Improved methods for identification of MHC Class I-associated peptides

Keira Erol Mahoney
Jamestown, Rhode Island

Bachelor of Science in Chemistry
University of Rhode Island, 2016

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 2022

Acknowledgments

I would first like to thank my advisor, Dr. Donald Hunt, for providing me with the opportunity to research in his lab. I truly feel that I couldn't have had a better graduate school experience in any other lab. I would also like to thank Dr. Jeffrey Shabanowitz for his unending guidance, support, and friendship. I could not have done any of this without his help, and I can only hope to one day have his level of expertise. Dina Bai was also a great source of support, and I cannot thank her enough for helping me to learn C#, which has become an invaluable tool in my research.

I would also like to thank my mother and grandmother, without whom I would never would have made it through. Additionally, thank you to Judy Norris for putting me in touch with Donald Hunt in the first place.

Countless current and former Hunt lab members were also invaluable sources of knowledge, labor, and support. I appreciate Mark Ross, Ben Barnhill, and Stephanie Lehman for training me in the ways of the lab. Additionally, thanks to Paisley Myers and Jenn Abelin, who helped me to learn IMAC and immunopurifications respectively. A particular thanks to Xi Peng, Negin Ghafourian, Emily Henry, and Maria Panepinto, all of whom helped test protocols from this dissertation. Thanks to Elisabeth Duselis, Rob D'Ippolito, and Maria Panepinto for their friendship and willingness to listen to my rants. And of course a massive thanks to Stacy Malaker for her invaluable support throughout the writing process.

Many of my samples were generated by collaborators, including Nico Buettner, Jamie Heather, Katie Cascino, Amanda Lulu, Kara Cummings, and Jennifer Sokolowski. Thanks to them as well and their principal investigators for providing me with material for my research

Thank you to Drs. David Cafiso, Jill Venton, Stacy Malaker, and Jennifer Sokolowski for serving on my committee. I know it was short notice, and I greatly appreciate it. Thank you to Drs. David Cafiso, Jill Venton, and Ken Hsu for previously serving on my candidacy committee.

And finally, an additional thank you for understanding that if I forgot to include someone, it is not due to a lack of appreciation, I'm just really, really tired of typing.

Table of Contents

List of Figures and Tables

List of Abbreviations

°C	degrees Celsius
●	radical species
β2m	beta-2 microglobulin
Å	angstrom
a-	atto- ($1x10^{-18}$)
AC	analytical column, alternating current
ACT	adoptive cell transfer
AD	Alzheimer's disease
AIDS	acquired immune deficiency syndrome
AIIP	phosphorylated angiotensin II phosphate (DRVyIHPF)
ACN	acetonitrile
ACT	adoptive cell transfer therapy
AGC	automatic gain control
AL	aldehyde
Ala, A	alanine
ALL	acute lymphoblastic lymphoma
AmBic	ammonium bicarbonate
AML	acute myeloid leukemia
amu	atomic mass unit
Angio	angiotensin I (DRYIHPFHL)
API	atmospheric pressure ionization
APC	antigen-presenting cell
Arg, R	arginine
Asn, N	asparagine
Asp, D	aspartic acid
ATP	adenosine triphosphate
AVL3	melanoma cell line
bar	100,000 Pascals; 14.5038 pounds per square inch

BLAST	Basic Local Alignment Search Tool
BP	base peak
BSA	bovine serum albumin
c-	centi- ($1x10^{-2}$)
C18	octadecyl
Ca^{+2}	calcium (II) cation
$CaCl_2$	calcium chloride
CAD	collision activated dissociation
CAR	chimeric antigen receptor
CD	cluster of differentiation
CD8+	presenting surface glycoprotein CD8
Cdk2	cyclin-dependent kinase 2
ceq	cell equivalents, number of cells used to generate
CHAPS	3- [(3-cholamidopropyl) dimethylammonio]-1-propane sulfonate
CIP2A	cancerous inhibitor of PP2A
CLL	chronic lymphocytic leukemia
CML	chronic myelogenous leukemia
cm	centimeter
CO_2	carbon dioxide
CRM	charged residue model
C-term	C-terminus of peptide
CTL	cytotoxic T lymphocyte
Cys, C	cysteine
Da	Dalton, 1 amu
DC	direct current
DMEM	Dulbecco's Modified Eagle Medium
DNA	deoxyribonucleic acid
EBV	Epstein-Barr virus
EDTA	ethylenediaminetetraacetic acid

EGF	epidermal growth factor
ELISpot	enzyme-linked immunospot assay
ER	endoplasmic reticulum
ESI	electrospray ionization
ETD	electron transfer dissociation
eV	electron volt, 1.602×10^{-19} J
f-	femto- (1×10^{-15})
FaDu	hypopharyngeal squamous cell carcinoma cell line
FBS	fetal bovine serum
Fe^{2+}	iron (II) cation
Fe^{3+}	iron (III) cation
$FeCl_3$	iron (III) chloride
FETD	front-end electron transfer dissociation
FHIOSE	Simian virus 40 transformed nontumorigenic ovarian cell line
FT	Fourier transform or high resolution
FTMS	Fourier transform mass spectrometry
FWHM	full width at half maximum
g	gram
Gln, Q	glutamine
Glu, E	glutamic acid
Gly, G	glycine
h, hr	hour
HBV	hepatitis B virus
HCV	hepatitis C virus
HCC	hepatocellular carcinoma
HCD	higher-energy collisional dissociation
HCl	hydrochloric acid
HD	healthy donor
HEK293	transformed human embryonic kidney cell line
HeLa	cervical adenocarcinoma cell line

Hex	Hexose
HexNAc	N-Acetyl hexosamine
HILIC	hydrophilic interaction liquid chromatography
His, H	histidine
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
HPLC	high performance/pressure liquid chromatography
HPV	human papillomavirus
I2PP2A	inhibitor 2 of PP2A (protein SET)
ICR	ion cyclotron resonance
i.d.	inner diameter
IDA	iminodiacetic acid
IDO	indoleamine 2,3-dioxygenase
IEM	ion evaporation model
IL	interleukin
Ile, I	isoleucine
IMAC	immobilized metal affinity chromatography
IFN-γ	interferon gamma
IRM	ion routing multipole
IT	ion trap or low resolution
ITMS	ion-trap mass spectrometer
JEG3	placental choriocarcinoma cell line
JY	Epstein-Barr virus immortalized B-cell line
K562	blast phase CML cell line
KaSil	potassium silicate solution
kDa	kilodalton ($1x10^3$ Da)
K_{eq}	association constant
kV	kilovolt ($1x10^3$ V)
L	liter
LC	liquid chromatography/ chromatograph

Leu, L	leucine
LN	lymph node
LOD	limit of detection
LTQ	linear ion trap quadrupole
Lys, K	lysine
m	mass, meter
M	molar (moles/liter)
m-	milli- ($1x10^{-3}$)
µ-	micro- ($1x10^{-6}$)
m/z	mass to charge ratio
M+nH	molecular ion with n charges
mAb	monoclonal antibody
MC	metal-chelating
MeCN	acetonitrile
MeOH	methanol
Met, M	methionine
Mg^{2+}	magnesium (II) cation
$MgCl_2$	magnesium chloride
MHC	major histocompatibility complex
min	minute
MOAC	metal oxide affinity chromatography
mol	mole, $6.022x10^{23}$ molecules
MS	mass spectrometry
MS1	full mass spectrum
MS2	tandem mass spectrum
MW	molecular weight
MWCO	molecular weight cut off
n-	nano- ($1x10^{-9}$)
NA	not applicable
NaCl	sodium chloride

nCE	normalized collision energy
ND	not detected, neurodegenerative
NHS	N-hydroxysuccinimide
NK	natural killer cells
NL	normalized
NSB	non-specific binding
NTA	nitrilotriacetic acid
N-term	amino-terminus of peptide
o.d.	outer diameter
O-GlcNAc	O-linked β-N-acetylglucosamine
OT	Orbitrap, high resolution
p-	pico- ($1x10^{-12}$)
p53	tumor protein 53
PBMC	peripheral blood mononucleated cells
PBS	phosphate-buffered saline
PC	pre-column
PD	Proteome Discoverer
PEEK	polyether ether ketone
PEG	polyethylene glycol
PHA	phytohemagglutinin
Phe, F	phenylalanine
PHEA	polyhydroxy ethyl aspartamide
pKa	acid dissociation constant
PMSF	phenylmethylsulphonyl fluoride
PP1	protein phosphatase 1
PP2A	protein phosphatase 2A
ppm	parts per million
pRb	retinoblastoma protein
Pro, P	proline
PRTC	Pierce retention time calibration mixture

psi	pounds per square inch
PSM	peptide spectral match
pStds	phosphorylated standard peptides
PTFE	polytetrafluoroethylene
PTM	post-translational modification
PVDF	polyvinylidene fluoride
QMF	quadrupole mass filter
RB	retinoblastoma
x g	relative centrifugal force
RF	radio frequency
RNA	ribonucleic acid
RP	reverse phase
RT	room temperature, retention time
s, sec	seconds
S/N	signal to noise
sHLA	soluble human leukocyte antigen
Ser, S	serine
SET	protein SET
SPE	solid-phase extraction
STAGE	stop and go extraction
STCU	STAGE tip cleanup
TAP	transporter associated with antigen processing
TCR	T-cell receptor
TDO	tryptophan 2,3-dioxygenase
TFA	trifluoroacetic acid
TGFβ	transforming growth factor β
THP1	AML cell line
TBS	Tris-buffered saline
Thr, T	threonine
TIC	total ion current

TIL	tumor-infiltrating lymphocytes
TNFα	tumor necrosis factor α
T_{reg}	regulatory T-cells
Tris	tris(hydroxymethyl)aminomethane
Trp, W	tryptophan
Tyr, Y	tyrosine
U	units of activity
UV	ultraviolet
V	volt
Val, V	valine
Vaso/VIP	vasoactive intestinal peptide (HSDAVFTDNYTR)
w/v	weight to volume
w/w	weight to weight
WGA	wheat germ agglutinin
X	any amino acid
XIC	extracted ion chromatogram
z	charge

Dissertation Abstract

This dissertation presents three projects, each focusing on the identification of peptides presented to the immune system through the major histocompatibility complex class I pathway. This pathway is responsible for digesting intracellular proteins into peptides and presenting these peptides to circulating CD8+ T cells, providing all nucleated cells with a mechanism to communicate their health to the immune system. Often, in transformed or diseased cells, dysregulated cellular signaling leads to altered protein expression and/or upregulation in post-translational modifications. In particular, increased phosphorylation is a hallmark of cancer and translates to an increase in phosphopeptide presentation via the MHC class I processing pathway. In many cases, we have observed the same phosphorylated peptides presented across different cancerous tissues. Thus, the first project explores the reason for this, and presents early evidence that the same phosphorylated peptides may be presented by other chronic diseases. This information would allow for immunotherapeutics targeting phosphopeptide antigens to be used for a variety of different diseases.

In order to detect phosphopeptides by mass spectrometry, it is necessary to perform enrichment for phosphorylation prior to MS analysis. Unfortunately, the enrichment procedure used in the Hunt laboratory is extremely laborious and commonly involves issues that prevent the final analysis of phosphopeptides. In the second project presented here, we significantly improved our phosphopeptide enrichment protocol. These improvements maintain high sensitivity and specificity while reducing the length and difficulty of the process. Another issue regarding phosphopeptide detection and enrichment

is the losses of hydrophilic peptides during the steps used to remove contaminants from the samples prior to analysis. Thus, the third project explores a hydrophilic interaction-based protocol that I developed to remove contaminants from our samples has since been modified and adopted to a variety of other applications. Prior to the improvements described within, the most phosphopeptides identified in a single sample was 234 using two enrichments and 1 billion cells. After these advances, we identified 522 phosphorylated peptides using 100 million cells, thus demonstrating the utility of the protocol improvements. Together, this work presents significant advances in immunoproteomics, particularly phosphopeptide analysis.

Chapter 1. Introduction to the dissertation

1.1 Introduction to major histocompatibility complex class I

The major histocompatibility (MHC) class I processing pathway is used by all nucleated cells as a way to express health status to the immune system (**Figure 1.1**).(1) First, cellular proteins are marked for degradation by ubiquitin. These proteins are then unfolded and fed through proteasomes, where they are digested into short fragments. These fragments may be recycled into the cytosol or transferred to transporters associated with antigen processing (TAPs). TAPs funnel peptides into the endoplasmic reticulum (ER), where the peptides are further shortened to 8-12 amino acid peptides. Peptide binding stabilizes the MHC molecule, allowing for the MHC-peptide complex to be transported via the Golgi to the cell surface, where the peptides are displayed to circulating immune cells.

The human version of MHC class I is called the human leukocyte antigen (HLA) class I.(2) The classic HLA alleles (HLA-A, B, and C) are all encoded by a highly polymorphic genomic region on chromosome 6. An individual can express up to 6 different allotypes (two each of A,

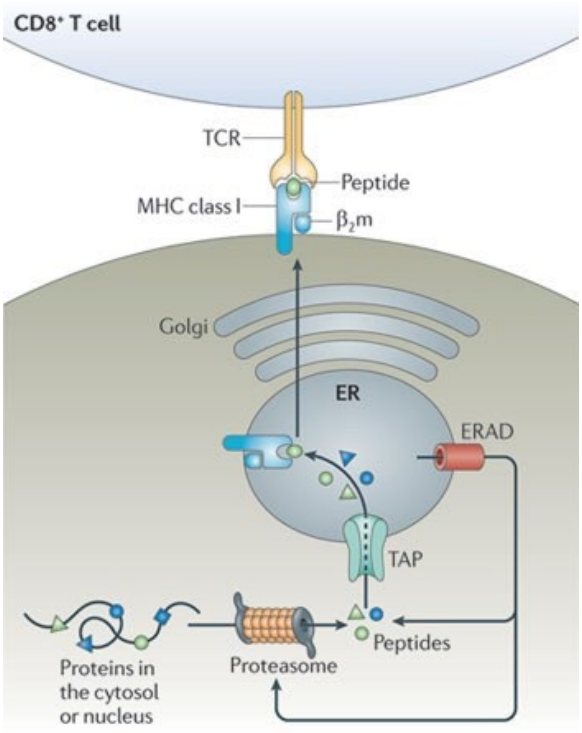


Figure 1.1 MHC Class I presentation pathway
In all nucleated cells, cytosolic and nuclear proteins are degraded by proteasomes, transported to the ER to bind MHC molecules, and then expressed on the cell surface for recognition by CD8+ T-cells.

B, C), as one of each type is inherited from each parent. Class I molecules are heterodimers consisting of a transmembrane heavy α-chain associated with a smaller β-2-microglobulin protein (β2m). Variability between the different alleles occurs within the peptide binding groove of the α-chain, which results in diverse peptide binding motifs. The HLA molecule relies on interactions near the beginning and end of the peptide that anchor the peptide in the binding groove. At the beginning of the peptide, the second or third residue is the most common anchor site. The C-terminal anchor is usually the final residue of the peptide. The center residues of the HLA-associated peptide are solvent exposed and capable of forming interactions with circulating CD8+ cytotoxic T-cells (CTLs).

CTLs are trained to differentiate antigenic peptides from healthy cellular expression through thymic education. Briefly, progenitor T-cells, or thymocytes, are transported to the thymus.(2) Upon arrival in the thymus, they undergo a differentiation phase, wherein variable regions of genes encoding for T-cell receptor proteins are rearranged to produce different binding affinities.(3) After differentiation, they undergo a phase of intense proliferation. A period of positive selection follows, wherein only thymocytes that bind weakly to complexes of self-peptides and self-MHC are allowed to progress. This ensures that the receptors are able to recognize and associate with MHC complexes.

Table 1.1 Common HLA types and anchor residues
Most common HLA types in the United States, as well as their prevalence when adjusted for the world population. Peptides bind at position 2 or 3 and the final residue of the peptide

			Position		
Allele	**USA**	**World**	**2**	**3**	**9**
A*01:01	20%	10%		DE	Y
A*02:01	39%	22%	LM		VL
A*03:01	20%	8%	LVM		KYF
A*11:01	14%	27%	(VIFY)	(MLFYIA)	KR
A*24:02	14%	25%	YF		ILF
B*07:02	16%	5%	P		L
B*08:01	14%	3%		K	L
B*35:01	11%	8%	P		YFMLI
B*40:01	9%	9%	E		L
B*44:02	14%	<1%	E		FY
C*03:04	16%	15%	A		LM
C*04:01	23%	15%	(YFW)	D	FM
C*05:01	13%	1%		D	MFLIV
C*06:02	16%	9%	R		VILM
C*07:01	24%	6%	R		FYLM
C*07:02	22%	25%	RY		FYML

During this process, the thymocytes will mature and begin to express higher levels of TCR on the cell surface. At this point, thymocytes that bind too strongly to the self-peptide self-MHC complexes will be removed by negative selection. This prevents an autoimmune response to self-peptides. The remaining thymocytes express Approximately 98% of thymocytes are removed during thymic education, while the remaining 2-4% (about one-two million daily) are released to the peripheral immune system as naïve T-cells.(2)

Naïve T-cells then circulate through lymphoid organs and blood via the lymphatic system.(4) Here, they are allowed to test their binding affinity with a variety of MHC-associated peptides. This primarily occurs in the lymph nodes with antigen presenting cells (APCs). When the naïve CD8+ T-cell associates strongly with an antigen presented on an APC, the APC releases costimulatory cytokines that will prompt its conversion to mature effector T-cell and begin rapid proliferation. These effector T-cells then begin to migrate through the body in search of antigen positive cells. Upon encountering the antigen, the CTL will release lytic vesicles containing perforin and granzyme into the synaptic region, causing lysis of the target cell.(5)

After the primary response, 90-95% of the antigen-specific CTLs die off.(5) The remaining antigen-specific CTLs are differentiated into memory CD8+ T-cells that retain the ability to respond to the antigen in case of reinfection. These memory cells are able to generate a rapid response upon reinfection, as they are able to migrate more widely than naïve T-cells,(6) are less dependent on costimulatory molecules for activation,(5) and can generate a response with a much lower quantity of expressed antigen.(7)

Despite the limitations placed on peptide binding by the allelic motifs, each class I molecule can bind tens of thousands of different peptide sequences.(8) Since each cell can

present anywhere from thousands to hundreds of thousands of HLA molecules, each cell can display an extremely heterogeneous population of peptides on its surface. The expression level of each peptide on the cell surface can vary dramatically. Only a single copy of a peptide may be present, while others may have hundreds or thousands of copies on the cell surface.

Most diseases that remain undetected by the immune system only generate minor changes in MHC peptide expression. Therefore, especially in immune suppressive environments, it may not be possible to generate enough of an antigen for the body to make the transition from a naïve T-cell to an effector T-cell. However, if the body can be primed to generate an immune response to a peptide that is expressed in diseased cells, then it is more likely for the immune system to be able to recognize and eliminate the cells itself. These immunotherapeutics have two main branches: antigen-based vaccines and adoptive T-cell therapy (ACT).(9) Antigen based vaccines are designed to prime the immune system by providing APCs with additional antigen for presentation, making it more likely that a naïve T-cell will be activated and generate an immune response. ACT produces antigen-specific T-cells in vitro before transferring them to the patient. Either method allows for the immune system to perform its own treatment, which should not produce the severe side effects of chemotherapeutics. A variety of these immunotherapeutics are in clinical trials and have had promising early results.(9-11)

The field of immunopeptidomics is developed around identification of MHC-associated peptides for development of such immunotherapeutics. Most of these efforts focus on identification of mutated neoantigens caused by increased mutagenesis in

cancerous cells or identification of peptides from pathogen proteins.(12-14) However, this limits the applications to the relevant disease, or sometimes to individual patients.(9, 15)

The ability of diseases to successfully evade the immune system is dependent on dysregulated cell signaling, which is very tightly controlled in normal cells and relies on post-translational modifications (PTMs) to mediate transcription and translation of cell cycle proteins.(16, 17) In diseased cells, various PTMs are upregulated, occurring more often and/or on amino acids they may not normally modify. These modified peptides can be presented on the cell surface, resulting in an abnormal repertoire of modified peptides on the surface of the tumor cell. Further, proteins containing PTMs may undergo antigen processing differently, so that the resulting HLA- peptide may be different from that tolerized during thymic education. However, peptides with PTMs are estimated to be present at approximately 0.1-1% of the total HLA-associated peptide population.

The overarching purpose of this work is to identify disease-associated post-translationally modified peptides for use in immunotherapeutics. This idea is corroborated by the fact that many post-translationally modified peptides have demonstrated an ability to stimulate a T-cell response that is specific for the modification.(18, 19) Without central tolerance, the immune system should mount a substantial response against the modified peptides. Together, this leads us to believe that disease-associated modified peptides are excellent candidates for immunotherapy.

1.2 Introduction to separation methods

1.2.1 Immunoprecipitation for HLA-associated peptides

To identify HLA-associated peptides, cells from cell lines or tissues must be harvested and lysed. Then, HLA class I molecules are immunoaffinity-purified from cell lines or tissues and their associated peptides are extracted. First, the pan-HLA antibody W6/32 is covalently linked to N-hydroxy-succinimide (NHS) derivatized agarose beads. Membranes are dissolved by rotation in a non-denaturing detergent, allowing for MHC molecules to be extracted from the cell membrane in conditions that allow peptides to remain within the HLA binding groove. After immunoprecipitation, the beads are rinsed, then eluted using 10% acetic acid. This disrupts non-covalent interactions, eluting the protein from the antibody and the peptide from the binding pocket.

1.2.2 Liquid chromatography

Samples are loaded onto a capillary precolumn (360-µm outer diameter, 75-µm inner diameter) packed with 8 cm of 10µm C18 material using one of two methods.

The first is by placing the sample into a pressure vessel and the precolumn into the sample volume. The sample is then pushed directly into the precolumn by using helium to increase the pressure inside of the vessel. The precolumn is then washed on the HPLC before connection to an analytical capillary column (360-µm outer diameter, 75-µm inner diameter) packed with 3 µm C18 material. An emitter tip is pulled on the analytical column using a laser puller to produce an opening of 2-5 µm. The solvent line which introduces the mobile phase to the HPLC column is split to reduce flow rate from 200 $\mu Lmin^{-1}$ at the chromatograph to 100 $nLmin^{-1}$ flow through the column. A 2 kV voltage is applied to the waste line, which provides the electrospray emitter tip with an electric potential. Since most

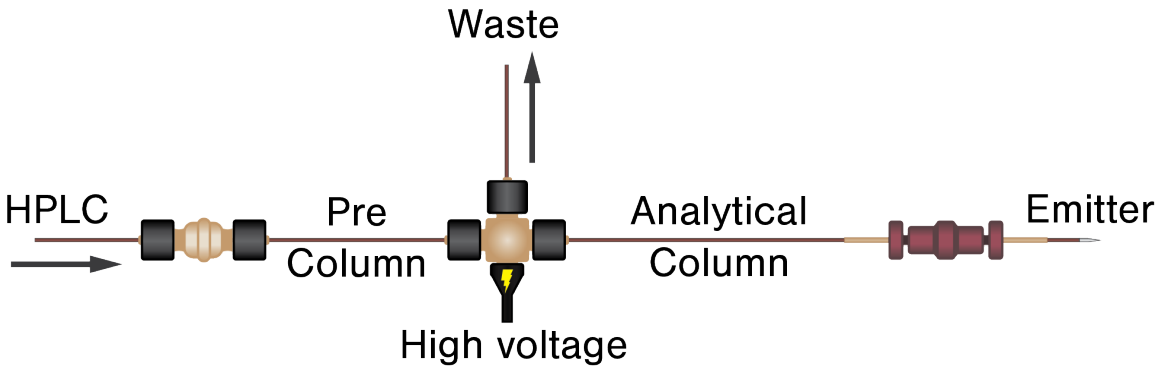


Figure 1.2 Schematic for in-line separation with Easy nLC 1200
For UHPLC separations the flow is not split and the voltage is applied in-line with the emitter tip. The sample is also loaded using an autosampler rather than directly onto the column. This provides a more convenient but less sensitive option for sample processing but will have problems with losses for low-level samples.

flow travels away from the instrument, any artifacts caused by applying voltage to the solvents travel to waste rather than into the instrument. When using this method, 100 mM acetic acid is used as the mobile phase additive.

The second is through use of an autosampler attached with a different chromatography setup (**Figure 1.2**). Here, the sample is placed into an injection vial, sampled by an injector needle, and injected onto the precolumn. The waste line is open during sample loading, so the precolumn is loaded and rinsed similarly to above. The waste line then closes and the flow proceeds through the analytical column (15 cm, 1.9 µm diameter C18 material) at a rate of 200 $nLmin^{-1}$. Here the voltage is applied at the tee connection between the precolumn and analytical column. However, since the voltage is not on the waste line it more commonly produces artifacts that are visible in the MS1 spectra. Due to this we instead must use 0.1% formic acid as the mobile phase additive. The change in additive drops signal by 2-5x, but application of the voltage in-line with the column results in large background ions to be generated from acetic acid conjugates, while formic acid conjugates are below the scan range. Despite its lower sensitivity, this setup is

useful as it provides a more convenient way to run large numbers of samples, since the samples are injected and data are collected without user oversight.

After loading onto the precolumn, peptides are separated by reverse-phase HPLC, which retains based on hydrophobic groups. Hydrophobic peptides interact strongly with the stationary phase and are retained longer than hydrophilic peptides over the course of a reversed phase gradient. The precolumn is then washed with aqueous solvent to remove any salts and firmly load the sample before connecting the precolumn to a C18 (1.9 or 3 μm diameter material) analytical capillary column (360-μm outer diameter, 75-μm inner diameter) with a laser-pulled glass emitter. Peptides are gradient eluted by an increasing concentration of organic solvent from 0 to 40% acetonitrile for 60- to 120-minutes.

When relative quantitation was desired, commercially available peptides (angiotensin I and vasoactive intestinal peptide) are added at known concentrations as internal standards. The peak areas of these standard peptides are compared with peptide peak areas to obtain a relative measure of the amount of peptide present.

1.3 Introduction to mass spectrometry

1.3.1 Electrospray ionization

During electrospray ionization, peptides are converted from liquid phase to the gas phase for ensuing detection by the mass spectrometer. HPLC columns used in the experiments presented in this dissertation were prepared from fused silica with laser pulled tips that are a few μm in diameter, through which peptides elute in charged droplets.(20) There is a ~2-3 kV potential between the inlet of the mass spectrometer and the column, causing the surface of the droplet to be positively charged. As the droplet traverses the length of the heated capillary, the solvent evaporates, shrinking the droplet and increasing the density of the charge on the surface Although the mechanism of complete ionization is disputed, there are two main theories of how the ions then convert to the gas phase (**Figure 1.3**). The most common theory for peptides is the charge residue model (CRM), which

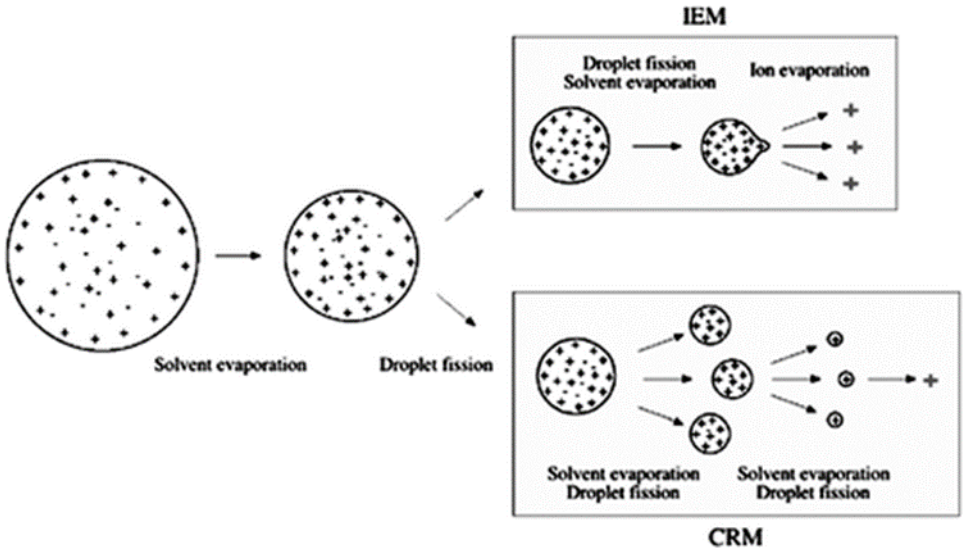


Figure 1.3 Proposed mechanisms for electrospray ionization
Two common theories for how analytes desolvate during electrospray ionization. The charged residue model (bottom) proposes a series of coulombic fissions as solvent evaporates and charge density becomes greater than the Rayleigh limit, eventually becoming individual charged ions in droplets that are then desolvated. The ion evaporation model proposes that the gaseous ions escape from the surface of the droplet when the charge density is high. Reproduced from Rohner et al. (2004) with permission from the PCCP Owner Societies.

suggests that the droplets begin to shrink due to solvent evaporation until the charge density becomes too high.(21) When the repulsion between charges exceeds the surface tension the droplet undergoes a Coulombic explosion into smaller droplets. This continues until a single molecule occupies the droplet. Then the molecule is desolvated, becoming gaseous while retaining the positive charge(s) imparted upon it. The other mechanism is the ion evaporation model (IEM), which posits that as the charge density at the surface increases, charged molecules are desorbed from the droplet surface.(22) In reality, some combination of the two models is most likely occurring, with the balance determined by the solvent system and analyte.

1.3.2 Mass analyzers

1.3.2.1 Quadrupole mass filter

The simplest type of mass analyzer is the quadrupole mass filter (QMF), which consists of four round or hyperbolic rods set parallel to each other.(23) A radio frequency (RF) voltage is applied to the rods, with adjacent rods 180° out of phase. This results in adjacent rods taking opposite charges. Additionally, a positive DC voltage is applied to one pair of rods and an equal negative voltage applied to the other set. An ionized analyte in gas phase will be attracted to the rods with the opposite charge and repelled by the ones with the same charge. When the rods alternate charges, the ion will change direction to approach the new set of rods.

The acceleration of this direction change is dependent on the molecule's mass to charge ratio (*m/z*). Molecules with larger *m/z* values will accelerate more slowly, causing them to travel in a cloud closer to the center of the quadrupole, while molecules with small *m/z* values accelerate more quickly and will travel closer to the outside of the quadrupole.

As the RF voltage is increased or its frequency is decreased, smaller ions will begin to travel in a large enough orbit to collide with the rods, neutralizing them and preventing transmission. The second (DC) voltage can be instead used to prevent transmission of larger *m/z* values. The larger *m/z* values are minimally affected by the RF voltage, but the application of a continuous DC voltage causes them to continually drift towards the oppositely charged rods. If they are insufficiently defocused by the RF voltage, they will collide with the rods, neutralizing them and preventing transmission.

Therefore, these voltages can be manipulated to prevent transmission of ions above and below desired *m/z* values. This is sometimes used to obtain a wide mass range, such as for an MS1 scan (removing ions below 300 *m/z* and above 1500 *m/z*). It may also be used to obtain a small *m/z* window, such as isolating a precursor for fragmentation (±1 Da).

1.3.2.2 Linear ion trap

Linear ion traps work with many of the same principles as quadrupoles, but they contain three segments, each composed of four hyperbolic rods (**Figure 1.4**).(24-26) Upon transmission of the ions into the center segment, the front and back segments are set to the opposite charge of the analyte, trapping the analyte. The trap can then be used to fragment ions and/or as a mass analyzer.

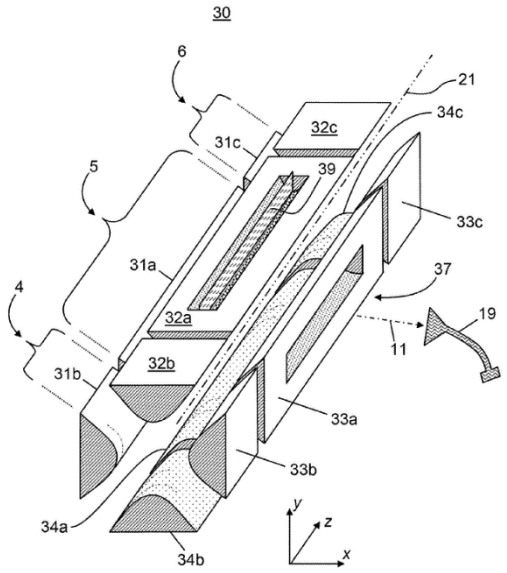


Figure 1.4 Diagram of a 2D linear ion trap
Linear ion traps consist of three sequential quadrupole mass filters. Two opposing rods have longitudinal slits to allow for ejection of ions for detection using electron multipliers. Image reproduced from patent #US9117646B2.

For use as a mass analyzer, ions are ejected through slits in two opposing rods by scanning the RF voltage. This causes ions to be ejected from the trap based on their *m/z* value, with the smallest *m/z* ions being ejected first. Instead of hitting the rods and becoming neutralized, they are ejected through a slit and attracted towards a conversion dynode held at a high voltage with the opposite polarity from the ion. When the ion strikes the conversion dynode, it causes the dynode to release oppositely charged particles and/or electrons. These particles are then further amplified through a channel electron multiplier, with more electrons emerging from each impact. After amplification, the accumulated charge is measured as a voltage pulse. This measurement corresponds to the abundance of the ions that left the trap, while the time that they arrive corresponds to the *m/z* value.

This method of detection is low-resolution, approximately unit resolution with a rapid scan rate. However, this method uses direct detection and generates spectra without significant noise, which provides significantly higher sensitivity than higher resolution methods using indirect detection. Fragmentation spectra from peptides present at the attomole level can provide enough information for a confident identification.

1.3.2.3 Orbitrap

In Orbitrap mass analysis, peptide ions travel to the C-trap for storage prior to injection into the Orbitrap mass analyzer (**Figure 1.5**). Storage of ions in the C-trap ensures that the entering packet of ions is compact, which is necessary for their

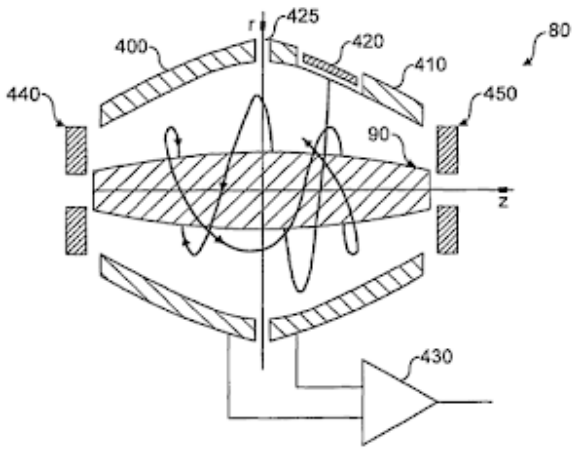


Figure 1.5 Diagram of Orbitrap mass analyzer
Orbitrap mass analyzers generate high resolution mass spectra through detection and Fourier transform of an image current between two outer electrodes. Image reproduced from patent #US7714283.

stability and coherency. Ion packets are radially ejected from the C-trap into the Orbitrap where they become trapped in an electrostatic field traveling around a central electrode. The ions begin to orbit axially and distribute into oscillating rings, depending on their *m/z* ratio. This movement induces an image current between the ions and the two outer electrodes that can be measured and fast Fourier transformed into a high-resolution mass spectrum.

This method of detection provides significantly higher resolution and accuracy, but the sensitivity is much lower than ion trap detection in addition to requiring a larger number of ions. Though confident identifications can be made using fragmentation spectra from peptides present at the 100 attomole level, this is mostly due to the increased confidence that the high resolution provides as it will usually provide lower sequence coverage. Therefore, obtaining high resolution fragmentation spectra is primarily useful for reducing the need for sequence verification in complex samples with plenty of sample material rather than for detection of low-level analytes.

1.3.2.4 Ion cyclotron resonance

Fourier transform ion cyclotron resonance (FT-ICR) mass analysis involves trapping ions in a magnetic field, in which their trajectories are inherently curved. Ions will travel in a circular path with a frequency dependent on their specific *m/z* ratio. To accomplish this, a uniform, electric field oscillating at the frequency of a given ion is applied, causing all ions with a given *m/z* ratio to be accelerated to the same radius, forming a spatially coherent packet. The packet can then induce an image current with a pair of electrodes, which is subsequently Fourier transformed into a high-resolution mass spectrum.

1.3.3 Fragmentation

1.3.3.1 Collisional dissociation

Collisional dissociation is enacted through collision of the peptide with inert gas molecules. The kinetic energy from the peptide's movement is converted into energy to break the peptide. Bonds that require the smallest amount of energy to break tend to fragment preferentially. For backbone cleavages, collision-based methods primarily fragment the carboxyl-amino bond. When working with labile PTMs, the weakest bond is usually the modification loss, so collision induced dissociation is primarily useful for sequence identification rather than modification localization.

1.3.3.1.1 Collision-activated dissociation

In collision-activated fragmentation (CAD, trap-type CID), the precursor is collected in a linear ion trap. The frequency of the quadrupole RF is adjusted to the secular frequency of a precursor *m/z*, causing the precursor to accelerate and undergo hundreds of low-energy collisions with helium atoms. For each collision, the kinetic energy is converted to translational energy to the peptide in a stepwise fashion, allowing protons to travel along the peptide backbone.(27) The translational energy is converted to vibrational

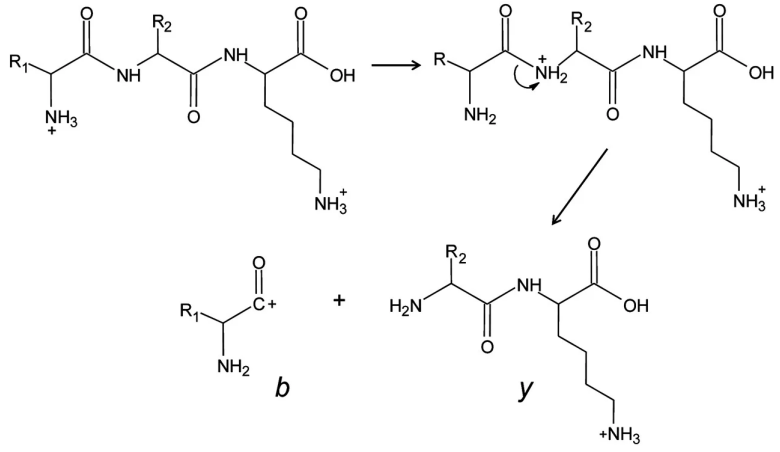


Figure 1.6 Proposed mechanism for collisional dissociation

energy, which is dispersed throughout all covalent bonds. When the vibrational energy exceeds the activation energy necessary to break the weakest bond, fragmentation occurs (**Figure 1.6)**. After fragmentation, the fragment ions no longer have the same secular frequency and therefore not accelerated, are collisionally cooled by the helium and cease fragmentation.

A peptide sequence can be determined by assignment of complementary b- (N-terminal) and y- (C-terminal) type ions. CAD works best for short (<15 residues), low charge (z<+3) peptides. Because CAD is dependent on random backbone protonation, longer peptides with higher charge states tend to display suboptimal fragmentation since areas of high charge limit proton transfer along the backbone.

Another downside of CID is that adjusting the RF frequency to match the precursor's secular frequency causes ions below 28% of the precursor *m/z* value to collide with the rods of the trap. MHC I peptides with 2-3 charges usually have *m/z* values between 300 and 600, so the majority of unstable ions will be < 150 Da and not much sequence information is lost. However, when investigating singly charged MHC I peptides with *m/z* values from 800-1300 Da, significantly more information is lost. Therefore, a separate collisional fragmentation is used for singly charged peptides.

1.3.3.1.2 Higher energy collisional dissociation

During higher energy collisional dissociation, peptides are accelerated through an ion routing multipole (IRM) filled with nitrogen. Since nitrogen is larger than helium, the translational energy imparted upon impact is much larger. This means that fragments can be generated after few or single collisions. Additionally, not just the precursor is accelerated through the cell, so fragments will continue to fragment. Since b-ions are

fragile, sequence informative ions are primarily y-type.(28) In addition to b- and y-type ions, internal fragments, immonium ions, and a-type ions are more common in HCD spectra. HCD usually provides more sequence information for charge dense peptides and peptides with high precursor *m/z* than CAD. However, due to the large number of different fragment types, it requires the higher certainty that high-resolution mass spectra provide. Therefore, it is suboptimal for identification of extremely low-level analytes.

1.3.3.2 Electron transfer dissociation

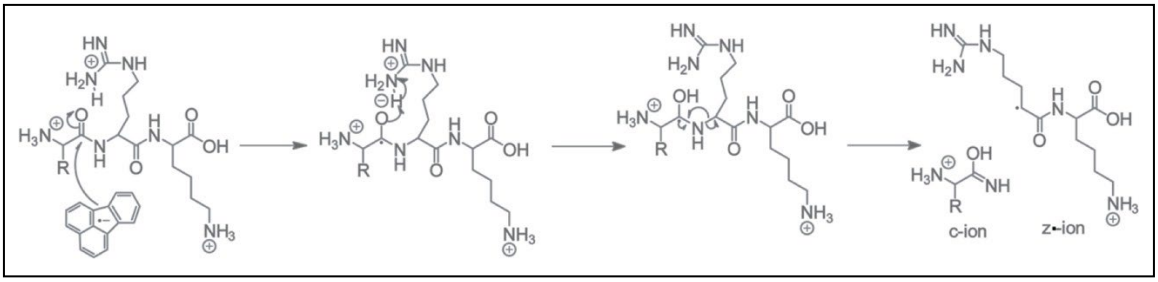


Figure 1.7 Electron transfer dissociation reaction mechanism
Mechanism for peptide reaction with a radical anion causing cleavage along the N-Cα bond.

Electron transfer dissociation (ETD), developed in the Hunt Laboratory, fragments peptides by transferring an electron from a radical anion to a protonated peptide.(29) Capture of this thermal electron (<0.2 eV) is exothermic by ~6 eV and causes fragmentation of the peptide backbone at the N-Cα bond. This reaction creates fragments analogous to b/y type ions, but instead named c- (N-terminal) and z•- (C-terminal) type ions, which then allow assignment of peptide sequences (**Figure 1.7**). The ETD mechanism causes more backbone cleavages than CAD, resulting in higher sequence coverage of longer, highly charged peptides. Also, since ETD maintains labile bonds, PTMs are preserved, and site-localization of the modified residue can be confidently assigned.

1.3.4 Data processing

In manual interpretation, we can use information from different fragmentation types, charge states, averaging only MS2s with high signal to noise, spectral matching from other files, and various other tricks to obtain a confident assignment. However, generating identifications using only manual identification has become incredibly difficult as instrumentation improves, since more MS2 spectra are collected from each sample. Orbitrap instruments released a decade ago collect around 10,000 scans in an hour long gradient, while modern Tribrid instruments can collect 30,000 or more scans in that time. When analyzing MHC phosphopeptides, we may be able to identify anywhere from tens to hundreds of peptides in a sample depending on the size, purity, enrichment yield, and MHC expression on the cells. The median number of MHC phosphopeptides identified using an IMAC enrichment is 25. However, tens of thousands of fragmentation spectra are collected in each file and there is no simple way to eliminate noisy or irrelevant spectra. Therefore, we use database search algorithms to identify potential spectra for analysis.

Database search algorithms compare the data acquired by the mass spectrometer to list of peptide sequences generated based on user input. The user chooses a list of protein sequences to search, how those proteins can be processed into peptides (digest specificity), and what modifications are allowed. Based on these parameters, the algorithm will generate a list of potential peptides to compare against the acquired data. Spectra are then compared against the expected fragmentation spectra of the generated peptide list, and an identification is based on how well the spectrum matches the expected spectrum. However, this is a complicated process and based on the way that different algorithms generate

expected spectra, what qualities they consider most important, and how they filter the matches, different algorithms can produce extremely varied results.

In general, search algorithms are developed and tested using tryptic digests, high-resolution fragmentation spectra, and predominantly unmodified peptides. While this may cover the majority of proteomics users, the results tend to differ widely for other applications. This is likely because some algorithms were written without certain considerations. For example, while ETD can produce c, $z^{\bullet}$, $a^{\bullet}$, and y type ions, it produces predominantly c and $z^{\bullet}$ type ions. If an algorithm does not consider both the expected ion types and their behavior, it can make an identification based solely on supposed a or y ions. Another example is labile modifications like phosphorylation and glycosylation. If the algorithm attempts to identify the intact fragment mass rather than the neutral loss from the fragment it will not identify the correct masses.

These differences in algorithms have more of an effect when using a larger search space. That is, if the search algorithm is provided more potential peptide sequences though different digestion specificity, increased number of modifications, or wider search tolerances, algorithms are more likely to make incorrect identifications. Sometimes, an algorithm may match the sequence to an incorrect database sequence while another algorithm may make the correct identification. Otherwise, the peptide may happen to have enough in common with a peptide that does not fall within the search parameters (**Figure 1.8**).

Many MHC peptides do not have a basic residue at the C-terminus, as they are generated by nonspecific proteases in the proteasome. Peptides lacking this positive charge at the C-terminus generate drastically different spectra than tryptic peptides. For example, ETD usually generates ideal spectra for identification of MHC peptides with three charges. However, only $z^{\bullet}$ ions that contain a protonated residue will be observed. Consequently, when algorithms develop their scoring metrics using tryptic digests, identification of $z^{\bullet}$ ions that do not contain a protonated residue would be considered beneficial, when it should instead detract from the confidence of the identification. Additionally, searching nonspecific termini and labile modifications against the human proteome creates a very large search space, allowing for much more variability in identifications between different

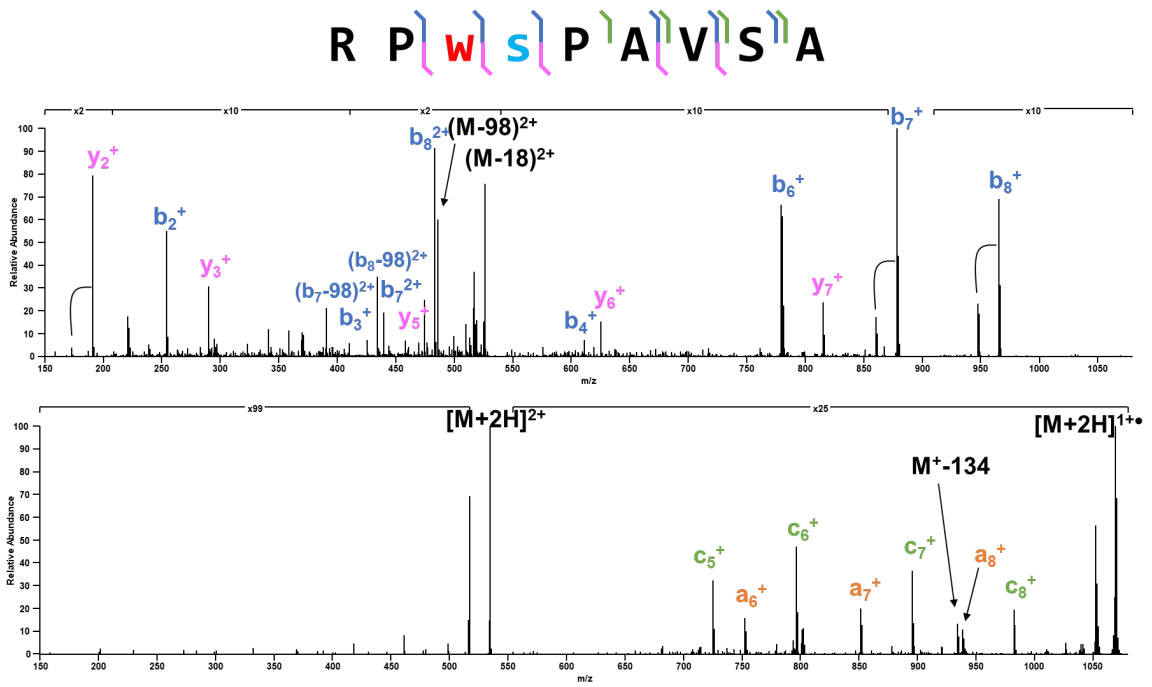


Figure 1.8 Correct assignment of a false positive search algorithm identification
Search algorithms can only search within the parameters provided, so any peptides with unexpected modifications or mutations can be incorrectly assigned. This was identified as TSNPVsLVSA, but is correctly assigned as RPwsPAVSA with a kynurenine modification to Trp. Though previously unidentified as a modification on MHC associated peptides, we have since identified 144 class I peptides and 268 class II peptides with kynurenine modifications.

database search algorithms based on how they perform scoring for peptides. To further complicate matters, we sometimes collect low-resolution MS2 spectra, which allows for identification of sub attomole levels of peptides but further increases the uncertainty of identifications. When collecting low-resolution spectra, it is more likely that a peak will be within the allowed tolerance window by chance, making it more likely that an identification will be made from incorrect assignments.

Due to the large search space, there is a high chance of identifying sequences incorrectly. However, there is also a high chance of sequences not being identified or being identified with very low confidence. A common way to "validate" search results in proteomics is by searching with two algorithms and reporting the overlapping identifications. However, depending on how the algorithm was developed, groups of algorithms may be more likely to have the same false positive identifications. In some cases, the same base code or concept will be used, but additional considerations (e.g. labile modifications) may be added. Alternately, new algorithms may be written by the same developer, and are likely to have the same style of peptide identification. Other times, the algorithm may be developed or trained by comparison against the results of an existing algorithm. This can lead to multiple algorithms making the same, incorrect identification.

We primarily use the search algorithm Byonic™ (Protein Metrics), which is a proprietary software that predicts fragment abundance based on sequence information and considers the predicted spectrum when making identifications. It was developed by Marshall Bern and first published in 2007 as a combination *de novo* and database search algorithm.(30) That is, it will consider fragmentation rules that are useful in *de novo* identification such as production of strong fragments N-terminal to proline or C-terminal

to acidic residues in CAD and HCD. Additionally, it uses stronger peptide identifications to support the assignment of weaker spectra, assuming that if a protein identification is strong based on some spectra, finding peptides from other parts of the protein is more likely. This allows for more complete protein sequence coverage, which is useful when working with protein digests of complex mixtures. It may also help to identify weaker spectra of modified peptides based on identification of their unmodified counterparts. However, this feature is less useful for MHC class I peptide identification since there are usually not many sequences from a given protein that are both phosphorylated and have strong binding affinity for a given HLA type, so it is uncommon for multiple peptides to be derived from the same protein.

After identifications are made by Byonic, they are manually validated or filtered for high confidence identifications. For unmodified analyses (**Chapter 4**), we collected most data in high resolution and the individual identities of the peptides was less important than the number of peptides present. Due to the number of identifications, confident identifications were determined by algorithm-defined score and probability score. For identification of modified peptides, both CAD and ETD spectra are collected for each peptide. To make a confident identification of a peptide that has not previously been observed as an MHC associated antigen, nearly full coverage of the sequence is required.

1.4 References

1. Neefjes J, Jongsma MLM, Paul P, & Bakke O (2011) Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature Reviews Immunology* 11(12):823-836.
2. Murphy K, Travers P, & Walport M (2008) *Janeway's immunobiology* (Garland Science, New York, NY) 7 Ed.
3. Kimura N, Toyonaga B, Yoshikai Y, Du RP, & Mak TW (1987) Sequences and repertoire of the human T cell receptor α and β chain variable region genes in thymocytes. *European journal of immunology* 17(3):375-383.
4. van den Broek T, Borghans JAM, & van Wijk F (2018) The full spectrum of human naive T cells. *Nature Reviews Immunology* 18(6):363-373.
5. Pennock ND, White JT, Cross EW, Cheney EE, Tamburini BA, & Kedl RM (2013) T cell responses: naive to memory and everything in between. *Adv Physiol Educ* 37(4):273-283.
6. Woodland DL & Kohlmeier JE (2009) Migration, maintenance and recall of memory T cells in peripheral tissues. *Nature Reviews Immunology* 9(3):153-161.
7. Grayson JM, Harrington LE, Lanier JG, Wherry EJ, & Ahmed R (2002) Differential sensitivity of naive and memory CD8+ T cells to apoptosis in vivo. *The Journal of Immunology* 169(7):3760-3770.
8. Marsh SGE, Parham P, & Barber LD (2000) *The HLA FactsBook* (Academic Press, San Diego, CA).

9. Jiang T, Shi T, Zhang H, Hu J, Song Y, Wei J, . . . Zhou C (2019) Tumor neoantigens: From basic research to clinical applications. *Journal of Hematology and Oncology* 12(1):1-13.

10. Keskin DB, Anandappa AJ, Sun J, Tirosh I, Mathewson ND, Li S, . . . Reardon DA (2019) Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. *Nature* 565(7738):234-239.

11. Engelhard VH, Obeng RC, Cummings KL, Petroni GR, Ambakhutwala AL, Chianese-Bullock KA, . . . Jr (2020) MHC-restricted phosphopeptide antigens: preclinical validation and first-in-humans clinical trial in participants with high-risk melanoma. *Journal for Immunotherapy of Cancer* 8(1):e000262.

12. Bassani-Sternberg M, Bräunlein E, Klar R, Engleitner T, Sinitcyn P, Audehm S, . . . Krackhardt AM (2016) Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nature Communications* 7(1):1-16.

13. Han XJ, Ma XL, Yang L, Wei YQ, Peng Y, & Wei XW (2020) Progress in Neoantigen Targeted Cancer Immunotherapies. *Frontiers in Cell and Developmental Biology* 8:728.

14. Lyngaa R, Pedersen NW, Schrama D, Thrue CA, Ibrani D, Met z, . . . Hadrup SR (2014) T-cell responses to oncogenic Merkel cell polyomavirus proteins distinguish patients with Merkel cell carcinoma from healthy donors. *Clinical Cancer Research* 20(7):1768-1778.

15. Ott PA, Hu-Lieskovan S, Chmielowski B, Govindan R, Naing A, Bhardwaj N, . . . Srinivasan L (2020) A Phase Ib Trial of Personalized Neoantigen Therapy Plus Anti-

PD-1 in Patients with Advanced Melanoma, Non-small Cell Lung Cancer, or Bladder Cancer. *Cell* 183(2):347-362.e324.

16. Ardito F, Giuliani M, Perrone D, Troiano G, & Muzio LL (2017) The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy. *International journal of molecular medicine* 40(2):271-280.
17. Sever R & Brugge JS (2015) Signal transduction in cancer. *Cold Spring Harbor: Perspectives in Medicine* 5(4):a006098.
18. Andersen MH, Bonfill JE, Neisig A, Arsequell G, Sondergaard I, Valencia G, . . . Haurum JS (1999) Phosphorylated peptides can be transported by TAP molecules, presented by class I MHC molecules, and recognized by phosphopeptide-specific CTL. *Journal of immunology* 163(7):3812-3818.
19. Mohammed F, Stones DH, Zarling AL, Willcox CR, Shabanowitz J, Cummings KL, . . . Willcox BE (2017) The antigenic identity of human class I MHC phosphopeptides is critically dependent upon phosphorylation status. *Oncotarget* 8(33):54160-54172.
20. Martin SE, Shabanowitz J, Hunt DF, & Marto JA (2000) Subfemtomole MS and MS/MS Peptide Sequence Analysis Using Nano-HPLC Micro-ESI Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Analytical Chemistry* 72(18):4266-4274.
21. Wilm M (2011) Principles of electrospray ionization. *Molecular & cellular proteomics* 10(7).
22. Rohner TC, Lion N, & Girault HH (2004) Electrochemical and theoretical aspects of electrospray ionisation. *Physical chemistry chemical physics* 6(12):3056-3068.

23. Miller PE & Denton MB (1986) The quadrupole mass filter: basic operating concepts. *Journal of chemical education* 63(7):617.

24. Schwartz JC, Senko MW, & Syka JEP (2002) A two-dimensional quadrupole ion trap mass spectrometer. *Journal of the American Society for Mass Spectrometry* 13(6):659-669.

25. Douglas DJ, Frank AJ, & Mao D (2005) Linear ion traps in mass spectrometry. *Mass Spectrometry Reviews* 24(1):1-29.

26. Schwartz JC (2015) Method and apparatus for a combined linear ion trap and quadrupole mass filter US9117646B2.

27. Wysocki VH, Tsaprailis G, Smith LL, & Breci LA (2000) Mobile and localized protons: a framework for understanding peptide dissociation. *Journal of Mass Spectrometry* 35(12):1399-1406.

28. Harrison AG (2009) To b or not to b: the ongoing saga of peptide b ions. *Mass spectrometry reviews* 28(4):640-654.

29. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, & Hunt DF (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* 101(26):9528-9533.

30. Bern M, Cai Y, & Goldberg D (2007) Lookup peaks: a hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry. *Analytical chemistry* 79(4):1393-1400.

Chapter 2. Identification of phosphorylated MHC Class I antigens caused by exogenous or disease-associated phosphatase inhibition

2.1 Introduction

Nearly 30 years have passed since mass spectrometry (MS) was first used to understand the MHC Class I presentation process. After almost a year of trial and error to generate useful samples, the first HLA-associated peptide sequences were finally characterized by tandem mass spectrometry in 1992.(1) Since then, both instrument sensitivity and sample preparation techniques have improved significantly, allowing for identification of far more candidate peptides while decreasing analysis and validation times. Despite these advances, the promise of using peptides presented by the body's own immune system as a widespread immunotherapy has encountered many challenges.

Sequence-based (mutated) neoantigens have emerged at the forefront of these efforts.(2-5) However, finding previously undiscovered mutations is difficult and generally requires prior knowledge from RNA exome sequencing.(4-6) The difference in MHC expression between healthy and diseased cells is small, and neoantigens are expressed at much lower levels than canonical peptides.(3-5, 7) While mutated and post-translationally modified antigens were discovered concurrently and have similar reactivity,(8) modified antigens are comparatively understudied.(9, 10) However, disease-associated post-translational modifications are caused by dysregulated signaling pathways that are common across different individuals, cancers, and even other diseases.(10-12) Many cancer treatments already take advantage of dysregulated signaling pathways, particularly

those associated with phosphorylation, to target cancerous cells over healthy ones, but immunopeptidomics remains focused largely on mutated antigens.(10, 13-15) With the increases in cellular phosphorylation caused by dysregulated signaling, we see a similar increase in presentation of phosphorylated MHC peptides. Targeting these modified peptides should allow for development of immunotherapeutics that apply to an entire HLA type instead of requiring personalized treatments.

The first phosphorylated MHC I peptide identified by mass spectrometry was published in 1998.(8) That peptide (RVAsPTSGV) has since been found in 22 samples across 9 types of cancer and two HLA types (A*02:01 and A*68:02). An elongated version, RVAsPTSGVK, has been found in 16 samples, 4 cancers, and two HLA types (A*03:01 and A*11:01).[1] Since 1998, our lab has identified more than 2500 potential phosphorylated targets across fifteen types of cancer; approximately 1000 of these have been found in multiple cancer types.[1](16) Some phosphorylated antigens have been found in up to 13 types of cancer and ~40 samples. Our current protocols for phosphopeptide enrichment and analysis can produce anywhere from tens to hundreds of MHC I phosphopeptide identifications using 100-500 mg of crude cancerous tissue.(17, 18) Studies looking for mutated antigens use similar amounts of tissue and generally identify under 10 mutated antigens, most of which are specific to the tissue and patient.(7, 19) Conversely, more than 50% of the phosphopeptide identifications for any given sample are already identified in another sample (depending on the prevalence of the HLA types).[1] Approximately 80% of phosphopeptides tested can generate central memory T-cells

[1] Hunt lab phosphopeptide list (2021) [Unpublished data]

responses from healthy donors.[1] This response indicates that their immune system considers these peptides to be antigenic, and that the response has overcome central tolerance.(20, 21) Therefore, peptides that generate a healthy donor response are promising targets for immunotherapy of multiple cancers, as well as any other disease that causes their expression.(20, 21)

Many class I MHC phosphopeptides are uniquely expressed on diseased cells and are potential targets for the immunotherapy of hepatocellular cancer,(22) breast cancer,(23) melanoma,(24, 25) colorectal cancer,(16) leukemias,(20) and other cancers.(26-28) Of particular note are results obtained in a recent preclinical trial that used two of our class I MHC phosphopeptides to treat high-risk melanomas.(24) T-cell responses to one peptide were observed for 5/12 patients. Another peptide generated T-cell responses for 2/12 patients. Adverse effects were minimal. Two additional clinical trials that use multiple class I MHC phosphopeptides to treat acute myeloid leukemia (AML) and colorectal cancer are in the planning stages.

Two important questions are posed by these findings: (a) why do dysregulated cell signaling pathways generate the same phosphopeptide antigens in multiple types of cancer and (b) why do healthy blood donors, with no sign of autoimmune disease, have central memory T-cells that recognize and kill cells that present the same class I MHC phosphopeptide antigens? Answers to both questions likely involve three major tumor suppressor proteins: protein phosphatase 2A (PP2A),(29-36) retinoblastoma protein (pRb),(37) and tumor protein 53 (p53).(38)

2.1.1 Oncogenic proteins and phosphorylation

Phosphopeptide antigens are derived from dysregulated cell signaling pathways common to many different cancers. In normal cells, phosphorylation is a brief event, cycling on and off a particular protein several times during the lifetime of that protein. However, dysregulation in cancerous cells extends the length of phosphorylation, allowing the proteasome sufficient time to degrade the phosphorylated protein forms and present phosphopeptide antigens on Class I MHC molecules. Since the same cell signaling pathways are also dysregulated during many viral and bacterial infections, many of the same phosphopeptide antigens should be presented. Therefore, these antigens could be employed for immunotherapy of, or vaccination against, both infectious agents and cancer.

Protein phosphatase 2A is the most abundant Serine/Threonine phosphatase in the cell (comprising ~0.1-1% of all cellular proteins) and plays a major role in the regulation of many important cell signaling pathways (e.g. Wnt, P13K/Akt, MAPK, and c-Myc) that control cell proliferation, transformation, and apoptosis.(29-36) PP2A exists as a heterotrimeric complex consisting of a scaffolding subunit (A), regulatory subunit (B), and a catalytic subunit (C). There are 2 isoforms for subunit A, 2 isoforms for subunit C, and 4 classes of subunit B, each of which has 2-5 isoforms and additional splice variants. The regulatory subunits allow for cellular localization and impart specificity toward different substrates.(35) As the most prolific and abundant phosphatase, altering the activity of PP2A should have the greatest effect on phosphorylation states.(32, 33)

A similar phosphatase, protein phosphatase 1 (PP1), is the next most prolific phosphatase. PP1 lacks a scaffolding subunit, but has 4 isoforms of the catalytic subunit and about 60 known regulatory subunits.(39, 40) The catalytic subunits of the two

phosphatases have a high degree of sequence homology, but have different substrate binding motifs.(41) Both are also reversibly phosphorylated near the C-terminus to control activation.(42, 43) Since the catalytic subunits are so similar, inhibitors that target the PP2A catalytic subunit or phosphate frequently have a significant effect on PP1 as well.(30, 43)

PP1 and PP2A together are responsible for the vast majority of dephosphorylation in the cell, so dysregulation of one or both of them would have a significant effect on cellular phosphorylation.(44, 45) Many phosphorylation sites may be dephosphorylated by multiple phosphatases, but a reduction in PP2A activity would cause a large reduction in overall phosphatase activity. Therefore, even if there are redundant phosphatases for a particular phosphorylation site, a large reduction in total available phosphatase should cause an extended phosphorylation event. While kinases are also dysregulated in cancer, each one is responsible for a far smaller subset of sites, meaning they individually have a much smaller effect on cellular phosphorylation. For example, all fifteen members of the protein kinase C family, the most prolific kinase family, are together predicted to be responsible for ~20% of phosphorylation sites.(46, 47) Based on their relative levels of control over phosphorylation, inhibition of PP2A and/or PP1 should increase cellular phosphorylation much more than overactivation of any kinase or kinase family.

PP2A inhibition is accepted as a prerequisite for cellular transformation and is likely responsible for many of the phosphorylated antigens that we have observed across cancers.(48-51) Accordingly, a number of cancer treatments are in development or released that directly target the phosphatase. PP2A can be directly inactivated by somatic mutation and phosphorylation or demethylation near the C-terminus of the catalytic subunit.(35, 42)

Somatic mutations of PP2A are found in about 8% of cancer cases, most commonly on the scaffolding subunit.(35) Mutation of the scaffolding subunit can cause an inability to bind regulatory subunits, thereby preventing the phosphatase from localizing or associating with its targets.(35) While these direct inhibitions are present in some cancers, PP2A is more commonly inactivated by upregulation of endogenous PP2A inhibitors: protein SET (also known as inhibitor 2 of PP2A) and cancerous inhibitor of phosphatase 2A (CIP2A). CIP2A, the more heavily studied of the two, is overexpressed in ~40-90% of patients across different cancer types.(52-54) Both CIP2A and SET appear to inhibit PP2A activity by binding to the catalytic subunit, preventing it from associating with the rest of the holoenzyme.(55-57) High CIP2A and/or SET levels have repeatedly been associated with worse prognosis across cancers.(58-62) Overexpression of these endogenous inhibitors is commonly caused by inhibition of another major tumor suppressor, pRb.(32, 55)

The retinoblastoma protein was proposed by Alfred Knudson in 1971 as a ubiquitous mutation in retinoblastoma patients.(63) It is one of the earliest discovered oncogenic proteins and its dysregulation has since been observed in multiple other cancers. pRb is a major regulator of the cell cycle and is mutated or functionally inactivated in most human cancers.(37, 64) The retinoblastoma protein represses gene transcription required for cell growth by directly binding to the transactivation domains of the E2 family transcription factors (E2F). pRb also represses transcription by binding to proteins that are involved in nucleosome remodeling, histone acetylation/deacetylation, and methylation. pRb becomes inactive when it is phosphorylated by cyclin dependent kinases. Hyperphosphorylated pRb dissociates from E2F and allows the cell to enter S-phase with high levels of transcription. Synthesis of endogenous PP2A inhibitors cancerous inhibitor

of CIP2A and SET is increased when pRb is dissociated from E2F. Since reactivation requires dephosphorylation by PP2A, reactivation becomes very difficult with even a small amount of dysregulation.

Tumor protein 53, perhaps the most heavily studied oncogenic protein, is a sequence-specific DNA-binding protein that regulates transcription and promotes cell cycle arrest and apoptosis.(38, 64) It can activate DNA repair proteins when DNA has sustained damage, arrest growth by holding the cell cycle at the G1/S regulation point to give the repair machinery time to fix DNA sequence errors, and initiate programmed cell death if the damage proves to be irreparable. p53 is activated by phosphorylation and degraded in the proteasome when it is ubiquitinated by the ubiquitin ligase, MDM2.(65) PP2A is responsible for a dephosphorylation at Thr55 that activates p53, and for a number of phosphorylation sites on MDM2 that allow for appropriate p53 regulation.(65-67) Mutation of p53 also occurs in most human cancers.(38)

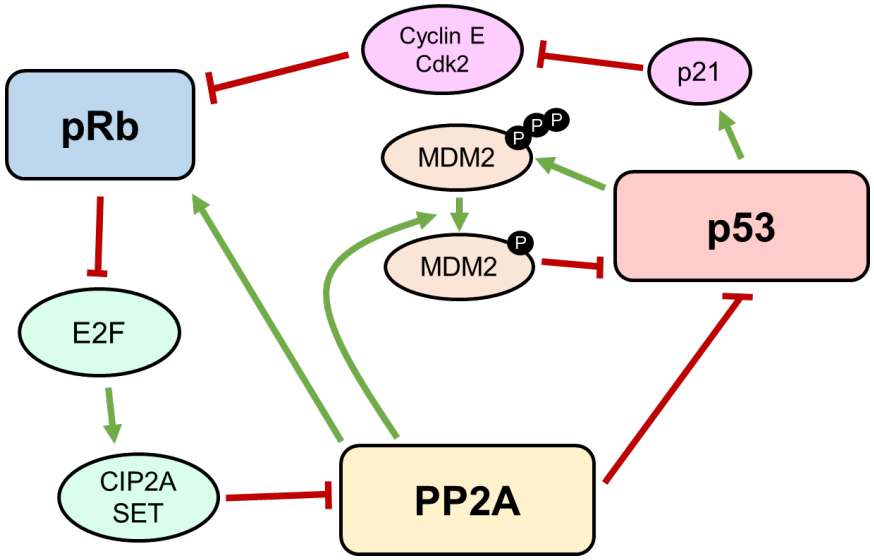


Figure 2.1 Signaling interplay between PP2A, pRb, and p53
Inhibition of any one of tumor suppressor proteins pRb, p53, or PP2A can cause inhibition of the other tumor suppressors, either directly or through dysregulation of connected pathways. This should ultimately result in PP2A inhibition and increased presentation of phosphorylated MHC I peptides.

PP2A is responsible for activating both pRb and p53 through dephosphorylation.(36) Thus, its inhibition causes a decrease in the essential functions of pRb and p53 in maintaining cell homeostasis.(64, 66) Inactivation of pRb increases transcription of CIP2A and SET, leading to PP2A inhibition. Inactivation of p53 prevents cell cycle arrest after DNA damage and lowers production of p21. The lack of p21 increases cyclin dependent kinase 2 (Cdk2) activity, resulting in pRb hyperphosphorylation and dissociation from E2F.(68) E2F is then active and able to increase transcription of CIP2A and SET. Inactivation of p53 also allows for mutations to continue unchecked, increasing the chance of mutating further tumor suppressor proteins.

Many cell functions have redundant pathways, allowing individual proteins to malfunction without causing overall cell malfunction. However, these three proteins have limited redundancy, so interruption of their function leads to cellular dysfunction and, in many cases, oncogenesis. Furthermore, interference with one of these proteins often negatively affects the other two, causing progressively greater changes to cellular functions that increase cell proliferation and immune evasion.

Due to their relationships, interference with any of these proteins should also cause reduced activity of the most prolific phosphatase. Disruption of PP2A causes an extended phosphorylation event, allowing enough time for expression of phosphorylated MHC peptides. Therefore, disruption that reduces PP2A activity allows us to find cancer-specific MHC phosphopeptides expressed across multiple cancer and/or tissue types. Based on the phosphorylation sites it is responsible for modifying and the ubiquity of its inhibition, PP2A (perhaps in combination with PP1) is the most likely cause of increased cellular phosphorylation as well as phospho-antigen presentation.

Since p53, pRb, and PP2A are important in pathways that are responsible for any disease's success (proliferation and avoidance of immune detection), they are also common targets for some of the most chronic and difficult to treat diseases, regardless of the pathogen.(36, 39, 64) If this reasoning is valid, any disease that targets any combination of p53, pRb, or PP2A has the potential to produce the same MHC-associated phosphopeptides as cancerous cells. In other words, an immunotherapy that uses MHC-associated phosphopeptides should work against diseases that target any of these three proteins. Since inhibition of these proteins should lead to extended protein phosphorylation through phosphatase inhibition, it should also lead to the presentation of phosphorylated MHC I peptides.

2.1.2 Inhibition of pRb, PP2A, and p53 in disease

2.1.2.1 Hepatitis viruses

Worldwide, 140 million people are infected with hepatitis C virus (HCV) and more than 250 million people with hepatitis B virus (HBV).(69) Both viruses can be prevented by vaccines but are difficult or impossible to treat once infected.(70) Over time, both viruses are oncogenic and can lead to development of hepatocellular carcinomas.

HCV consists of a single stranded RNA (9600 nucleotide bases) surrounded by a protected shell of proteins. The viral RNA codes for a single polyprotein (~3,000 AA) that is post-translationally cleaved into two highly glycosylated structural proteins, E1 and E2, a transmembrane protein p7, and six nonstructural, accessory proteins: NS2, NS3, NS4A, NS4B, NS5A and NS5B. HCV does not integrate its genome into the host chromosomal DNA.(71) It exhibits a high mutational rate and deregulates many cellular processes. One of the accessory proteins, NS5B, forms a complex with the retinoblastoma tumor

suppressor protein that is then targeted for degradation in the proteasome following ubiquitination by the E6-associated protein (E6AP).(72) Expression of another member of the pRb family, p130, is down-regulated by the HCV core protein that triggers hyper-methylation of the promoter region of the corresponding gene.(73) Another accessory protein, NS2, sequesters p53 to the cytoplasm and prevents it from monitoring DNA damage and triggering cell apoptosis.(74) The expected result is high levels of gene transcription, likely including production of CIP2A, and uncontrolled cell division.

The hepatitis B virus (HBV) is a partially double-stranded DNA virus that replicates via reverse transcription. The two DNA chains contain ~3200 and ~2300 nucleotides, respectively. The genome contains four overlapping reading frames that code for the viral coat protein (capsid), surface proteins (envelope), reverse transcriptase, and the small (17.4 kDa) regulatory oncoprotein, HBx.(75) In hepatocellular carcinoma (HCC) cases caused by HBV, the virus frequently (86.4%) is integrated into the host genome.(76) HBx activates the E2F1 group of transcription factors by upregulating kinases that phosphorylate and inactivate pRb.(77) HBx is also known to block apoptosis of HBV infected cells by several different mechanisms.(75)

2.1.2.2 Neurodegenerative disease

Abnormal hyperphosphorylation of protein Tau is the main hallmark of a number of neurodegenerative disorders called tauopathies. The most studied of these is Alzheimer's disease (AD), but they include a number of other dementias, and neurodegeneration following traumatic brain injury or encephalopathy.(78) In these diseases, Tau is 3-4 times more phosphorylated than it is in healthy cells. This hyperphosphorylation leads to microtubule disorganization, protein aggregation, and cell death.(79) PP2A is responsible

for 71% of phosphatase activity for protein Tau and has become a major topic of study in the field.(80) Inhibition of PP2A using okadaic acid has been shown to induce p-Tau formation as well as other characteristics of tauopathies including cognitive impairment, protein aggregation, and cell death.(79) Increased pRb phosphorylation and decreased PP2A methylation have been observed in AD, both of which should ultimately cause PP2A inhibition.(31) Increased levels of both CIP2A and SET colocalize with increased p-Tau levels in AD tissue.(81)

2.2 Materials

2.2.1 Reagents

Aldrich Chemical Co. (Milwaukee, WI)

- Ethylenediaminetetraacetic acid, 99.99% purity

Bio-Rad Labs (Hercules, CA)

- Tris base >99.8% pure

Fisher Chemical (Ward Hill, NJ)

- Methanol, Optima™ LC/MS grade

Honeywell (Charlotte, NC)

- Acetonitrile, B&J Brand™ LC-MS, for LC-MS and HPLC, >99.9%

Oxford Glyco Systems (Oxfordshire, England)

- Tris HCl, Electrophoresis grade

PQ Corporation (Malvern, PA)

- KaSil 1624 potassium silicate solution

Roche (Basel, Switzerland)

- cOmplete™, Mini Protease inhibitor tablet, EDTA-free

Sigma Aldrich (St Louis, MO)

- Acetic acid, glacial, ≥99.99% trace metals basis
- L-Ascorbic Acid, ACS reagent, ≥99%
- Azulene
- CHAPS, ≥98%
- Iron (III) chloride, anhydrous, powder, ≥99.99% trace metals basis

- Formamide, ≥99.5% (GC), BioReagent for molecular biology
- Hydrochloric Acid, ACS reagent, 37%
- Sodium Chloride, ACS reagent, ≥99.0%
- Sodium hydroxide, reagent grade, 97%, powder
- PhosSTOP™ phosphatase inhibitor tablets
- Triton X-100

Supelco (Bellefonte, PA)

- Hydrogen chloride solution, 3M in methanol, for GC derivatization, LiChropur™

Thermo Scientific (Waltham, MA)

- Pierce™ Formic Acid, LC-MS grade
- Halt™ Protease and phosphatase inhibitor cocktail (100x)
- Pierce™ Water, LC-MS grade

2.2.2 Proteins and Peptides

Anaspec (Fremont, CA)

- Angiotensin II Phosphate >95% HPLC
- Vasoactive Intestinal Peptide (VIP) 1-12 human, porcine, rat >95% HPLC

Atlantic Peptides (Concord, NH)

- Custom phosphopeptide standards
    - ST(pS)LVGGR
    - FLG(pT)PIAKV

Biolegend (San Diego, CA)

- Ultra-LEAF™ Purified anti-human HLA-A, B,C Antibody

Sigma Aldrich (St Louis, MO)

- Angiotensin I human acetate salt hydrate >90% HPLC

Thermo Scientific (Waltham, MA)

- Pierce™ Peptide Retention Time Calibration Mixture
- Pierce™ HeLa Protein Digest Standard
- Custom Phosphopeptide Standards
    - RVK(pS)PLFQF
    - RTH(pS)LLLLG

2.2.3 Chromatography Resins

Applied Biosystems TM (Carlsbad, CA)

- POROS® MC 20 metal chelating packing material, 20 µm diameter

Cytiva Life Sciences (Marlborough, MA)/GE Healthcare (Boston, MA)

- NHS-Activated Sepharose 4 Fast Flow

Dr Maisch GMBH (Ammerbuch, Germany)

- Reprosil Pur 120 C18 AQ 3 µm
- Reprosil Pur 120 C18 AQ 10 µm

QIAGEN (Hilden, Germany)

- Ni-NTA spin columns

2.2.4 Equipment and Instrumentation

Agilent Technologies (Palo Alto, CA)

- 1100 Agilent high performance liquid chromatograph

Boekel Scientific (Feasterville-Trevose, PA)

- Digital Incubator 133000

Branson (Danbury, CT)

- Branson 1200 Ultrasonic Bath

Cadence Science (Cranston, RI)

- Blunt end pipetting needle (16G)

Eppendorf (Enfield, CT/Hamburg, Germany)

- Model 5424 Centrifuge
- Protein LoBind Microcentrifuge tubes (1.5 mL)

Fisherbrand (Pittsburgh, PA)

- Model 50 Sonic Dismembrator with CL-18 Probe
- Mini Tube Rotator
- Digital Vortex
- Vortex Genie 2

GL Sciences (Tokyo, Japan)

- Centrifuge adaptor 5010-21514

Labconco Corp. (Kansas City, MO)

- Centrivap centrifugal vacuum concentrator 7810016

Millipore (Billerica, MA)

- Amicon ultra, 10 kDa regenerated cellulose spin filter
- Amicon ultra, 3kDa regenerated cellulose spin filter

New Objective (Littleton, MA)

- PicoClear 360 Union

PolyMicro Technologies, Inc. (Phoenix, AZ)

- 360 μm o.d. x 75 μm i.d. polyimide coated fused silica capillary
- 360 μm o.d. x 100 μm i.d. polyimide coated fused silica capillary
- 360 μm o.d. x 150 μm i.d. polyimide coated fused silica capillary

Sarstedt Inc (Newton, NC)

- 1.5 mL screw top microtubes

Supelco (Bellefonte, PA)

- Empore octadecyl C18 extraction disks
- Screw top clear glass vials with green thermoset caps and PTFE liners, 2 mL

Sutter Instrument Co. (Novato, CA)

- P-2000 microcapillary laser puller with fused silica adapter

Thermo-Fisher Scientific (San Jose, CA/Bremen, Germany)

- Orbitrap Fusion Tribrid mass spectrometer with commercial front-end ETD
- LTQ mass spectrometer with commercial back-end ETD
- LTQ FT-ICR hybrid mass spectrometer (custom modified with front-end ETD)
- LTQ-Orbitrap mass spectrometer (custom modified with front-end ETD)

Zeus Industrial Products, Inc. (Orangeburg, SC)

- Teflon tubing, 0.012-inch i.d. x 0.060-inch o.d.

2.3 Methods

2.3.1 PBMC collection, activation, and treatment with calyculin

Healthy donor peripheral blood mononuclear cells (PBMC) were collected from two donors (VA-AH and VA-DH). The cells were activated with anti-CD3 antibodies and interleukin 2 to stimulate cell proliferation. When the PBMC samples were large enough to analyze (>1 billion cells), half of each sample was treated with 10nM calyculin A for six hours. The samples were then lysed, immunopurified, and sent to our lab.

2.3.2 Collection of hepatocellular carcinoma margin tissue

Human liver tissue was obtained by Nico Buettner from patients undergoing liver transplantation at the Queen Elizabeth Hospital, Birmingham, (Local Research Ethics Committee Reference Number: 06/Q2702/61).

2.3.3 Immunopurification

MHC-bound peptides were isolated as described previously, described here in brief.(20) First, the pan-human class I antibody W6/32 was conjugated to NHS-Sepharose beads. For each gram of tissue (or 1E9 cells), 3 mg of antibody conjugated to 300 µL of beads were used. Beads were washed twice in PBS and incubated rotating overnight with antibody. The following day beads were pelleted, blocked for 1 hour with 100 mM Tris-HCl, then washed twice in alternating solutions of 100 mM ammonium acetate pH 5 with 500 mM NaCl and 100 mM Tris buffer pH 8. Beads were resuspended in 20 mM Tris, 150 mM NaCl up to 1 mg/ml remaining antibody and stored at 4°C until use.

Cells were lysed in a buffer of 20 mM Tris-HCl, 150 mM NaCl, 1% CHAPS, pH 8 supplemented with protease and phosphatase inhibitors. Buffer was added to the cells at 5

mL per $1x10^9$ cells or 1 gram of tissue. Frozen cells were resuspended in lysis buffer before rotating at 4°C for 1-2 hours. Lysed cells were ultracentrifuged at 36,800 rpm for 1 hour at 4°C. The supernatant was then mixed with the antibody-bead conjugates and rotated at 4°C overnight to allow MHC binding.

The following day, pelleted beads were washed in 10 mL lysis buffer, resuspended in ~500 µL 20 mM Tris-HCl and transferred to a microcentrifuge tube or polypropylene column. Beads are then washed with: 2x 20 mM Tris-HCL, 150 mM NaCl; 2x 20 mM Tris-HCl, 1 M NaCl; 3x 1 ml 20 mM Tris-HCl. Washed beads were then spun through a pre-wet 3 kDa Amicon Ultra centrifugal filter column and all liquid removed by centrifugation. Columns were covered with Parafilm and stored at -80°C for shipment.

2.3.4 Sample cleanup

Stop and go extraction (STAGE) tips were fabricated with two cores as previously described.(82) Samples larger than 3E8 ceq were processed on multiple tips to reduce clogging. STAGE tips were equilibrated using the following wash steps: two washes of 100 µL of methanol, one wash of 100 µL 80% acetonitrile/0.1% acetic acid at 1500 *x g*, and two washes of 100 µL 1% acetic acid at 3500 *x g*. Thawed beads were transferred from the filter to a low-protein binding tube using two transfers in 200 µL water. After centrifugation, the supernatant was removed and set aside.

To elute the peptides from HLA molecules bound to beads, 150 µL of 10% acetic acid was added to the tube and placed in a shaker for 5 minutes at room temperature. The beads were centrifuged at 300 *x g* for 1 minute and the supernatant transferred to a low protein binding tube. This process was repeated to ensure complete elution of peptides from HLA. Two internal phosphopeptide standards were added to the 10% acetic acid elution to

determine recovery from the cleanup process. The elution supernatant was loaded onto the STAGE tips in 150 µL aliquots, centrifuging at 3500 *x g* until the entire volume had passed through. STAGE tips were washed twice using 100 µL of 1% acetic acid. Peptides were eluted from the STAGE tips using a stepped acetonitrile gradient: 20 µL of 20% acetonitrile/0.1% acetic acid, 20 µL of 40% acetonitrile/0.1% acetic acid, and 20 µL of 60% acetonitrile/0.1% acetic acid. Eluted peptide fractions were dried to completion using a Centrivap, reconstituted in 0.1% acetic acid to a concentration of 1E7 cell equivalents per µL (or 1 mg/ µL for tissue), and stored at -35°C.

2.3.5 Sample screening

Before performing an enrichment, the unenriched sample was analyzed to verify successful immunoprecipitation and determine the level of peptidic content in the samples. A precolumn was attached to an HPLC and equilibrated with 100 mM acetic acid for 5-10 minutes at approximately 5 $\mu L min^{-1}$ (30 bar). One microliter of the reconstituted sample was added to 10 µL of 0.1% acetic acid and loaded by pressure vessel onto the precolumn at 1 $\mu L min^{-1}$. After loading, the sample was desalted on the HPLC for 5-10 minutes at approximately 5 $\mu L min^{-1}$. After desalting, 100 fmol of angiotensin I and vasoactive intestinal peptide was loaded via pressure vessel, rinsed for 30 seconds on the HPLC, and the column was reattached to the analytical column. The column was allowed to equilibrate for 15 minutes at approximately 40 bar (~100 $nL min^{-1}$) before beginning data collection on the mass spectrometer. A 40-minute gradient was run from 0-40% acetonitrile with 100mM acetic acid. MS1 scans were collected at 60,000 resolution in the Orbitrap of an LTQ-Orbitrap or the ICR cell of an FT-ICR. Both CAD and ETD MS2 scans were acquired in the ion trap, fragmenting the 4 most abundant ions with charge states of +2 or +3. After

2 sets of MS2 scans were taken in 10 seconds, masses were excluded from selection for 10 seconds.

The collected data were searched against the human proteome using Byonic (Protein Metrics) with MS1 tolerance at 10 ppm and MS2 tolerance at 0.35 Da, allowing 2 variable modifications of oxidation (Met), phosphorylation (Ser, Thr, Tyr), or cysteinylation (Cys). The results were filtered to only include identifications with a score greater than 300 and a |logProb| greater than 3. Relative abundances of peptides were calculated by comparing the peptide peak area to the average peak areas of two internal standard peptides. The amount of material used for an enrichment was then decided based on a combination of the peptide content in the screen and the amount of sample available.

2.3.6 Fischer-Speier Esterification

The selected amount of sample was moved to a new 1.5 mL microcentrifuge tube and dried with a vacuum concentrator. To remove any residual water, the sample was dried twice more after additions of 50 µL methanol. Then, 80 µL of 3M anhydrous hydrogen chloride in methanol was added to the sample, vortexed, bath sonicated for 10 seconds, centrifuged, and allowed to react for 1 hour. The process was then repeated after drying once with methanol.

2.3.7 Phosphopeptide enrichment

2.3.7.1 Preparing activated IDA column

A fused silica microcapillary column (360 µm o.d. x 75 µm i.d.) equipped with a 2 mm Kasil frit was packed with 5 cm of POROS MC 20 iminodiacetate resin. The IMAC column was pressure rinsed at a flow rate of 20 $\mu L min^{-1}$ with the following steps: a 20-

minute water rinse, a 10-minute 50 mM EDTA rinse, and a 10-minute water rinse. The column was activated using filtered 100 mM $FeCl_3$ for 5 minutes at a flow rate of 20 $\mu L min^{-1}$ followed by a 3-minute incubation period in which the iron was allowed to sit on the column to diffuse into resin particles. This process was repeated three times to ensure complete activation. The activated column was equilibrated with 25 µL of 0.01% acetic acid at a flow rate of 0.5 $\mu L min^{-1}$.

2.3.7.2 Preparing activated NTA column

For the preparation of Fe-NTA resin using a 10 kDa molecular weight cutoff filter, all forward spins were performed at 14,000 *x g* for 5 minutes and all spins with the filter inverted (reverse spins) at 1,000 *x g* for 1 minute. A 10 kDa spin filter was rinsed twice with 500 µL of 0.01% acetic using one forward and one reverse spin during each round. Ni-NTA resin was removed from a Qiagen Ni-NTA spin column, reconstituted in water (1 mg/mL), and added to the spin filter in three aliquots, each followed by a forward spin. When adding solutions to the spin filter, the mixture was aspirated to remove any resin that had accumulated on the rim of the filter. The resin was rinsed using the following steps: two rinses with 450 µL of water using two forward spins, two rinses with 450 µL of 50 mM EDTA using two forward spins, and two rinses with 450 µL of water using two forward spins. The resin was activated by three rounds of 450 µL of filtered 100 mM FeCl3 using three forward spins. The activated resin was washed using the following steps: one rinse with 450 µL of 0.01% acetic acid using one forward spin, two rinses of 450 µL of 15% acetonitrile in 0.01% acetic acid using two forward spins, and one rinse with 450 µL of 0.01% acetic acid using one forward spin. The prepared Fe-NTA resin was stored in 0.01% acetic acid at 4 ℃ for up to 1 month.

A fused silica microcapillary column (360 µm o.d. x 150 µm i.d.) equipped with a 2 mm Kasil frit was packed with 2.5 cm of previously prepared Fe-NTA resin. The column was re-activated by pressure loading 100 mM $FeCl_3$ for 10 minutes at a flow rate of 20 $\mu Lmin^{-1}$ followed by a 3-minute incubation period in which the iron was allowed to sit on the column. This process was repeated two times to ensure complete activation.

2.3.7.3 Enrichment protocol

The activated column was equilibrated with 25 µL of 0.01% acetic acid at a flow rate of 0.5 $\mu Lmin^{-1}$. The dried, esterified peptide sample was reconstituted in 50 µL of 1:1:1 (methanol: acetonitrile: 0.01% acetic acid (vol/vol)) and pressure loaded onto the activated IMAC column at a flow rate of 0.5 $\mu Lmin^{-1}$. Following sample loading, 25 µL of 1:1:1 was added to the sample tube and loaded onto the IMAC column at a flow rate of 0.5 $\mu Lmin^{-1}$. A final rinse with 15 µL of 0.1% acetic acid at a flow rate of 0.5 $\mu Lmin^{-1}$ was completed. The sample flow through, 1:1:1 rinse, and 0.1% acetic acid rinse were collected in an Eppendorf tube and stored at -35 °C.

At this point, a PC was rinsed on a HPLC with solvent A at 30 bar for 10 minutes. The PC was connected to the end of the IMAC column with a Teflon sleeve and the IMAC-PC column was rinsed with 0.01% acetic acid for 10 minutes at a flow rate of 0.5 $\mu Lmin^{-1}$ to ensure that no leaks were present. Phosphopeptides were eluted directly onto the PC by pressure loading 15 µL fresh 250 mM L-ascorbic acid in water (pH 2) at a flow rate of 0.5-1 $\mu Lmin^{-1}$. The column was then rinsed with 5 µL 0.01% acetic acid before disconnecting from the precolumn.

The precolumn was rinsed on the HPLC for 20 minutes at a pressure of 30 bar. 100 fmol of standard peptides were then loaded on the precolumn and the column was rinsed

for an additional 30 seconds on the HPLC before connecting to the analytical column (10-12 cm of 3 μm Dr. Maisch Reprosil Pur AQ in 360 o.d. 75 i.d. fused silica). The connected columns were equilibrated on the HPLC for 15 minutes at 45 bar before beginning recording. Peptides were separated on a gradient from 0-40% MeCN in 100 mM acetic acid at 45 bar with a flow rate of ~150 $nLmin^{-1}$.

2.3.8 Mass spectrometric methods

Samples were analyzed using an Orbitrap Fusion Tribrid equipped with ETD. MS1 scans were collected in the Orbitrap with a scan range of 300-1500 and a resolution of 60,000. MS2 scans were collected starting with the most abundant precursors at top speed for a 3 second duty cycle. Precursors with charges of 2-3 were selected for fragmentation. A pair of low resolution MS2 scans were taken at a rapid scan rate of each precursor, one with CAD fragmentation and the other with a 45 ms ETD reaction. If more than 3 sets of scans were collected in 10 seconds, the precursor was excluded from selection for fragmentation for 15 seconds. Samples were recorded for approximately 80 minutes or until the end of the gradient was visible.

2.3.9 Analysis

Data files were searched using the Byonic (Protein Metrics) search algorithm against the human database (SwissProt with isoforms) and against an in-house database of previously identified phosphorylated MHC antigens. An MS1 tolerance of 10 ppm and an MS2 tolerance of 0.35 Da were allowed. A fixed methylation was set on the C-terminus, aspartate, and glutamate with one variable demethylation set as a rare modification. Two oxidized methionine and two phosphorylated serine, threonine, or tyrosine were allowed as common variable modifications. Homocysteinylated cysteine was allowed as a single

rare modification. A maximum of one rare modification and three common modifications were allowed. The search algorithm identifications were then manually validated and verified by a second party.

2.4 Results and Discussion

2.4.1 Phosphorylation in cells treated with phosphatase inhibitor

To investigate the effects of phosphatase inhibition on MHC class I peptide phosphorylation, we analyzed healthy donor PBMC cells treated with calyculin A, a potent phosphatase inhibitor isolated from marine sponges.(43) Two samples were collected, one donor expressing HLA-A*02, A*31, B*08, and B*44 and the other expressing HLA-A*02, B*14, and B*44. Between the two sets of samples, we identified 201 phosphorylated MHC class I peptides, 135 of which were only in the treated cells (**Figure 2.2**). Additionally, many of the peptides from overlapping HLA types HLA-A*02 and B*44 were found expressed on both the samples, with 61 peptides identified on both samples and exclusively in the inhibited sample.

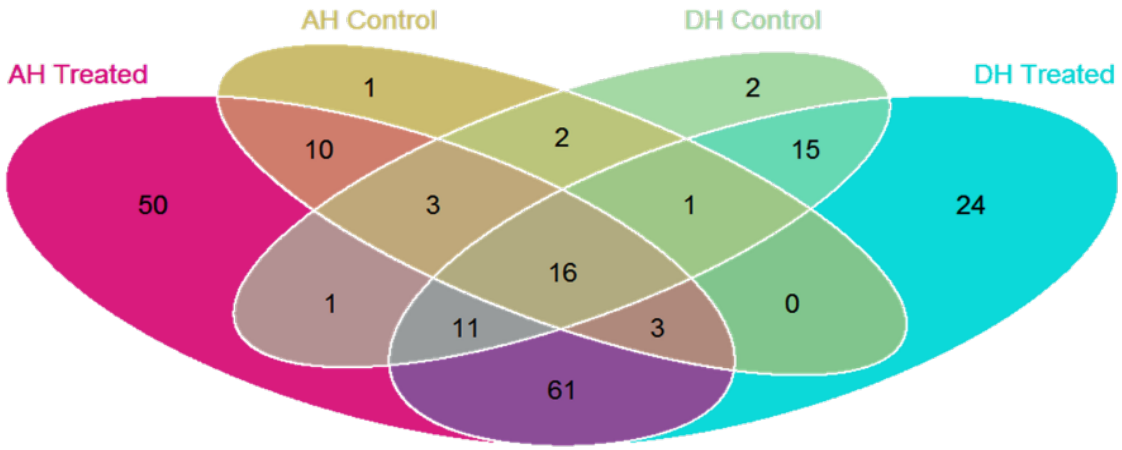


Figure 2.2 Venn diagram of phosphopeptide identifications in samples before and after phosphatase inhibition
Sample overlap between VAAH control (yellow) and treated (pink) with VADH control (green) and treated (blue). The majority of identifications are in the treated samples, and many are found in both treated samples.

Approximately 50% of these peptides have previously been seen in cancerous cells, supporting that phosphatase inhibition is a major cause of MHC class I-associated peptide phosphorylation in cancer.

2.4.2 Phosphorylation in margin tissue from hepatocellular carcinoma

When removing tumors from liver cancer patients, surgeons also remove a margin around the tumor to make sure that all cancerous tissue is removed. Our theory that viral infection should lead to increased phosphorylation of MHC peptides on the surface is supported by analysis of this margin tissue from HCC patients. Previously, we analyzed phosphorylation on hepatocellular carcinoma tumors against their surrounding margin tissue for the identification of tumor specific antigens. In most cases, some phosphorylated antigens will be expressed on both the margin tissue and the tumor tissue. This is

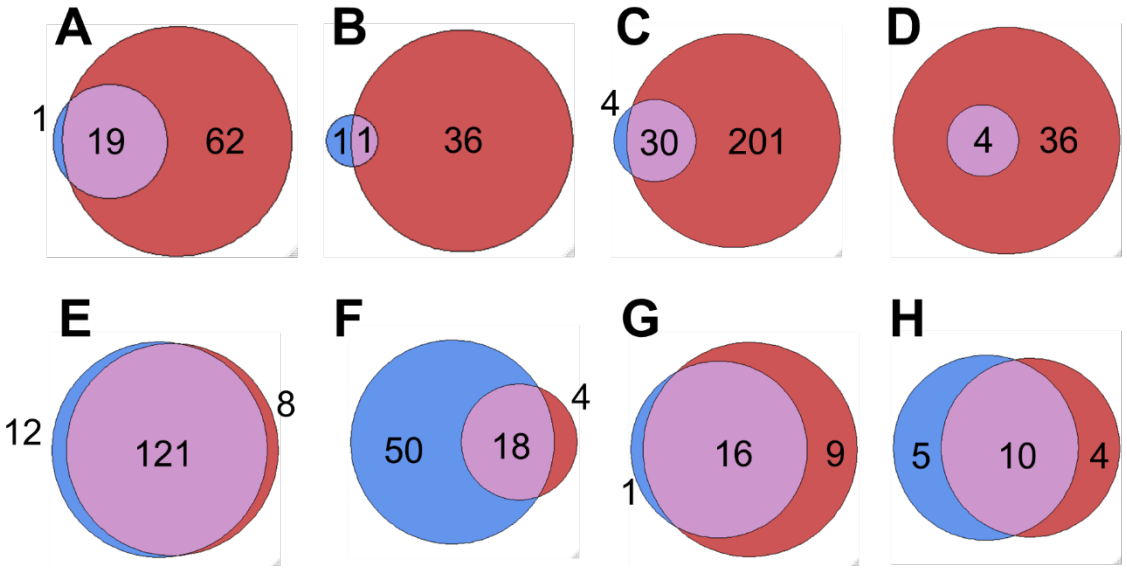


Figure 2.3 Overlap of phosphopeptide identifications in tumor tissue and margin tissue of HCC patients

Overlap between MHC phosphopeptide expression on tumor tissue (red) compared to the surrounding tissue (blue). Margin tissue can express some phosphopeptides due to cellular changes prior to transformation. However, the overlap between tumor and margin tissue is much more pronounced in tumors caused by hepatitis B (E,H) and/or hepatitis C (F,G,H) than it is in samples without a known cause (A,B), or adenoma transformation (C,D).

unsurprising, as margin tissue is generally compromised in some way, either because the margin and tumor were not cleanly separated or because the tissue surrounding the tumor suffers from the same conditions that caused the tumor (e.g. alcoholic liver disease).

Margin tissue from patients that had HCC with an unknown etiology (**Figure 2.3a-b**) or caused by adenoma transformation (**Figure 2.3c-d**) expressed, on average, 11% (0-25%) of the phosphorylated MHC peptides detected on the corresponding tumor tissue.(22) In contrast, margin tissue from patients with HBV and/or HCV shows far more phosphorylation, ranging from 55-675% of the corresponding tumor's expression (**Figure 2.3e-h**). This supports the theory that hepatitis infection causes increased phosphopeptide expression prior to transformation. It also suggests that the same phosphopeptide immunotherapeutics could work as a treatment or preventative for hepatitis as well as cancer.

2.4.3 Phosphorylation in neurodegenerative disorders

The last attempt to characterize the brain immunopeptidome was over a decade ago and used 50 grams of tissue for each sample, finding between 6 and 38 unmodified MHC class I peptides per sample.(83) Most obviously, the brain is not a common operation site and is unlikely to be donated. Additionally, brain tissue has low MHC expression, about 70 times lower than most other tissue types.(84).

Table 2.1 Sequences of two phosphorylated peptides identified in neurodegenerative tissue
Both have been found in multiple cancer types and have been shown to generate an immune response in healthy donors (HD)

Peptide	Uniprot	Tumor	HD	Source protein
KAFsPVRSV	Q02363	B C E H L N O Pi V	Yes	DNA-binding protein inhibitor ID-2
SIMsPEIQL	Q96RK0	B C H N O	Yes	Protein capicua homolog

However, I was able to identify two phosphorylated MHC I peptides (**Table 2.1**) from 400 mg of cadaverous neurodegenerative sample. Both peptides had previously been found in multiple cancers and generate responses from healthy donor memory T-cells. While these peptides were low level, previous observations of the peptides improves confidence in the identification (**Figure 2.4**). The ratio of fragments is characteristic of a given peptide sequence, along with the retention time. Therefore, the presence of these features in combination with the spectrum provides additional confidence.

Later work has identified additional phosphorylated MHC antigens on neurodegenerative tissue.(85) We have not identified any phosphorylated antigens on cadaverous tissues from healthy donors.

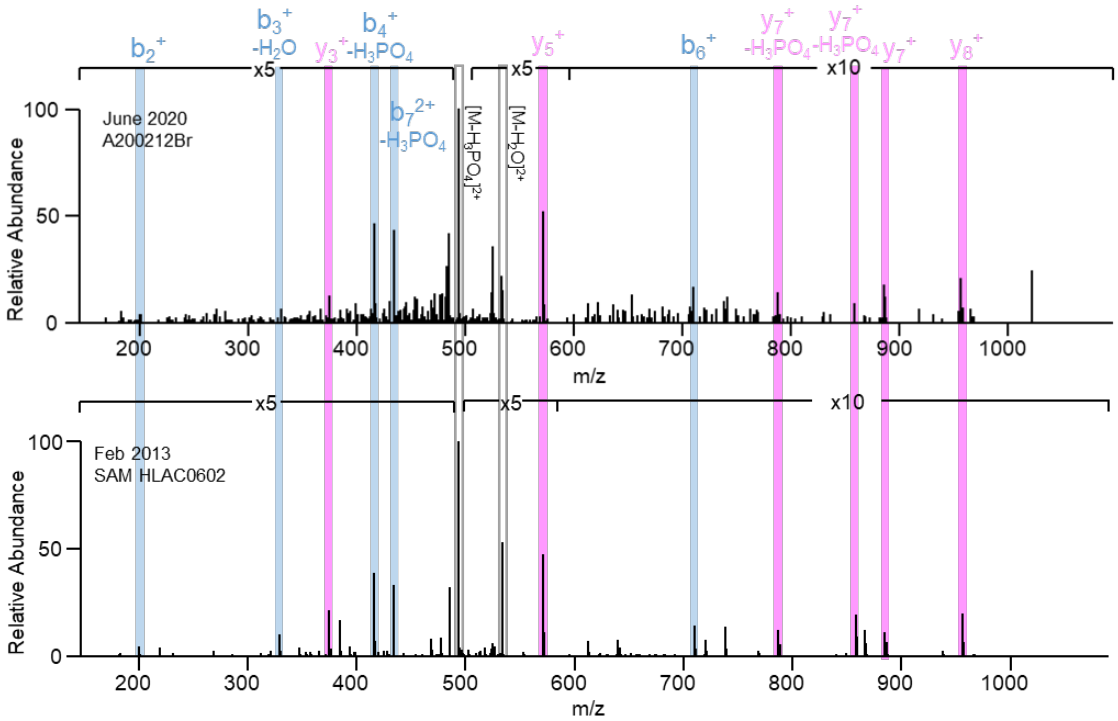


Figure 2.4 Spectral matching allows for higher confidence identification of phosphorylated peptide in neurodegenerative tissue

Though the spectrum for KAFsPVRSV is weak in the enriched sample, comparison to previous identifications shows the same fragment intensity distribution for b ions (blue) and y ions (pink) and allows for confident identification despite incomplete coverage.

2.5 Conclusions and Future Work

From our studies over the past 20 years, we have observed that phosphorylated MHC peptides are expressed primarily on diseased tissue and across multiple cancer types. Since PP2A is prolific and its inhibition is heavily linked with cancer, it is the most likely cause of the increased phosphorylated antigen expression seen in cancer tissues. If PP2A inhibition caused this increase, then treatments developed targeting these phosphopeptides could apply to any disease with the same dysregulation (**Figure 2.5**). Expression of phosphorylated MHC antigens by cells infected with these pathogens would explain why healthy donors with no prior cancer diagnosis have memory T-cells that recognize these antigens. That is, common viruses such as EBV may give the immune system an opportunity to recognize these peptides as antigenic, resulting in memory T cells.

Even though the evidence we have collected thus far is promising, far more research is needed to confirm our theories. Generating the quantities of cells needed to confirm

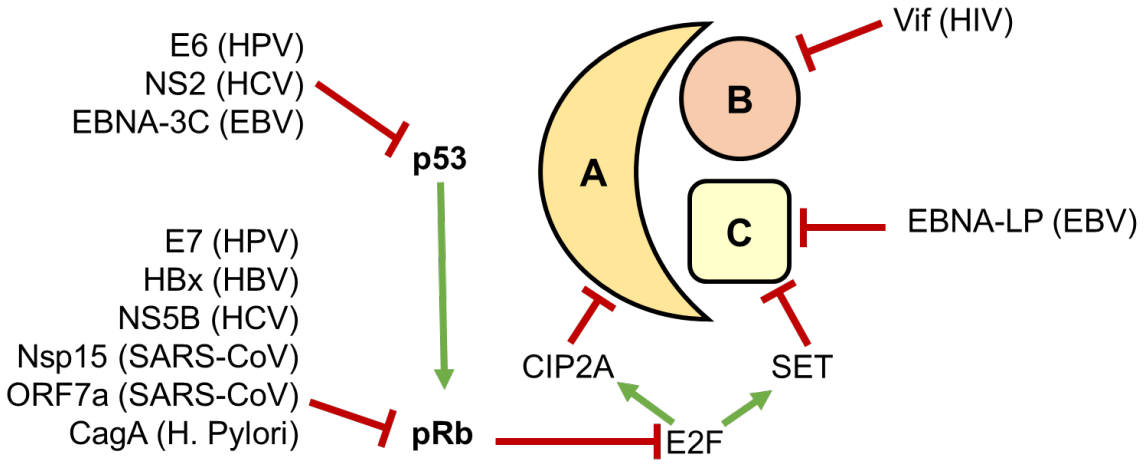


Figure 2.5 Viral and bacterial proteins that inhibit PP2A, pRb, or p53
Since a variety of diseases target these proteins for immune evasion and increased growth, the same phosphorylated peptides should be presented in cells infected with any of these diseases

whether dysregulated cell signaling in other diseases cause expression of the same cancer-

linked phosphorylated antigens is difficult. While infection of cancerous cell lines is commonplace, efficient infection of high quantities of healthy cells is difficult. Additionally, confirmation of this hypothesis requires cells that are infected but not cancerous. Since many of these viruses are oncogenic, the line between infection and cancer may be uncertain. For example, EBV is frequently used to immortalize cells, so it is difficult to define when to consider the cells transformed rather than merely infected.

Despite these challenges, modified antigens are appealing targets for immunotherapeutic treatments for many reasons. Not only are they expressed across many different cancers and individuals, but they also have the potential to allow treatment of many other chronic and debilitating diseases. Since this is a less common approach for neoantigen identification, ample space exists for discovery of novel antigens or diseases. The concept could be further expanded to study other post-translational modifications, such as methylation and glycosylation, which are also dysregulated across many cancers and diseases.

2.6 References

1. Hunt DF, Henderson RA, Shabanowitz J, Sakaguchi K, Michel H, Sevilir N, . . . Engelhard VH (1992) Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* 255(5049):1261-1263.
2. Castle JC, Uduman M, Pabla S, Stein RB, & Buell JS (2019) Mutation-derived neoantigens for cancer immunotherapy. *Frontiers in Immunology* 10:1856.
3. Schumacher TN, Scheper W, & Kvistborg P (2019) Cancer Neoantigens. *Annual review of immunology* 37:173-200.
4. Han XJ, Ma XL, Yang L, Wei YQ, Peng Y, & Wei XW (2020) Progress in Neoantigen Targeted Cancer Immunotherapies. *Frontiers in Cell and Developmental Biology* 8:728.
5. Jiang T, Shi T, Zhang H, Hu J, Song Y, Wei J, . . . Zhou C (2019) Tumor neoantigens: From basic research to clinical applications. *Journal of Hematology and Oncology* 12(1):1-13.
6. Sahin U, Derhovanessian E, Miller M, Kloke BP, Simon P, Löwer M, . . . Türeci Ö (2017) Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* 547(7662):222-226.
7. Kote S, Pirog A, Bedran G, Alfaro J, & Dapic I (2020) Mass spectrometry-based identification of MHC-associated peptides. *Cancers* 12(3):535.
8. Hogan KT, Eisinger DP, Cupp SB, Lekstrom KJ, Deacon DD, Shabanowitz J, . . . Ross MM (1998) The peptide recognized by HLA-A68.2-restricted, squamous cell carcinoma of the lung-specific cytotoxic T lymphocytes is derived from a mutated elongation factor 2 gene. *Cancer Research* 58(22):5144-5150.

9. Solleder M, Guillaume P, Racle J, Michaux J, Pak HS, Müller M, . . . Gfeller D (2020) Mass spectrometry based immunopeptidomics leads to robust predictions of phosphorylated HLA class I ligands. *Molecular and Cellular Proteomics* 19(2):390-404.

10. Zeneyedpour L, Sten-van `t Hoff J, & Luider T (2020) Using phosphoproteomics and next generation sequencing to discover novel therapeutic targets in patient antibodies. *Expert Review of Proteomics* 17(9):675-684.

11. Berg JM, Tymoczko JL, & Stryer L (2007) Defects in Signal-Transduction Pathways Can Lead to Cancer and Other Diseases. *Biochemistry*, (W.H. Freeman, New York), 6th Ed, pp 400-405.

12. Sever R & Brugge JS (2015) Signal transduction in cancer. *Cold Spring Harbor: Perspectives in Medicine* 5(4):a006098.

13. Ardito F, Giuliani M, Perrone D, Troiano G, & Muzio LL (2017) The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy. *International journal of molecular medicine* 40(2):271-280.

14. Mazhar S, Taylor SE, Sangodkar J, & Narla G (2019) Targeting PP2A in cancer: Combination therapies. *Biochimica et Biophysica Acta: Molecular Cell Research* 1866(1):51-63.

15. Zhang Q & Claret FX (2012) Phosphatases: The new brakes for cancer development? *Enzyme Research* 2012:659649.

16. Hunt DF, Shabanowitz J, Myers PT, Cobbold M, Büttner N, Malaker SA, & Penny S (2019) Target peptides for cancer therapy and diagnostics 16/098,634.

17. Abelin JG, Trantham PD, Penny SA, Patterson AM, Ward ST, Hildebrand WH, . . . Hunt DF (2015) Complementary IMAC enrichment methods for HLA-associated phosphopeptide identification by mass spectrometry. *Nature protocols* 10(9):1308-1318.

18. Heather JM, Myers PT, Shi F, Aziz-Zanjani MO, Mahoney KE, Perez M, . . . Hunt DF (2019) Murine xenograft bioreactors for human immunopeptidome discovery. *Scientific Reports* 9(1):1-15.

19. Bassani-Sternberg M, Bräunlein E, Klar R, Engleitner T, Sinitcyn P, Audehm S, . . . Krackhardt AM (2016) Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nature Communications* 7(1):1-16.

20. Cobbold M, De La Peña H, Norris A, Polefrone JM, Qian J, English AM, . . . Engelhard VH (2013) MHC class I–associated phosphopeptides are the targets of memory-like immunity in leukemia. *Science translational medicine* 5(203):203ra125-203ra125.

21. Henderson RA, Cox AL, Sakaguchi K, Appella E, Shabanowitz J, Hunt DF, & Engelhard VH (1993) Direct identification of an endogenous peptide recognized by multiple HLA-A2.1-specific cytotoxic T cells. *Proceedings of the National Academy of Sciences of the United States of America* 90(21):10275-10279.

22. Buettner N, Yaron TM, Myers PM, Krimmer M, Cohen BM, Gruetzmann K, . . . Cobbold M (2021) The evolution of immunity against phosphorylated tumor antigens in hepatocellular carcinoma (HCC). *Cancer Discovery* in revisio.

23. Hunt DF, Norris A, English AM, Shabanowitz J, Hilderbrand WH, & Hawkins OE (2020) Identification of MHC class I phospho-peptide antigens from breast cancer utilizing sHLA technology and complementary enrichment strategies 10/640,535.

24. Engelhard VH, Obeng RC, Cummings KL, Petroni GR, Ambakhutwala AL, Chianese-Bullock KA, . . . Jr (2020) MHC-restricted phosphopeptide antigens: preclinical validation and first-in-humans clinical trial in participants with high-risk melanoma. *Journal for Immunotherapy of Cancer* 8(1):e000262.

25. Topalian SL, Depontieu FA, Hunt DF, Shabanowitz J, Qian J, Engelhard VH, & Zarling AL (2016) Phosphopeptides as melanoma vaccines 9/279,011.

26. Hunt DF, Shabanowitz J, Cottine J, English AM, Norris A, Engelhard VH, . . . Obeng RC (2013) Class I MHC phosphopeptides for cancer immunotherapy and diagnosis 13/699,563.

27. Mohammed F, Cobbold M, Zarling AL, Salim M, Barrett-Wilt GA, Shabanowitz J, . . . Willcox BE (2008) Phosphorylation-dependent interaction between antigenic peptides and MHC class I: a molecular basis for the presentation of transformed self. *Nature immunology* 9(11):1236-1243.

28. Zarling AL, Polefrone JM, Evans AM, Mikesh LM, Shabanowitz J, Lewis ST, . . . Hunt DF (2006) Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. *Proceedings of the National Academy of Sciences* 103(40):14889-14894.

29. Arriazu E, Pippa R, & Odero MD (2016) Protein Phosphatase 2A as a therapeutic target in acute myeloid leukemia. *Frontiers in oncology* 6:78.

30. Eto M & Brautigan DL (2012) Endogenous inhibitor proteins that connect Ser/Thr kinases and phosphatases in cell signaling. *IUBMB Life* 64(9):732-739.

31. Guergnon J, Godet AN, Galioot A, Falanga PB, Colle JH, Cayla X, & Garcia A (2011) PP2A targeting by viral proteins: A widespread biological strategy from DNA/RNA tumor viruses to HIV-1. *Biochimica et Biophysica Acta: Molecular Basis of Disease* 1812(11):1498-1507.

32. Haesen D, Sents W, Ivanova E, Lambrecht C, & Janssens V (2012) Cellular inhibitors of protein phosphatase PP2A in cancer. *Biomedical Research* 23(SI):197-211.

33. Perrotti D & Neviani P (2013) Protein phosphatase 2A: A target for anticancer therapy. *The Lancet Oncology* 14(6):e229-e238.

34. Ruvolo PP (2016) The broken "Off" switch in cancer signaling: PP2A as a regulator of tumorigenesis, drug resistance, and immune surveillance. *BBA Clinical* 6:87-99.

35. Sangodkar J, Farrington CC, McClinch K, Galsky MD, Kastrinsky DB, & Narla G (2016) All roads lead to PP2A: Exploiting the therapeutic potential of this phosphatase. *FEBS Journal* 283(6):1004-1024.

36. Wlodarchak N & Xing Y (2016) PP2A as a master regulator of the cell cycle. *Critical reviews in biochemistry and molecular biology* 51(3):162-184.

37. Vélez-Cruz R & Johnson DG (2017) The retinoblastoma (RB) tumor suppressor: pushing back against genome instability on multiple fronts. *International Journal of Molecular Sciences* 18(8):1776.

38. Kastenhuber ER & Lowe SW (2017) Putting p53 in context. *Cell* 170(6):1062-1078.

39. Garcia A, Cayla X, Guergnon J, Dessauge F, Hospital V, Rebollo MP, . . . Rebollo A (2003) Serine/threonine protein phosphatases PP1 and PP2A are key players in apoptosis. *Biochimie* 85(8):721-726.

40. Virshup DM & Shenolikar S (2009) From promiscuity to precision: protein phosphatases get a makeover. *Molecular Cell* 33(5):537-545.

41. Smith RJ, Cordeiro MH, Davey NE, Vallardi G, Ciliberto A, Gross F, & Saurin AT (2019) PP1 and PP2A use opposite phospho-dependencies to control distinct processes at the kinetochore. *Cell Reports* 28(8):2206-2219.

42. Janssens V, Longin S, & Goris J (2008) PP2A holoenzyme assembly: in cauda venenum (the sting is in the tail). *Trends in Biochemical Sciences* 33(3):113-121.

43. Favre B, Turowski P, & Hemmings BA (1997) Differential inhibition and posttranslational modification of protein phosphatase 1 and 2A in MCF7 cells treated with calyculin-A, okadaic acid, and tautomycin. *The Journal of biological chemistry* 272(21):13856-13863.

44. Eichhorn PJA, Creyghton MP, & Bernards R (2009) Protein phosphatase 2A regulatory subunits and cancer. *Biochimica et biophysica acta* 1795(1):1-15.

45. Kiely M & Kiely PA (2015) PP2A: The wolf in sheep's clothing? *Cancers* 7(2):648-669.

46. Duong-Ly KC & Peterson JR (2013) The human kinome and kinase inhibition. *Current Protocols in Pharmacology* 2(2.9).

47. Li T, Du P, & Xu N (2010) Identifying human kinase-specific protein phosphorylation sites by integrating heterogeneous information from various sources. *PLoS ONE* 5(11):e15411.

48. Hahn WC, Counter CM, Lundberg AS, Beijersbergen RL, Brooks MW, & Weinberg RA (1999) Creation of human tumour cells with defined genetic elements. *Nature* 400(6743):464-468.

49. Janssens V, Goris J, & Van Hoof C (2005) PP2A: The expected tumor suppressor. *Current Opinion in Genetics and Development* 15(1):34-41.

50. Rangarajan A, Hong SJ, Gifford A, & Weinberg RA (2004) Species- and cell type-specific requirements for cellular transformation. *Cancer Cell* 6(2):171-183.

51. Wang J, Okkeri J, Pavic K, Wang Z, Kauko O, Halonen T, . . . Westermarck J (2017) Oncoprotein CIP2A is stabilized via interaction with tumor suppressor PP2A/B56. *EMBO reports* 18(3):437-450.

52. Böckelman C, Lassus H, Hemmes A, Leminen A, Westermarck J, Haglund C, . . . Ristimäki A (2011) Prognostic role of CIP2A expression in serous ovarian cancer. *British Journal of Cancer* 105(7):989-995.

53. Côme C, Laine A, Chanrion M, Edgren H, Mattila E, Liu X, . . . Westermarck J (2009) CIP2A is associated with human breast cancer aggressivity. *Clinical Cancer Research* 15(16):5092-5100.

54. Dong QZ, Wang Y, Dong XJ, Li ZX, Tang ZP, Cui QZ, & Wang EH (2011) CIP2A is overexpressed in non-small cell lung cancer and correlates with poor prognosis. *Annals of Surgical Oncology* 18(3):857-865.

55. Junttila MR, Puustinen P, Niemelä M, Ahola R, Arnold H, Böttzauw T, . . . Westermarck J (2007) CIP2A inhibits PP2A in human malignancies. *Cell* 130(1):51-62.

56. Li M, Makkinje A, & Damuni Z (1996) The myeloid leukemia-associated protein SET is a potent inhibitor of protein phosphatase 2A. *Journal of Biological Chemistry* 271(19):11059-11062.

57. Li M, Makkinje A, & Damuni Z (1996) Molecular identification of I1PP2A, a novel potent heat-stable inhibitor protein of protein phosphatase 2A. *Biochemistry* 35(22):6998-7002.

58. Cristóbal I, Rincón R, Manso R, Caramés C, Zazo S, Madoz-Gúrpide J, . . . García-Foncillas J (2015) Deregulation of the PP2A inhibitor SET shows promising therapeutic implications and determines poor clinical outcome in patients with metastatic colorectal cancer. *Clinical Cancer Research* 21(2):347-356.

59. Kauko O, Imanishi SY, Kulesskiy E, Yetukuri L, Laajala TD, Sharma M, . . . Jumppanen M (2020) Phosphoproteome and drug-response effects mediated by the three protein phosphatase 2A inhibitor proteins CIP2A, SET, and PME-1. *Journal of Biological Chemistry* 295(13):4194-4211.

60. Liu J, Wang X, Zhou G, Wang H, Xiang L, Cheng Y, . . . Zhao W (2011) Cancerous inhibitor of protein phosphatase 2A is overexpressed in cervical cancer and upregulated by human papillomavirus 16 E7 oncoprotein. *Gynecologic Oncology* 122(2):430-436.

61. Lucas CM, Harris RJ, Giannoudis A, Copland M, Slupsky JR, & Clark RE (2011) Cancerous inhibitor of PP2A (CIP2A) at diagnosis of chronic myeloid leukemia is a critical determinant of disease progression. *Blood* 117(24):6660-6668.

62. Westermarck J (2018) Targeted therapies don't work for a reason; the neglected tumor suppressor phosphatase PP2A strikes back. *FEBS Journal* 285(22):4139-4145.

63. Knudson AG (1971) Mutation and cancer: statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America* 68(4):820-823.

64. Hickman ES, Moroni MC, & Helin K (2002) The role of p53 and pRB in apoptosis and cancer. *Current opinion in genetics & development* 12(1):60-66.

65. Zhao L, Samuels T, Winckler S, Korgaonkar C, Tompkins V, Horne MC, & Quelle DE (2003) Cyclin G1 has growth inhibitory activity linked to the ARF-Mdm2-p53 and pRb tumor suppressor pathways. *Molecular Cancer Research* 1(3):195-206.

66. Shimada M & Nakanishi M (2013) Response to DNA damage: why do we need to focus on protein phosphatases? *Frontiers in Oncology* 3(January):1-14.

67. Li HH, Cai X, Shouse GP, Piluso LG, & Liu X (2007) A specific PP2A regulatory subunit, B56γ, mediates DNA damage-induced dephosphorylation of p53 at Thr55. *EMBO Journal* 26(2):402-411.

68. Kasten MM & Giordano A (1998) pRb and the Cdks in apoptosis and the cell cycle. *Cell death and differentiation* 5(2):132-140.

69. Ringelhan M, McKeating JA, & Protzer U (2017) Viral hepatitis and liver cancer. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 372(1732):20160274-20160274.

70. Lunn RM, Jahnke GD, & Rabkin CS (2017) Tumour virus epidemiology. *Philosophical Transactions of the Royal Society B: Biological Sciences* 372(1732):20160266-20160266.

71. Ringelhan M, McKeating JA, & Protzer U (2017) Viral hepatitis and liver cancer. *Philosophical Transactions of the Royal Society B* 372(1732):20160274.

72. Munakata T, Liang Y, Kim S, McGivern DR, Huibregtse J, Nomoto A, & Lemon SM (2007) Hepatitis C virus induces E6AP-dependent degradation of the retinoblastoma protein. *PLoS Pathogens* 3(9):1335-1347.

73. Mileo AM, Mattarocci S, Matarrese P, Anticoli S, Abbruzzese C, Catone S, . . . Ruggieri A (2015) Hepatitis C virus core protein modulates pRb2/p130 expression in human hepatocellular carcinoma cell lines through promoter methylation. *Journal of Experimental and Clinical Cancer Research* 34(1):140-140.

74. Bittar C, Shrivastava S, Bhanja Chowdhury J, Rahal P, & Ray RB (2013) Hepatitis C virus NS2 protein inhibits DNA damage pathway by sequestering p53 to the cytoplasm. *PLoS ONE* 8(4):e62581-e62581.

75. Lin S & Zhang YJ (2017) Interference of apoptosis by hepatitis B virus. *Viruses* 9(8):230-230.

76. Ringelhan M, O'Connor T, Protzer U, & Heikenwalder M (2015) The direct and indirect roles of HBV in liver cancer: Prospective markers for HCC screening and potential therapeutic targets. *Journal of Pathology* 235(2):355-367.

77. Jin KJ, Arora P, Pagano JS, & Kyung LJ (2007) Expression of DNA methyltransferase 1 is activated by hepatitis B virus X protein via a regulatory circuit involving the p16INK4a-cyclinD1-CDK 4/6-pRb-E2F1 pathway. *Cancer Research* 67(12):5771-5778.

78. Arif M, Wei J, Zhang Q, Liu F, Basurto-Islas G, Grundke-lqbal I, & Iqbal K (2014) Cytoplasmic retention of protein phosphatase 2A inhibitor 2 (I2PP2A) induces alzheimer-like abnormal hyperphosphorylation of Tau. *Journal of Biological Chemistry* 289(40):27677-27691.

79. Sontag J-M & Sontag E (2014) Protein phosphatase 2A dysfunction in Alzheimer's disease. *Frontiers in Molecular Neuroscience* 7(March):1-10.

80. Liu F, Grundke-Iqbal I, Iqbal K, & Gong CX (2005) Contributions of protein phosphatases PP1, PP2A, PP2B and PP5 to the regulation of tau phosphorylation. *European Journal of Neuroscience* 22(8):1942-1950.

81. Tanimukai H, Grundke-Iqbal I, & Iqbal K (2005) Up-regulation of inhibitors of protein phosphatase-2A in Alzheimer's disease. *American Journal of Pathology* 166(6):1761-1771.

82. Rappsilber J, Mann M, & Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature protocols* 2(8):1896-1896.

83. Fissolo N, Haag S, De Graaf KL, Drews O, Stevanovic S, Rammensee HG, & Weissert R (2009) Naturally presented peptides on major histocompatibility complex I and II molecules eluted from central nervous system of multiple sclerosis patients. *Molecular and Cellular Proteomics* 8(9):2090-2101.

84. Lampson LA & Hickey WF (1986) Monoclonal antibody analysis of MHC expression in human brain biopsies: Tissue ranging from 'histologically normal' to that showing different levels of glial tumor involvement. *Journal of immunology* 136(11):4054-4062.

85. Peng X (2022) Identification of MHC-Associated Peptides and Phosphopeptides Presented in Various Diseased Cells by Tandem Mass Spectrometry. PhD (University of Virginia).

Chapter 3. Improved methods for enrichment of phosphorylated MHC Class I associated peptides

3.1 Introduction

Phosphorylation is the most common post-translational modification (PTM), as nearly one-third of proteins are phosphorylated in their lifetime.(1) It is a major component of signaling pathways due to its reversible and transient nature.(2) Kinases phosphorylate proteins on serine, threonine, and tyrosine residues, whereas phosphatases reverse the reaction. The balance between these enzymes is crucial to maintain normal cell functions, including transcription, differentiation, cell cycle regulation, migration, and metabolism, among others.(3) Aberrant phosphorylation results in dysregulation of these cell signaling pathways, which promotes disease progression in cancer and other diseases.

Peptides with PTMs are estimated to be present at approximately 0.1-1% of the total HLA-associated peptide population.(4) Due to their extremely low abundance coupled with the complexity and quantity of unmodified peptides present on the surface of cells, enrichment is necessary identify any low abundance phosphorylated MHC antigens. The two main avenues for performing this enrichment are immobilized metal affinity chromatography (IMAC) and metal oxide affinity chromatography (MOAC).

3.1.1 Phosphopeptide enrichment

Phosphopeptide enrichment with IMAC is performed through conjugation of a chelator such as iminodiacetic acid (IDA) or nitrilotriacetic acid (NTA) with a trivalent metal ion, most commonly iron (III). Negatively charged phosphorylation associates with

the positively charged metal. However, depending on the pH of the loading solvent, the peptide can carry additional negative charges at acidic residues and the C-terminus leading to significant nonspecific binding. To counteract this, the pH of the loading solvent can be decreased, but this also lowers the affinity of the phosphate group for the metal. Our protocols use esterification to circumvent this issue. Fischer esterification adds a methyl group to carboxylic acids in the peptide but does not esterify the phosphate. This allows for the enrichment to be performed at higher pH (~5), which means that the phosphate will carry two negative charges while the rest of the peptide has zero negative charges, allowing for a highly specific enrichment.

Elution from Fe-IMAC is most commonly enacted through competition with phosphate buffer. This is undesirable for our purposes, as phosphate is not volatile and requires additional desalting before MS analysis. For MHC analyses, while we can desalt before enrichment, the total amount of phosphopeptide present after enrichment is usually in the femtomole level, with most identifications being present at less than 5 attomoles. The losses from subjecting such a small, dilute sample to any additional processing will be much more noticeable than when there are unmodified peptides and HLA molecules acting as carriers. That is, for each additional step in a protocol, a certain amount of analyte will be lost. When only the phosphopeptides are present, all those losses are taken from the phosphopeptides. If the step is added before enrichment, this loss is spread between the phosphopeptides and unmodified peptides, having a much lower impact on the phosphopeptide yield. Therefore, we elute from the Fe-IMAC columns using ascorbic acid, which reduces iron (III) to iron (II), causing it to lose affinity for the phosphate. Ascorbic acid is sufficiently volatile that we can attach a precolumn to the IMAC column

and elute directly on to the column. The sample is then rinsed on the same column used for analysis. This minimizes the steps necessary when the enriched sample is at the highest risk for losses.

More recently, MOAC has become the most commonly used method for phosphopeptide enrichment in proteomics.(5) Titanium oxide is the most common support, but zirconium can also be used. The protocols are straightforward, quick, and require little instruction for new users. Additionally, the titanium maintains its affinity for phosphate at low pH, allowing for a more specific enrichment. However, it tends to suffer for low-level phosphopeptides, requiring hundreds of micrograms of lysate for successful enrichment. This is mostly due to its high affinity of TiO_2 for phosphate, which makes it very difficult to elute phosphopeptides from the column. Therefore, both alkaline conditions (~ pH 11) as well as competitors are required for sufficient elution. The increased pH can cause a high level of contaminating polymers. While the effects of contamination are minimal with large sample amounts, contaminant ions do not decrease as sample amount decreases, making it difficult to scale the procedure for smaller samples. The sample also requires additional desalting steps before analysis, which presents the same problems discussed above. Therefore, despite the appeal for other fields of proteomics, metal oxide affinity is not preferred for low-level MHC phosphopeptide enrichment.

While our current protocol allows for identification of tens to hundreds of phosphopeptides with less starting material than is generally used in the field for unmodified analysis, we identified several potential areas for improvement. Previous work in our lab indicated that Fe-IDA IMAC and Fe-NTA IMAC enriched complementary subsets of phosphopeptides.(6) My reanalysis indicates that Fe-NTA does not enrich a

complementary subset of phosphopeptides, but instead is a superior enrichment technique. Another drawback of our protocol was the length of the enrichment, ranging from 6 to 12 hours of active time on a single day. I have made several alterations to the protocol that shorten this process to approximately 4-6 hours. Ultimately, this optimization has resulted in a greatly improved enrichment technique that is more accessible and has enhanced our ability to detect phosphopeptides via mass spectrometry.

3.2 Materials

3.2.1 Reagents

Aldrich Chemical Co. (Milwaukee, WI)

- Ethylenediaminetetraacetic acid, 99.99% purity

EMD Millipore (Burlington, MA)

- Benzonase® Nuclease, Purity >90%

Fisher Chemical (Ward Hill, NJ)

- Methanol, Optima™ LC/MS grade

Honeywell (Charlotte, NC)

- Acetonitrile, B&J Brand™ LC-MS, for LC-MS and HPLC, >99.9%

PQ Corporation (Malvern, PA)

- KaSil 1624 potassium silicate solution

Sigma Aldrich (St Louis, MO)

- Acetic acid, glacial, ≥99.99% trace metals basis
- L-Ascorbic Acid, ACS reagent, ≥99%
- Iron (III) chloride, anhydrous, powder, ≥99.99% trace metals basis
- Formamide, ≥99.5% (GC), BioReagent for molecular biology
- Magnesium Chloride hexahydrate

Supelco (Bellefonte, PA)

- Hydrogen chloride solution, 3M in methanol, for GC derivatization, LiChropur™

Thermo Scientific (Waltham, MA)

- Pierce™ Water, LC-MS grade

3.2.2 Proteins and Peptides

Anaspec (Fremont, CA)

- Angiotensin II Phosphate >95% HPLC
- Vasoactive Intestinal Peptide (VIP) 1-12 human, porcine, rat >95% HPLC

Atlantic Peptides (Concord, NH)

- Custom phosphopeptide standards
    - ST(pS)LVGGR
    - FLG(pT)PIAKV

Sigma Aldrich (St Louis, MO)

- Angiotensin I human acetate salt hydrate >90% HPLC

Thermo Scientific (Waltham, MA)

- Pierce™ Peptide Retention Time Calibration Mixture
- Pierce™ HeLa Protein Digest Standard
- Custom Phosphopeptide Standards
    - RVK(pS)PLFQF
    - RTH(pS)LLLLG

3.2.3 Chromatography Resins

Applied Biosystems TM (Carlsbad, CA)

- POROS® MC 20 metal chelating packing material, 20 μm diameter

Dr Maisch GMBH (Ammerbuch, Germany)

- Reprosil Pur 120 C18 AQ 3 μm
- Reprosil Pur 120 C18 AQ 10 μm

QIAGEN (Hilden, Germany)

- Ni-NTA spin columns

3.2.4 Equipment and Instrumentation

Agilent Technologies (Palo Alto, CA)

- 1100 Agilent high performance liquid chromatograph

Boekel Scientific (Feasterville-Trevose, PA)

- Digital Incubator 133000

Branson (Danbury, CT)

- Branson 1200 Ultrasonic Bath

Cadence Science (Cranston, RI)

- Blunt end pipetting needle (16G)

Eppendorf (Enfield, CT/Hamburg, Germany)

- Model 5424 Centrifuge
- Protein LoBind Microcentrifuge tubes (1.5 mL)

Fisherbrand (Pittsburgh, PA)

- Model 50 Sonic Dismembrator with CL-18 Probe
- Digital Vortex
- Vortex Genie 2

Labconco Corp. (Kansas City, MO)

- Centrivap centrifugal vacuum concentrator 7810016

New Objective (Littleton, MA)

- PicoClear 360 Union

PolyMicro Technologies, Inc. (Phoenix, AZ)

- 360 μm o.d. x 75 μm i.d. polyimide coated fused silica capillary
- 360 μm o.d. x 100 μm i.d. polyimide coated fused silica capillary
- 360 μm o.d. x 150 μm i.d. polyimide coated fused silica capillary

Sarstedt Inc (Newton, NC)

- 1.5 mL screw top microtubes

Supelco (Bellefonte, PA)

- Screw top clear glass vials with green thermoset caps and PTFE liners, 2 mL

Sutter Instrument Co. (Novato, CA)

- P-2000 microcapillary laser puller with fused silica adapter

Thermo-Fisher Scientific (San Jose, CA/Bremen, Germany)

- Orbitrap Fusion Tribrid mass spectrometer with commercial front-end ETD
- LTQ mass spectrometer with commercial back-end ETD
- LTQ FT-ICR hybrid mass spectrometer (custom modified with front-end ETD)
- LTQ-Orbitrap mass spectrometer (custom modified with front-end ETD)

Zeus Industrial Products, Inc. (Orangeburg, SC)

Teflon tubing, 0.012-inch i.d. x 0.060-inch o.d.

3.3 Methods

Unless otherwise mentioned, analysis followed protocols described in section 2.3.

3.3.1 Activation of NTA resin with iron (III)

Resin was removed from a Qiagen Ni-NTA spin column by pushing the frits out using a paperclip and tapping the material into a 1.5 mL centrifuge tube. Approximately half of the material was separated into another tube and the rest was stored at 4 ℃ until the next activation. The resin was activated by adding 500 µL of each of the following solutions, shaking, and removing the supernatant. For the additions of EDTA and $FeCl_3$, the resin was allowed to react for 5 minutes on a shaker before removing the supernatant. Each step was repeated twice; water, 50 mM EDTA, water, 100 mM $FeCl_3$, 0.1% acetic acid, and 15% MeCN/ 85% 0.1% acetic acid. The resin was then transferred into a 1 mL glass vial in 500 µL of 0.1% acetic acid and stored for up to one month at 4 ℃.

3.3.2 Removal of nucleic acid contaminants

The sample was reconstituted in 100 µL of 20 mM Tris pH8 with 2mM $MgCl_2$. Benzonase® nuclease (5 units) were added, and the sample was incubated at 37 ℃ for 1 hour. The sample was then desalted on STAGE tips as previously described.

3.3.3 Fischer esterification

The selected amount of sample was moved to a new 1.5 mL microcentrifuge tube and dried with a vacuum concentrator. To remove any residual water, the sample was dried twice more after additions of 50 µL methanol. Using a filter tip, 80µL of 3M anhydrous hydrogen chloride in methanol was added to the sample, vortexed, bath sonicated for 10 seconds, centrifuged, and allowed to react for 15 minutes. The process was then repeated

after drying once with methanol. An additional 80 µL of 3M anhydrous hydrogen chloride in methanol was added to the sample, vortexed, bath sonicated for 10 seconds, centrifuged, and dried immediately. Methanol was then added and dried once more to ensure complete removal of the acid.

3.3.4 Phosphopeptide enrichment

A 150 µm i.d. piece of fused silica was fitted with a frit with the same method used for precolumn construction (**Appendix A**). The frit was cut to 2 mm in length, rinsed upside down on a pressure vessel, reversed, and the bottom of the column was trimmed. Unless otherwise mentioned, the flow rates were kept at approximately 1.5-2 µLmin-1. After equilibration with 0.01% acetic acid, activated NTA-agarose resin was packed to a bed length of 2.5 cm. Any leaching during storage was addressed by flowing 100 mM iron (III) chloride through the column for a few seconds before stopping flow for 3-5 minutes. The column was then rinsed with 30 µL of 0.01% acetic acid, and the pH was checked to ensure it was between 3 and 3.5. The column was then equilibrated with 10 µL of a solution consisting of equal parts 0.01% acetic acid in H_2O with 60% acetonitrile, and the flow adjusted to be between 0.5-1 $\mu Lmin^{-1}$ before sample loading.

The dried sample was reconstituted in 50 µL of a solution consisting of equal parts 0.01% acetic acid, 60% acetonitrile in H_2O. The pH was checked to ensure a pH of 3.5-5. If the pH was lower, the sample was dried and reconstituted again. The sample was then loaded at a flow rate of 0.5-1 $\mu Lmin^{-1}$ and the flow through was collected. If the sample loaded in under one hour, the flow through was loaded a second time through the column. After loading, 25 µL of the loading solvent was used to rinse the tube and was also loaded through the column. After binding, the column was rinsed with 25 µL of 0.01% acetic acid

at 1 μLmin^{-1}. The column was then attached directly to a precolumn (8 cm of 10 μm Dr. Maisch Reprosil Pur AQ in 360 o.d. 100 i.d. fused silica) that was previously equilibrated for 10 minutes on an Agilent 1100 HPLC. The connected columns were rinsed with 5 μL of 0.01% acetic acid before elution with 15 μL of 25 mM ascorbic acid in water.

The precolumn was then rinsed with 5 μL 0.01% acetic acid before disconnecting from the precolumn. The precolumn was then rinsed on the HPLC for 20 minutes at a pressure of 30 bar. 100 fmol of standard peptides were loaded on the precolumn and the column was rinsed for an additional 30 seconds on the HPLC before connecting to the analytical column (10-12 cm of 3 μm Dr. Maisch Reprosil Pur AQ in 360 o.d. 75 i.d. fused silica). The connected columns were equilibrated on the HPLC for 15 minutes at 45 bar before beginning recording. Peptides were separated on a gradient from 0-40% MeCN in 100 mM acetic acid at 45 bar with a flow rate of ~150 nLmin^{-1}.

3.4 Results and Discussion

3.4.1 Phosphopeptide enrichment resin selection

Iminodiacetic acid was the original material used for metal immobilization for phosphopeptide enrichment.(7) However, after its introduction in the late 1980s, nitrilotriacetic acid quickly because a more popular support due to its higher specificity.(8, 9) This higher specificity is achieved because the NTA has an additional ligand arm (**Figure 3.1**), which imparts approximately 1.5E5 times the binding affinity for iron (III) (**Table 3.1**).(10) The pKa of each ligand arm is around 2, below which, affinity will decrease dramatically. Therefore, it is impossible to load at a low enough pH to protonate glutamate (pKa=4.1), aspartate (pKa=3.9), and the C-terminus (pKa~2) while keeping the

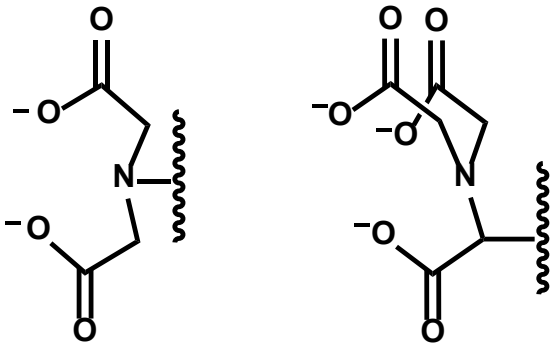


Figure 3.1 Structures of chelating ligands for Fe-IMAC enrichment
Iminodiacetic acid (left) has two acetate ligand arms, while nitrilotriacetic acid (right) has three.

ligand fully deprotonated, causing iron (III) to leach from the column, thus decreasing phosphopeptide retention. The additional chelating arm on the NTA gives an additional negative charge, which should decrease leaching at low pH. Using esterification, this issue is circumvented, as the acidic regions are esterified rather than protonated. This allows samples to be loaded at a higher pH, which should both decrease leaching in addition to increasing the affinity of the

Table 3.1 Table of stability constants of iron complexed with different ligands

Metal	IDA	NTA	Phosphate
Fe (III)	10.72	15.9	8.3
Fe (II)	5.54	8.33	3.6

phosphate for iron by providing an additional partial negative charge on the phosphate group. However, iron can still leach from the column, so the greater affinity of NTA should reduce or prevent this leaching.

Previous comparison of the two chelating ligands indicated that they enrich complementary phosphopeptides.(6, 11) An ovarian cancer cell line (FHIOSE) and a colorectal cancer patient sample (CRCLM03T) were compared using an LTQ-Orbitrap instrument (**Figure 3.2**).(6)

However, these comparisons only considered the results from search algorithms (i.e., identifications from fragmentation spectra), as opposed to investigating the presence of the intact mass in the MS1. While it remains a useful instrument, the LTQ Orbitrap is not able to collect MS2 scans as quickly as the newer Orbitrap Fusion Tribrid instruments. This makes it much more common for a peptide to be present in a full mass spectrum but not be selected for fragmentation by the mass spectrometer. Therefore, it is necessary to determine whether peptides were enriched in both samples by extracting the ion chromatogram (i.e., abundance of a particular mass in each MS1 spectrum) in any files where the peptide was

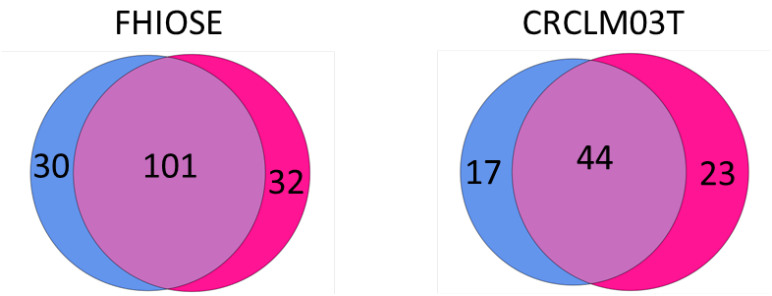


Figure 3.2 Comparison of identifications using NTA or IDA resin and an LTQ Orbitrap
Previous analyses indicated that phosphopeptides enriched by IDA (blue) were complementary to those enriched by NTA (pink). Data from Abelin et al. (2015).

not identified. Unfortunately, programs that could perform this automatically are not able to handle data files generated from the LTQ Orbitrap due to its ETD being built in-house,

so this must be performed manually. The lack of this manual comparison could account for many of the identifications that were only found using one resin.

To avoid this type of analysis, sometimes technical and biological replicates are used. For FHIOSE, two technical replicates were collected for each enrichment type. For CRCLM03T, two technical replicates were collected for IDA and only one for NTA. However, without seeing the overlap between identifications using a single resin it is difficult to make conclusions about the amount of overlap attributable to the resin compared to random chance.

Later, a hepatocellular carcinoma tumor sample (LL4857T) was analyzed using both resins on an Orbitrap Fusion Tribrid mass spectrometer.(11) LL4857T was enriched once with each technique. Here, the NTA enriched nearly double the number of phosphopeptides when compared to IDA. All of the peptides identified only using IDA were present at less than a femtomole, as were the majority of the peptides only found in NTA. Therefore, the variation is likely more attributable to the stochasticity of mass spectrometry than to IDA and NTA being complementary enrichment techniques. However, that work still concluded that the techniques were complementary.(11)

Given this discovery, we reasoned that revisiting the complementarity of the two techniques was warranted. In addition to further investigation of the earlier enrichments of LL4857T, I made comparisons using enrichments of a colorectal carcinoma cell line grown in a mouse (colo205m) and an additional hepatocellular carcinoma clinical sample (LL5176T).

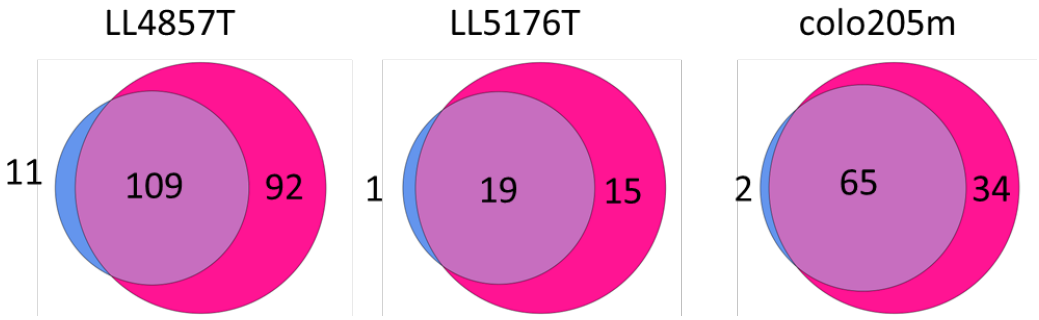


Figure 3.3 Comparison of MHC phosphopeptide identifications using Fe-NTA or Fe-IDA on an Orbitrap Fusion Tribrid
Euler plots showing overlap of peptides identified by IDA (blue) and NTA (pink) for three MHC class I phosphopeptide enrichments analyzed using an Orbitrap Fusion Tribrid mass spectrometer.

The enrichment using NTA allowed for more identifications than IDA in all cases (**Figure 3.3**). The major difference between the two resins was in the identification of peptides present at an abundance of less than 1 femtomole. Between the three samples, 55% were identified using both resins, 41% only by NTA, and 4% only by IDA. The majority of peptides identified only in IDA enrichments were low-level peptides that were present at a similar or higher abundance in the NTA chromatograms and did not have assigned charge states, eliminating them from selection for fragmentation. Only one peptide identified solely using IDA was not present in the corresponding NTA enrichment by mass and retention time (**Figure 3.4**). However, when we take the same consideration for the IDA enrichment, only an additional 11 peptides were detected.

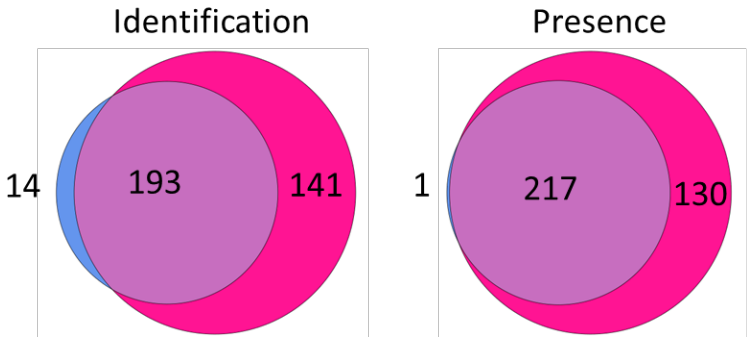


Figure 3.4 Comparison of identification and presence of peptides between enrichment techniques
Of the peptides identified only using IDA resin, 13/14 peptides were present in the NTA enrichment. For those identified only in the NTA enrichment, only 11/141 were present in the IDA enrichment.

Therefore, based on this continued analysis, NTA appeared to be a superior enrichment technique rather than a complementary technique. The peptides only present in samples enriched with NTA were predominantly low-level peptides, less than 1 femtomole, which suggests that the

3.4.2 Addressing issues with NTA

Despite previously determining that NTA is a superior resin, the lab returned to using IDA after some difficulties. The major issue was that, as mentioned previously, the fused silica microcolumn IMAC protocols are very sensitive, but also very temperamental. One aspect of this is column clogging, wherein solvent flow through the IMAC column is slowed or stopped completely. These clogs can happen in one of two ways: the first is that some particulate is in the sample (e.g. dust, fibers), in which case the column flow can be reversed and the particulate can be very carefully pushed out of the back of the column. The second can occur when something in the sample binds strongly to the resin and is abundant enough to build up on the column and prevent flow. In this case, it cannot be cleared by reversing flow.

While this type of column blockage happens equally often with IDA and NTA, IDA had the benefit that the column rarely completely stopped flowing, so the sample could still be loaded, albeit sometimes over the course of 8 hours. Often, if the pressure was increased, the column would suddenly increase solvent flow, after which the pressure could be reduced and sample loading could proceed. Likely, this unclogging was the result of the iron dissociating from the resin along with the contaminant, since with NTA these clogs cannot be removed or reduced.

But the benefits provided by the NTA protocol warranted investigating the problem. Theories regarding causation of these clogs varied, but I hypothesized that they were caused by DNA nonspecifically bound to the NHS beads during immunopurification, as I frequently identified peptides from nuclear proteins as contaminants in unmodified peptide analyses. That said, immunopeptidomic samples are often severely limited in material; we often do not receive enough material to perform multiple enrichments. Thus, testing the hypothesis required a sample that was impossible to analyze without addressing the issue.

When I encountered a sample that clogged irreversibly within the first microliter of sample loading, I tested my theory that the contaminant was DNA that nonspecifically bound to the enrichment beads. The sample was dried again to remove the organic loading solvent, then resuspended in Tris buffer with 2mM $MgCl_2$ for treatment with Benzonase nuclease. Benzonase nuclease digests both RNA and DNA into small 1-5 base pair fragments. After this digestion, the sample was desalted again to remove the nonvolatile Tris and $MgCl_2$ as well as most of the small, hydrophilic nucleic acid fragments. The subsequent enrichment was straightforward. Interestingly, in the MS analysis, we noticed proteins eluting towards the end of the gradient (**Figure 3.5**). We were able to identify these proteins as fragments from histone H2B and H4, further confirming the hypothesis. Phosphopeptide standards were recovered at 39% and 34 MHC-associated phosphopeptides were identified.

The presence of contaminant proteins (e.g. histones) interferes with peptide identification, so we determined that it is best to perform Benzonase treatment during cell lysis so the contaminant proteins are not present in the immunopurification. In following

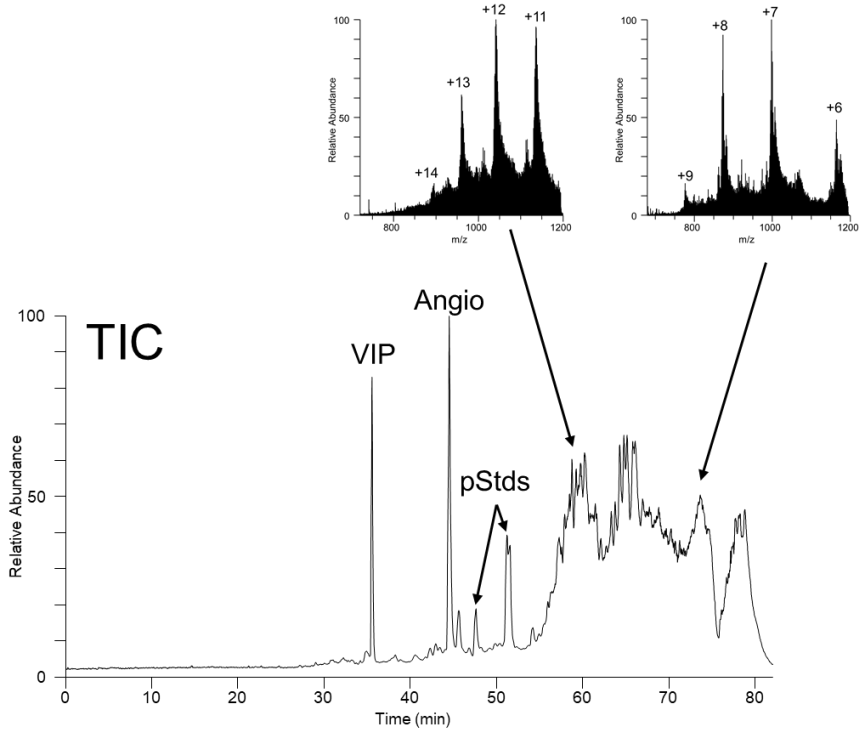


Figure 3.5 Chromatogram of HEK293Tm after IMAC following Benzonase treatment
After HEK293Tm clogging the IMAC column and subsequent treatment with Benzonase, the chromatogram consists of standards and phosphostandards followed by elution of large, modified species. Upon investigation, these species were primarily histones and histone fragments. The TIC (bottom) depicts the summed abundance of all peaks present in each MS1. After elution of standards as peaks between 30-55 minutes, a sharp increase in the TIC is observed. These chromatographic peaks corresponded to proteins and protein fragments from 7-13 kDa, some of which were identified as being histone fragments.

immunopurifications, I found that samples lysed and immunopurified in a Tris buffer including 100 U/mL Benzonase nuclease and 2 mM $MgCl_2$ did not have problems with clogging, thus allowing NTA to be used more effortlessly. Additionally, fewer contaminant peptides were observed in the initial sample screening.

3.4.3 Decreased length of esterification

Prior to IMAC, our protocol calls for a Fischer esterification. This has previously been performed by adding anhydrous 3M HCl in methanol, reacting for one hour, drying (~1 hour), and repeating. However, Fischer esterification should be a fast reaction, especially when the acid and alcohol are present in such excess. I found that reacting for

15 minutes instead of one hour reaches the same level of derivatization (~99%) and shortens the drying step to ~20 minutes. This is likely due to the generation of water as a side product of the reaction, as it is less volatile and takes longer to evaporate than the methanolic HCl. The removal of water more rapidly also reduced the generation of side reactions, such as deamidation of asparagine and glutamine, as well as the possibility of reversing reactions. This brings the esterification protocol from 4 hours to ~1.5 hours without sacrificing completeness.

3.4.4 Decreasing length of enrichment protocol

Table 3.2 Comparison of IMAC enrichment protocol times in minutes

Enrichment Step	Old	New
Removal of Nickel	40	0
Conjugation with Iron	25	5
Rinse	50	20
Sample load	60-100+	
Rinses	90	50
Elution	25	10
Rinse	10	5
Total	**300**	**145**

Using the previous protocol, IMAC resin was activated on the same day as the enrichment. The NTA resin can be activated before the enrichment and appears to retain its activity for at least one month. This reduces the amount of time needed overall, as the activated resin can be used for multiple enrichments. But more importantly, it reduces the amount of time required on the day of enrichment. In theory, the previous enrichment protocol takes five hours to complete (**Table 3.2**), followed by rinsing the precolumn on the HPLC and LC-MS analysis. Therefore, the day of enrichment is guaranteed to be at least seven hours. However, as discussed above, this is rarely the case and various issues often arise such as the sample loading slowly, loading too quickly and needing be reloaded,

or having to pack multiple columns prior to loading the sample. Therefore, enrichment days can easily become 10- or 12-hour days with no warning, and the protocol does not have an ideal stopping point to resume the following day.

Due to the stronger affinity of iron for NTA, I theorized that the resin would retain iron even with an increased flow rate. Previously, most steps were carried out at a flow rate of 0.5 $\mu Lmin^{-1}$. I increased the flow rate prior to sample loading to 2 $\mu Lmin^{-1}$ and after sample loading to 1 $\mu Lmin^{-1}$. However, for optimal enrichment the sample needs to be exposed to the resin for a sufficient time for association, so I used the same flow rate during sample loading. This change to the protocol showed no noticeable changes in recovery of phosphopeptide standards compared to the previous protocol.

Between the increased flow rates and bulk activation prior to the day of enrichment, this protocol theoretically takes 2.5 hours to complete (**Table 3.2**). With the addition of the LC-MS analysis, it is often possible to finish in just 4.5 hours. Since the Benzonase prevents most sample-related clogs, sample loading times tend to be more predictable, so much of the variability between samples is removed as the flow rate remains constant rather than having to be adjusted to address slowed or increased flow as the column develops and releases clogs. Overall, these improvements have led to more straightforward and reproducible phosphopeptide enrichment.

3.5 Conclusions and Future Work

Despite the greater sensitivity of the Hunt laboratory IMAC enrichment protocol, it is a lengthy and temperamental process. However, by altering the resin used for iron immobilization we were able to obtain even higher sensitivity while altering the lysis and flow rates to decrease the procedure's variability and length. This not only improved reproducibility through eliminating contaminants, but also allowed a simpler protocol with fewer steps and active user time.

The improvements to the protocol are significant. Using the previous method, the maximum number of phosphopeptides identified in a single sample was 234 (LL4857T using IDA and NTA). Currently, the maximum phosphopeptide identifications in a single IMAC enrichment is 522 (incorporating the changes discussed here as well as those in Chapter 4) using 100 million cells and 273 (incorporating only these changes) using 300 million cells. Using some or all of the improvements discussed here, the median number of phosphopeptides identified per sample was 52 compared with 25 prior to these improvements.

3.6 References

1. Ardito F, Giuliani M, Perrone D, Troiano G, & Muzio LL (2017) The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy. *International journal of molecular medicine* 40(2):271-280.
2. Cohen P (2002) The origins of protein phosphorylation. *Nature Cell Biology* 4(5):E127-E130.
3. Cohen P (2001) The role of protein phosphorylation in human health and disease. *The FEBS Journal* 268(19):5001-5010.
4. Engelhard VH, Altrich-Vanlith M, Ostankovitch M, & Zarling AL (2006) Post-translational modifications of naturally processed MHC-binding epitopes. *Current opinion in immunology* 18(1):92-97.
5. Leitner A (2010) Phosphopeptide enrichment using metal oxide affinity chromatography. *TrAC Trends in Analytical Chemistry* 29(2):177-185.
6. Abelin JG, Trantham PD, Penny SA, Patterson AM, Ward ST, Hildebrand WH, . . . Hunt DF (2015) Complementary IMAC enrichment methods for HLA-associated phosphopeptide identification by mass spectrometry. *Nature protocols* 10(9):1308-1318.
7. Holmes LD & Schiller MR (1997) Immobilized iron (III) metal affinity chromatography for the separation of phosphorylated macromolecules: Ligands and applications. *Journal of liquid chromatography & related technologies* 20(1):123-142.
8. Nilsson CL (2012) Advances in quantitative phosphoproteomics. *Analytical chemistry* 84(2):735-746.

9. Dunn JD, Reid GE, & Bruening ML (2010) Techniques for phosphopeptide enrichment prior to analysis by mass spectrometry. *Mass spectrometry reviews* 29(1):29-54.

10. Martell AE & Smith RM (1974) *Critical stability constants* (Springer).

11. Myers PT (2016) Enrichment and Identification of HLA-Associated Phosphopeptides for the Development of Novel Cancer Immunotherapeutics. PhD (University of Virginia).

Chapter 4. Improved MHC class I peptide identification using offline hydrophilic interaction spin tips

4.1 Introduction

The major histocompatibility complex (MHC) class I presentation pathway is used by all nucleated cells to communicate their health status to CD8+ T cells.(1) Intracellular proteins are degraded and presented on MHC molecules, allowing T cells to recognize peptides that are not presented by healthy cells. These abnormal peptides can come from proteins that are not normally in the human proteome (such as peptides from bacterial or viral proteins), mutated human proteins, or abnormally modified proteins. Since a major percentage of peptides are the same between healthy and diseased cells, identification of these antigenic peptides can be problematic.

To identify these peptides, MHC molecules are immunopurified from a cell lysate, acid eluted from MHC antibody-associated beads, and prepared for analysis by mass spectrometry.(2) An important step of this preparation is a "cleaning" procedure to remove non-peptidic impurities. These contaminants can include any number of species: proteins, lipids, salts, DNA, detergents, and countless more. Such impurities can cause a variety of issues which may interfere with the relevant sample analysis but can also ruin an analytical column or subsequent samples.

4.1.1 Expected contaminants in MHC samples

When MHC peptides are eluted from immunoaffinity beads, the elution will also, at minimum, elute the MHC molecules as well, which are comprised of HLA alpha chains

(~50 kDa) and beta-2-microglobulin (β2m, 12 kDa). While our lab now prefers covalently conjugating antibodies to beads, beads with immobilized Protein A are more commonly used for antibody immobilization. In those protocols, the primary antibody will elute from the beads alongside MHC molecules and peptides. While these contaminants are unavoidable, many more contaminants inevitably make it into the elution. The lysate also contains protease (and sometimes phosphatase) inhibitors, which are a mix of small molecules and proteins. These should be removed during rinses after incubation but are difficult or impossible to remove completely through rinsing. Nucleic acids, lipids, and abundant cellular proteins can also be present in the eluate. Most salts used in lysis buffers are nonvolatile and need to be removed before sample analysis. Additionally, some level of polymer contamination is inevitable when performing a multi-step process.

4.1.2 Contamination impacts on sample analysis

Contaminants are an important consideration in a chromatographic separation because their presence will affect the ability to detect low abundance peptides. As with any analytical technique, the main way that contaminants interfere with analysis is by reducing the ratio of signal to noise.

Lowering peak abundance can interfere with identification in several ways. First, if peptide abundance is significantly lower than other co-eluting peptides, it is possible that it will not be selected for fragmentation via data-dependent acquisition (DDA), since DDA methods select species in order from highest to lowest abundance. Alternatively, a weakly abundant peptide that is selected for fragmentation may not have enough signal intensity to generate useful MS2 spectra; here, the fragments are lost to noise or are not intense enough to detect.

Increasing the number of ions eluting from the column at once increases the complexity of the MS1 spectrum. This can also cause issues with DDA, as there are more ions to select for fragmentation and lower-level ions may not be selected for fragmentation. Additionally, having more ions eluting at once increases the chance of multiple species being co-isolated for fragmentation. Co-isolation will generate fragment ions from both species. These extra ions can produce ambiguous or false assignments, either by adding in new ions or incompletely resolving from the peptide fragment ions. Secondly, a portion of the ion target is taken by the contaminant, lowering the ion current in the spectrum that is attributable to the ion of interest, thus weakening and reducing the quality of the fragment peaks.

4.1.3 Contamination by hydrophobic species

Contaminants that elute after the gradient has completed can cause a variety of issues.(3) Larger contaminants, like proteins, lipids, and undigested DNA often fall into this category. MHC-peptide analyses are overwhelmingly performed using C18 material with 12 nm pores, because this material provides ideal retention and resolution for peptides but can encounter problems when larger molecules (i.e. intact proteins) are contaminating a sample. The pore size is ideal for maximum resolution of MHC peptides, which are average approximately 1 kDa for class I, but can be up to 5 kDa for class II. Larger molecules (> 10 kDa) understandably require larger (20 or 30 nm) pores for proper separation, but some will still be able to elute through 12 nm material (albeit slowly and with poor chromatography). However, many proteins are close to or larger than the pore sizes, preventing them from traveling through pores at all. Proteins also require less hydrophobic stationary phases such as C4, C8, or PRP materials, to elute properly. Intact

protein contaminants can sometimes be removed from a column using a high percentage of organic solvents, but larger or more hydrophobic proteins can associate with the C18 material so strongly that they cannot be removed.

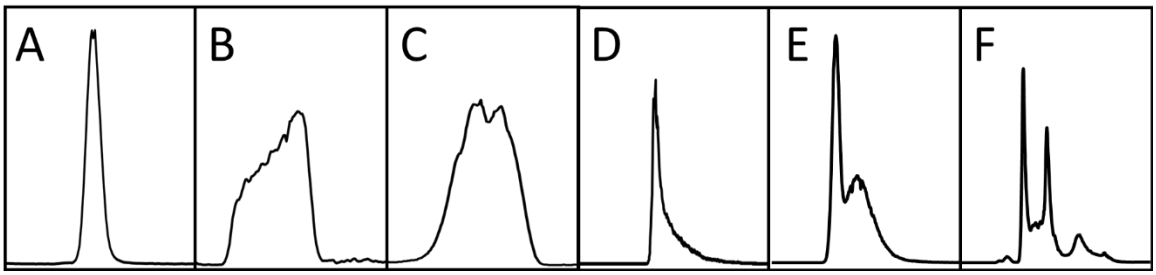


Figure 4.1 Chromatographic peak shapes
While ideally peaks elute as a sharp gaussian shape (A), contamination can lead to a loss of symmetry at the front of the peak (B) as they are pushed forward by overloading, peak broadening (C) caused by different interactions. If the analyte interacts with something other than the packing material, this can cause loss of symmetry towards the back of the peak (D). As tailing gets more severe (E), it can progress to forming multiple peaks (F).

For in-line separations, altered peak shapes can result from contamination, as seen in **Figure 4.1**. As the column approaches capacity, the earliest-eluting peptides are the first to lose symmetry. Here, peak symmetry is lost as the peptide is pushed towards the start of the gradient (fronting**, B**).(3) If the peptide is sufficiently hydrophilic or the contamination is significant, these peptides can elute during equilibration and no longer be observed during the gradient (bumping). In severe cases, peptides eluting late in the gradient can be observed moving earlier in retention time. Finally, analytes can associate with hydrophobic contaminants on the column, causing additional retention that is not based on the column media. This can cause broad peaks (**C**) and/or a loss of symmetry at the back of the peak (tailing, **D**). Significant tailing (**E**) can progress to double peaking (**F**). Regardless of how the peak loses symmetry, in all cases abnormal chromatography leads to wider, less abundant peaks. This in turn increases the complexity of MS1 spectra, making it less likely for a given analyte to be chosen.

Depending on the sample size, frequency of use, and cleaning regimen of a column, these contaminants are likely to build up over time. Irreversible accumulation of contaminants, also known as “column fouling”, leads to a variety of chromatographic issues.(3) These issues are caused by interaction of the analytes of interest with accumulated contaminants so that separation is no longer solely dependent on the separation media. Mild cases, such as an overloaded sample or moderate accumulation, may result in some peptides losing symmetry in the back of the peak (tailing) due to interaction with impurities. However, as contaminants accumulate chromatographic abnormalities become more varied, including altered retention times, asymmetrical peak shape, and sometimes totally preventing elution. This in turn leads to a lack of reproducibility, which prevents comparison across files.

In addition to accumulation of contaminants on a column over time, contaminants can instantly clog a column. Sometimes, cell debris from lysis may be present in the final sample. However, this is avoidable by putting the lysate supernatant through a 0.2 μm filter before adding it to the beads. Another cause of instantaneous clogging is differential solubility of molecules in lysis versus loading buffers. Solubility could be altered for a few reasons: the lysis buffer has higher salt content and is ~pH 8 instead of ~pH 3. However, the most significant difference in this case is likely the removal of detergent. Detergent allows more hydrophobic molecules to solubilize in aqueous solution. An obvious example would be lipids, but a more germane example is membrane proteins. Since HLA molecules have low solubility and are guaranteed to be in any successful MHC pulldown, they are likely culprits for clogging columns. While it may be possible to remove completely insoluble compounds from the sample by centrifugation, if weakly soluble compounds

build up on an HPLC column, they can precipitate and block flow. Column clogs are frequently irreversible, preventing sample recovery, but they can also overpressure the system or eject a column from the HPLC.

Removing contaminants that will clog columns or interfere with chromatography is necessary for reproducible and comparable results. Therefore, MHC isolation protocols overwhelmingly use C18 material for an offline cleanup prior to inline analytical separation. However, while column clogging or fouling are legitimate concerns that are addressed with the current methodology, issues caused by other contaminants, such as polyethylene glycol (PEG) polymer, NP-40, Triton-X, etc. are not addressed with the current protocols. Using the same material for preparation and analysis can only remove contaminants that elute outside of the gradient, ignoring the effects of contaminants that coelute with analyte ions.

4.1.4 Contaminants that coelute with analytes

Contaminants that elute during the gradient mainly interfere with ionization of the analytes rather than chromatography. The most recognizable way that contaminants can decrease analyte signal is through ion suppression. Ion suppression occurs during the electrospray process when protons are distributed between species entering the gas phase. However, when the number of potential protonation sites in a droplet exceed the number of protons available, only a subset of the sites will be protonated. When this occurs, low abundance peptides can easily drop below the limit of detection, and the charge state can decrease for even abundant peptides. It is possible for this to take place between different analytes in a sample when a very abundant peptide elutes or if there are several peptides eluting simultaneously, but it is more commonly seen when a contaminant is present. Even

contaminants that have low proton affinity can suppress signal if they are abundant enough. In **Figure 4.2**, the extracted ion chromatograms of two HLA-A*01 phosphopeptides shows a low-abundance (attomole level) peptide, SELRsPRISY, begin to elute from the column. When the much more abundant ITQGtPLKY begins to elute, SELRsPRISY drops below the limit of detection. As the abundance of the larger peak decreases, signal from the less abundant peptide returns. However, if the suppressing peak(s) elute the entire time that the lower-level peptide is eluting, there is way to know what is being suppressed since it cannot be identified. While suppression between analytes can be problematic, the more common issue is suppression by contaminants.

Suppressing contaminants often dominate the MS1 spectrum. Small molecules (>300 Da) and polymers are frequently seen as a large ion with an individual charge as the most abundant ion in the MS1. Larger contaminants such as proteins can be recognized by a marked increase in the total ion current without an increase in the base peak abundance.

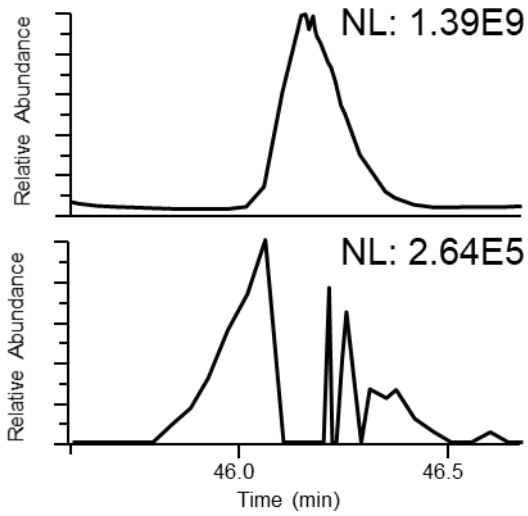


Figure 4.2 Ion suppression of chromatographic peaks
Extracted ion chromatogram showing the loss of signal from ion suppression of a lower-abundance peptide (bottom) when an abundant peptide (top) begins to elute. After the more abundant peak elutes, ion signal returns for the less abundant peak

Regardless of the contaminant type, all this requires is that the coeluting contaminant has a combination of proton affinity and abundance that allows it to commandeer protons that would otherwise be distributed to the analytes during electrospray ionization.

Interfering with ionization is not the only way that ion abundance can be altered by coeluting contaminants. Salts can form adducts with the analytes, splitting the signal between several peaks, which both decreases analyte signal and increased complexity. Additionally, contaminants with multiple charges (z >=2) can be selected for fragmentation before analyte ions, preventing identification of the analyte without affecting analyte abundance. This is particularly true of contaminants that have similar properties to the analytes of interest (i.e. charge and molecular weight), so even limited selection criteria cannot fix the problem. Another way that identifications can be hindered is when contaminants have similar mass to charge ratios as an analyte ion, causing them to be collected alongside the target ion for fragmentation, thus complicating the fragmentation spectrum.

4.1.5 Current methods for contaminant removal from MHC samples

Despite the wide range of effects that can be caused by contamination, only a few are addressed by current protocols in the field. Commonly, C18 cleanups will remove a specific set of contaminants that can cause chromatographic problems.(4) Theoretically, it is possible to perform multiple clean-up steps on a particular sample; however, each cleanup will cause substantial sample loss. Additionally, each cleanup will remove contaminants, but will also inevitably add some new contaminants (e.g. polymer, keratin). For these reasons, clean ups are usually limited to 1-2 rounds. By far the most common is a reverse phase cleanup using microscale solid phase extraction (SPE) devices like C18

STAGE tips, ZipTips, or SepPaks. This step is sometimes preceded by a molecular weight cutoff (MWCO) filter. Using a MWCO removes many problematic elements, most notably HLA molecules. However, while it works well for large quantities of material, it is known to cause major losses when working with smaller samples. Additionally, the filters tend to add polymer to samples, which would again be more noticeable with small samples.

4.1.5.1 STAGE tip extraction

Many reverse phase techniques are used for cleanup in immunopeptidomics. However, most offline separations are intended for large amounts of material and can produce significant losses for less abundant, more precious samples. But because these columns are often commercially available, these cleanup columns are commonly used, even despite being ill suited for immunopeptidomics. To address the issues involved with handling smaller sample concentrations (i.e. sub microgram amounts), stop and go extraction (or STAGE) tips were recently developed. For STAGE tip cleanups, a piece of a solid phase extraction disk is forced into the bottom of a pipette tip, liquids are loaded on top of the disk, and centrifuged through. After activation, samples can be loaded on to the disk, rinsed, and eluted using appropriate solvents. As with other reverse phase methods, these STAGE tips associate better with hydrophobic moieties. Therefore, proteins are eluted by increasing the organic concentration of the eluent using 30-50% acetonitrile.

I observed significant losses of hydrophilic peptides using this technique. Class I peptides tend to elute relatively early in gradients due to their small size. HLA type and modifications of interest can also affect their hydrophilicity. I began working towards changing the standard STAGE tip protocol to minimize these losses: limiting rinse volume, changing solvent pH and ionic strength, and the speed of loading/eluting samples. While

some of these had minor effects, losses were still significant. Since sample sizes are generally limited for this work and it is uncommon that there is enough sample available for multiple analyses, these losses can majorly affect the quality of results. Given these issues and the limited sample size, I began to investigate other options for sample cleanup

4.1.6 Hydrophilic interaction liquid chromatography

Unlike reversed-phase chromatography, hydrophilic interaction liquid chromatography (HILIC) retains peptides based on the number of hydrophilic residues or modifications (**Figure 4.3)**. Unlike normal phase chromatography, HILIC uses aqueous-miscible solvents. Since it uses similar solvents to reverse phase, but with the opposite gradient, it is sometimes referred to as "reverse reversed phase". HILIC associates with

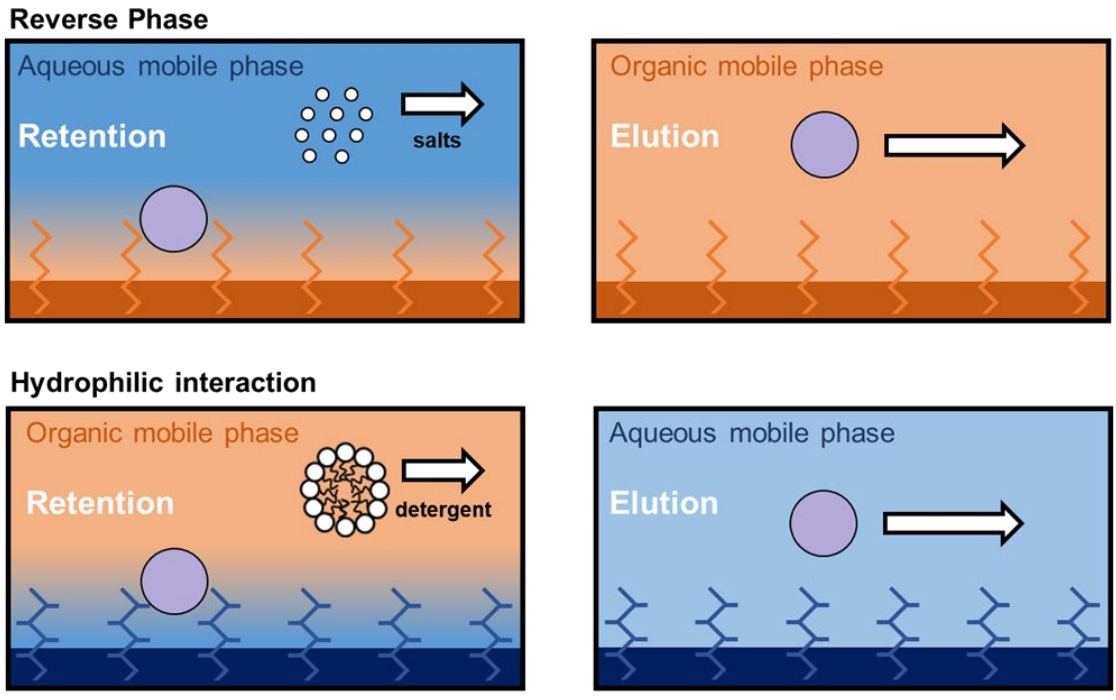


Figure 4.3 Retention of compounds on reverse phase and hydrophilic interaction materials
On reverse phase materials, hydrophobic species are partitioned into a hydrophobic stationary phase while leading sample in an aqueous mobile phase. Elution is then enacted by increasing the organic content of the mobile phase. In HILIC, species partition into a hydrophilic stationary phase and are eluted with an aqueous mobile phase.

hydrophilic residues while reverse phase associates with hydrophobic residues. Since a peptide can contain both hydrophilic and hydrophobic regions, HILIC provides an orthogonal separation to C18 reversed phase techniques. Therefore, using HILIC for a cleanup procedure allows for removal of some contaminants that coelute with analyte ions.

The use of HILIC for the separation of peptides was introduced in 1990 by Andrew Alpert using a polyhydroxy ethyl aspartamide (PHEA, Poly LC) based separation media (**Figure 4.4**).(5) The material was frequently used in the early 1990s but was quickly replaced by other materials for in-line separations using mass spectrometry. Two main problems are frequently listed as a reason for using other materials: 1) a lack of reproducibility as the column ages, and 2) poor retention and peak shapes with buffer systems and concentrations that are MS-compatible.(6, 7) However, these problems are circumvented if the separation is performed offline prior to in-line separation with a reversed phase analytical column. Microscale cleanup devices are usually not used multiple times, so there is no concern about column aging. Additionally, volatile salts provide adequate retention and separation on PHEA when used at higher concentrations that would be incompatible with electrospray ionization.(8) When performing an offline separation, these salts can be removed using a vacuum concentrator prior to reconstitution and analysis. Additionally, peak width and symmetry is not important for offline preparative techniques.

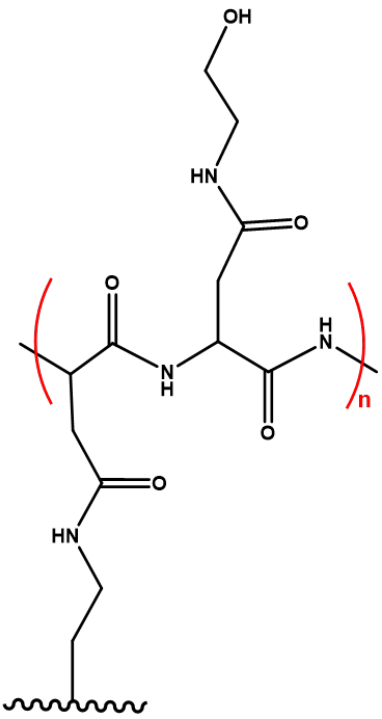


Figure 4.4 Structure of polyhydroxy ethyl aspartamide stationary phase

To remove contaminants while retaining hydrophilic peptides, I designed a spin-tip HILIC cleanup method using PHEA beads packed on a glass fiber filter. Unlike current cleanup methods in the field, this material can remove polymers and detergents from the sample, allowing better ionization and analysis of antigenic peptides. Proteins and peptides are eluted using separate solvents, allowing proteins to be removed from the sample but still collected. The protocol takes approximately 5-10 minutes, costs less than other methods, and shows significant improvements over the C18 cleanup method in both analyte retention and contaminant removal.

4.2 Materials

4.2.1 Reagents

Alfa Aesar (Haverhill, MA)

- Ammonium formate, >98%

Bio-Rad Labs (Hercules, CA)

- Tris base >99.8% pure

EMD Millipore (Burlington, MA)

- Benzonase® Nuclease, Purity >90%
- N-octyl-β-D-glucopyranoside

Fisher Chemical (Ward Hill, NJ)

- Ammonium Acetate (Crystalline/Certified ACS)
- Methanol, Optima™ LC/MS grade

Honeywell (Charlotte, NC)

- Acetonitrile, B&J Brand™ LC-MS, for LC-MS and HPLC, >99.9%

Oxford Glyco Systems (Oxfordshire, England)

- Tris HCl, Electrophoresis grade

PQ Corporation (Malvern, PA)

- KaSil 1624 potassium silicate solution

Roche (Basel, Switzerland)

- cOmplete™, Mini Protease inhibitor tablet, EDTA-free

Sigma Aldrich (St Louis, MO)

- Acetic acid, glacial, ≥99.99% trace metals basis

- Aprotinin
- L-Ascorbic Acid, ACS reagent, ≥99%
- Azulene
- Formamide, ≥99.5% (GC), BioReagent for molecular biology
- Hydrochloric Acid, ACS reagent, 37%
- Magnesium Chloride hexahydrate
- Pepstatin A
- Phenylmethylsulfonyl fluoride (PMSF)
- PhosSTOP™ phosphatase inhibitor tablets
- Sodium Chloride, ACS reagent, ≥99.0%
- Sodium hydroxide, reagent grade, 97%, powder
- Dulbecco's Phosphate Buffered Saline, without calcium chloride and magnesium chloride, liquid, sterile filtered
- Triton X-100

Thermo Scientific (Waltham, MA)

- Pierce™ Formic Acid, LC-MS grade
- Halt™ Protease and phosphatase inhibitor cocktail (100x)
- Pierce™ Water, LC-MS grade

4.2.2 Proteins and Peptides

Anaspec (Fremont, CA)

- Angiotensin II Phosphate >95% HPLC
- Vasoactive Intestinal Peptide (VIP) 1-12 human, porcine, rat >95% HPLC

Biolegend (San Diego, CA)

- Ultra-LEAF™ Purified anti-human HLA-A,B,C Antibody

Sigma Aldrich (St Louis, MO)

- Angiotensin I human acetate salt hydrate >90% HPLC

Thermo Scientific (Waltham, MA)

- Pierce™ Peptide Retention Time Calibration Mixture
- Pierce™ HeLa Protein Digest Standard

4.2.2.1 Chromatography Resins

Cytiva Life Sciences (Marlborough, MA)/GE Healthcare (Boston, MA)

- NHS-Activated Sepharose 4 Fast Flow

Dr Maisch GMBH (Ammerbuch, Germany)

- Reprosil Pur 120 C18 AQ 1.9 μm
- Reprosil Pur 120 C18 AQ 3 μm
- Reprosil Pur 120 C18 AQ 10 μm

PolyLC (Colombia, MD)

- Polyhydroxy ethyl aspartamide, 12 μm, 200 Å
- Polyhydroxy ethyl aspartamide, 20 μm, 200 Å

4.2.3 Equipment and Instrumentation

Agilent Technologies (Palo Alto, CA)

- 1100 Agilent high performance liquid chromatograph

Boekel Scientific (Feasterville-Trevose, PA)

- Digital Incubator 133000

Branson (Danbury, CT)

- Branson 1200 Ultrasonic Bath

Cadence Science (Cranston, RI)

- Blunt end pipetting needle (16G)

Eppendorf (Enfield, CT/Hamburg, Germany)

- Model 5424 Centrifuge
- Protein LoBind Microcentrifuge tubes (1.5 mL)

Fisherbrand (Pittsburgh, PA)

- Model 50 Sonic Dismembrator with CL-18 Probe
- Mini Tube Rotator
- Digital Vortex
- Vortex Genie 2

GL Sciences (Tokyo, Japan)

- Centrifuge adaptor 5010-21514

Labconco Corp. (Kansas City, MO)

- Centrivap centrifugal vacuum concentrator 7810016

New Objective (Littleton, MA)

- PicoClear 360 Union

PolyMicro Technologies, Inc. (Phoenix, AZ)

- 360 µm o.d. x 75 µm i.d. polyimide coated fused silica capillary
- 360 µm o.d. x 100 µm i.d. polyimide coated fused silica capillary

Supelco (Bellefonte, PA)

- Screw top clear glass vials with green thermoset caps and PTFE liners, 2 mL

Sutter Instrument Co. (Novato, CA)

- P-2000 microcapillary laser puller with fused silica adapter

Thermo-Fisher Scientific (San Jose, CA/Bremen, Germany)

- Easy-nLC 1200 liquid chromatograph
- Orbitrap Fusion Lumos Tribrid Mass Spectrometer with commercial front-end ETD
- Orbitrap Fusion Tribrid mass spectrometer with commercial front-end ETD
- LTQ mass spectrometer with commercial back-end ETD
- LTQ FT-ICR hybrid mass spectrometer (custom modified with front-end ETD)
- LTQ-Orbitrap mass spectrometer (custom modified with front-end ETD)

Whatman (Buckinghamshire, United Kingdom)

- Glass microfiber filters grade GF/F

Zeus Industrial Products, Inc. (Orangeburg, SC)

- Teflon tubing, 0.012-inch i.d. x 0.060-inch o.d.

4.3 Methods

4.3.1 Tissue collection

Samples were collected by Dr. Jennifer Sokolowski from cadavers at the University of Virginia Medical Center.

4.3.2 Immunopurification

MHC-bound peptides were isolated as described previously, described here in brief.(9) First, the pan-human class I antibody W6/32 was conjugated to NHS-Sepharose beads. 3 mg of antibody is prepared per $1x10^9$ cells or 1 g of tissue, whichever was greater. 100 µL of beads were used per 1 mg of input antibody. Beads were washed twice in PBS and incubated rotating overnight with antibody. The following day beads were pelleted, blocked for 1 hour with 100 mM Tris-HCl, then washed twice in alternating solutions of 100 mM ammonium acetate pH 5 with 500 mM NaCl and 100 mM Tris buffer pH 8. Beads were resuspended in 20 mM Tris, 150 mM NaCl up to 1 mg/ml remaining antibody and stored at 4°C until use.

Cells were lysed in a buffer of 20 mM Tris-HCl, 150 mM NaCl, 1% CHAPS, pH 8 supplemented with protease and phosphatase inhibitors. Buffer was added to the cells at 5 ml per $1x10^9$ cells or 1 gram of tissue. Cells are lysed from frozen, rotating in buffer at 4°C for 1-2 hours. Lysed cells were ultracentrifuged at 36,800 rpm for 1 hour at 4°C, and supernatants are then mixed with the antibody-bead conjugates prepared earlier and rotated at 4°C overnight to allow MHC binding.

The following day, pelleted beads were washed in 10 mL lysis buffer, resuspended in ~500 µL 20 mM Tris-HCL, and transferred to a microcentrifuge tube or Poly-prop

column. Being kept at 4°C, beads are then washed with: 2x 20 mM Tris-HCL, 150 mM NaCl; 2x 20 mM Tris-HCl, 1 M NaCl; 3x 20 mM Tris-HCl.

If the sample was prepared at the University of Virginia, it was then kept on ice for transport, then processing continued immediately. Otherwise, the washed beads were then spun through a pre-wet 3 kDa Amicon Ultra centrifugal filter column and all liquid removed by centrifugation. Columns were covered with Parafilm and stored at -80°C for shipment.

4.3.2.1 VMM39, AVL3, and JY-A2 (performed by Kara Cummings)

Samples were immunopurified as discussed previously, but after rinses, the beads were transferred to a 5 kDa MWCO filter and eluted from the HLA class I molecules with 10% acetic acid. Extracted peptides were stored at -80ºC before shipment to the University of Virginia for analysis.

4.3.3 HLA-associated peptide elution

If elution was not performed by the collaborator, thawed beads were transferred from the filter to a low-protein binding tube using two transfers in 200 µL of water. After centrifugation, the supernatant was removed and set aside. To elute the peptides from HLA molecules bound to beads, 150 µL of 10% acetic acid was added to the tube and shaken for 5 minutes at room temperature. The beads were centrifuged at 1000 *x g* for 1 minute and the supernatant transferred to a low protein binding tube. This process was repeated to ensure complete elution of peptides from HLA molecules and the supernatant added to previous elution. Two internal phosphopeptide standards were spiked into the 10% acetic acid elution to determine recovery from the cleanup process.

This elution was dried using a vacuum concentrator and reconstituted at 1E7 ceq per µL in 0.1 % acetic acid. The desired sample amount for the cleanup was removed and dried in separate tubes prior to the cleanups.

4.3.4 Reversed phase SPE cleanup

STAGE tips were fabricated with two cores as previously described.(10) STAGE tips were equilibrated using the following wash steps: two washes of 100 µL of methanol, one wash of 100 µL 80% acetonitrile/0.1% acetic acid at 1500 *x g*, and two washes of 100 µL 1% acetic acid at 3500 *x g*.

The dried sample was reconstituted in 100 µL 0.1% acetic acid. This was loaded onto the STAGE tip, spinning at 3500 *x g* until the entire volume had passed through. The tubes were rinsed once using 100 µL of 0.1% acetic acid and the rinse was put through the STAGE tip. The tip was then rinsed once using 100 µL 1% acetic acid. Peptides were eluted from the STAGE tips using a stepped acetonitrile gradient: 20 µL of 20% acetonitrile/0.1% acetic acid, 20 µL of 40% acetonitrile/0.1% acetic acid, and 20 µL of 60% acetonitrile/0.1% acetic acid. Eluted peptide fractions were dried to completion using a Centrivap.

4.3.5 HILIC cleanup

Spin tips were fabricated by using a 16-gauge blunt tip needle to cut a core from a Whatman glass fiber filter and blowing the filter into the tip of a 200 µL pipette tip. A slurry was then made, using approximately 10 µL of PHEA material (12 µm or 20 µm diameter, 200 Å pores) and 200 µL of 200 mM ammonium acetate (pH6) per tip. The material was added to the tip and centrifuged at 350 *x g* for one minute. The 10 µL of material created a packed bed length of approximately 5 mm over the glass fiber filter. The material was then rinsed using 100 µL of the following solvents and spun for 1-2 minutes

at 300-400 *x g*: 0.5% formic acid, 90% acetonitrile 20mM ammonium formate (pH 3), 200 mM ammonium formate (pH3), and water. The columns were then equilibrated twice using 90% acetonitrile with 20mM ammonium formate. Dried samples were reconstituted in 10 µL of 200 mM ammonium formate before adding 90 µL of acetonitrile and using a vortex mixer. The reconstituted sample was then added to the tip. Then 100 µL of 90% acetonitrile 20 mM ammonium formate was used to rinse the sample tube and added to the column to rinse. An additional 100 µL can be added to the column to further rinse polymeric contaminants without major sample losses. Peptides were then eluted by adding 100 µL of 50% acetonitrile in 0.2% acetic acid to the column twice. Proteins can optionally be eluted for further analysis using 200 µL of 0.5% formic acid. After cleanup, samples were dried down and reconstituted in 0.25 µL glacial acetic acid and diluted with 15 µL LCMS grade water.

4.3.6 LCMS analysis

JYA2 and HeLa were separated using an Easy nLC 1200 with a 100-minute gradient from 0-40% acetonitrile with 0.1% formic acid. The analytical column was packed to 18 cm with 1.9 µm Dr Maisch C18 material, and the precolumn had 8 cm of 10 µm Dr. Maisch C18. Using the autosampler, 18 µL was injected on to the precolumn and rinsed with 20 µL 0.1% formic acid before beginning the gradient.

Otherwise, samples were separated in-line using an Agilent 1100 LC system with a 60-minute gradient from 0-40% acetonitrile with 100 mM acetic acid. Analytical columns were packed with 12 cm of 3 µm Dr. Maisch C18 material and precolumns were packed with 8 cm of 10 µm Dr. Maisch C18 material. Samples were loaded onto the precolumn using a pressure vessel at a flow rate of 1 $\mu L min^{-1}$ before desalting on the HPLC at 30-40

bar for 5-10 minutes. After rinsing, 100 fmol of angiotensin and VIP and/or 10 fmol of PRTC were loaded onto the column as internal standards. The precolumn was then rinsed for an additional 30 seconds before reconnection to the analytical column. Once connected, the column was allowed to equilibrate for 15 minutes at 45 bar (approximately 150 $nLmin^{-1}$) before beginning data collection. Once stable spray was established, we began recording scans and started the gradient.

Analyses of THP1 and VMM39 were performed on an LTQ Orbitrap XL and LTQ FT Ultra respectively, both equipped with front-end ETD sources. High resolution MS1s were taken at 60,000 resolution with a scan range from 300-1500 *m/z*. Data were collected in a data dependent manner, with the 5 most abundant ions selected for fragmentation with an isolation window of 3 Da. After being selected twice for fragmentation in 10 seconds, the masses were put on an exclusion list for ten seconds. Only peptides with a charge of 2 or 3 were selected. For each precursor selected for fragmentation, low resolution MS2s were collected at a normal scan rate. A pair of MS2s were collected for each precursor: CAD fragmentation at 35% nCE and ETD fragmentation with a reaction time of 50 ms.

Analyses of HeLa and JYA2 were performed using an Orbitrap Fusion Lumos Tribrid mass spectrometer. All other samples were analyzed using the Orbitrap Fusion Tribrid mass spectrometer. MS1 scans were collected in the Orbitrap at a resolution of 60,000 and a scan range of 300-1500 *m/z*. Data were collected in a data-dependent manner, with the precursors selected for fragmentation in order of decreasing intensity using a top speed method with a 3s cycle. Precursors were isolated by quadrupole isolation with a window of 1.8 Da. Singly charged precursors were isolated with an offset of 0.4 Da. Based on charge state, precursors were subjected to any of three fragmentation types: (1)

Fragmentation with CAD using 30% nCE collected at 7500 resolution with an inject target of 100,000 ions and maximum inject time of 120 ms, (2) fragmentation with HCD using 27% nCE but otherwise the same as CAD, and/or (3) fragmentation with ETD collected at a rapid scan rate in the ion trap with a target of 30,000 ions with 50 ms maximum inject time. Singly charged precursors were only selected if their precursor mass was greater than 850 *m/z* and were subjected to HCD fragmentation. Doubly charged precursors between 350 and 950 *m/z* were subjected to CAD fragmentation and ETD fragmentation with 60 ms reaction time. Precursors with charge states of 3 or 4 were selected between 300 and 650 *m/z* and underwent HCD fragmentation as well as ETD fragmentation with a charge-calibrated reaction time.

4.3.7 Data analysis

Data were searched using the Byonic node of Proteome Discoverer 3.3 against the human proteome. In most cases, to reduce false positives only variable modifications of common chemical modifications such as oxidation, deamidation, and cysteinylation were allowed. If additional variable mods of post-translational modifications were searched (e.g. dimethylated arginine, phosphorylation, glycosylation), they were validated manually and not included in the overall reported identifications.

Identifications were considered confident if they had a Byonic score greater than 300, |logProb| greater than 3, and Proteome Discoverer considered the identification to be high confidence. Peptide quantification and chromatographic alignment were performed by Proteome Discoverer.

4.4 Results and Discussion

HLA molecules have different affinities for peptides based on the conformation of their binding groove (**Table 4.1**). Usually, the peptides associate with the alpha chain at residue 2 or 3 and the final residue of the peptide. Based on the identity of these amino acids, it is sometimes possible to determine the HLA type of the sample. Even outside of the anchor residues, MHC molecules have preferred residues in the surrounding areas. Since HLA class I peptides are usually under 12 residues, peptides that bind a certain HLA type have similar properties, including hydrophobicity. This can cause certain HLA types to have noticeably different elution profiles.

Researchers in the United States generally focus on HLA-A*02:01 as an initial target for development of immunotherapeutics due to its prevalence in the United States. Therefore, most of these cleanup techniques are primarily tested on HLA-A*02 positive samples. As seen in **Figure 4.5A**, HLA-A*02 is markedly hydrophobic compared to most

Table 4.1 Common HLA types and anchor residues
Most common HLA types in the United States, as well as their prevalence when adjusted for the world population. Peptides bind at position 2 or 3 and the final residue of the peptide sequence and have preferred amino acids at those positions.

			Position		
Allele	**USA**	**World**	**2**	**3**	**9**
A*01:01	20%	10%		DE	Y
A*02:01	39%	22%	LM		VL
A*03:01	20%	8%	LVM		KYF
A*11:01	14%	27%	(VIFY)	(MLFYIA)	KR
A*24:02	14%	25%	YF		ILF
B*07:02	16%	5%	P		L
B*08:01	14%	3%		K	L
B*35:01	11%	8%	P		YFMLI
B*40:01	9%	9%	E		L
B*44:02	14%	<1%	E		FY
C*03:04	16%	15%	A		LM
C*04:01	23%	15%	(YFW)	D	FM
C*05:01	13%	1%		D	MFLIV
C*06:02	16%	9%	R		VILM
C*07:01	24%	6%	R		FYLM
C*07:02	22%	25%	RY		FYML

other motifs, with few peptides eluting before ~10-15% acetonitrile. Therefore, techniques that test primarily with HLA-A*02 peptides would be unlikely to have noticeable losses due to bumping (i.e. column overload causing peptides to elute during equilibration). A*01:01 and B*35:01 are similarly hydrophobic, with few peptides eluting early in the

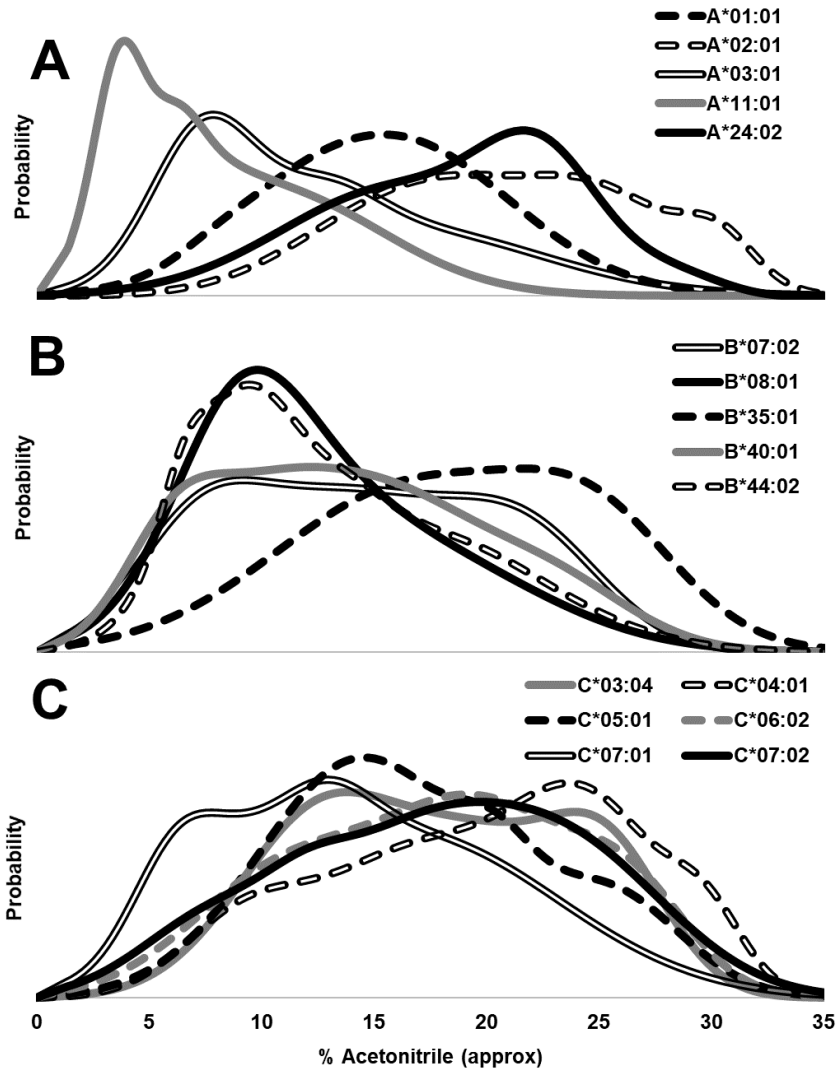


Figure 4.5 Elution profiles of peptides expressed on common HLA types
Since MHC class I peptides are small, the preferred anchor residues have a strong effect on the retention profiles of the HLA type. Identification of peptides from HLA types with hydrophilic anchor residues, such as HLA-A*11:01 (grey line, top graph) and HLA-A*03:01 (while line, top graph) would be most affected by losses in the early gradient. Alternately, peptides associated with HLA types that have hydrophobic anchors, such as HLA-A*02:01 (top graph, white dotted line), would be more affected by losses from the end of the gradient. **Appendix B**: contains information about data collection and graph generation.

gradient. However, while A*02:01 is by far the most common HLA type in the United States due to large Caucasian and Hispanic populations, the most common HLA type worldwide should be HLA-A*11:01 due to its higher prevalence in Asian demographics (**Table 4.1**). While A*02:01 tends to produce hydrophobic peptides, A*11:01 was by far the most hydrophilic motif, with the majority of peptides eluting before 10% acetonitrile. HLA-A*03:01, B*08:01, and B*44:02 would also have significant losses if the most hydrophilic moieties are removed. B*07:02, B*40:01, and C*07:01 would have fewer but still notable losses.

Previous work investigating cleanup methods indicate that cleanup conditions can show strong preferences for specific HLA types.(4, 11) Mostly, the work focused on losses of HLA-A*02 when eluting with low organic percentages from C18 to remove HLA chains from the sample. However, it highlights the need to investigate several samples with a variety of HLA types to ensure consistent results.

4.4.1 Comparison of MHC Class I peptide retention

The goal of a cleanup method is to retain all MHC antigens while removing all other species from the sample. The motif that would be most likely to have losses with PHEA would be HLA-A*02. As seen in **Table 4.1**, its motif consists of aliphatic residues, and its antigens tend to elute later than any other common allele (**Figure 4.5**). Additionally, the associated peptides often are less charge dense than most MHC-I peptides, with over 40% of identifications lacking both arginine and lysine residues. Since these peptides tend to contain few hydrophilic residues, we expect to see more losses from PHEA. Additionally, we expect C18 to perform better than PHEA, since A*02 lacks many of the hydrophilic peptides that caused issues with other alleles. To compare this, I evaluated a

sample of the JY lymphoblastoid line immunopurified with BB7.2, an antibody specific for HLA-A*02 molecules. In **Figure 4.6** there is a clear decrease in retention of peptides towards the end of a gradient in the PHEA cleanup (blue line) compared to the C18 cleanup (red line) and original sample (green dotted line). There is a more notable drop in peptide area attributable to identifications (top) compared to identifications per minute (bottom). However, while PHEA shows losses towards the end of the gradient, the C18 cleanup shows losses towards the beginning of the gradient, mostly in identifications. These losses are primarily low abundance, somewhat hydrophilic A*02 peptides. Based on this sample,

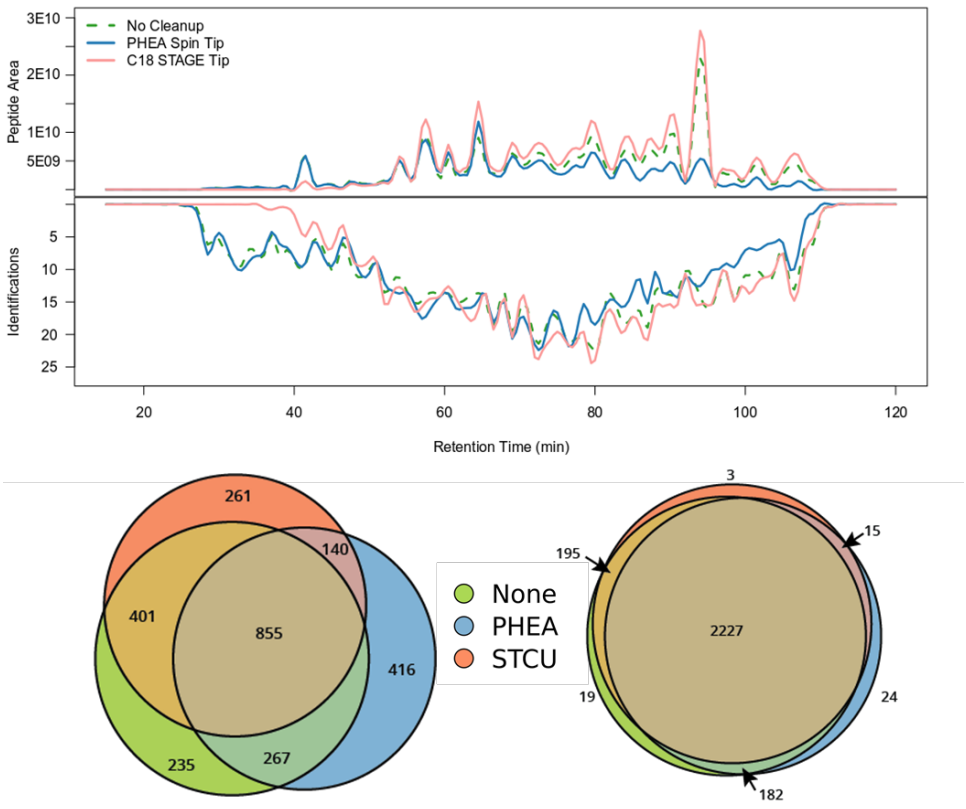


Figure 4.6 Comparison of cleanup techniques for a hydrophobic HLA type
The top panel includes a rolling average of the chromatographic area assignable to MHC I peptides (top), as well as the number of identifications per minute (bottom) for the sample JYA2. The left Euler plot shows the overlap in confident identifications between the three techniques, while the right shows the overlap when the peptides were found by mass and retention time but not identified. Green is used to show identifications without a cleanup, red for STAGE tip cleanups (STCU), and blue for PHEA spin tip.

PHEA and STAGE tips appear to have comparable results when working with hydrophobic motifs, but both have fewer identifications than not performing a cleanup.

However, when investigating identifications in samples that include more hydrophilic HLA types, the loss inherent to the C18 cleanups is apparent. While PHEA has similar levels of peptide retention and identification compared to the original sample, C18 shows significant losses in the early gradient, shown in results from another sample, AVL3 (**Figure 4.7**). AVL3 is a melanoma cell line that expresses HLA-A*01, A*02, B*27, B*44, C*05, and C*07. The sample used here had the HLA-A*02 molecules depleted by BB7.2 antibody. The remaining HLA types for this sample contain hydrophilic residues as at least

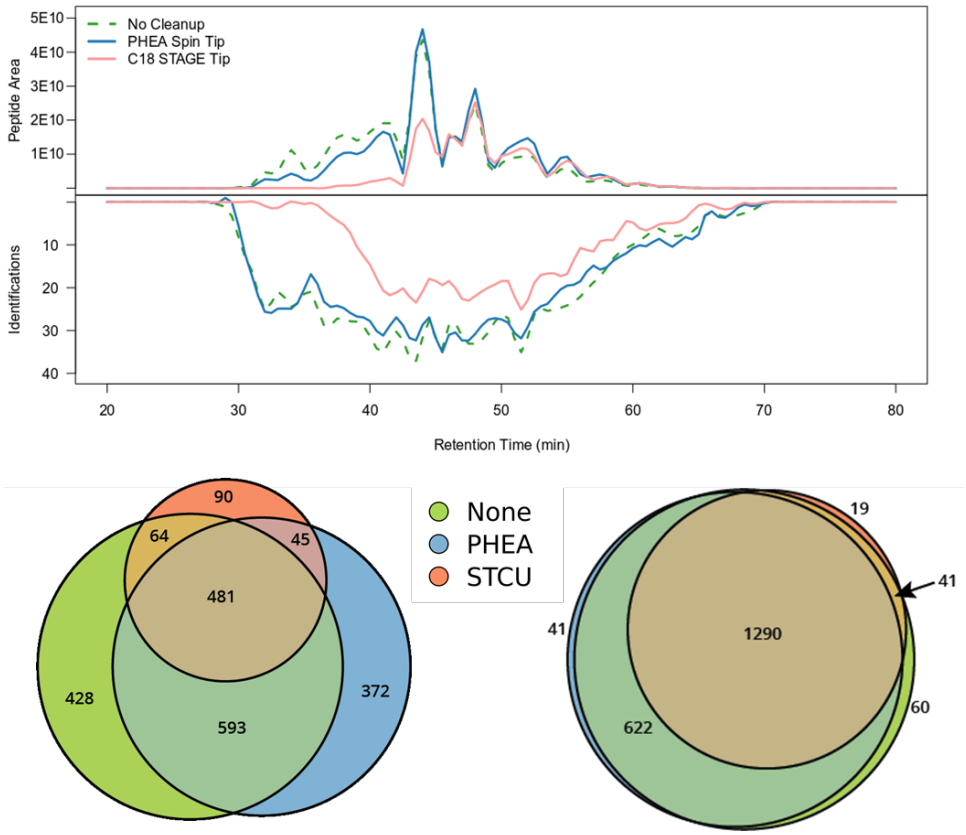


Figure 4.7 Comparison of cleanup techniques for a sample with mixed HLA types
Identifications and peptide area for melanoma cell line AVL3. Here, the loss of hydrophilic peptides is more pronounced using the C18 STAGE tip (red line), while the PHEA cleanup follows the same elution profile as the sample with no cleanup. Most peptides identified in only the STAGE tip cleanup were oxidations of methionine.

one anchor point. A*01, B*44 and C*05 anchor at an acidic residue, C*07 anchors at a basic residue, and B*27 has two basic anchors. In **Figure 4.7**, there is a sizable drop in both identifications and peptide area for the first 15 minutes that peptides elute following C18 STAGE tip cleanup, while PHEA and the original sample have nearly identical elution profiles.

This is likely due in part to an overloaded column capacity. The amount of contamination in a sample is difficult to determine, especially prior to analysis. Therefore, the assumed content of the sample is usually based on the quantity of the desired sample material (MHC peptides) rather than the actual sample content (MHC peptides, molecules, contaminants). However, while the MHC molecules would still contribute to the binding capacity for PHEA, many other contaminants do not.

Of the 2073 total identifications for this sample, 1566 (75%) were identified without a cleanup, 1491 (73%) were identified by PHEA, and only 680 (33%) were identified using the STAGE tip cleanup. Of the 90 confident identifications only made using a STAGE tip, 61 were oxidations of methionine with the unoxidized version identified by the other two methods. Similarly, for peptides found by mass and retention time, 97% were found with no cleanup, 94% using PHEA, and 65% by STAGE tip. Of the 19 exclusively identified in STAGE tip cleanups, 15 were oxidations of methionine. This indicates that not only do the STAGE tips have significant losses, they also allow for a greater degree of methionine oxidation, which will split ion signal and lower the abundance of the unoxidized moiety, decreasing the chance of making a confident identification.

4.4.2 Removal of coeluting polymeric contaminants

Another advantage of using PHEA is its ability to remove polymers, small molecules (e.g. protease inhibitors), and hydrophobic molecules (e.g. detergents) that cannot be removed using C18. Removal of these interferences from a sample has two clear benefits. Firstly, the ion current is no longer diverted to the contaminant, so the peptide signal increases. Secondly, the instrument can sample lower-abundance peptides since it is not preoccupied with sampling the contaminants. Therefore, contaminated samples have apparent losses due to signal suppression, and can have fewer identifications when ions are not selected for fragmentation, even though the peptide is still present in the sample. This

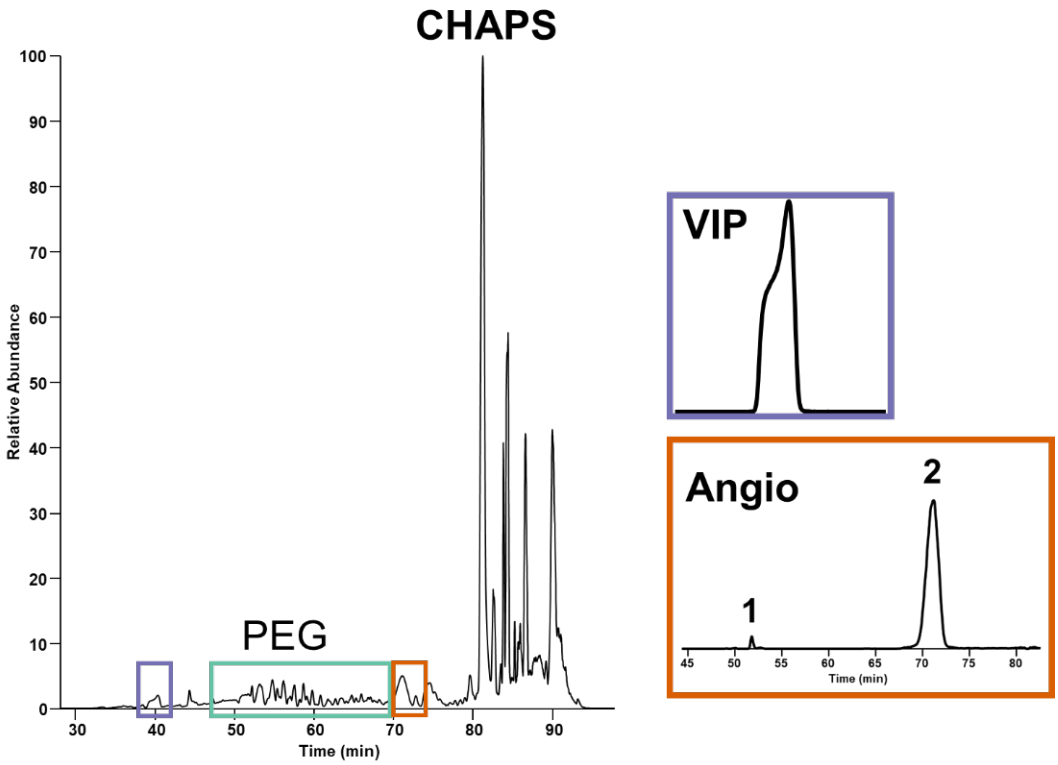


Figure 4.8 Base peak chromatogram of VMM39 contaminated with detergent and polymer
Prior to cleanup, the chromatogram is primarily composed of PEG (aqua box) eluting during the gradient and CHAPS eluting towards the end of the gradient. If we extract the ion chromatograms for the internal standards, altered chromatography is significant. The earlier-eluting VIP (purple box) has some fronting while Angio (orange) is double peaking. The correct elution time for Angio is peak 1 at ~52 minutes. However, secondary interactions with contaminants cause it to elute primarily nearly 20 minutes later.

is particularly important when using instrumentation that scans more slowly, as they are less likely to reach the suppressed peptides of interest.

In this situation, cleanup with PHEA shows drastic improvement over both C18 and the control samples. A melanoma cell line (VMM39) was heavily contaminated with polymers and detergent peaks (**Figure 4.8**). This sample expresses HLA-A*02, A*03, B*07, B*44, C*05, and C*07. As seen in **Figure 4.9**, PHEA has noticeably more identifications than either the control or C18, particularly in the portions with polymer contamination. For this sample, PHEA had 130% recovery of total peptide signal, while C18 had 41%. While legitimate recoveries over 100% are not possible, it is a frequent occurrence after PHEA cleanup, likely due to removal of suppressive elements. Thus, the

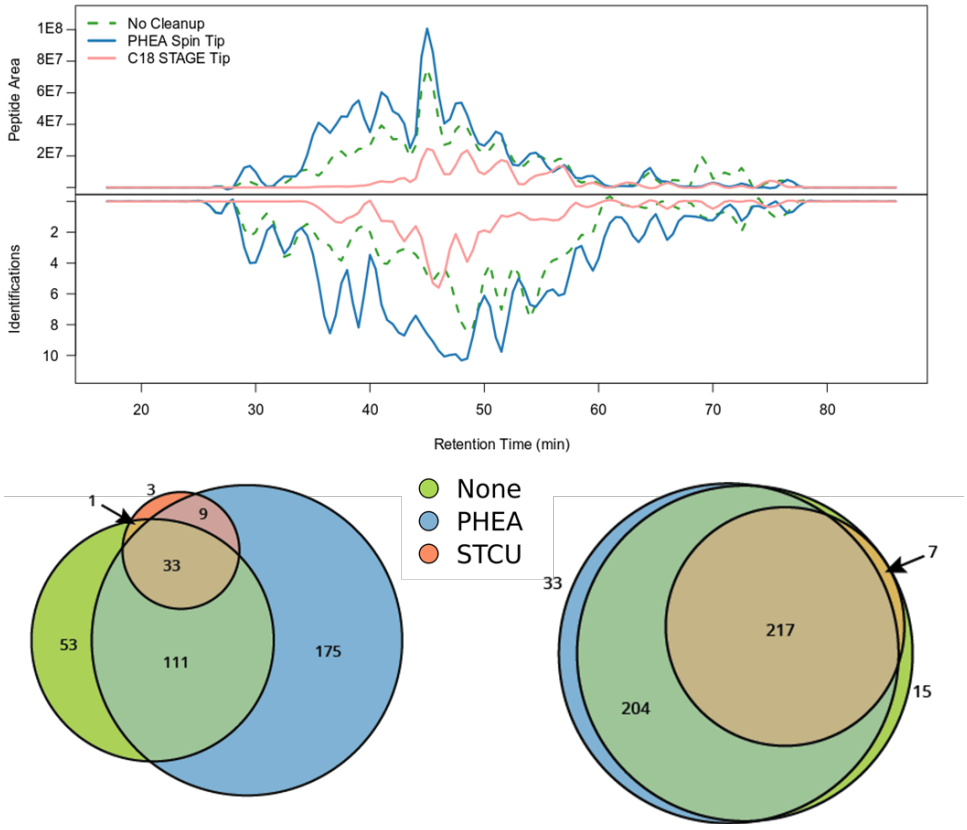


Figure 4.9 Comparison of cleanup techniques for a sample with polymeric contamination
PHEA cleanup shows significant increases in identifications when used on a contaminated sample (VMM39) and a non-hybrid instrument with a slower scanning speed.

apparent recovery is over 100%, but the true recovery is unknown. A PHEA cleanup also allowed for identification of 402 peptides, compared to 277 identified without a cleanup and the 131 detected after a C18 stage tip.

A leukemia cell line (THP1) was previously analyzed following a STAGE tip protocol, but only five peptides were confidently identified due to severe contamination. A subsequent cleanup using PHEA on the eluate of this cleanup removed significant contamination, allowing confident identification of 270 unique peptides (**Figure 4.10)**. Prior to HILIC cleanup, polymers coelute with the peptides which both decreases the

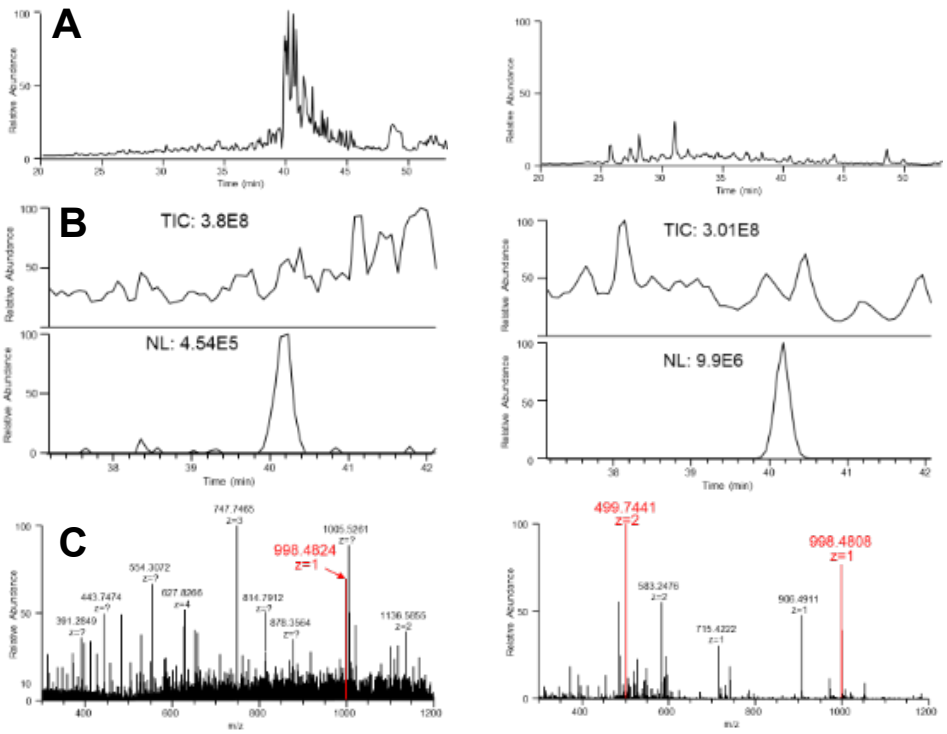


Figure 4.10 Ion suppression in a contaminated sample is reduced by HILIC cleanup
Comparison of chromatograms and spectra after a STAGE tip cleanup (left) and subsequent HILIC cleanup (right) for THP1 cell line.
A) Total ion current for the chromatographic gradient showing the removal of β2m (eluting at ~40 minutes prior to HILIC cleanup).
B) TIC (top) and XIC for peptide SLPDFGISY (bottom) showing 20x signal intensity for the peptide after PHEA cleanup due to removal of suppressants.
C) MS1 spectra at the time the peak elutes shows significant noise that is removed after PHEA cleanup. With removal of the coeluting contaminants, the peptide is allowed to take an additional charge, which allows it to be selected for fragmentation. The peptide is now the most abundant species eluting at the time.

chance of a peptide being selected for fragmentation and suppresses the signal of the peptide. This not only causes a decreased abundance of the peptide, as seen in the extracted ion chromatogram (**Figure 4.10b**) before PHEA cleanup (left) and after cleanup (right). The suppression also causes a lower charge distribution, visible in the MS1 spectra (**Figure 4.10c**) before cleanup with the peptide only taking a single charge and therefore not being selected for fragmentation, compared to after cleanup (**Figure 4.10d**) where the doubly charged precursor is the most abundant ion of the MS1 spectrum. Subsequent selection and fragmentation of the precursor then allows a clear, unambiguous identification where there was previously no spectrum available to identify (**Figure 4.11**).

However, it is not only the abundance and charge increase that aids identification, but the removal of contaminants with similar *m/z* values. **Figure 4.12** shows an MS2 spectrum before (top) and after (bottom) PHEA cleanup. When isolating a peak for fragmentation, any peaks within a set isolation window (generally 2 Da) will also be fragmented alongside the peptide target. Before cleanup, the peak is co-isolated with a

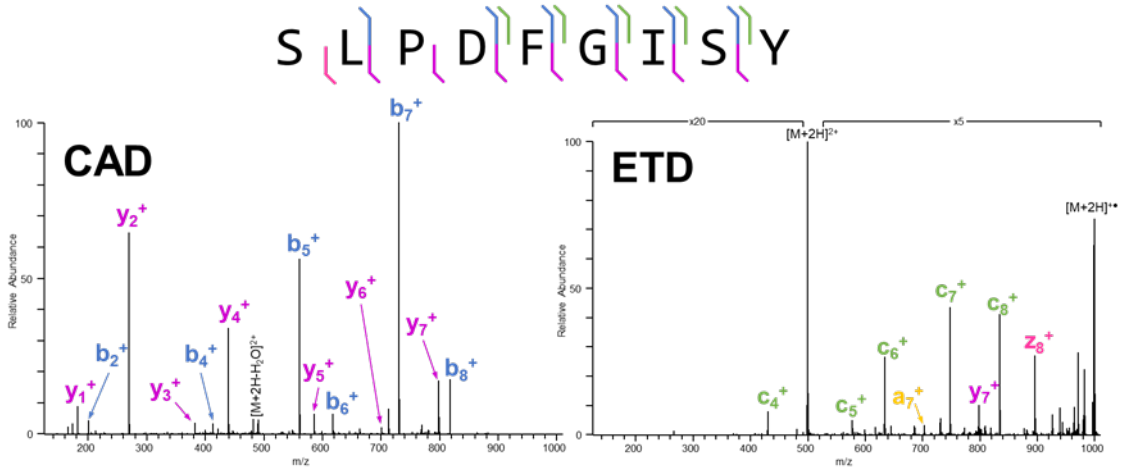


Figure 4.11 Fragmentation spectra of previously suppressed peptide allows unambiguous sequence identification
Despite previously being suppressed enough that it could not be selected for fragmentations, after HILIC cleanup the increased charge and abundance allow for a confident identification to be made, demonstrating an avenue for increased identifications after HILIC cleanup.

contaminant for fragmentation (left). While the correct peptide identification has similar coverage before and after the cleanup, the assignment is more ambiguous before cleanup due to peaks that cannot be explained by the sequence (red). After the cleanup, there is minimal signal that cannot be assigned, providing a clearer, more confident identification.

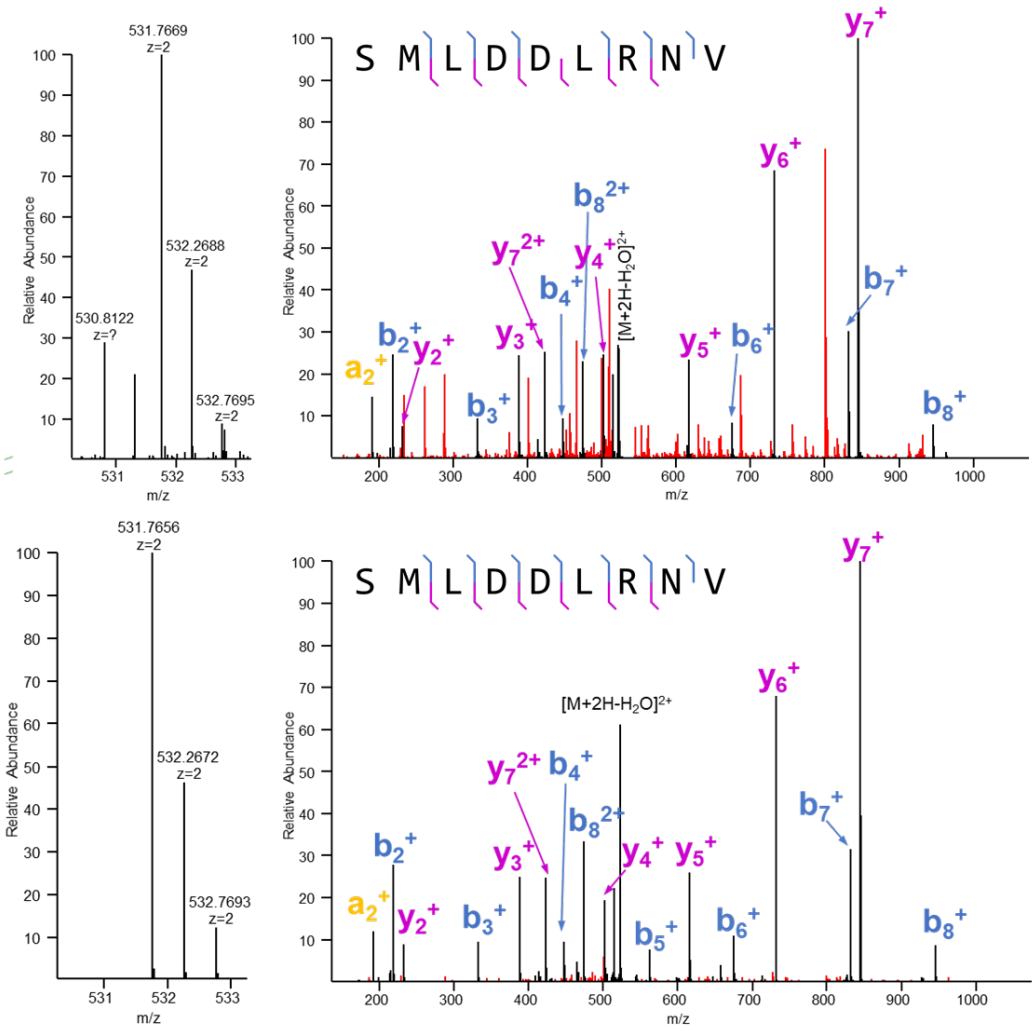


Figure 4.12 Improved fragmentation spectra due to removal of coisolated contaminants
A contaminant close to the molecular weight of HLA-A*02-associated peptide SMLDDLRNV complicates the fragmentation spectrum, with the unexplained fragments (red) from the contaminant making assignment ambiguous despite having acceptable coverage of the peptide sequence. After PHEA cleanup the contaminant is removed and the identification is more confident despite having the same coverage.

4.4.3 Removal of contaminant proteins

Most reverse phase cleanup protocols attempt to remove detergents, β2m, and other proteins by eluting from a C18 cleanup with ~30-40% acetonitrile. PHEA can elute proteins and peptides separately. In this protocol, 50% acetonitrile with 10mM ammonium formate or 0.1% acetic acid (~pH 3) is used to elute peptides from the spin tip. Proteins remain

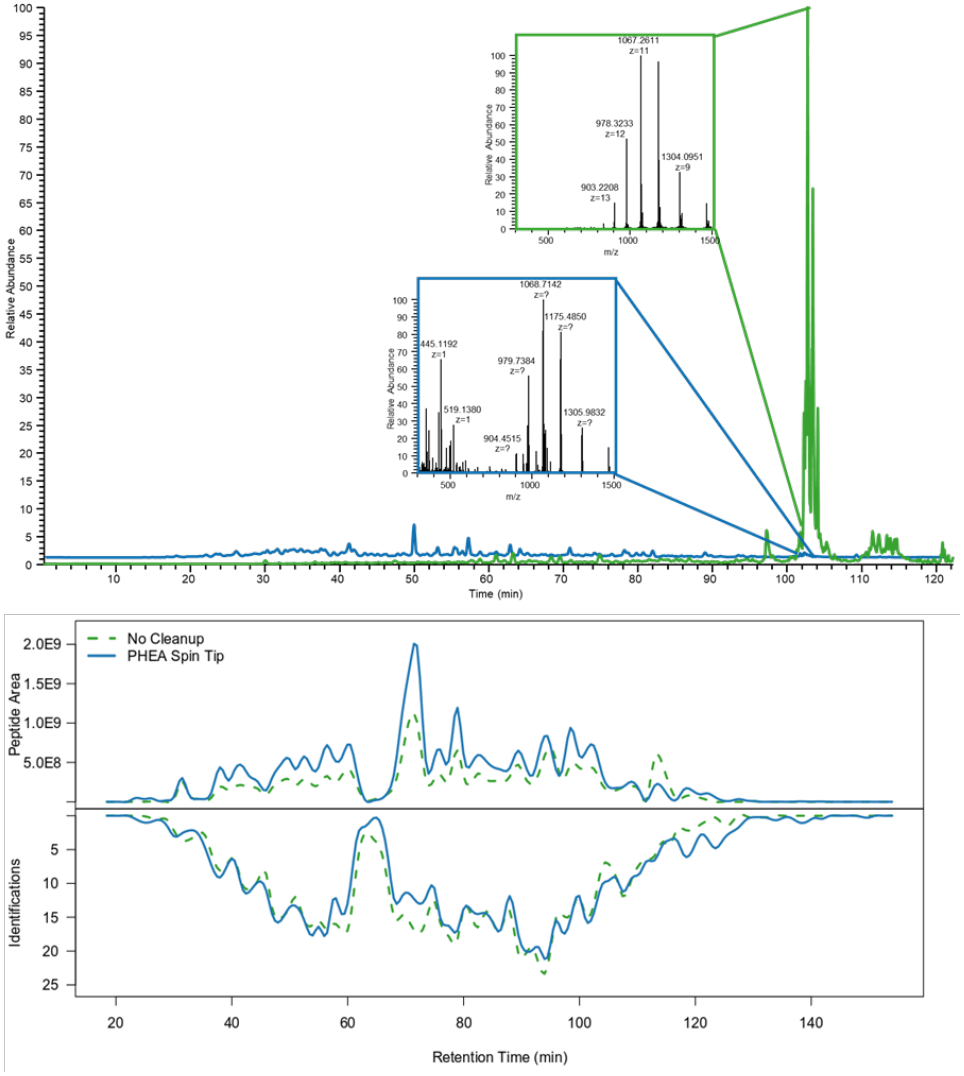


Figure 4.13 Selective removal of beta-2-microglobulin from HLA-I sample
Unlike reverse phase removal of β2m from samples, PHEA is able to retain peptides that elute after the protein despite its preference for hydrophilic molecules.
Top: TIC before (green) and after (blue) PHEA cleanup of a HeLa sample, with the PHEA run having about 1% abundance of the contaminating protein.
Bottom: Despite removal of β2m, which elutes at ~104 minutes, PHEA continues to have identifications at 140 minutes, indicating selective separation of proteins from peptides.

bound to the material and require chaotropic conditions to elute from the column. The tip can then be moved to a new collection tube before eluting the proteins with unbuffered 0.2% formic acid. This separation allows for removal of proteins as low as 8 kDa while retaining these late-eluting peptides for analysis.

While PHEA has some losses of hydrophobic peptides, it would not have the same level of loss that has been observed when attempting to remove proteins using reverse phase methods by eluting with low percentages of organic.(11) As seen in **Figure 4.13**, an MHC I enrichment of cervical cancer cell line HeLa, β2m elutes at ~105 minutes, while peptide identifications continue to ~140 mins. However, β2m dominates the chromatogram (top, green) and is both abundant and large enough to cause problems for C18 packing materials with small pore sizes. Therefore, it is better to be able to remove it. Reversed phase methods for removal of β2m and HLA alpha chains would result in removal of all peptides that elute after β2m, as has been previously observed.(11) In hydrophobic motifs such as A*02, approximately 15-20% of the peptide identifications elute after β2m, so this method of separation is not ideal. One workaround for this is to use molecular weight cutoff filters to remove proteins in a size dependent manner rather than by hydrophobicity, but those incur significant losses when working with small sample sizes. In contrast, the PHEA cleanup still provides this benefit when working with much smaller samples.

4.4.4 Cadaverous lymph and brain tissue

Generally, MHC peptide analysis requires large amounts of tissue, from hundreds to thousands of milligrams. However, there has been a continuous drive to decrease sample sizes while obtaining similar results. A major part of avoiding sample loss is to minimize sample handling and the number of steps in a protocol, attempting to avoid steps that would

cause losses. With small or poorly expressing samples, a moderate amount of contamination can cause a significant decrease or complete lack of peptide identifications, and small samples can still just as easily clog a column. Therefore, while avoiding a cleanup step overall would be ideal for small samples, it is frequently still a necessary step in the process.

We were developing a protocol for analysis of 5-25 mg (~1-5 million cells) of cadaverous lymph tissue (**Figure 4.14**). Due to the extremely small sample amount and ensuing losses in the STAGE tip protocol, the cleanup was removed altogether from the sample preparation. In our initial examination, the samples had significant contamination, primarily large, unresolved protein as well as residual protease inhibitor.

PHEA cleanup increased identifications from 321 to 466 peptides. Between the two conditions, 543 peptides were identified in total, with 98% of these identifications present by mass after PHEA cleanup. Comparing only peptides that were present at a quantifiable

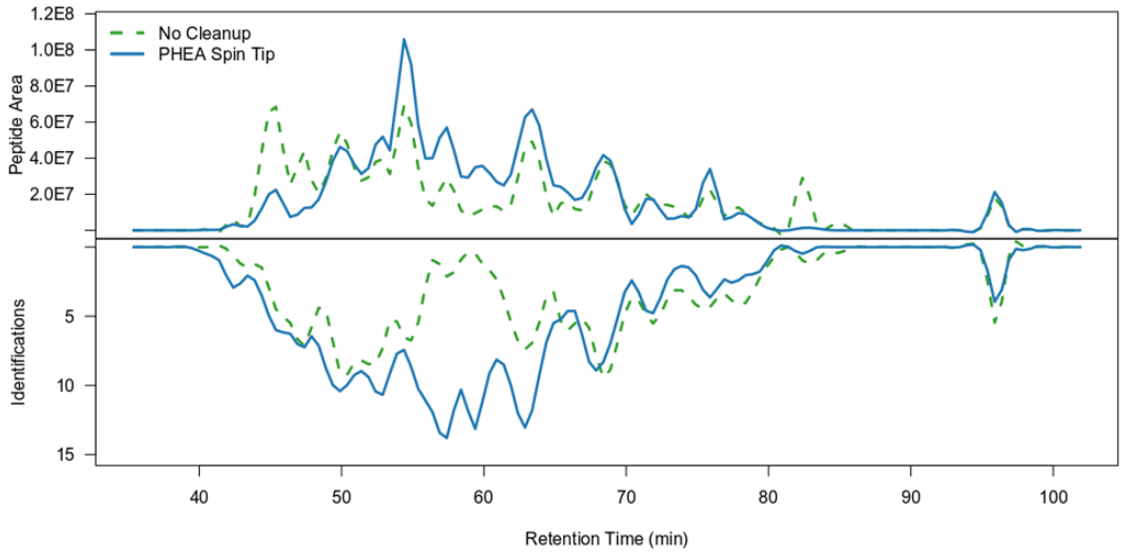


Figure 4.14 Identification of MHC I peptides in small samples still improves identifications
Even when working with small samples, PHEA shows few losses. Here, ~5 million ceq were used (~1% the amount used for most MHC analyses), and PHEA was able to remove contaminant protein eluting at 50-70 min without significant losses from additional processing steps on such a small sample.

level in the control run, the cleanup obtained 130% median recovery (200% mean). This, as mentioned earlier, is likely due to removal of ion suppressants; in this case, there appeared to be an unresolved protein eluting throughout the gradient, as well as significant protease inhibitor remnants.

It has also allowed for identification of antigens expressed on cadaverous brain tissue. Brain tissue expresses MHC class I antigens at a much lower level than most tissues and can generally only generate a response if the immune system has previously encountered an antigen. The last attempt to characterize the brain immunopeptidome was over a decade ago and used 50 grams of tissue for each sample, finding between 6 and 38 MHC class I peptides per sample.(12) With less than one percent the amount of starting material, we have been able to regularly identify 100-200 peptides per sample.

4.4.5 Identification of peptides with hydrophilic modifications

The peptides that the HILIC medium does not retain usually contain no residues that have a charge at pH3, and are predominantly aliphatic residues (e.g. MLLSVPLLL, ALWLVSPLL, ALSVIELLL). Retention of these peptides may be important for unmodified MHC analyses, but if the goal is identification of sequences with hydrophilic modifications, the modification should provide enough hydrophilicity for association to the resin despite a lack of other hydrophilic regions.

As discussed in previous chapters, our lab primarily focuses on identification of phosphorylated MHC antigens. In general, hydrophilic modifications do not affect elution time on reversed phase material significantly, usually only affecting elution time less than 5 minutes. As seen in **Figure 4.15**, unmodified peptide identifications distribute evenly around the standards in AVL3 before cleanup. This same elution profile can be seen in the

IMAC of AVL3 after a PHEA cleanup, indicating that phosphopeptides follow the same elution profile as the unmodified peptides and that PHEA maintains this profile. In the AVL3 sample, we manually confirmed 521 MHC phosphopeptides, the most phosphopeptides we have identified in a single sample. This can be partially attributed to the high MHC expression of the AVL3 cell line. However, using internal standards VIP and angiotensin to compare the distribution of identifications between files, the phosphorylated peptides identified in the IMAC were distributed similarly to the unmodified identifications.

Since phosphorylation was not the main focus of this method development, I did not perform a direct comparison of the same sample with each cleanup after IMAC

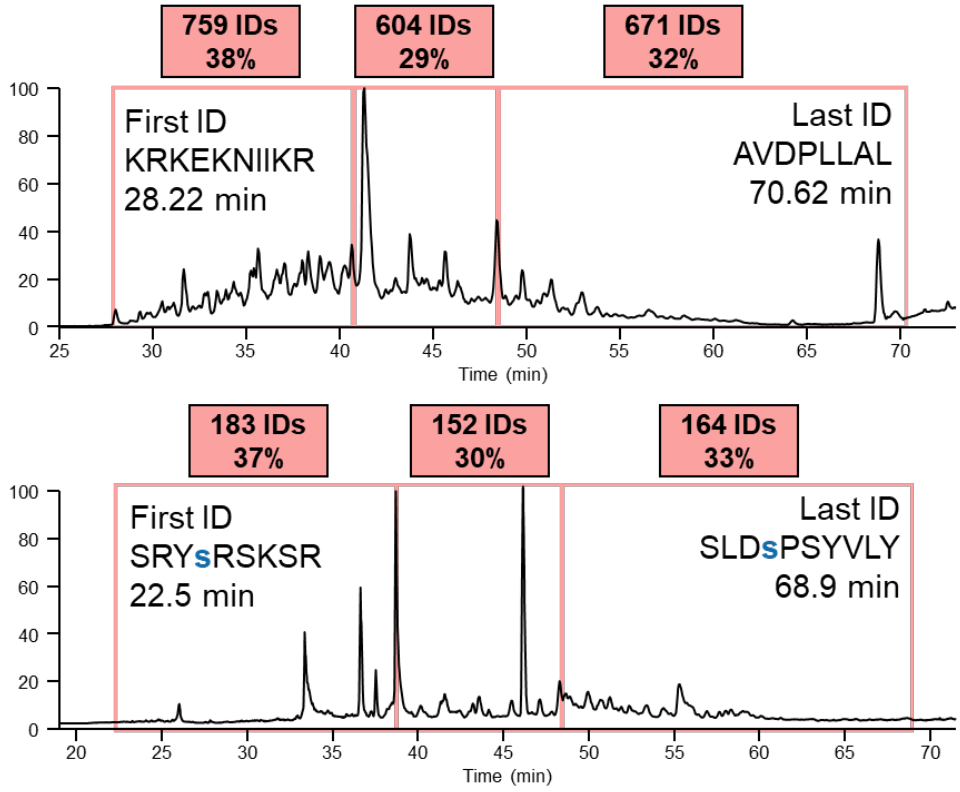


Figure 4.15 Peptide distribution for unmodified and IMAC files
Before IMAC or cleanup (top) 38% of peptides elute before VIP, 29% elute between the standards, and 32% elute after angiotensin I. This distribution is upheld in the IMAC file after PHEA cleanup.

enrichment, though the direct comparisons have since been performed.(13) However, we can infer the losses in the early gradient through looking at the elution profiles of IMAC enriched samples expressing multiple HLA types. Comparing the 5 samples with the most phosphopeptide identifications (105-280) after STAGE tip and IMAC, on average 12% (5-19%) of identifications are before VIP, 29% (16-35%) are between VIP and angiotensin, and 59% (46-78%) elute after angiotensin. Comparing the elution times of different HLA types in unmodified peptides with no cleanup (**Figure 4.16**), even in the most hydrophobic alleles at least 20% of identifications occur before VIP elutes and at maximum 50% elute after angiotensin. In contrast, in the 3 IMAC files collected after PHEA cleanup, 44% (28-67%) elute before VIP, 29% (20-38%) elute between the standards, and 26% (12-34%) elute after angiotensin. This elution was much more in line with what would be expected by combining the elution profiles of multiple HLA types.

Additionally, O-GlcNAc and extended O-GlcNAc have previously been observed on MHC associated peptides.(14) However, we have never found a glycosylated peptide after STAGE tip cleanup. Interestingly, even glycosylated peptides that elute late in the gradient are no longer present after STAGE tip cleanup. It is difficult to say why late eluting

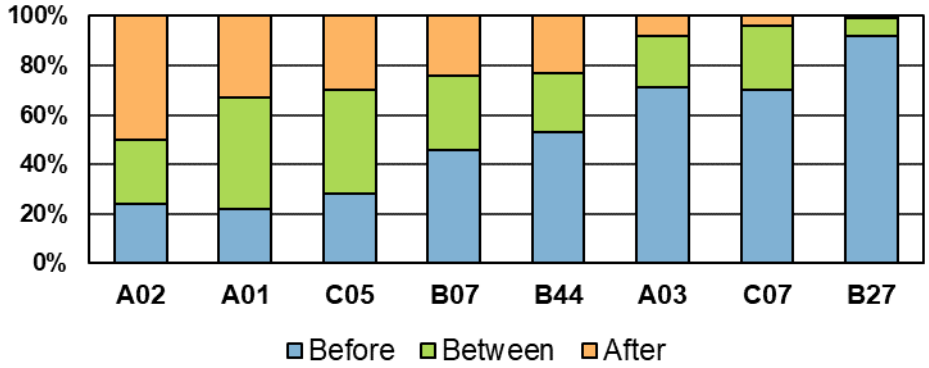


Figure 4.16 Elution profiles of different HLA types compared to elution times of standard peptides
The standard peptides frequently divide mixed HLA type samples evenly, with approximately 1/3 eluting before vasoactive intestinal peptide (blue), 1/3 between the standards (green), and 1/3 eluting after angiotensin I (orange).

glycopeptides are removed, but it may be that the small amount of contamination caused by adding any additional processing steps is enough to suppress ionization. Regardless of the reason, PHEA cleanups allow identification of these glycosylated peptides. After PHEA cleanups I have identified sequences modified O-GlcNAc and extended O-GlcNAc, as well as O-hexose on serine and threonine, and O-galactose on lysine.

4.4.6 Limitations

Like most samples analyzed by mass spectrometry, the usage of nonvolatile salts has minimized our preparatory protocol. When analyzing MHC peptides, nonvolatile salts are last used when rinsing the immunopurification beads. The buffer used in that step is 10-20 mM Tris buffer followed by a rinse with water. Therefore, the amount of salt in the loaded sample does not merit specific desalting and does not have a major effect on the cleanup. However, when extending this cleanup for other protocols, chaotropes and anything that may significantly affect the pH would have to be removed prior to performing the HILIC cleanup.

Another problem that can occur when only performing a HILIC cleanup is column fouling by highly hydrophobic molecules with a hydrophilic region that allows for association with HILIC material. It may be possible to optimize PHEA rinses or eluents to remove such contaminants, but it is also possible that a C18 cleanup would always be necessary for their complete removal.

4.4.7 Cost and convenience

A less obvious benefit of using HILIC for cleanup is related to the cleanup column clogging. Like other membrane proteins, MHC molecules are not fully soluble in aqueous solution without addition of detergent. While detergent is used during cell lysis, it must be

removed before putting it in-line with mass spectrometer, so it is no longer a solution after immunoprecipitation. Not only can these semi-insoluble HLA alpha chains clog analytical columns but can clog or slow the flow on spin tips and large prep columns, generally not unclogging until 60-80% acetonitrile is forced through. While the protocol should theoretically take 15-20 minutes, many samples slow the flow and cause the procedure to take an hour or more (or in one particularly traumatic case, a full 17 hours). However, PHEA spin tips have not had any similar clogging issues. This is likely because many proteins that are not soluble in aqueous solution are more soluble in the high organic solvents used for separation. Therefore, this protocol reliably takes somewhere between 5 and 10 minutes. Both protocols employ a drying step after elution, which takes approximately half an hour to an hour depending how many samples are processed.

A major appeal of using STAGE tips over commercial separation columns is their low cost. The PHEA spin tip protocol is slightly less expensive than STAGE tips, costing less than a quarter per cleanup. However, the initial cost for materials is about 1/3 the cost of Empore disks. In addition to price, the other appeal of STAGE tips over commercially produced devices is the ability to run multiple samples in parallel. This benefit is maintained using PHEA tips.

4.5 Conclusions and future directions

While more optimization and additional testing is needed, this HILIC based cleanup already shows significant advantages over the current methods for MHC associated peptide sample cleanup. The protocol has been used successfully by other students for MHC analyses of cancerous tissue, with and without subsequent enrichment for phosphorylation. In unmodified analyses, it reliably results in an increased number of identifications in samples with polymer, small molecules, or proteins that coelute with MHC peptides. This benefit is further increased when using instrumentation with slower scan rates.

In samples with multiple HLA types the PHEA cleanup vastly outperforms C18 STAGE tip cleanups but has moderate losses of the more hydrophobic peptides presented on HLA types with hydrophobic binding motifs. It may be possible to reduce these losses by modifying the loading solvent. The concentration of 10 mM ammonium formate pH3 was chosen because it was sufficiently volatile for use in-line with the mass spectrometers during initial experimentation. However, PHEA binds peptides more strongly with higher concentrations of buffer that are insufficiently volatile for use as an LCMS buffer, but may allow for better retention of these hydrophobic peptides.(8) I have since found that retention is maximized at 40mM ammonium formate pH3. While this is too concentrated for in-line analysis, the sample must be dried multiple times or desalted to completely remove the ammonium ion, which will otherwise adduct peptides that elute early in the gradient. However, retention of HLA-A*02 associated peptides may be improved using this higher buffer concentration.

While the ability to remove proteins from peptide material is not a common problem in most proteomics applications, the ability to tailor solvents to appeal to different

separation modes of PHEA makes the protocol modifiable to address a variety of issues. By choosing different loads, rinses, and eluents, I have adapted protocols for diverse issues including detergent removal after membrane protein digestion,(15) isolation of peptides after probe modification,(16) and coarse fractionation for O-glycopeptide identification.(17)

While HILIC-based techniques are frequently used for N-glycopeptide enrichment, they tend to suffer for O-linked glycopeptide identification.(18, 19) This is due to the larger N-glycan trees having sufficient hydrophilicity to be retained on HILIC material even with harsh rinses. However, O-linked glycosylation can be much smaller and does not provide enough hydrophilicity to outcompete more hydrophilic non-glycosylated sequences. Through tailoring elution buffers, it is possible to separate contaminants from the peptides in a sample, then have a weak elution to remove most non-glycosylated peptides while still increasing identifications of small O-glycosylated peptides. In **Figure 4.17**, we performed

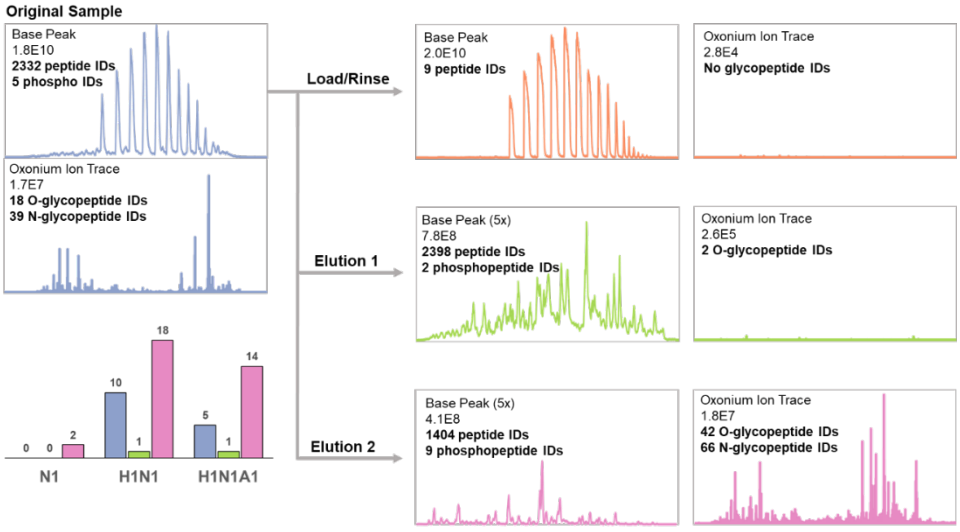


Figure 4.17 Using PHEA for coarse fractionation of polymers, unmodified peptides, and peptides with hydrophilic modifications
Before cleanup, the sample is contaminated with PEG but some glycosylation is still present (purple), but collecting the load (orange), rinsing with 80% MeCN, 40mM ammonium formate pH 3 (green), and 0.5% formic (pink) nearly doubles identifications of glycopeptides and phosphopeptides. Small glycans are also enriched (bottom left), which is uncommon for HILIC enrichments.

a PHEA fractionation on a sample enriched with wheat germ agglutinin that has been heavily contaminated with PEG. Without enrichment, 5 phosphopeptides, 18 O-linked glycopeptides, and 39 N-linked glycopeptides were confidently identified. Of these, 12 peptides had small O-glycans (HexNAc (N1) or HexNAc-Hex (H1N1)). By loading the sample in 90% acetonitrile 20mM ammonium formate, the PEG can be separated from the peptide signal (orange). While some peptides are identified in this flow through, all of them were also found at a higher abundance in the first elution. The first elution is strong enough to elute most unmodified peptides but retain peptides with more hydrophilic regions. The second elution then has decreased signal suppression of lower-level peptides containing hydrophilic modifications, allowing identification of 9 phosphopeptides, 42 O-glycopeptides, and 66 N-glycopeptides. Eight additional sequences modified with small O-glycans were identified, indicating that the first elution is sufficiently weak to allow small O-glycans to contribute enough hydrophilicity for retention but strong enough to improve their ionization by removal of many unmodified peptides.

Overall, the ability to select solvents on PHEA to alter binding based on multiple connected modes of separation allows for it to be adapted to a variety of applications. However, the ability to separate both polymer and proteins from peptides make this technique uniquely appealing for MHC analysis. It appears to outperform current methods for cleanup of MHC class I peptides and has minimal drawbacks.

4.6 References

1. Murphy K, Travers P, & Walport M (2008) *Janeway's immunobiology* (Garland Science, New York, NY) 7 Ed.
2. Kote S, Pirog A, Bedran G, Alfaro J, & Dapic I (2020) Mass spectrometry-based identification of MHC-associated peptides. *Cancers* 12(3):535.
3. Anonymous (1991) *High-performance liquid chromatography of peptides and proteins: Separation, analysis, and confromation* (CRC Press, Boca Raton, FL).
4. Pandey K, Mifsud NA, Lim Kam Sian TCC, Ayala R, Ternette N, Ramarathinam SH, & Purcell AW (2020) In-depth mining of the immunopeptidome of an acute myeloid leukemia cell line using complementary ligand enrichment and data acquisition strategies. *Molecular Immunology* 123:7-17.
5. Alpert AJ (1990) Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *Journal of Chromatography A* 499:177-196.
6. Van Dorpe S, Vergote V, Pezeshki A, Burvenich C, Peremans K, & De Spiegeleer B (2010) Hydrophilic interaction LC of peptides: Columns comparison and clustering. *Journal of separation science* 33(6-7):728-739.
7. Yoshida T (2004) Peptide separation by hydrophilic-interaction chromatography: a review. *Journal of biochemical and biophysical methods* 60(3):265-280.
8. Alpert AJ (2018) Effect of salts on retention in hydrophilic interaction chromatography. *Journal of Chromatography A* 1538:45-53.

9. Heather JM, Myers PT, Shi F, Aziz-Zanjani MO, Mahoney KE, Perez M, . . . Hunt DF (2019) Murine xenograft bioreactors for human immunopeptidome discovery. *Scientific Reports* 9(1):1-15.

10. Rappsilber J, Mann M, & Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature protocols* 2(8):1896-1896.

11. Nicastri A, Liao H, Muller J, Purcell AW, & Ternette N (2020) The Choice of HLA-Associated Peptide Enrichment and Purification Strategy Affects Peptide Yields and Creates a Bias in Detected Sequence Repertoire. *Proteomics* 20(12):1900401.

12. Fissolo N, Haag S, De Graaf KL, Drews O, Stevanovic S, Rammensee HG, & Weissert R (2009) Naturally presented peptides on major histocompatibility complex I and II molecules eluted from central nervous system of multiple sclerosis patients. *Molecular and Cellular Proteomics* 8(9):2090-2101.

13. Peng X (2022) Identification of MHC-Associated Peptides and Phosphopeptides Presented in Various Diseased Cells by Tandem Mass Spectrometry. PhD (University of Virginia).

14. Malaker SA, Penny SA, Steadman LG, Myers PT, Loke JC, Raghavan M, . . . Cobbold M (2017) Identification of Glycopeptides as Posttranslationally Modified Neoantigens in Leukemia. *Cancer immunology research* 5(5):376-384.

15. Panepinto M (2022) Leveraging Proteomic Strategies to Analyze Post-Translational Modifications and Membrane Proteins. Ph.D. (University of Vrginia, Charlottesville, VA).

16. McCloud RL, Yuan K, Mahoney KE, Bai DL, Shabanowitz J, Ross MM, . . . Hsu K-L (2021) Direct target site identification of a sulfonyl–triazole covalent kinase probe by LC-MS chemical proteomics. *Analytical chemistry* 93(35):11946-11955.
17. Rangel-Angarita V, Mahoney KE, Ince D, & Malaker SA (2022) A systematic comparison of current bioinformatic tools for glycoproteomics data. *bioRxiv*:2022.2003.2015.484528.
18. Riley NM, Bertozzi CR, & Pitteri SJ (2021) A Pragmatic Guide to Enrichment Strategies for Mass Spectrometry–Based Glycoproteomics. *Molecular & Cellular Proteomics* 20.
19. Ahmad Izaham AR, Ang C-S, Nie S, Bird LE, Williamson NA, & Scott NE (2020) What are we missing by using hydrophilic enrichment? Improving bacterial glycoproteome coverage using total proteome and FAIMS analyses. *Journal of Proteome Research* 20(1):599-612.

Chapter 5. Conclusions and Future Directions

The research presented in this dissertation focuses on enrichment and identification of MHC class I associated post-translationally modified peptides for the development of immunotherapeutics. These peptides are paramount immunotherapeutic candidates for a number of reasons. Cancer and other chronic diseases can evade detection by the immune system in part by dysregulation of cellular signaling. This also leads to aberrant post-translational modifications, which may then be displayed via the MHC class I pathway. Additionally, proteins containing PTMs may undergo antigen processing differently, so that the resulting peptide may be different from that tolerized during thymic education. This idea is corroborated by the fact that these post-translationally modified peptides generate a different immune response than their unmodified counterparts.

Previously, only cancerous tissue was investigated for identification of these phosphorylated MHC I antigens. Protein phosphatase 2A (PP2A) and protein phosphatase 1 (PP1) are frequently inhibited in cancers and are responsible for the vast majority of dephosphorylation events in the cell. Due to their wide substrate specificities, we theorized that the increased phosphorylation we observe in cancerous cells is caused by inhibition of PP2A and/or PP1 rather than by upregulation of kinases. Since PP2A is inhibited by other diseases, these cancer-associated phosphorylated peptides should also be presented. Investigation of margin tissue from resected hepatocellular carcinoma indicated that margin tissue from tumors caused by hepatitis B and C expressed similar levels of phosphorylated peptides as tumor tissue, while margin tissue from other etiologies expressed very few phosphorylated antigens. Additionally, two phosphopeptides were

identified in a sample of tissue with neurodegenerative disease. This indicates it may be possible to leverage cancer immunotherapeutics that target phosphorylated antigens as immunotherapeutics for a variety of chronic and debilitating diseases.

The phosphopeptide enrichment protocol used by the Hunt laboratory provides an enrichment that is much more specific and sensitive than commercially available methods. However, the previous protocol was lengthy and frequently presented issues that could prevent the ability to analyze a sample. Changing the metal chelating ligand allowed for bulk activation and increased flow rates during enrichment, which decreases the active time for the enrichment by half. Additionally, treatment of samples with nuclease prior to immunoprecipitation decreased contaminants present in the sample as well as preventing nucleic acid contaminants from slowing or preventing flow through the microcolumn during enrichment. This allows for a faster and more consistent sample loading time as well as reducing contamination from DNA and DNA-binding proteins.

The final chapter describes a hydrophilic interaction-based method for cleanup prior to mass spectrometric analysis. It allows for superior retention of hydrophilic peptides than current methods, allows for removal of HLA proteins from the sample, and removes polymeric and detergent contaminants. Through eliminating ion-suppressive and/or coisolated contaminants the cleanup provides more and clearer identifications. Prior to these improvements, the most phosphopeptides identified in a single sample was 234 using two enrichments and 1 billion cells. After these advances, we identified 522 phosphorylated peptides using 100 million cells, demonstrating the utility of the protocol improvements for identification of phosphorylated antigens. Furthermore, the improvements in contaminant removal allow for superior identification of unmodified

peptides as well as other post-translationally modified peptides, particularly for tissues with low MHC expression and/or less starting material.

5.1 Other disease-associated post translational modifications

The signaling dysregulation seen in cancer does not only cause increased expression of phosphorylation, but also a variety of other post-translational modifications. Previously, we have observed O-linked glycosylation and methylation of arginine or lysine on MHC-associated peptides. I continued to identify these modifications, finding an additional 24 GlcNAc, an extended GlcNAc, and 60 methylated arginines and lysines.

In addition to expanding the identifications of peptides modified with known modifications, I have identified multiple modifications not previously identified on endogenous MHC class I peptides through further investigation of incorrect assignments by search algorithms. I identified 58 peptides modified with hexose and 2 peptides with O-galactosylated lysine. I also identified peptides containing a kynurenine modification to tryptophan (**Figure 5.1**). I have since found 144 peptides containing this modification presented on MHC class I. Some of these peptides have already been identified on up to 5 samples and 4 cancer types.

Kynurenine is a metabolite of tryptophan produced by indoleamine 2,3-dioxygenase (IDO) and tryptophan dioxygenase (TDO). Kynurenine has been shown to suppress immune response by binding to aryl hydrocarbon receptors on T cells and natural killer cells.(1-3) Increased IDO/TDO activity and resultant high kynurenine levels are known causes of tumoral immune resistance.(4, 5) Some approved cancer therapeutics work

partially through IDO/TDO inhibition,(6, 7) but there are many emerging cancer therapeutics that inhibit IDO/TDO or directly reduce kynurenine levels.(8-13) Dysregulation of the kynurenine pathway is linked with disease progression in cancer and HIV, but is primarily studied in connection with neurodegenerative and neurological disorders.(1, 4, 14) This dysregulation usually takes the form of increased activity of enzymes indoleamine 2,3 dioxygenase (IDO) and/or tryptophan 2,3 dioxygenase (TDO), resulting in increased conversion of free tryptophan to kynurenine (9-11) as well as post-translational modification of tryptophan in proteins.(15)

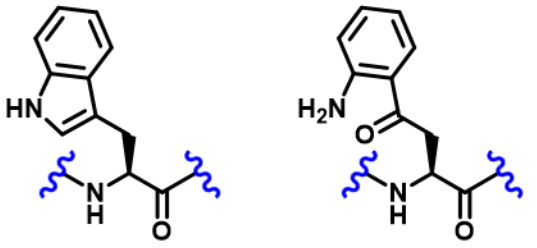


Figure 5.1 Structures of tryptophan and kynurenine
Tryptophan (left) is converted to kynurenine (right) in the kynurenine pathway. While this modification is primarily performed on free tryptophan, increased activity of converting enzymes IDO and TDO allows for their conversion in the peptide backbone.

An MHC associated peptide from the HIV envelope protein has previously been observed with a kynurenine modification, and seen to generate a slightly increased response compared to the native peptide.(16) Therefore, these MHC peptides containing kynurenine may be antigenic and potential targets for development of immunotherapeutics. However, this modification may also occur as an oxidative artifact during in-gel enzymatic digestion.(17) However, the abundance of these modifications does not correlate with the abundance of other oxidative modifications (e.g. methionine, cysteine), indicating that it is not a purely chemical modification. Additionally, some peptides are exclusively presented in their modified form. Most notably, peptide LwESPSLAI was identified in 5 samples and 4 cancer types. All samples were HLA-B*07:02 or HLA-B*35:01 positive. Both HLA types prefer proline as the second residue. This suggests that a post-translationally modified

kynurenine may be able to associate with certain HLA types similarly to a proline, allowing for presentation of new peptide sequences in addition to altering the level of the modification expressed on the cell surface.

All of these post-translational modifications Testing whether these peptides are antigenic would be the natural direction for this work to proceed. This is most commonly performed by ELIspot assay. In short, healthy donor peripheral blood mononuclear cells (PBMCs) are exposed to synthetic peptide antigens in microplates pre-coated with anti-IFN-γ antibody. The synthetic peptide antigens can then associate with empty MHC molecules caused by dissociation of weakly-associated MHC-peptide complexes. After incubation to allow for cytotoxic T-lymphocytes (CTLs) in the PBMCs to recognize the antigen and proliferate, cells are restimulated with synthetic antigen before reaction with a detection reagent. Based on IFN-γ production compared with positive and negative controls, it is possible to determine if the healthy donor had memory CD8+ T-cells against the antigen. If they do, that indicates that (1) the peptide is non-self, since T-cells made it through thymic education and have been used to previously clear an infection and (2) the donor has previously encountered the antigen. While we postulate the reason that phosphorylated peptides generate this response in Chapter 2, it is difficult to anticipate whether it will translate to other post translational modifications.

Diseases interfere with different signaling pathways, resulting in a variety of different post translational modifications to be upregulated, even when phosphorylation is not. Identifying these dysregulated modifications would allow immunotherapeutics to be developed for a variety of other diseases. Phosphorylated MHC-associated peptides are producing promising results in clinical trials,(18) suggesting that targeting of dysregulated

modifications is a viable approach for immunotherapeutics. Therefore, the clearest future work will determine the antigenicity of alternate modified peptides that have been identified in a variety of samples and investigating their therapeutic potential.

5.2 References

1. Mándi Y & Vécsei L (2011) The kynurenine system and immunoregulation. *Journal of Neural Transmission* 119(2):197-209.
2. Mezrich JD, Fechner JH, Zhang X, Johnson BP, Burlingham WJ, & Bradfield CA (2010) An Interaction between Kynurenine and the Aryl Hydrocarbon Receptor Can Generate Regulatory T Cells. *The Journal of Immunology* 185(6):3190-3198.
3. Opitz CA, Litzenburger UM, Sahm F, Ott M, Tritschler I, Trump S, . . . Platten M (2011) An endogenous tumour-promoting ligand of the human aryl hydrocarbon receptor. *Nature* 478(7368):197-203.
4. Baren NV & Van Den Eynde BJ (2015) Tumoral immune resistance mediated by enzymes that degrade tryptophan. *Cancer Immunology Research* 3(9):978-985.
5. Uyttenhove C, Pilotte L, Théate I, Stroobant V, Colau D, Parmentier N, . . . Van den Eynde BJ (2003) Evidence for a tumoral immune resistance mechanism based on tryptophan degradation by indoleamine 2,3-dioxygenase. *Nature Medicine* 9(10):1269-1274.
6. Riess C, Schneider B, Kehnscherper H, Gesche J, Irmscher N, Shokraie F, . . . Maletzki C (2020) Activation of the Kynurenine Pathway in Human Malignancies Can Be Suppressed by the Cyclin-Dependent Kinase Inhibitor Dinaciclib. *Frontiers in Immunology* 0:55-55.
7. Balachandran VP, Cavnar MJ, Zeng S, Bamboat ZM, Ocuin LM, Obaid H, . . . DeMatteo RP (2011) Imatinib potentiates antitumor T cell responses in gastrointestinal stromal tumor through the inhibition of Ido. *Nature Medicine 2011 17:9* 17(9):1094-1100.

8. Stone TW & Darlington LG (2002) Endogenous kynurenines as targets for drug discovery and development. *Nature Reviews Drug Discovery 2002 1:8* 1(8):609-620.
9. Triplett TA, Garrison KC, Marshall N, Donkor M, Blazeck J, Lamb C, . . . Georgiou G (2018) Reversal of indoleamine 2,3-dioxygenase–mediated cancer immune suppression by systemic kynurenine depletion with a therapeutic enzyme. *Nature Biotechnology* 36(8):758-764.
10. Je Cheong LS (2018) Targeting the IDO1–TDO2–KYN–AHR pathway for cancer immunotherapy—challenges and opportunities. *Trends in pharmacological sciences* 39(3):307-325.
11. Muller AJ, DuHadaway JB, Donover PS, Sutanto-Ward E, & Prendergast GC (2005) Inhibition of indoleamine 2,3-dioxygenase, an immunoregulatory target of the cancer suppression gene Bin1, potentiates cancer chemotherapy. *Nature Medicine 2005 11:3* 11(3):312-319.
12. Schramme F, Crosignani S, Frederix K, Hoffmann D, Pilotte L, Stroobant V, . . . van den Eynde BJ (2020) Inhibition of tryptophan-dioxygenase activity increases the antitumor efficacy of immune checkpoint inhibitors. *Cancer Immunology Research* 8(1):32-45.
13. Naour JL, Galluzzi L, Zitvogel L, Kroemer G, & Vacchelli E (2020) Trial watch: IDO inhibitors in cancer therapy. *Oncoimmunology* 9(1):1777625.
14. Lehtinen M, Tervo S, Halme E, Tiilikainen L, Pohjola K, Honkanen T, . . . Paavonen T (2015) 46 Indoleamine 2,3-dioxygenase and cancerous inhibitor of pp2A expression and inflammation in lingual hyperplasia and tongue cancer. *Oral Oncology* 51(5):e40-e41.

15. Taylor SW, Fahy E, Murray J, Capaldi RA, & Ghosh SS (2003) Oxidative Post-translational Modification of Tryptophan Residues in Cardiac Mitochondrial Proteins *. *Journal of Biological Chemistry* 278(22):19587-19590.

16. Ramarathinam SH, Gras S, Alcantara S, Yeung AWS, Mifsud NA, Sonza S, . . . Purcell AW (2018) Identification of Native and Posttranslationally Modified HLA-B*57:01-Restricted HIV Envelope Derived Epitopes Using Immunoproteomics. *PROTEOMICS* 18(12):1700253-1700253.

17. Perdivara I, Deterding LJ, Przybylski M, & Tomer KB (2010) Mass spectrometric identification of oxidative modifications of tryptophan residues in proteins: chemical artifact or post-translational modification? *Journal of the American Society for Mass Spectrometry* 21(7):1114-1117.

18. Engelhard VH, Obeng RC, Cummings KL, Petroni GR, Ambakhutwala AL, Chianese-Bullock KA, . . . Jr (2020) MHC-restricted phosphopeptide antigens: preclinical validation and first-in-humans clinical trial in participants with high-risk melanoma. *Journal for Immunotherapy of Cancer* 8(1):e000262.

Appendix A Additional Methods

A1. HPLC column construction

For precolumn construction, fused silica (100 μm i.d.) was cut into 20 cm lengths. A vortexed solution of 3:1 KaSil 1624: formamide was briefly put in contact with one end of the fused silica, allowing about 1 cm of the mixture to draw in, after which the solution was polymerized at 65°C for 3 hours to overnight. After polymerization, a column scribe was used to cut the frit to ~1 mm in length. A pressure vessel was used to flow methanol in reverse over the column, then the column was flipped and packed with a slurry of 10 μm Reprosil Pur AQ C18 in methanol with constant stirring. Once the bed reached a length of 6cm, the pressure was slowly released, and methanol was run over the column for ~1 minute to pack any remaining material into the bed. Once the bed was packed, the back of the column was trimmed to leave ~5 cm behind the bed, and 3x10 minute gradients from 100 mM acetic acid to 100mM acetic acid in 66% MeCN were run at 50 bar to settle the material and rinse the frit before attachment to the analytical column.

For analytical column construction, fused silica (75 μm i.d.) was cut into 30 cm lengths. The same KaSil: formamide solution was drawn into the end, cut to ~2 mm by crunching the silica using a scribe. After checking to be sure that an air bubble was formed by the cutting technique, water was then used to push the unpolymerized solution ~4-5 cm from the end of the fused silica. The frit was then allowed to polymerize 6 hours to overnight. After flowing in reverse with methanol, the column was packed with 10 cm 3 μm Reprosil Pur AQ C18 in methanol and settled by flowing 15 minutes at 500-800 psi with methanol. One 10-minute gradient was run at 50 bar, then liquid was pushed out by pushing helium through the column at 500 psi for 10-15 minutes. A window was burned

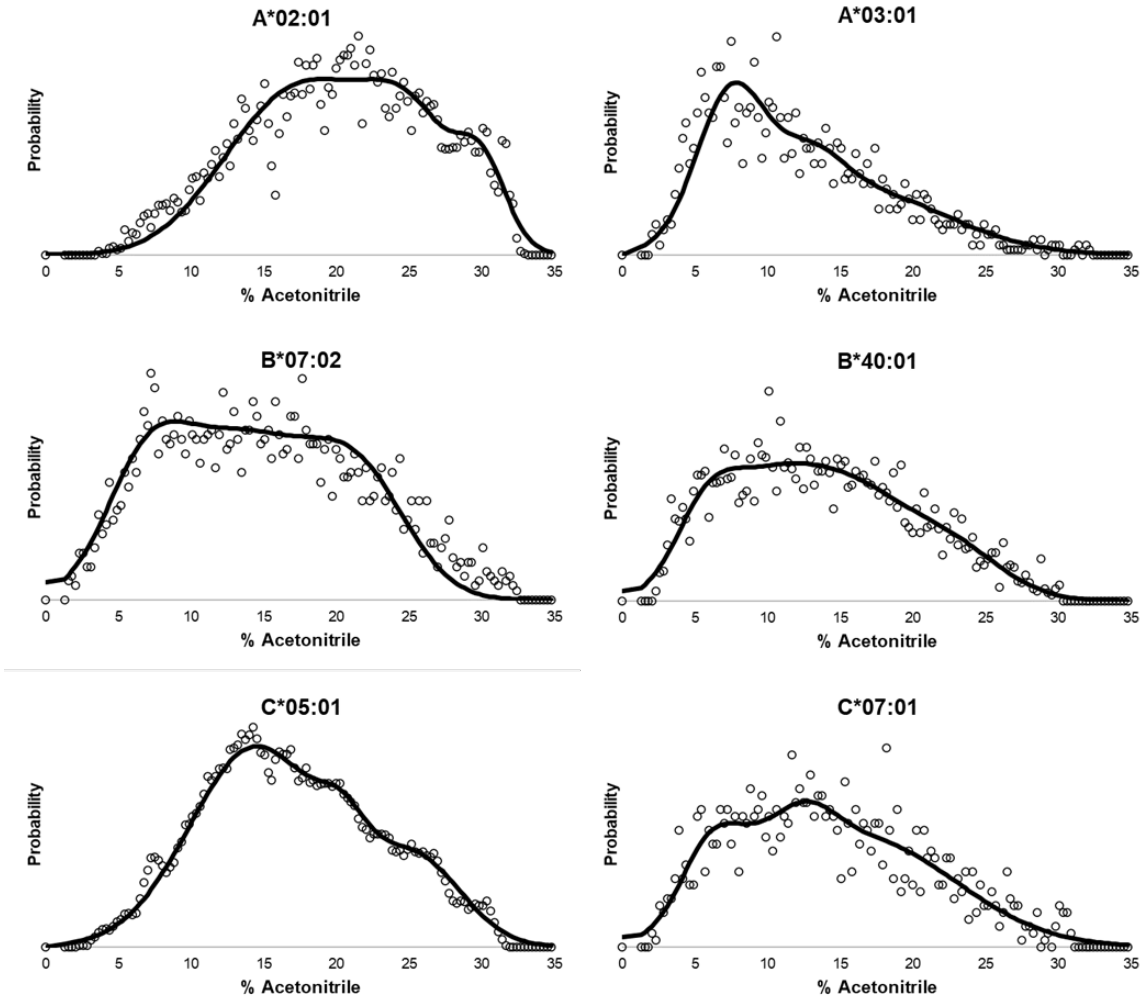
in the polyimide coating using a Bic lighter a few mm in front of the frit, and the burnt coating was removed using a methanol soaked Kimwipe. A ~ 5 μm tip was then pulled within the clear window using a laser puller. A Teflon sleeve was then pushed over the back of the column to ~2 cm from the tip, and the back of the column was trimmed to leave 2-3 cm behind the packed bed. The analytical column was then rehydrated in 0.1% acetic acid at 800 psi for 10-15 minutes or by a 10-minute gradient on the HPLC.

The analytical and precolumn were then connected using a PicoClear 360 union and two 20-minute gradients were run over the connected columns. The precolumn was then disconnected, loaded with 0.1 μg HeLa digest, reconnected to the analytical column, and a 40-minute gradient from 0-40% acetonitrile in 100 mM acetic acid was recorded to test the column's quality.

Appendix B Supplementary Information

B1. Examples of curve fitting to raw data on MHC I elution by motif.

Coefficients of determination for all types are in table. Motifs were assigned based on predicted affinity by netMHCpan4.0 assignments to known sample types. Peptides were collected from 2-5 samples and included at minimum 300 peptides for each



HLA Type	R^2	# peptides
A*01:01	0.96	1579
A*02:01	0.96	7271
A*03:01	0.94	1599
A*11:01	0.94	1316
A*24:02	0.95	2563

HLA Type	R^2	# peptides
B*07:02	0.95	2873
B*08:01	0.92	1338
B*35:01	0.93	2510
B*40:01	0.97	4877
B*44:02	0.96	869

HLA Type	R^2	# peptides
C*03:04	0.92	303
C*04:01	0.96	1224
C*05:01	0.99	2221
C*06:02	0.93	1549
C*07:01	0.95	1420
C*07:02	0.95	1999