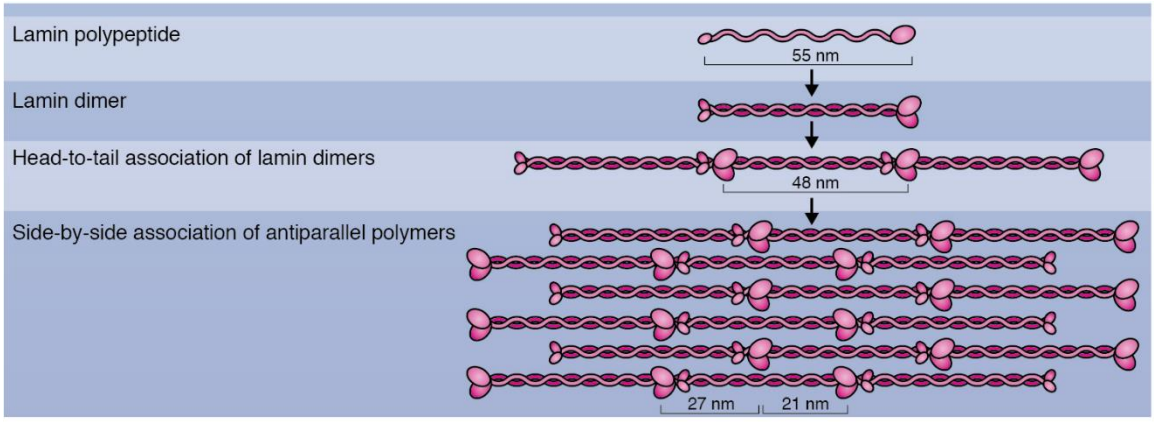

Figure 5.3. Assembly of lamins into filaments.(2)

Lamins have a multitude of functions ranging from structural support of the nucleus to facilitating chromatin organization, gene regulation, and DNA repair.(2) A-type lamins are responsible for determining the shape and mechanical stiffness of the nucleus. Lamins also enable the organization of chromatin and are mediators of transcriptional activities.(2) Lamins provide binding sites for numerous regulatory proteins, and, therefore, contribute to the control of cell proliferation.(12) Alternatively, lamins are directly and indirectly involved in factors regulating programmed cell death or apoptosis.(5) Lamins have over fifty binding partners; however, the process by which lamin associations with these binding partners are regulated is unknown.(6) Lamins are known to be decorated with several post translational modifications including phosphorylation, O-GlcNAcylation, oxidation, SUMOylation (small ubiquitin-like modifier), acetylation, and ubiquitination.(6) These modifications are likely relatively rare and transient, and not much is known about the nature, regulation, or consequences of these modifications.(6) These are just a few examples of lamin influence and regulation of cellular functions. It is clear that cellular processes can be disturbed if the lamins are disrupted in any way. Improper post translational processing of lamin A is a contributing factor to many laminopathies.

Laminopathies are diseases caused by mutations in genes encoding lamin proteins. There have been over 400 different disease-causing mutations identified in the *LMNA* gene.(6) These mutations cause several different types of diseases, which are broadly classified into four groups: diseases of striated muscle, peripheral neuropathy, lipodystrophy syndromes, and accelerated aging disorders.(4) Thus far, over 20 diseases have been classified as laminopathies, some of which are listed in **Table 5.1**.

Laminopathies wreak havoc on the normal functions of lamin A causing nuclear fragility, altered nuclear positioning, and affected gene-expression.(1)

Table 5.1. Diseases caused by mutations in the *LMNA* gene.(8)

Laminopathies
Diseases of Striated Muscle
Autosomal dominant Emery-Dreifuss muscular dystrophy
Autosomal recessive Emery-Dreifuss muscular dystrophy
Autosomal dominant cardiomyopathy dilated 1A
Autosomal dominant limb girdle muscular dystrophy type 1B
Peripheral Neuropathy
Autosomal recessive Charcot-Marie-Tooth disorder type 1B
Lipodystrophy Syndromes
Autosomal dominant Dunnigan-type familial partial lipodystrophy
Autosomal dominant lipoatrophy with diabetes, hepatic steatosis, hypertrophic Cardiomyopathy and leukomelanodermic papules
Autosomal recessive mandibuloacral dysplasia
Accelerated Aging Disorders
Autosomal dominant atypical Werner Syndrome
Autosomal dominant Hutchinson-Gilford progeria syndrome
Autosomal dominant restrictive dermopathy lethal

Of particular interest is the accelerated aging laminopathy, Hutchinson-Gilford progeria syndrome (HGPS). HGPS is a rare, fatal genetic disorder that causes premature aging in children. The disease typically manifests around 24 months of age and is characterized by the loss of subcutaneous fat, hair loss, bone deformations, osteoporosis, delayed dentition, severe growth retardation, progressive arteriosclerosis, hip dislocations, and sclerodermatous areas.(3, 13, 14) Patients with HGPS have an aged appearance, in addition to hypertension, angina, and dilated hearts. HPS afflicted children typically die of myocardial infarction or cerebrovascular accident around age 13.(13) HGPS is caused by a single *de novo* point mutation, GGC>GGT, in the *LMNA*

gene.(7) This point mutation does not alter the amino acid sequence (G608G) of the protein, but does create a cryptic splice site in exon 11, shown in **Figure 5.2B**.(3) This mutant allele produces an alternatively spliced truncated variant of the lamin A mRNA lacking the 3' terminal 150 nucleotides of exon 11.(3) This produces a curtailed protein lacking residues 607-656 of prelamin A, retaining the C-terminal CaaX box.(14) This truncated version of lamin A is called progerin. The post translational processing of progerin is shown in **Figure 5.1B**.(8)

Nuclei in HGPS patients suffer dramatic changes in nuclear morphology.(3) Progerin does not properly integrate into the lamina, which disrupts the scaffold structure and leads to significant disfigurement of the nucleus characterized by lobular shape.(14) The lobular shape in patients with HGPS causes significantly transformed mechanical properties and organization.(3) Alterations in lamina architecture in HGPS nuclei have been shown to alter epigenetic modifications.(3) **Figure 5.4** highlights the differences in the shape of the nucleus in normal cells, which lack progerin, and in HGPS cells, which contain progerin.(14) The severe alterations observed in HGPS nuclear shape are a hallmark for multiple laminopathies and have deleterious effects on important nuclear functions.(13)

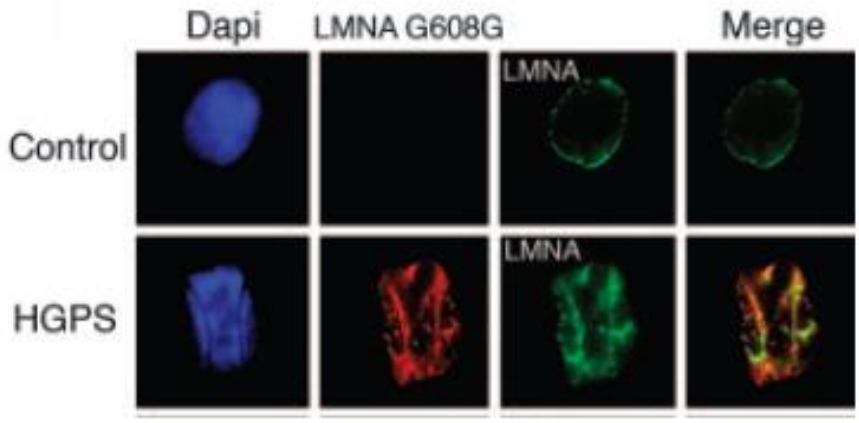

Figure 5.4. HGPS and control cells labeled with anti-LMNA G608G and anti-lamin A reveal the differences in nuclear shape. Chromatin was stained with DAPI.(13)

The fifty amino acid deletion in HGPS patients causes a disruption in the post translational modification of lamin A. A 2010 study by Wang and colleagues identified two O-GlcNAcylated sites on lamin A, S612 and T643.(15) This study did not investigate the functionality of these sites, and their roles have not yet been elucidated. Both of these residues are deleted in progerin, and could play a role in HGPS. For this reason, we and our collaborators wanted to investigate *in vitro* O-GlcNAcylation of the C-terminal tail of mature lamin A and a form of progerin, as well as a double mutant removing the two known O-GlcNAc sites, S612A/T643A. The lamin A tails were analyzed by mass spectrometry to determine sites of O-GlcNAcylation.

5.2 Materials

5.2.1 Reagents

Fisher Scientific, Waltham, MA

Methanol

Honeywell, Morristown, NJ:

Acetonitrile, HPLC grade, ≥99.8% purity

Pierce, Rockford, IL:

LC-MS grade water

Roche Diagnostics, Indianapolis, IN

Endoproteinase AspN

Chymotrypsin

Sigma-Aldrich, St. Louis, MO:

Acetic Acid (AcOH, glacial, 99.99%)

Ammonium bicarbonate

Angiotensin I human acetate hydrate, ≥90% purity (Angio)

Azulene

Dithiothreitol

Formamide

Iodoacetamide

Vasoactive intestinal peptide fragment 1-12, human ≥97% (Vaso)

Thermo-Fisher Scientific, Waltham, MA

Hydrochloric acid (HCl)

5.2.2 Column Making Materials

Polymicro Technologies, Inc., Phoenix, AZ:

Polyimide coated fused silica capillary

360 μm o.d. × 50 μm i.d.

360 μm o.d. × 75 μm i.d.

360 μm o.d. × 150 μm i.d.

Zeus Industrial Products, Inc., Orangeburg, SC:

Teflon tubing, 0.012" i.d. × 0.060" o.d.

PQ Corporation, Valley Forge, PA:

Kasil-Potassium silicate solution

YMC Corp., Morris Plains, NJ (available through Waters Corp.):

ODS-AQ, C-18, 5 μm spherical silica particles, 120 Å pore size

ODS-AQ, C-18, 5-20 μm spherical silica particles, 300 Å pore size

5.2.3 Equipment

Agilent Technologies, Palo Alto, CA:

1100 Agilent HPLC

Eppendorf, Hamburg, Germany:

Centrifuge 5424

Labconco, Kansas City, MO:

CentriVap Vacuum Concentrator System

Sutter Instrument Co., Novato, CA:

P-2000 laser puller with fused silica adapter

Thermo Fisher Scientific, Waltham, MA:

LTQ mass spectrometer (custom modified with backend ETD)

LTQ FT-ICR hybrid mass spectrometer (custom modified with front-end ETD)

LTQ-Orbitrap mass spectrometer (custom modified with front-end ETD)

VWR, Radnor, PA:

Mini vortexer

5.3 Methods

Lamin A C-terminal tail samples were prepared by Dr. Dan Simon of the Wilson Lab at Johns Hopkins University.

5.3.1 Proteolytic Digestion

Each of the samples were dried on a vacuum concentrator to a volume of approximately 10 µL. Disulfide bonds were reduced with 0.1 mM dithiothreitol for 1 hr at room temperature, pH 8. Cysteines were carbamidomethylated with 0.25 mM iodoacetamide in the dark for 1 hr at room temperature, pH 8. Each sample was proteolytically digested using endoproteinase AspN at an enzyme to substrate ratio of 1:20 *w/w* in 100 mM ammonium bicarbonate, pH 8, at room temperature for 8 hrs. After 8 hrs the digestion was quenched by acidifying the sample to pH 3.5 with glacial acetic acid. A portion of each digested sample was subdigested with Chymotrypsin using the same conditions as the AspN digestion.

5.3.2 Assembly of Capillary HPLC Columns

Precolumns (PC) were constructed by creating a Kasil frit in the end of a fused-silica capillary column (360 μm o.d. × 75 μm i.d.). Kasil frits were formed by dipping the end of a capillary column into a mixture of Kasil:formamide (3:1) and heating in an oven overnight at 70˚C. The fritted column was packed with 5-8 cm of irregular, 5-20 μm diameter, 300 Å pore size, C18 reverse-phase resin using a pressure bomb (500 psi). Analytical columns (AC) were constructed with integrated electrospray emitter tips.(16) A Kasil frit was created in a short section of the fused silica (360 μm o.d. × 50 μm i.d.). Columns were then packed with 6-9 cm of 5 μm diameter, 120 Å pore size, C18 reverse-phase resin using a pressure bomb (500 psi). A micropipette laser puller was used to generate a tip (2 μm i.d.) approximately 5 mm from the Kasil frit. Analytical columns were butt-connected to the precolumns using Teflon tubing (0.012" i.d. × 0.060" o.d.).

5.3.3 Loading and Rinsing of Samples

Samples were pressure loaded (~200 psi, ~1 μL/min.) on a PC for each mass analysis. Approximately 100 fmol each of two standard peptides (angiotensin I and vasoactive intestinal peptide), were also loaded onto the column for quantitation purposes. The precolumn was then rinsed on the HPLC with solvent A to remove salts and other small molecules that may interfere with mass spectrometric analysis (30 bar, 30-40 min.). The analytical column was then connected to the precolumn using Teflon tubing (0.012" i.d. × 0.060" o.d.) and rehydrated with solvent A for ~15 min.

5.3.4 Mass Spectrometric Analysis

Samples were analyzed by nanoflow-HPLC/microelectrospray ionization coupled directly to a Thermo Orbitrap Classic or FT-ICR mass spectrometer with a home built front-end electron transfer dissociation (FETD) source.(17) After loading the sample on the PC and reattaching the PC to the AC the sample was gradient eluted through a laser-pulled electrospray tip directly into the mass spectrometer with an Agilent 1100 series binary LC pump at a flow rate of ~60 nL/min. The elution gradient was as follows: 0-60% B in 60 minutes, 60-100% B in 3 minutes, hold at 100% B for 3 minutes, 100-0% B in 2 minutes, and hold at 0% B for 20 minutes.

For standard peptide analysis the instrument was operated in a data-dependent mode, performing a high resolution MS1 scan followed by acquisition of MS2 spectra in the linear ion trap on the most abundant peak, first for CAD then ETD. The ETD reagent used was azulene and ETD reaction times were either 30 or 35 ms. For peptide analysis the instrument was operated in a data-dependent mode similar to the standard peptide analysis selecting the top five most abundant, non-excluded ions in the MS1 for fragmentation by both CAD and ETD. Targeted runs were performed on the WT and S612A/T643A lamin A tails to obtain MS2 spectra of multiply O-GlcNAcylated species.

5.3.5 Data Analysis

MS/MS spectral data were searched against the lamin A protein sequence using the Open Mass Spectrometry Search Algorithm (OMSSA) software. (18) OMSSA search tolerances were ± 0.01 Da and ± 0.35 Da for precursor and product ion masses, respectively. For ETD spectra, search parameters were set to exclude reduced charge species from peak lists prior to searching. Data were searched using variable modifications of O-GlcNAcylation of serine and threonine residues, and oxidation of methionine, with 2048 variable modifications per peptide being allowed. Other search parameters utilized were: +2 charge state products allowed for peptides of charge +3 and above. Peptide sequences were manually validated, and O-GlcNAc sites were manually assigned. Additionally, in house software was used to generate a list of spectra with an ion at m/z 204, O-GlcNAc oxonium signature peak, and the charge reduced product ion with the loss of 203 amu to identify potential O-GlcNAcylated peptides

5.4 Results and Discussion

5.4.1 Evidence of O-GlcNAcylation on Lamin A Tails

The purpose of this study was to identify O-GlcNAc modifications of wild type (WT) versus mutated lamin A C-terminal tail domains. To evaluate this, our collaborators performed *in vitro* O-GlcNAcylation reactions on purified lamin A C-terminal tails by incubating the protein with UDP-GlcNAc, the donor substrate for O-GlcNAcylation, calf intestinal phosphatase, and OGT for 2 hrs at 22-24°C, then overnight at 4°C. The western blot in **Figure 5.5** shows the results of the reacted lamin A tails

probed for O-GlcNAcylation using the CTD 110.6 antibody, which is specific for O-GlcNAc. The western blot revealed that the WT, S612A, T643, and S612A/T643A lamin

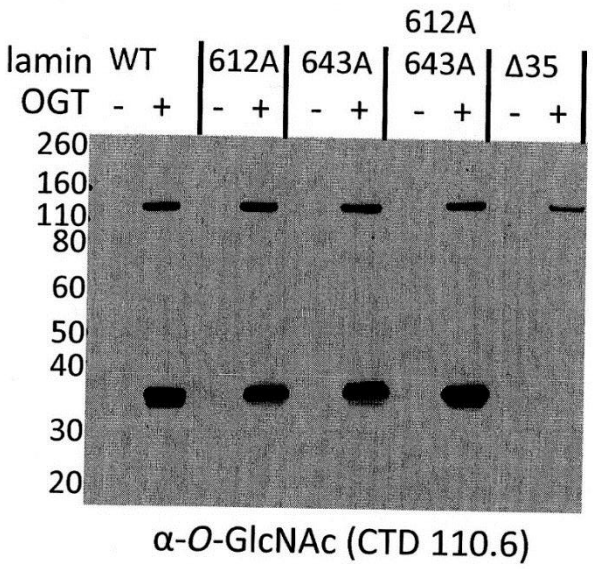

Figure 5.5. Western blot of lamin A tails with the GlcNAc specific antibody CTD 110.6

A tails were robustly O-GlcNAcylated. In contrast, the HGPS-associated Δ35 deletion does not appear to be O-GlcNAcylated. However, upon a longer exposure time, a faint band was present indicating O-GlcNAcylation of the Δ35 lamin A tail does occur (data not shown).

5.4.2 Mass Spectrometric Analysis

Typically, proteins are analyzed by mass spectrometry using a bottom up approach, in which the proteins are proteolytically digested prior to MS analysis. There are several proteolytic enzymes available, and, it is important to select an appropriate enzyme. *In silico* digestions were utilized to determine the protease best suited. Since

the goal of this project was to site-map O-GlcNAc sites, the enzyme chosen provided the highest theoretical coverage of Ser and Thr residues, and produced peptides with higher charge states. AspN, which hydrolyzes bonds on the N-terminal side of aspartic acid residues, was selected to digest lamin A tails. **Table 5.2** lists the results from the *in silico* digest of the WT and S612A/T643A lamin A tails by AspN, and **Table 5.3** lists the results from the *in silico* digest of the Δ35 lamin A tail by AspN. A portion of each digested sample was sub-digested with chymotrypsin, which cleaves C-terminal of Trp, Tyr, Phe, Leu, and Met.

Table 5.2. *In silico* AspN digest of WT and S612A/T643A lamin A tails.

Mass	Position	Theoretical Peptide
1254.5388	T7 tag	*masmtggqqmgr*
6837.4947	T7 tag, 385-447	*dp* ERLRLSPSPTSQRSRGRASSHSSQTQGGGSVT KKRKLESTESRSSFSQHARTSGRVAVEEV
1820.9249	448-460	DEEGKFVRLRNKSNE
1661.7812	461-474	DQSMGNWQIKRQNG
134.0448	475	D
3708.99	476-510	DPLLTYRFPPKFTLKAGQVVTIWAAGAGATHSP PT
4531.3547	511-551	DLVWKAQNTWGCGNSLRTALINSTGEEVAMRK LVRSVTVVE
134.0448	552	D
263.0874	553-554	DE
263.0874	555-556	DE
191.0662	557-558	DG
134.0448	559	D
1610.6877	560-574	DLLHHHHGSHCSSSG
2251.0594	575-595	DPAEYNLRSRTVLCGTCGQPA
3863.839	596-638	DKASASGSGAQVGGPI**S**SGSSASSVTVTRSYRSV GGSGGGSFG
1959.9433	639-646, His tag	DNLV**T**RSY*lghhhhhh*

Lowercase italicized letters represent affinity tag sequences. Bolded uppercase letters denote the positions that are mutated A residues in the double mutant.

Table 5.3. *In silico* AspN digest of Δ35 lamin A tail.

Mass	Position	Theoretical Peptide
1254.5388	T7 tag	*masmtggqqmgr*
6837.4947	T7 tag, 385-447	*dp* ERLRLSPSPTSQRSRGRASSHSSQTQGGGSVT KKRKLESTESRSSFSQHARTSGRVAVEEV
1820.9249	448-460	DEEGKFVRLRNKSNE
1661.7812	461-474	DQSMGNWQIKRQNG
134.0448	475	D
3708.99	476-510	DPLLTYRFPPKFTLKAGQVVTIWAAGAGATHSP PT
4531.3547	511-551	DLVWKAQNTWGCGNSLRTALINSTGEEVAMRK LVRSVTVVE
134.0448	552	D
263.0874	553-554	DE
263.0874	555-556	DE
191.0662	557-558	DG
134.0448	559	D
1610.6877	560-574	DLLHHHHGSHCSSSG
2251.0594	575-595	DPAEYNLRSRTVLCGTCGQPA
4101.8747	596-664 (Δ622-657), His tag	DKASASGSGAQVGGPISSGSSASSVTSPQNCSIM *lghhhhhh*

Lowercase italicized letters represent affinity tag sequences. The underlined residues indicate where the Δ35 deletion occurs.

Mass analyses of the samples were performed using high resolution mass spectrometry and low resolution MS2 fragmentation by both CAD and ETD. The instrument was operated in a data-dependent method selecting the top 5 most abundant, non-excluded ions from the full MS scan. The data were searched against the lamin A sequence using OMSSA, which guided data analysis and interpretation. All spectra were manually validated. Additionally, in house software assisted in identifying spectra displaying the characteristic ions of an O-GlcNAcylated peptide. A CAD spectrum of an O-GlcNAcylated peptide is dominated by an ion at m/z 204, which corresponds to the

GlcNAc oxonium ion, and the corresponding charge reduced product ion with a loss of 203 amu.

The base peak chromatogram of WT lamin A tail was dominated by lamin A peptides indicating that any contaminants present did not appear to suppress the signal of the sample. The AspN-chymotrypsin mass spectral analysis resulted in 86% peptide coverage. The AspN-chymotrypsin digest resulted in numerous cleavage sites within the same peptide, since it targets multiple residues for cleavage. For example, the AspN peptide containing residues 639-646 produced six different peptides in the AspN-chymotrypsin digest. This results in a more complex data set.

Two peptides were found to be O-GlcNAcylated. The AspN peptide containing residues 639-646 confirmed the previously reported O-GlcNAc site on T643. Most interesting is the AspN peptide containing residues 596-638, which was found both unmodified and heavily modified with up to seven O-GlcNAc moieties. This peptide contains a poly S/T region that appears to be a 'hot spot' for O-GlcNAcylation. The extracted ion chromatogram (XIC) in **Figure 5.6** shows the elution profile of the multiple O-GlcNAcylated forms of the peptide containing residues 596-638. This peptide contains the previously reported O-GlcNAc site, S612. Interestingly, the di-GlcNAc version of this peptide was significantly more abundant than the mono-GlcNAc version, 35.8% compared to 9%. Also, the tri-GlcNAc version was more abundant than the mono-GlcNAc, 9% compared to 14%. The unmodified version was the most abundant, present at 37.3%. The penta-, hexa-, and hepta-GlcNAc species were present at a combined level of 3.9%. Abundances of each species are shown in **Figure 5.7**.

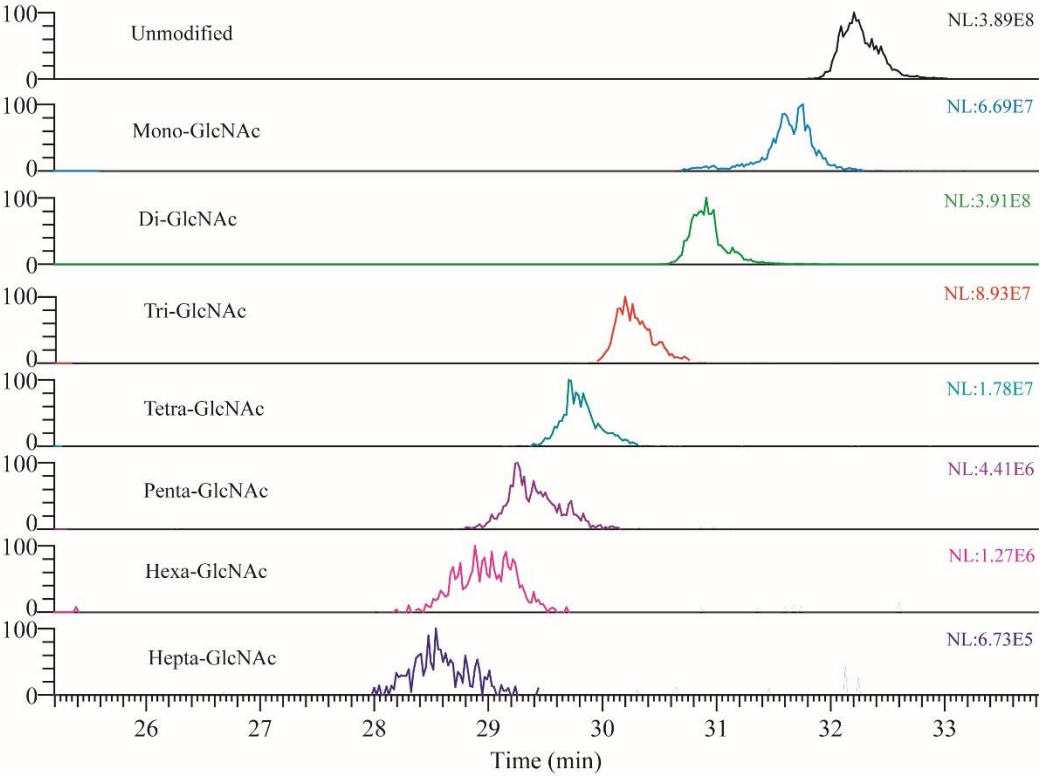

Figure 5.6. XIC's for the O-GlcNAcylated peptide of WT lamin A tail 596-638. XICs were generated using a 10ppm window around the calculated ^{13}C isotopic mass for each species of the peptide 596-638.

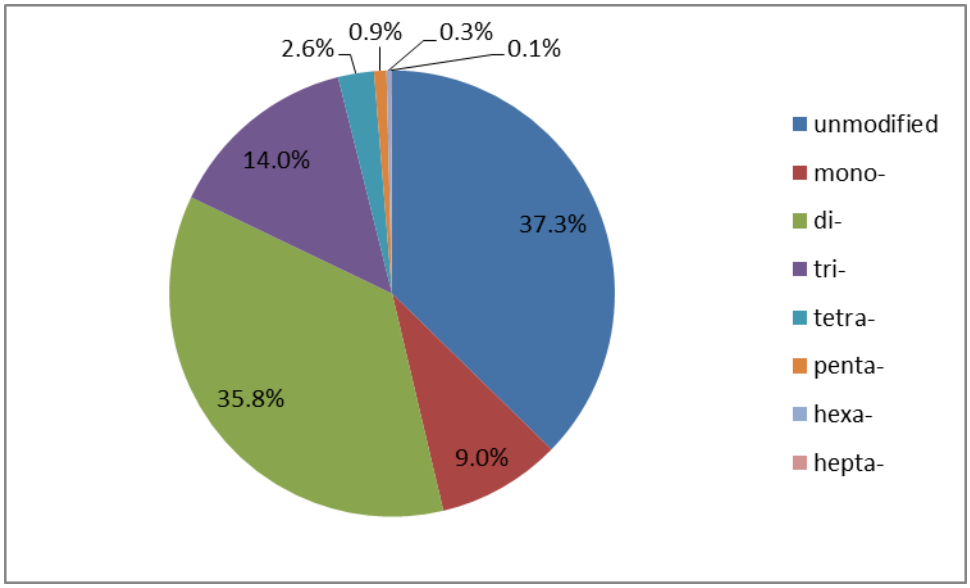

Figure 5.7. Abundance of O-GlcNAc modified species for the AspN peptide 596-638 for WT lamin A tail.

The penta-, hexa-, and hepta-GlcNAcylated versions of peptide 596-638 were low level and were not selected for MS2 fragmentation. Therefore, in order to site-map these versions a targeted run was performed. In a targeted run the user decides which masses will be selected for fragmentation. The AspN digested sample was used for the targeting experiment, because it contained the peptide with the most versions. The spectra of the O-GlcNAcylated peptides revealed that more residues were O-GlcNAcylated than were detected by accurate mass. For example the tri-GlcNAc spectrum displayed a mix of six different residues, with three being modified at any given time. In total twelve residues were found to be O-GlcNAcylated given all the modified versions of this peptide (596-638). There were three major sites: S612, S613, and S618; three minor sites: S603, S616, and T623; and six other low abundance sites: S601, S615, S619, T621, S625, and S628. **Table 5.4** summarizes the site assignments for the O-GlcNAcylated peptides.

Table 5.4. O-GlcNAc assignments for WT lamin A tail.

Peptide	**Mono-**	**Di-**	**Tri-**	**Tetra-**	**Penta-**	**Hexa-**	**Hepta-**
DKASASGSGAQVGGPIASGSSASSVTVTRSYRSVGGSGGGSFG	S612, S613, S603, T623	S612, S613, S603	S612, S613, S616, T623, S603	S612, S613, S616, T623, S619, S628, S601, S615, S618	S612, S613, S618, T621, S628, S615, S616, T623	S612, S613, S615, S618, T623, S601, S628, S616, S619, T621	S612, S613, S618, T621, S625, S616, T623
DNLVTRSYL	T643						

Analysis of the lamin A tail double mutant S612A/T643A yielded similar sequence coverage compared to the WT results. The bas peak chromatogram of the AspN-chymotrypsin digest was dominated by lamin A peptides. Multiple versions of a single AspN peptide were seen in the chymotrypsin subdigest as they were in the WT. Two peptides were observed to be O-GlcNAcylated. The same peptide containing the T643A mutation was O-GlcNAcylated, but was not selected for MS2 fragmentation. Therefore, the O-GlcNAc site could not be confirmed by MS2; however, there was only one residue, S645, that could potentially be O-GlcNAcylated.

Comparable to the WT, the peptide containing residues 596-638 was found to be heavily O-GlcNAcylated, although not to the same extent. The XIC in **Figure 5.8** shows the peptide is modified up to five times. The tetra- and penta- versions of this peptide were targeted in a separate run in order to obtain MS2 spectra of these species, since they were not selected for fragmentation in the data dependent run. Results showed the mono-GlcNAc version of this peptide is the most abundant with 50.1% being modified. The unmodified version was the next most abundant species with 24.5% modification. The di-GlcNAc species was present at 18.7% and the tetra- and penta-GlcNAc species were present at <2% each, as shown in **Figure 5.9**. In total, eleven residues in this peptide were found to be O-GlcNAcylated. There were three major sites: S613, S615, and S616; four minor sites: S618, S619, T621, and T623; and four other sites were also O-GlcNAcylated: S601, S603, S625, and S628. These results suggest OGT compensates for the loss of two modifiable residues in the double mutant by targeting S613 plus five nearby residues. **Table 5.5** summarizes the site assignments of the O-GlcNAcylated residues.

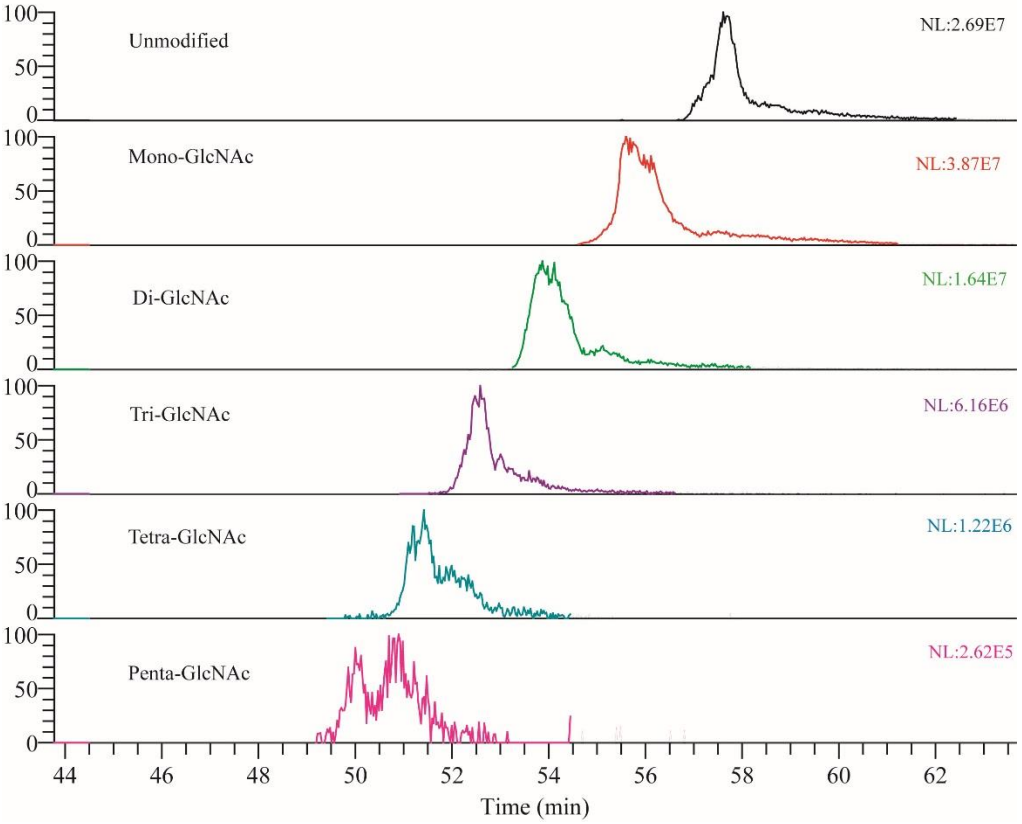

Figure 5.8. XIC's for the O-GlcNAcylated peptide of the S612A/T643A double mutant lamin A tail 596-638. XICs were generated using a 10ppm window around the calculated ^{13}C isotopic mass for each species of the

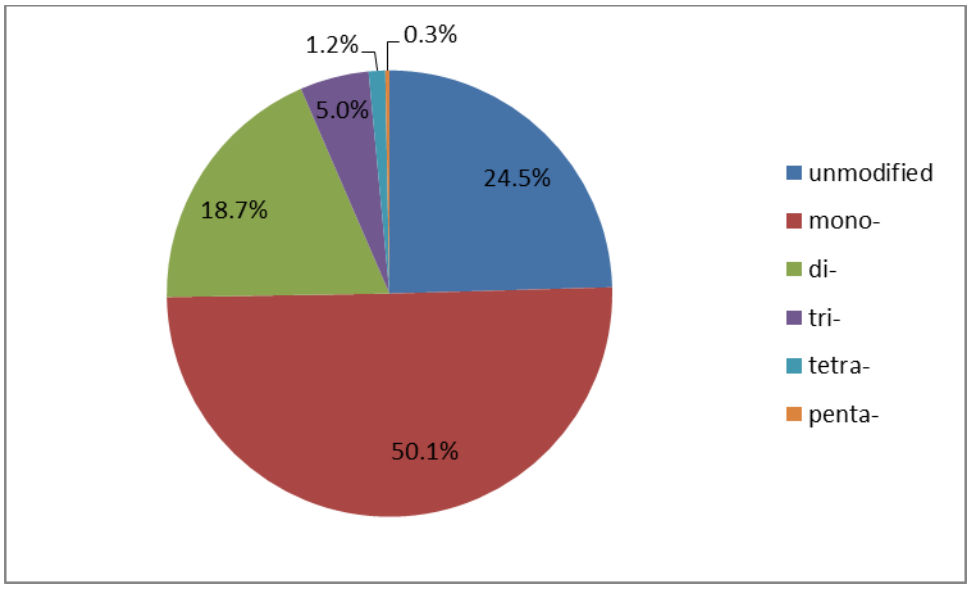

Figure 5.9. Abundance of O-GlcNAc modified species for the AspN peptide 596-638 for S612A/T643A double mutant lamin A tail.

Table 5.5. O-GlcNAc assignments for the S612A/T643A double mutant lamin A tail.

Peptide	Mono-	Di-	Tri-	Tetra-	Penta-
DKASASGSGAQVGGPIASGSSASSVTVTRSYRSVGGSGGGSFG	S613	S613, S615, S616, S618, S619	S613, S615, S616, T621, T623	S613, S615, S601, S603, T621, S619, T251, S628, S618	S603, S613, S615, S625, S218, S628, S601, S616, T623
DNLVARSYL	S645				

Mass spectral analysis of the Δ35 lamin A tail AspN run resulted in 52% sequence coverage, including the 'hot spot' poly S/T region. The base peak of the AspN run was dominated by lamin A peptides. Only two peptides were observed to be O-GlcNAcylated, the N-terminal tag and the peptide containing the 'hot spot' poly S/T region. However, this peptide was only detected to be mono-GlcNAcylated at a low efficiency of 2.25%. The mono-GlcNAc was most abundantly modified at S612 and less abundantly at S601 and S618, as shown in **Table 5.6**. The deficiency of O-GlcNAcylation of the HGPS related

Table 5.6. O-GlcNAc assignments for the Δ35 lamin A tail.

Peptide	Mono-
DKASASGSGAQVGGPISSGSSASSVTSPQNCSIMLGHHHHHH	S612, S601, S618

Δ35 tail indicates that OGT fails to recognize or modify S613 in this version of the protein although, it was found to be a major target in the previous two samples. The

deficiency of O-GlcNAcylation could have negative regulatory consequences detrimental to HGPS patients.

5.5 Conclusion

Mass spectral analysis revealed the tails of WT lamin A and the S612A/T643A double mutant lamin A were hyper-O-GlcNAcylated *in vitro*. The majority of the O-GlcNAcylation of these species occurred in a 'hot spot' region encompassing residues 601-628. Within this 27 residue region, there are 12 modifiable residues, and all were found to be O-GlcNAcylated to some degree. Conversely, analysis of the Δ35 lamin A tail exposed a deficiency in O-GlcNAcylation. The deleted amino acids in the Δ35 tail contains seven residues of the 'hot spot' found in the WT tail. Even with the deletion, there are still eight modifiable residues present, and only three of these were found to be O-GlcNAcylated, with only one modification occurring at any given time.

The modifications in the O-GlcNAc 'hot spot' discovered by this analysis could potentially cause dysregulation of lamin associated cellular processes in HGPS patients. Determining the biological implications of the hypo-O-GlcNAcylation in HGPS patients could reveal new treatment targets for this deadly laminopathy. Additionally, there is more work to be done to elucidate how hyper-O-GlcNAcylation affects normal cellular functions. Currently, the biological roles of lamin A O-GlcNAcylation are unknown indicating another area for future research. While lamin A O-GlcNAcylation is not genetically linked to disease, the modifications are located near residue mutations that do

cause laminopathies. It is possible that these mutations located near O-GlcNAc sites interfere with OGT preventing normal O-GlcNAcylation.

Finally, the relationship between phosphorylation and O-GlcNAcylation must be addressed. As discussed in the previous chapter, there is a dynamic interplay between phosphorylation and O-GlcNAcylation. In fact, all known O-GlcNAcylation sites are known phosphorylation sites.(19) Four of the identified O-GlcNAc sites are also known phosphorylation sites (S613, S615, S616, and S619).(6) Additionally, several O-GlcNAcylation sites are located in close proximity to phosphorylation sites, and the presence or lack of modification on one residue could affect the modification of a nearby residue. This dynamic interplay between these two modifications could have implications in disease and requires further investigation.

5.6 References

References

1. Burke B, Stewart CL. Life at the edge: The nuclear envelope andd human disease. *Nat Rev Mol Cell Bio.* 2002; 3: 575-585.

2. Ho CY, Lammerding J. Lamins at a glance. *J Cell Sci.* 2012; 125: 2087-2093.

3. Dahl KN, Scaffidi P, Islam MF, Yodh AG, Wilson KL, Misteli T. Distinct structural and mechanical properties of the nuclear lamina in hutchinson-gilford progeria syndrome. *Proc Nat Acad Sci.* 2006; 103(27): 10271-10276.

4. Worman HJ, Bonne G. "Laminopathies": A wide spectrum of human diseases. *Exp Cell Res.* 2007; 313: 2121-2133.

5. Burke B, Stewart CL. The laminopathies: The functional architecture of the nucleus and its contribution to disease. *Annu Rev Genomics Hum Genet.* 2006; 7: 369-405.

6. Simon DN, Wilson KL. Partners and post-translational modifications of nuclear lamins. *Chromosoma.* 2013; 122: 13-31.

7. Eriksson M, Brown WT, Gordon LB, Glynn MW, Singer J, Scott L, Erdos MR, Robbins CM, Moses TY, Berglund P, Dutra A, Pak E, Durkin S, Csoka AB, Boehnke M, Glover TW, Collins FS. Recurrent *de novo* point mutations in lamin A cause hutchinson-gilford progeria syndrome. *Nature.* 2003; 423: 293-297.

8. Baek J, McKenna J, Eriksson M. Hutchinson-gilford progeria syndrome. In: Puiu M, editor. Genetic Disorders. 3rd ed. InTech; 2013. p. 65-87.

9. Worman HJ. Nuclear lamins and laminopathies. *J Pathol.* 2012; 226: 316-325.

10. De Sandre-Giovannoli A, Bernard R, Cau P, Navarro C, Amiel J, Boccaccio I, Lyonnet S, Stewart CL, Munnich A, Le Merrer M, Levy N. Lamin A truncation in hutchinson-gilford progeria. *Science.* 2003; 300(5628): 2055.

11. Shumaker DK, Kuczmarski ER, Goldman RD. The nucleoskeleton: Lamins and actin are major players in essential nuclear functions. *Curr Opin Cell Bio.* 2003; 15: 358-366.

12. Wilson KL, Berk JM. The nuclear envelope at a glance. *J Cell Sci.* 2010; 123: 1973-1978.

13. Goldman RD, Shumaker DK, Erdos MR, Eriksson M, Goldman AE, Gordon LB, Gruenbaum Y, Khuon S, Mendez M, Varga R, Collins FS. Accumulation of mutant lamin A causes progressive changes in nuclear architecture in hutchinson-gilford progeria syndrome. *Proc Nat Acad Sci.* 2004; 101(24): 8963-8968.

14. McClintock D, Gordon LB, Djabali K. Hutchinson-gilford progeria mutant lamin A primarily targets human vascular cells as detected by an anti-lamin A G08G antibody. *Proc Nat Acad Sci.* 2006; 103(7): 2154-2159.

15. Wang Z, Udeshi ND, Slawson C, Compton PD, Kaoru S, Cheung WD, Shabanowitz J, Hunt DF, Hart GW. Extensive crosstalk between O-GlcNAcylation and phosphorylation regulated cytokinesis. *Sci Signal.* 2010; 3(104): 1-12.

16. Martin SE, Shabanowitz J, Hunt DF, Marto JA. Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem.* 2000; 72: 4266-4274.

17. Earley L, Anderson LC, Bai DL, Mullen C, Syka JEP, English AM, Dunyach J, Stafford GC, Shabanowitz J, Hunt DF, Compton PD. Fron-end electron transfer dissociation: A new ionization technique. *Anal Chem.* 2013; 85(17): 8385-8390.

18. Greer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, Bryant SH. Open mass spectrometry search algorithm. *J Prot Res.* 2004; 3: 958-964.

19. Hart GW, Housley MP, Slawson C. Cycling of O-linked β-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature.* 2007; 446: 1017-1022.