INVESTIGATION OF THE EFFECT OF GENETIC POLYMORPHISM ON THE SELECTION OF ANTIGENIC PEPTIDES IN MAJOR HISTOCOMPATIBILITY COMPLEXES

by

Chia-Wei Kuan

B.Sc., The University of British Columbia, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Genome Science and Technology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

March 2015

Abstract

Major histocompatibility complexes (MHCs) play a prominent role in the human adaptive immune system by presenting peptides derived from both host and foreign sources on the cell surface to T cells and eliciting appropriate immune responses during pathogenic invasions. MHC genes are highly polymorphic and the effect of polymorphism on the phenotype, known as an individual's immunopeptidome, is still unclear. In this thesis, two independent but complementary methods of research were conducted to better understand the interaction between MHC alleles and the identities of peptides presented. First, the antigen presentation machinery was reconstructed *in vitro* for class II MHCs. This was accomplished by cloning and expressing HLA-DM and HLA-DR in insect cells and purifying the proteins via affinity and size exclusion chromatography. While DM was successfully purified, DR was not. However, once established the *in vitro* system will offer a novel way to deduce the preferred binding residues for any MHC allele or combinations of alleles, information traditional immunoprecipitation experiments cannot obtain.

Next, in an effort to achieve higher confident assignments of class II MHC binding residues, a cell surface acid elution protocol was developed and performed on consanguineous B cell lines. Extracted peptides were identified using liquid chromatography tandem mass spectrometry. To verify that most surface peptides originated from MHCs, lentiviral shRNA was used to knock down HLA-A prior to acid elution, and the identities of peptides were compared to those obtained from the same cell line transduced with a non-targeting shRNA sequence. Results followed anticipated trends and validated the technique as a means to extract MHC peptides. Furthermore, the nature of consanguineous data sets allows for intra-experimental comparisons to decipher allele-specific peptides. Ultimately, these experiments present new ways to study the immunopeptidome and possess the potential to be applied to the vaccine development research field in the future.

Preface

This thesis is original work by the author. The bioinformatics component in chapter 3 is the work of lab member Dr. Q. Chan and the combination of our work is currently under preparation for manuscript submission. Ethics approval was not required for this research.

Table of contents

List of tables

List of figures

List of abbreviations

AGC	automatic gain control
b2m	β2-microglobulin
CV	column volume
ER	endoplasmic reticulum
ESI	electrospray ionization
FDR	false discovery rate
IP	immunoprecipitation
KD	knockdown
LC	liquid chromatography
LC-MS/MS	liquid chromatography tandem mass spectrometry
MALDI	matrix-assisted laser desorption ionization
MHC	major histocompatibility complex
MS	mass spectrometry
RT	retention time
SFM	serum-free medium
TAP	transporter associated with antigen processing
TBS	Tris-buffered saline
TCR	T cell receptor

Acknowledgements

For a science student, I probably shouldn't say things along the lines of "fate decided that…", but for the way I entered the lab there was no other explanation. Looking back now, I am glad I secured an undergrad thesis student position in Dr. Leonard Foster's lab, who I owe big thanks to for providing me with the opportunity to pursue my interest in biochemistry research. I am also very grateful for my committee members Drs. Wilfred Jefferies and Filip Van Petegem for their guidance and suggestions throughout my Master's project.

I also want to thank all the members of the Foster lab for helping me in one form or another throughout my journey. Special thanks to Hongbing for patiently teaching me all the molecular biology techniques and tricks, as well for his words of encouragement when seemingly all experiments yielded negative results. To Queenie, a million thank yous aren't enough for your help dating back to my final undergraduate year; thank you for making me a better scientist through high standards and expectations, and thank you for the life lessons to make me a better person. I will also miss telling you the tales of heroes of Warcraft. To Dr. Nick Scott (title strictly required), I will greatly miss our LC/MS theory chats as well as your dedication and passion for science, you truly are the Batman among post-docs. To Megan, I am glad Nat was gone for a month so I could volunteer you to help me validate the acid elution data and prove to myself that I can teach my own undergraduate student, much like when Queenie took me under her wing. A huge thank you to Amanda who surgically went through my thesis and gave me many helpful comments to polish my story, and for always lending a hand around the lab.

Lastly, I would like to thank my friends James W. and Gerard for helping me prepare for the defense with insightful questions, and James L. and Alice for constant moral support in exchange for poking fun at me. Your high expectations for me in areas outside of science – in games, on the basketball court, or with a microphone – kept my motors turning and reminded me to continue pushing my own limits in everything I do.

1. Introduction

1.1 Adaptive immune system and the immunopeptidome

The immune system is comprised of biological structures and mechanisms that defend our bodies against diseases. There are two main branches of the immune system in vertebrates – innate and adaptive. The innate immune system, also known as the non-specific immune system, is the body's first line of defense. It consists of anatomical barriers (e.g., skin, mucus in nasopharynx), cells responsible for inflammation (e.g., macrophages and dendritic cells), and the complement system, which aids in the removal of pathogens from the body via actions such as opsonization and agglutination. The adaptive, or acquired, immune system consists of cells that develop immunological memory, enabling an enhanced response which can be mounted against the same pathogen upon subsequent invasions into the body. This thesis focuses on the adaptive immune system so it is discussed here in more depth.

1.1.1 Components of the adaptive immune system

For the adaptive immune system to create immunological memory, antigens are first broken down into peptides and presented on proteins known as major histocompatibility complexes (MHCs). These transmembrane proteins are found on the surface of all nucleated cells. Antigens can be self or foreign (e.g., bacteria or virus-derived), and under normal circumstances only the foreign peptides elicit an immune response. MHC-peptide complexes interact with T cell receptors (TCRs) to form what is known as the immunological synapse, and T cells become activated if they recognize peptides as foreign. Differentiation between self and foreign peptides is established through positive and negative selection during maturation of T cells in the thymus; only those with TCRs

that bind to self-peptide with low affinity are allowed to mature and released into the lymphatic system[1].

The type of immunological response elicited in the adaptive immune system depends on the MHC class and the type of cell presenting a peptide. When a peptide is presented on class I MHC, CD8+ T cells, also known as cytotoxic T cells, release enzymes that destroy the presenting cell by producing reactive oxygen species and initiating apoptosis via the caspase and/or FasL/Fas pathways[1]. In contrast, when a peptide is presented on class II MHC, CD4+ T cells, or T helper cells, are activated. Under this circumstance, if a B cell is presenting peptides, then the humoural immunity pathway will be activated via production of antibodies; if other antigen presenting cells (e.g., macrophages and dendritic cells) are presenting peptides, then the cell-mediated immunity pathway will be activated, leading to inflammation and clearance of diseased cells[1].

Since MHC is directly involved in the antigen presentation process, it has been the subject of various studies including kinetics[2, 3], function[3], and peptide repertoire[4]. MHC genes possess the highest number of alleles in the human genome, and the relationship between polymorphism and the properties of peptides presented to T cells has been a key focus area in the immunology field[5, 6]. Before examining the characteristic of these peptides, it is important to first review our current understanding of MHC, with focus on its structure, function, and similarity and differences between classes.

1.1.2 Major histocompatibility complexes (MHCs) – structure, function, and classification

By definition, MHCs can be classified into three groups – class I, class II and class III. Of these, only the first two classes are involved in antigen presentation, while class III genes encode for components of the complement system, cytokines, and heat shock proteins[7]. Both class I and class II complexes are constituted by two polypeptide chains[1]. Class I MHCs are made of an α chain and an invariant protein β2-microglobulin (b2m),

whereas class II MHCs are made of an α chain and a β chain. While both class I and class II present antigenic peptides, there are moderate differences between how they are assembled, how peptides are loaded, what type of peptides can be loaded, and the pathway they take to end up on the cell surface for recognition by T cells.

Class I MHCs in humans (human leukocyte antigens, or HLAs) are expressed on most cells[1]. They can be subdivided into two categories – the classical, highly polymorphic, antigenic peptide presenting HLA-A, -B and -C, and the non-classical, less polymorphic HLA-E, -F, -G, -K, and -L. As this thesis is focused on the HLAs directly involved in the immunological synapse, only HLA-A, -B, and -C will be reviewed here for class I MHCs. Classical class I MHCs bind to peptides generated from degradation of cytosolic proteins by the proteasome[8, 9]. The antigenic peptide in the cytosol is translocated into the lumen of the endoplasmic reticulum (ER) by the transporter associated with antigen processing (TAP)[10]. With the help of tapasin, calreticulin, calnexin, and Erp57, it gets loaded onto a nascent class I MHC molecule[11]. From here, the MHC-peptide complex leaves the ER through the secretory pathway to reach the cell surface.

As class I MHCs present peptides derived from the cytosol, the majority of peptides presented are from the host cell itself. These peptides do not trigger an immune response from CD8+ T cells. In the case of a pathogenic invasion where a virus or a bacterium is producing proteins inside the host cell, these foreign proteins will also be broken down by the proteasome and have their peptides presented on class I MHCs[1]. Under this circumstance, cytotoxic T cells will recognize the infected cells and signal for apoptosis.

Class II MHCs in humans consist of the antigen presenting proteins HLA-DP, -DQ and -DR, as well as HLA-DM and -DO. Unlike class I, class II MHCs are only expressed by professional antigen presenting cells, which include B lymphocytes, macrophages, and dendritic cells[1]. It is worth noting however that certain cells can express class II MHCs upon stimulation by interferon-γ[12]. Class II MHCs bind to peptides that are derived from endocytosed proteins, usually from bacteria or dead cells[13]. Upon being taken into the cell, these exogenous entities are broken down into peptides in the lysosome, and eventually loaded onto class II MHCs.

Like all other proteins, class II MHCs are synthesized in the ER. Assembly of the α/β heterodimer is facilitated by a third protein, the invariant chain (Ii). This polypeptide occupies the binding pocket of class II MHCs and is subsequently removed as MHC travels from the ER to the Golgi apparatus, and finally into a vesicular compartment[13]. Upon fusion with a late endosome that contains exogenous antigenic peptides, the invariant chain is processed by enzymes called cathepsins to a short peptide known as CLIP, which still occupies the binding pocket of MHC[13]. In order to load antigenic peptides onto HLA-DP, -DQ, and -DR, HLA-DM facilitates the removal of CLIP[2]. HLA-DM has also been shown to enrich the presence of some peptides at cell surface while diminishing the presence of others[14, 15].

In addition to HLA-DM, B cells and thymic epithelial cells also express HLA-DO, which is thought to be the modulator of DM activity[16-18]. A recent crystallography study by Guce et al. agrees with the current model that DO acts as a substrate mimic and binds to DM, inhibiting its ability to catalyze the CLIP removal/peptide loading reaction in the endosome[3]. Interestingly, Poluektov et al. showed that some peptides have enhanced binding affinity in the presence of DO in an *in vitro* kinetics study[19]. In summary, the mechanism of action for DM is still not clear, but generally its functions are agreed upon. However, there is still some uncertainty as to how DO affects DM's ability to load peptides onto DP/DQ/DR.

1.2 The immunopeptidome and the tools to study it

The term "immunopeptidome" refers to the set of peptides presented by class I and II MHCs at the cell surface, including both self and foreign-derived epitopes. Since the differences can be confusing, it is important to note that an antigen is the whole molecule that binds to an antibody, whereas an epitope is the specific surface with which the antigen interacts with the antibody. Not all peptides derived from a protein are presented on MHCs. Additionally, the abundance of peptides presented on MHCs does not necessarily correlate to the abundance of the proteins from which they were derived

from. Given the high degree of polymorphisms in MHCs, it follows that each individual possesses distinctively different three-dimensional MHC structures that lead to different epitopes being presented for a given antigen. Interestingly, it has been shown that polymorphism in MHC do not occur on random residues, they are mostly located within the peptide-binding cleft[20]. In addition, it is important to acknowledge the phenomenon known as immunodominance, which limits the number of peptides that can potentially be recognized as epitopes on MHCs. The exact molecular mechanism for immunodominance remains unknown to date, and is another important field of study in antigen presentation[21-23].

There are two major approaches to study the immunopeptidome. The first is by reconstituting the antigen presentation complex *in vitro*. The second is via an *in vivo* approach where MHC peptides are isolated from human cell lines. Experiments for both were conducted in this thesis, and their purposes and applications are detailed below.

1.2.1 *in vitro* reconstitution of antigen presentation complex

Although MHCs are highly polymorphic, some alleles are less common than others. This means that if one were to use cell lines or primary cells to study the immunopeptidome, some alleles simply could not be studied. While studying common alleles seems to be the most logical and cost-effective approach, there is merit in discovering potential new epitopes from rarer alleles, as pathogens are less likely to adapt to them[24]. Another complication is that any cell system is going to contain all 6 classical MHC molecules that present peptides (HLA-A, -B, -C, -DP, -DQ, -DR), so one would also need to perform additional analyses to assign an epitope to a particular type of MHC.

One approach that would bypass both issues is *in vitro* reconstitution of the antigen presentation complex. This technology was recently developed for both class I[25] and II[26] MHC, but has been mostly used to study mechanisms of action[2, 19, 27]. The same technology, however, can be used as an assay to increase our understanding on the

peptide repertoire of individual MHC alleles, or to examine combinations of alleles to observe any potential interactions between different MHC sub-classes (i.e., does the abundance of a peptide change in the presence of a second MHC as a result of competition?). For both classes, baculovirus infection of insect cells has been the method of choice, because the post-translational modifications required are not available in bacterial systems, and higher protein production can be achieved per cell in an insect system than a mammalian system[25].

To reconstitute the class I antigen presentation system, in addition to the α chain of HLA-A, -B, or -C and b2m, chaperones and other proteins involved with proper folding, positioning, and peptide loading of class I MHC such as calreticulin, Erp57, and tapasin are required for mimicking peptide presentation[25]. Specifically, a total of three constructs are needed: co-expressed soluble α chain and b2m, calreticulin, and co-expressed Erp57 and soluble tapasin[25]. Although TAP participates in the peptide loading process *in vivo*, studies using soluble tapasin have shown that high-affinity peptide loading can proceed without TAP incorporation in an *in vitro* system[28-30]. Once all three constructs are successfully expressed and purified, one can evaluate the binding affinity of peptides by, for example, incubating the assembled peptide loading complex with peptides tagged with radioactive labels[25].

To reconstitute the class II antigen presentation system, soluble forms of a peptide-presenting MHC heterodimer (HLA-DP, -DQ, or -DR) and the editor HLA-DM are expressed and purified separately. Since the peptide-presenting MHCs are susceptible to proteolysis by cathepsins[31], MHCs are first incubated with the antigen to bind with highest affinity targets, then cathepsins are added for a shorter incubation period. Finally, the peptides can be examined by immunoprecipitation of MHC followed by elution in mild acid condition[32].

1.2.2 "*in vivo*" extraction of MHC peptides by immunoprecipitation and acid elution

While *in vitro* reconstitution of the antigen presentation complex can provide valuable insight on the peptide repertoire of individual MHC alleles, the caveat that always remains for such systems is whether the results are reflective of what really goes on *in vivo*. This section will examine two MHC peptide extraction methods from antigen presenting cells and weigh their respective pros and cons.

The first choice is to purify the MHC-peptide complex out of the cell by immunoprecipitation, then proceed to strip off the peptides[6]. The advantage of this technique is in the yield's purity, that is, all peptides eluted off should theoretically be an antigenic epitope. Immunoprecipitation also allows some individual MHC sub-classes to be purified, but suffers from not having a universal antibody that captures all MHCs in one experiment.

The second method is to simply elute off all surface peptides from the cell using mild acid conditions[32]. This technique, in contrast to immunoprecipitation, requires fewer steps and captures peptides presented by all MHCs. As a result, fewer cells are required to perform an acid elution. The major disadvantage to this method is that by eluting all peptides off the surface of a cell, one now has to demonstrate that the peptides are of MHC origin. Validation can be performed either through biological experiments such as gene knockdown or statistical methodologies, which can be used to filter out the less likely candidates[5].

1.3 Peptide identification by mass spectrometry

For large scale identification of peptides, liquid chromatography tandem mass spectrometry (LC-MS/MS) is usually employed[33]. Conventionally, LC-MS/MS is used to identify peptides derived from trypsin-digested proteins. In this approach, proteins of

interest (e.g., whole cell lysate, proteins from a subcellular compartment) are first incubated with dithiothreitol to reduce disulphide bonds between cysteine residues. This is followed by alkylation of these residues to prevent reverse oxidation. Finally, the proteins are treated with trypsin, resulting in peptides with C-terminal arginine or lysine residues[34]. The sample then undergoes a cleanup procedure where trypsin, other undigested proteins, and potential cell debris/membrane are removed, usually by a C18 cartridge or what is known as STAGE-tips[35]. The purified peptides are then eluted and injected into LC-MS/MS for identification and quantification.

A mass spectrometer measures the mass-to-charge ratio (m/z) of analytes, and tandem MS is employed in MS-based proteomics to determine the identity of the peptides and the proteins they were derived from[33]. Typically, two modes of scans are involved in tandem MS, known as MS1 and MS2. In MS1, the mass spectrometer records the ion abundance of all peptides, commonly referred to as precursor ions. It then selects and fragments peptides with the highest abundance one by one in a data-dependent manner. For each of these peptides, their fragments or product ions are recorded in subsequent snapshots by the mass spectrometer (MS2), and then the cycle repeats itself. Peptide fragmentation greatly enhances a database's ability to identify the correct peptide, as product ions usually provide information on amino acid sequences that are unobtainable with a MS1 scan alone. Modern Orbitrap instruments allow for scan rates on the order of 0.1second, enough time to yield multiple MS1/MS2 cycles in the duration of a chromatographic peak from LC[36].

1.3.1 Liquid chromatography-tandem mass spectrometry instrumentation

For identification of peptides by MS, the ideal front-end instrument is a reverse phase LC system[37, 38]. In such a system, samples are suspended in an acidic aqueous solution, typically known as Buffer A, and injected onto the column (the stationary phase), which is initially flushed with the same buffer (the mobile phase). The column is constituted of hydrocarbons, usually C18, with various modifications for resolving different classes of

compounds. Over time, the percentage of Buffer A in the mobile phase is decreased and replaced with an acidic organic solvent, usually designated as Buffer B. As this occurs, peptides bound to the stationary phase experience increasing competition in hydrophobic interaction with the mobile phase, eventually leading to their elution. The net result is a chronological profile of peptides ranked by increasing hydrophobicity. By separating peptides before they enter the mass spectrometer, the identification rate is dramatically increased compared to direct injection because the likelihood of co-eluting peptides with similar masses is greatly reduced.

There are many types of mass spectrometers, but all of them consist of the following essential parts: the ion source, which converts samples to gaseous ions; the mass analyzer, which discriminates one molecule from another based on their m/z; and the detector, which registers the amount of individual ions with a specific m/z. To analyze peptides by MS, they must be transferred into the mass analyzer without fragmenting during the ionization process. This is achieved by soft ionization techniques such as matrix-assisted laser desorption ionization (MALDI[39, 40]) or electrospray ionization (ESI[41]). These techniques create ions with low internal energy, such that they are very unlikely to fragment during the ionization process.

In MALDI, samples are dried and crystallized along with a matrix on a metal plate. The matrix is typically a small, crystallize-able weak organic acid that absorbs light strongly in the UV/Vis region and is used to assist the ionization of the sample. A laser is then pulsed at the sample/matrix mixture, causing an ablation at the surface and desorbing both materials into a plume. Many collisions occur inside the plume, leading to protonation (or a salt cation getting attached, forming an adduct) of the analyte and its detection by the mass analyzer[42]. Typically, MALDI is well suited for samples that contain a few species to be identified. Given the complexity of MHC peptides, the technique is not compatible for exploration of the immunopeptidome.

Electrospray ionization, on the other hand, does not require crystallization of sample prior to ionizing the analyte. In ESI, the analytes are dissolved in an aqueous solution. Prior to entering a mass analyzer, it is passed through a thin needle known as the emitter. A high electric potential (2-6 kV) is applied between the emitter and the inlet of

the mass analyzer component, which is located a short distance away orthogonally. This creates an electrically charged spray from solution, aerosolizing and ionizing the analyte in the process[33]. As the sample is in aqueous form, ESI can be easily coupled to an LC separation system, which offers reduction of sample complexity and a more streamlined methodology comparing to MALDI. As a result, ESI has become the most commonly used ionization technique to analyze complex biological samples[33, 43].

1.3.2 Database search

The most crucial step in any proteomics experiment is peptide identification and assignment to the protein it was derived from. In general, bioinformatics software uses either correlation-based or probability-based search algorithms to identify peptides and proteins[44]. In correlation-based algorithms, such as the one used by SEQUEST[45], potential peptide candidates are predicted based on MS2 scans. The peaks from product ions are given a summary score that takes into account the fragmentation method used (e.g., collision dissociation vs. electron transfer) and rewards for parameters such as consecutive residue matches and presence of immonium ions. This score is then mathematically cross-correlated with theoretical spectra, and the highest scored peptide is assigned to the experimental spectrum.

In probability-based approaches, as used by Andromeda[46], candidate peptides are also generated from fragment ion spectra. Much like SEQUEST, Andromeda creates a list of potential theoretical fragment ions based on the fragmentation method and the type of mass analyzer used. Peptide assignment is done by calculating the likelihood of candidate peptides generating the fragment spectra observed, and the one with the highest probability is assigned as the correct sequence.

Given the vast amount of potential peptides possible in a human proteome, validation of putative peptide matches and removal of false identifications are required before quantification. This is typically done through decoy searching[47], which involves searching against a peptide library generated by either scrambling or reversing the

peptide sequences in the proteome. The number of matches to these decoy peptides is used to estimate the false discovery rate (FDR), which is typically set at 1%. That is, the stringency of the search is set such that only 1% or less of the peptides assigned by the bioinformatics software are allowed to match to a sequence within the decoy database. After eliminating the false positives, the peptides that remain are now considered as true matches, and can be further analyzed for biological implications (e.g., quantification).

It is important to note that it is much harder to get the FDR below 1% for MHC peptides identification than it is for a conventional trypsin digest experiment, as MHC peptides can have any amino acid at their C-terminal. In this case, the database generates significantly higher number of peptides from the same number of proteins, and ultimately leads to less identification at the same FDR threshold.

1.3.3 Quantitative proteomics

The objective for quantitative proteomics is to compare the amount of proteins across two or more experimental conditions. There are two main types of quantification methods: differential isotopic labeling and label-free quantification. In the first approach, quantification is achieved by modifying each peptide such that a systematic mass shift can be observed between different experimental conditions, allowing the mass spectrometer to distinguish the same peptides in different samples. There are two common ways to carry out a differential isotopic labeling experiment. The first of which is by metabolically incorporating stable isotope amino acids. For example, for a trypsin digest proteome analysis of a cell line, normal culture medium is used in one experimental condition, while a culture medium containing carbon-13 substituted arginine and lysine is used in another experimental condition. Samples are mixed in equal proportions prior to digestion, and quantification is achieved by comparing the difference in peak intensities for each isotopic pair. The second differential isotopic labeling technique is chemical modification. This is performed by tagging specific amino

acids at the protein or peptide level and is usually done when metabolic incorporation is difficult or impossible. The principle for quantification is the same, however there are a wide selection of reagents which allow the mass difference to show at either the MS1 (e.g., isotope-coded affinity tags[48]) or MS2 level (e.g., tandem mass tags[49]), whereas quantification for metabolic labeling always result in mass shifts at the MS1 level. Chemical modification has the drawback in form of higher experimental variability, as the labeling step is performed later in the experimental protocol. Beside the cost and feasibility considerations, labeling peptides introduces more chemical species to be analyzed in any given LC/MS experiment. This can lead to reduction in coverage and an increase in experimental time frame when compared to label-free methods.

Label-free quantification methods can be broadly categorized into two approaches: spectral counting and ion abundance. Spectral counting is the less accurate of the two, where peptide abundance is calculated by simply counting the number of MS2 spectra assigned to that peptide, when the spectra are observed within the elution time of the peptide in question. This can greatly bias quantification results for larger proteins and longer peptides, for which more mass spectra are generated on average relative to smaller proteins and shorter peptides, assuming the same protein or peptide abundance in the sample. The second approach is more accurate and is done by measuring the extracted ion current of individual peptides at the MS1 level after identification at the MS2 level[50], as an extracted ion current represents the continuous signal of a peptide as it is recorded by the detector. Although label-free quantification methods are less accurate than their isotopic labeling counterparts, they offer a simpler work flow, higher dynamic range, and can be applied to more unconventional experiments, such as quantifying non-tryptic peptides, without having to worry about an efficient labeling strategy. Recent advances in high resolution MS1 instruments such as the Q-Exactive also lessen the accuracy disadvantage of label-free quantification[51].

1.4 Project outline

In this thesis, two methods are used to examine the relationship between MHC polymorphism and the peptide repertoire, with the working hypothesis that each individual's (excluding identical twins) MHC peptide repertoire is unique and that this uniqueness is shaped by the polymorphisms found in the MHC genes. First, the class II MHC antigen presentation system was reconstituted *in vitro*, covered in chapter two. Considerations and procedural schemes for cloning, transfection, and purification are described, and optimizations for purification protocols will be the focal point of discussion, as it was the stage where most troubleshooting experiments were performed.

The second approach to study the dynamics between MHC polymorphism and the immunopeptidome is by analyzing surface peptides extracted from consanguineous cell lines, covered in chapter three. The chapter first assesses the pros and cons of this novel method in contrast to the traditional immunoprecipitation experiments used to study MHC peptides. This is followed by a detailed peptide extraction protocol and a shRNA knockdown experiment to demonstrate the technique's validity for generating MHC peptide data through confirmation of expected biases.

2. *In vitro* reconstitution of the antigen presentation complex

2.1 Introduction

Deriving the epitopes for MHC alleles has been a major focus in the biomedical field for diseases that are not well understood, such as cancer[52-54]. A majority of these studies isolate MHC peptides via immunoprecipitation, in which MHCs are purified by affinity purification from cell lysate, and then the antigenic peptides are eluted off for identification. However, this approach is dependent on having access to a cell line with MHC allele(s) of interest, and given that some alleles are more common than others, this can be a limiting factor in studying less frequently occurring alleles. A method which can theoretically derive peptides from any allele is therefore of great interest, and this was accomplished in the Sadegh-Nasseri laboratory at Johns Hopkins University[26] for the MHC class II system. Briefly, soluble domains of MHC subunits α and β are cloned and expressed within a single vector, pFastBac Dual (Life Technologies; Burlington, ON), which can be readily converted into transfection-ready bacmids with the cell line DH10Bac (Life Technologies). Once this is complete, the bacmid can be used to transfect insect cell lines such as Sf9 and High Five cells. The proteins can then be purified from supernatant by affinity chromatography followed by size exclusion chromatography.

In this study we aim to further develop this technique into an assay in which any combination of MHC alleles can be used to examine the immunodominant peptides presented.

2.2 Experimental methods

2.2.1 Materials

Materials required for cloning, transfection, cell culture, and protein purification were obtained from the following sources:

- PCR purification kit, plasmid miniprep kit, DH10B and DH10Bac competent cells (Life Technologies; Burlington, ON)
- Gel extraction kit (Qiagen; Toronto, ON)
- Salts and chemicals of highest grade, e.g., Luria broth/agar (Sigma-Aldrich; Oakville, ON)
- BamHI, HindIII, XhoI, NheI, T4 DNA ligase and dNTP mix (New England Biolabs; Whitby, ON)
- Pfu Turbo DNA Polymerase (Agilent/Stratagene; Mississauga, ON)
- HLA-DMA*01:01:01:01 and HLA-DMB*01:01:01:01 were generously shared by the Sadegh-Nasseri laboratory
- HLA-DRA*01:01:01:01 (DNASU plasmid repository; Tempe, AZ)
- HLA-DRB1*15:01:01:01 (Origene; Rockville, MD)
- Sf9/High Five cells and associated media: Sf-900 II and Express Five serum free media; Cellfectin II transfection reagent; fetal bovine serum (Life Technologies)
- Antibodies: anti-FLAG tag (F1804, Sigma-Aldrich), anti-cMyc tag (05-724, EMD Millipore; Etobicoke, ON), anti-HLA-DM (ab55149, Abcam; Toronto, ON), and anti-HLA-DR (ab92511Abcam)
- Anti-FLAG M2 affinity gel (Sigma-Aldrich)
- Prep columns (Bio-Rad; Mississauga, ON)
- Bio-Sep SEC s4000 size exclusion column (Phenomenex; Torrence, CA)

2.2.2 Molecular cloning

For the reconstitution of the class II antigen presentation system, the expression of HLA-DM (the editor) and a peptide presenter (HLA-DP, -DQ, -DR) are required. In this thesis, the procedure for the production of HLA-DMA*01:01:01:01/HLA-DMB*01:01:01:01 (abbreviated as DMA/DMB from this point onward) and HLA-DRA*01:01:01:01/HLA-DRB1*15:01:01:01 (abbreviated as DRA/DRB from this point onward) will be detailed. When received from the Sadegh-Nasseri laboratory, DM subunits were already embedded within the pFastBac Dual vector. For DR subunits, soluble domains (bp 1-651 of ORF) of DRA was amplified with the addition of BamHI and HindIII restriction cut sites, and soluble domains (bp 1-681 of ORF) of DRB was amplified with the addition of XhoI and NheI restriction cut sites. For DMA/DRA, a FLAG tag was added to the 3' end of the gene, and for DMB, a c-Myc tag was added to the 3' end of the gene. A linker sequence GSGS was included between the C-terminal of DMA/DMB and their respective tags. **Table 1** summarizes the primers and PCR conditions used to obtain DM and DR constructs in pFastBac Dual vector.

Table 1 Primers and PCR conditions used to clone FLAG tagged HLA-DRA*01:01:01:01 and HLA-DRB1*15:01:01:01 into pFastBac Dual vector.

DRA - forward	GACTAGGATCCGCCACCATGGCCATAAGTGGAGTCCCT
DRA - reverse	CTGACAAGCTTTCACTTATCGTCGTCATCCTTGTAATCGTTCTCTGT AGTCTCTGGGAGA
DRB - forward	GTCTACTCGAGGCCACCATGGTGTGTCTGAAGCTCCCT
DRB - reverse	GTGAAGCTAGCTCACTTGCTCTGTGCAGATTCAGACCG
PCR conditions: 95°C for 1 min, 31 cycles of (94°C for 30 sec, 50°C for 30 sec, 72°C for 90 sec), 72°C for 10 min	

Note: primers were not required for DMA/DMB as the genes were already within the pFastBac Dual vector when received from the Sadegh-Nasseri laboratory.

DRA and DRB were cloned into the pFastBac Dual construct as described[55]. Briefly, 8 µg of pFastBac Dual plasmid DNA and PCR product of DR genes were subjected to 80 units of restriction enzyme digest for 2 hours at 37°C. The resulting DNA products were run on an agarose gel, and bands corresponding to the correct size of DNA were extracted with a gel extraction kit. Ligation was completed by incubating 250 ng plasmid and various amounts of PCR product ranging from 1:1 to 1:3 ratios at 16°C over night. The ligation products were then transformed into DH10B cells for propagation and bacmid generation after scale-up.

2.2.3 Bacmid generation

Transfection-ready bacmids for insect cells were prepared according to manufacturer's instructions[55]. Briefly, plasmids containing HLA-DM or DR genes were taken up by DH10Bac cells via heat shock at 42°C for 45 sec. Cells were then grown in S.O.C. medium at 225 rpm for 4 hr and selected with kanamycin, gentamicin, and tetracycline for 48 hr. Colonies containing recombinant bacmids appear white as the LacZα peptide sequence is replaced by the HLA genes inside DH10Bac cells. These colonies were grown, the bacmids isolated and verified with PCR/agarose gel electrophoresis (primers and PCR conditions shown in **Table 2**), and scaled up.

Table 2 Primers and PCR conditions used to verify recombinant bacmid constructs prior to transfection.

Bacmid - forward	CCCAGTCACGACGTTGTAAAACG
Bacmid - reverse	AGCGGATAACAATTTCACACAGG
PCR conditions: 95°C for 1 min, 31 cycles of (94°C for 30 sec, 50°C for 30 sec, 72°C for 180 sec), 72°C for 10 min	

2.2.4 Insect cell culture

Sf9 and High Five cells were maintained as described[56]. Both type of cells grow optimally at 27°C. Sf-900 II serum-free medium (SFM) was used to maintain Sf9 cells, whereas High Five cells were grown in Express Five SFM supplemented with 18 mM L-glutamine. Both cell lines were first grown in adherent culture; Sf9 cells required scraping to detach for passaging purposes whereas High Five cells could be sloughed with medium. After ~5-6 passages from frozen stock, cell number can be scaled up by moving to suspension culture, where the cells grow optimally at a density of 1-2X10^6 cells/mL. At this stage, no other action is required for Sf9 cells, but High Five cells tend to clump when transferred from adherent to suspension culture. Therefore, heparin was added, as recommended by the manufacturer at 10 units per mL cell culture, to alleviate this. Heparin concentration was decreased two-fold over each passage so it could be weaned out ~5 passages into suspension culture. This process is very important, as the presence of heparin, a highly negatively charged polysaccharide, negatively affects the efficiency of transfection[57].

2.2.5 Transfection of MHC constructs into insect cells

Since Sf9 cells are easier to transfect, and High Five cells produce more secreted proteins than Sf9 cells[58], bacmid transfection was done on Sf9 cells and High Five cells were used for viral infection and protein expression. Cell viability was verified to be >95% before either experiment was performed. Transfection was done in a 6-well plate where each well was seeded with 2 mL of cells at a density of 4X10^5 cells/mL. For each well, 3 µg of bacmid was diluted into 100 µL Sf-900 II SFM, and 8 µL of Cellfectin II was diluted 100 µL Sf-900 II SFM separately in preparation for transfection. These two mixtures were incubated at room temperature for 30 min, then combined and incubated for an additional 15 min. After washing the cells gently with Sf-900 II SFM and replacing the medium, the transfection mix was added drop by drop while swirling the plate slowly

to ensure maximal coverage. Finally, the plate was wrapped with Parafilm^{TM} and incubated at 27°C until late infection symptoms appear (~72 hr for Sf9 cells), as described by manufacturer's protocol[55]. Once the cells entered late infection phase, the supernatant was extracted and spun in centrifuge at 500 rcf for 5 min. The resulting mixture was used for infection of High Five cells. For virus storage, 2% fetal bovine serum (FBS) was added and the mixture was stored at 4°C away from light.

Viral amplification was performed by adding the supernatant harvested from Sf9 transfection to 1X10^6 High Five cells/mL at a 1 to 50 (vol/vol) ratio. The supernatant from High Five suspension culture, which contains amplified amount of virus and secreted MHC proteins, was harvested once cell viability dropped below 70%, typically at the 72 hr mark. 2% FBS was added to the supernatant, and the mixture was flash-frozen in liquid nitrogen before putting it in storage at -80°C.

2.2.6 Protein purification

The FLAG tag was exploited to purify both HLA-DM and HLA-DR out of the supernatant as described[26]. In summary, Tris-buffered saline (TBS), pH 7.4 was used to rinse an empty column prior to loading of affinity beads. 0.5 mL of the gel, equivalent to at least 0.3 mg binding capacity, was loaded onto the column to purify 80 mL supernatant, which should contain less than 0.1 mg protein (based on personal communication with Dr. Sadegh-Nasseri on protein yield). Before loading the sample, the beads were washed with 3 column volumes (CVs) of 0.1M glycine hydrochloride, pH 3.5, then neutralized with TBS. Each sample was passed through the beads 3 times, washed with 10 CVs of TBS, and FLAG-tagged proteins were eluted with 5 CVs of 100 µg/mL FLAG peptides in TBS. The column was regenerated by washing with 3 CV of 0.1M glycine hydrochloride, pH 3.5, and immediately followed by neutralization with TBS until effluent was at neutral pH.

In addition to purification via traditional gravity filtration, a batch-binding anti-FLAG affinity purification procedure was done by incubating the beads and supernatant inside

a 50 mL tube and rocking for 2 hr. The beads were then collected by centrifugation at 500 rcf for 5 min, loaded onto a gravity column, then washed and eluted as above.

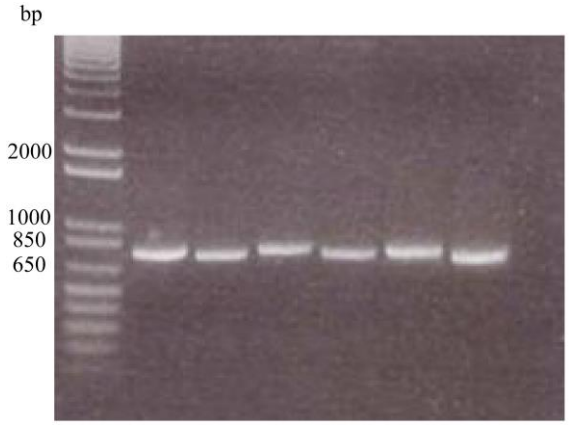
To prepare the eluted proteins for size exclusion chromatography, the sample was concentrated using a molecular weight cut-off filter at 10 kDa, acidified to pH 6.5 with citrate phosphate buffer, and then injected onto the size exclusion column (Bio-Sep 4000). Separation was done in isocratic mode over 30 min with buffer consisting of 50 mM sodium chloride, 50 mM sodium acetate, and 50 mM Tris base, pH adjusted to 7.2 with hydrochloric acid.

2.3 Results

2.3.1 Generation of HLA-DM and HLA-DR constructs

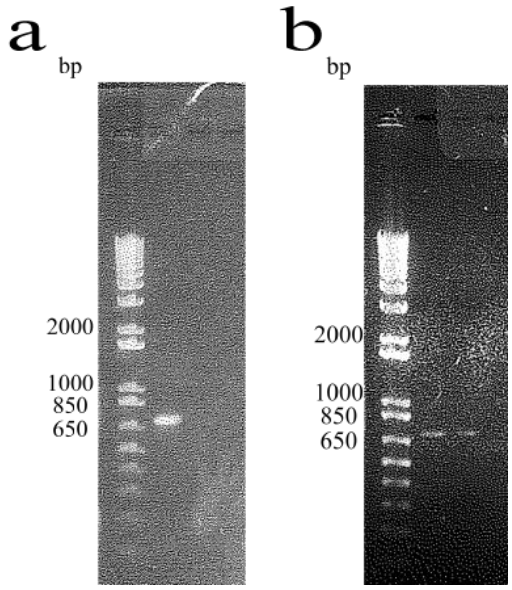
To capture antigen presentation in an *in vitro* setting, the class II MHCs HLA-DM and HLA-DR were cloned into a bacmid, which can in turn be used to transfect cells to produce proteins. The correct bacmid constructs should contain alpha and beta subunits of an HLA protein, complete with baculovirus promoters polyhedrin and p10. These were confirmed by a combination of agarose gel electrophoresis, DNA sequencing, and antibiotic selection. For HLA-DM, pFastBac Dual plasmid containing DMA and DMB (Figure 1) was directly transposed onto the bacmid in DH10Bac cells. Similarly, DRA and DRB genes from commercial vendor plasmids were amplified using PCR, cut with their respective restriction enzymes, then inserted into the pFastBac Dual vector. Gel images of the resulting construct are shown in Figure 2. Once this was done, the plasmid was transformed into DH10Bac cells for transposition, as above.

Figure 1 HLA-DM construct confirmation.



From left to right: lanes 1, 3, 5 - FLAG tagged HLA-DMA*01:01:01:01 (726 bp); lanes 2, 4, 6 - c-myc tagged HLA-DMB*01:01:01:01 (696 bp) in pFastBac Dual vector.

Figure 2 HLA-DR construct confirmation.



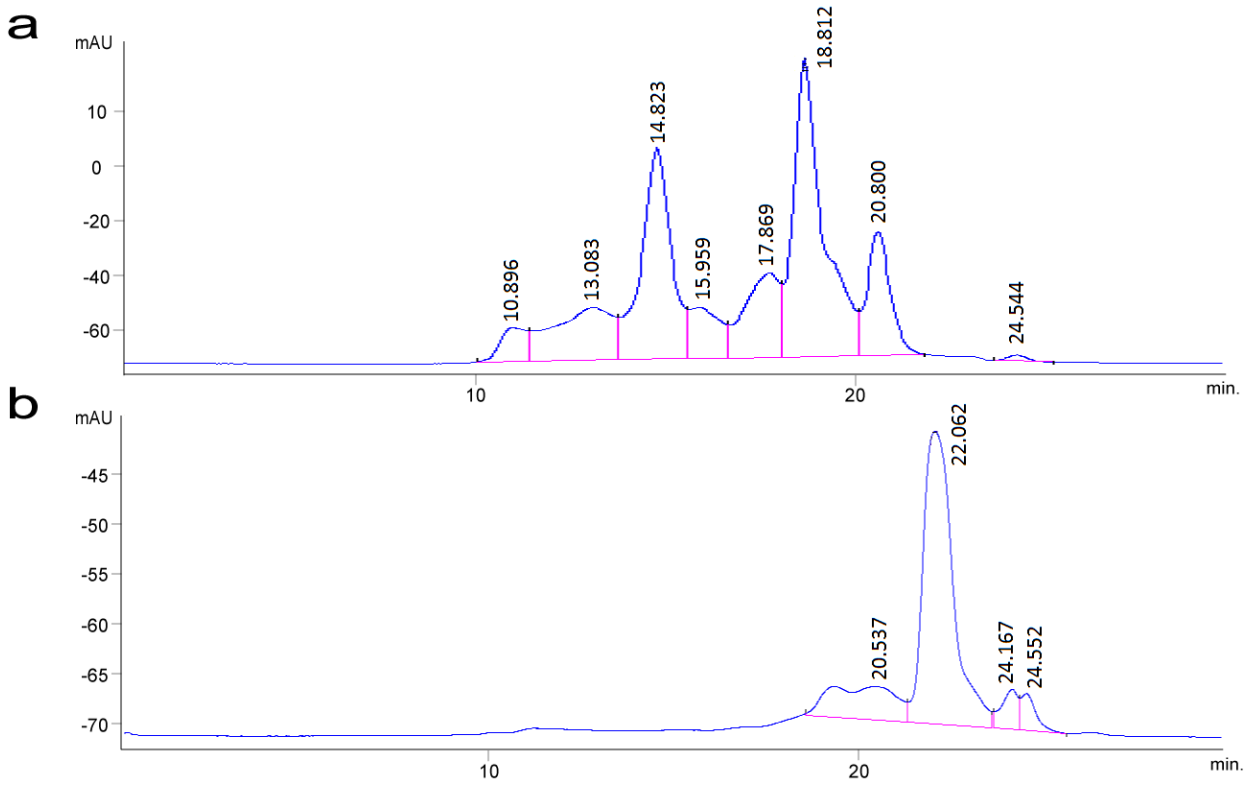
(a) FLAG tagged HLA-DRA*01:01:01:01(675 bp) (**b)** HLA-DRB1*15:01:01:01 (681 bp) in pFastBac Dual vector.

2.3.2 HLA-DM and HLA-DR purification

Following expression of HLA-DM and HLA-DR in High Five cells, the class II MHCs were purified from culture supernatant using a purification scheme of tandem affinity chromatography followed by size exclusion chromatography[26].

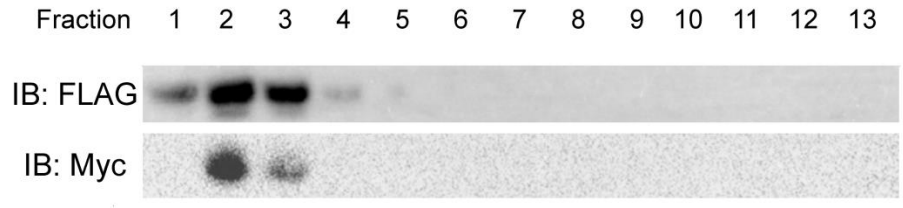
HLA-DM has a molecular weight of 53 kDa, which correlates to a retention time (RT) of approximately 19.5 min based on protein standards (Figure 3a). 280 nm chromatogram from the sample showed a small peak at 19-20 min (Figure 3b), and fractions were collected at 15/30-sec intervals from the chromatograph. Western blot analysis confirmed the proteins isolated from fractions 1 to 4, corresponding to RT 18.5-21 min, were HLA-DM proteins (Figure 4). Fractions were also collected along the process of tandem protein purification, and a western blot analysis showed that binding between DMA-FLAG and anti-FLAG beads was not very efficient (Figure 5).

Figure 3 Size exclusion chromatography profile of anti-FLAG purified HLA-DMA-FLAG/HLA-DMB-c-Myc.



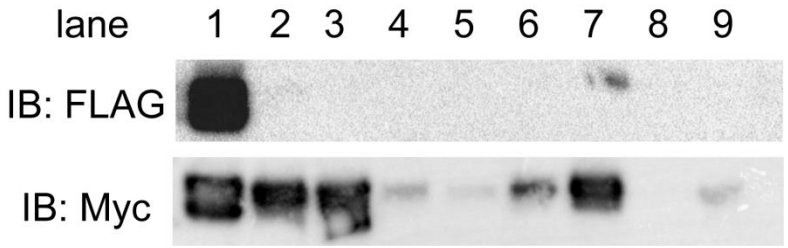
(a) 280 nm chromatogram of protein standards, including thyroglobulin (669 kDa; RT=14.823 min), β-amylase (200 kDa; RT slightly before 17.869 min), bovine serum albumin (66 kDa, dimerizes to 132 kDa; RT=18.812 and 17.869 min respectively), and carbonic anhydrase (29 kDa; RT=20.800 min). **(b)** 280 nm chromatogram of affinity purified HLA-DMA-FLAG/HLA-DMB-c-Myc (53.0 kDa).

Figure 4 Anti-FLAG and size exclusion chromatography successfully purified HLA-DMA-FLAG/HLA-DMB-c-Myc.



Fractions numbers reflect chromatographic retention time: 1 – 18.5-19 min, 2 – 19-19.5 min, 3 – 19.5-20 min, 4 – 20-21 min, 5 – 21.21.5 min, 6 – 21.5-22 min, 7 – 22-22.5 min, 8 – 22.5-22.75 min, 9 – 22.75-23.25 min, 10 – 23.25-23.75 min, 11 – 23.75-24 min, 12 – 24-24.25 min, 13 – 24.25-24.5 min.

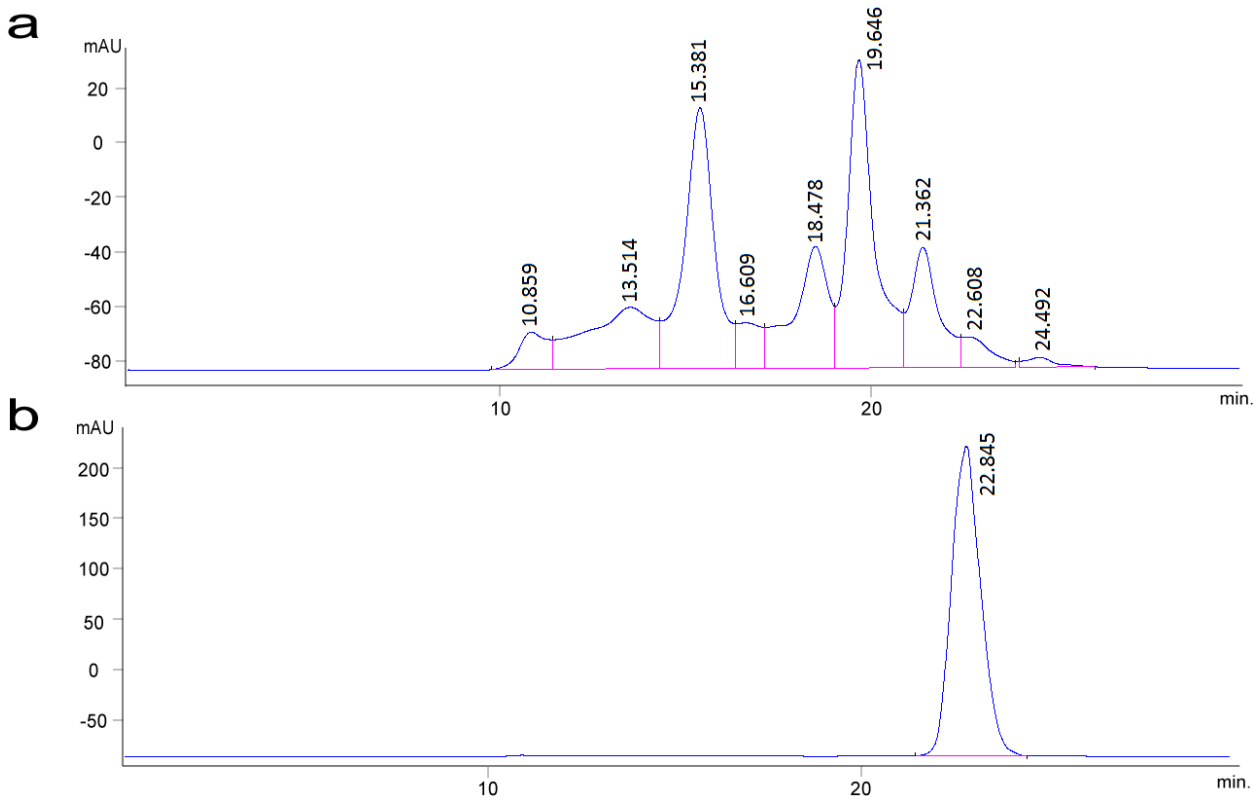
Figure 5 Affinity purification protocol of HLA-DMA-FLAG/HLA-DMB-c-Myc should be optimized for higher yield.



Lanes: 1. High Five cell lysate, 2. 2nd flow through, 3. 3rd flow through, 4. wash (2nd out of 10 CV), 5. wash (3rd out of 10 CV), 6. eluate, 7. acid regeneration (2nd out of 3 CV), 8. neutralization (1st mL), 9. neutralization (2nd mL).

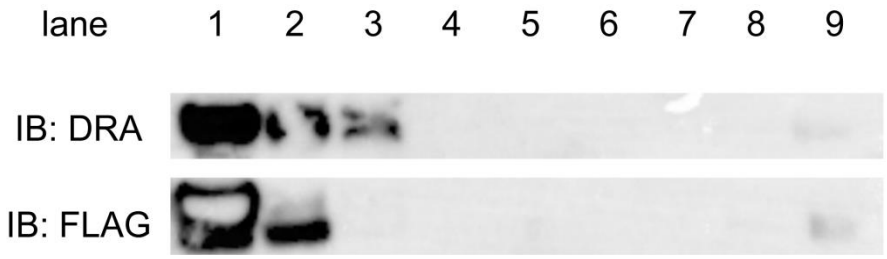
As the purification of HLA-DM did not perform at high efficiency, the protocol was modified to a batch binding method for HLA-DR purification using anti-FLAG beads. However, batch binding affinity purification appeared to perform at even lower efficiency, as none of the fractions from size exclusion chromatography showed the presence of HLA-DR (Figure 6), and western blot analysis of fractions collected along the purification process showed that all the proteins remained in the “flow-through” (i.e., none was bound onto the beads) (Figure 7).

Figure 6 Size exclusion chromatography profile of anti-FLAG purified HLA-DRA-FLAG/HLA-DRB.



(a) 280 nm chromatogram of protein standards, including thyroglobulin (669 kDa; RT=15.381 min), β-amylase (200 kDa; RT=slightly before 18.478 min), bovine serum albumin (66 kDa, dimerizes to 132 kDa; RT=19.646 and 18.478 min respectively), and carbonic anhydrase (29 kDa; RT=21.362 min). **(b)** 280 nm chromatogram of affinity purified HLA-DRA-FLAG/HLA-DRB (50.7 kDa).

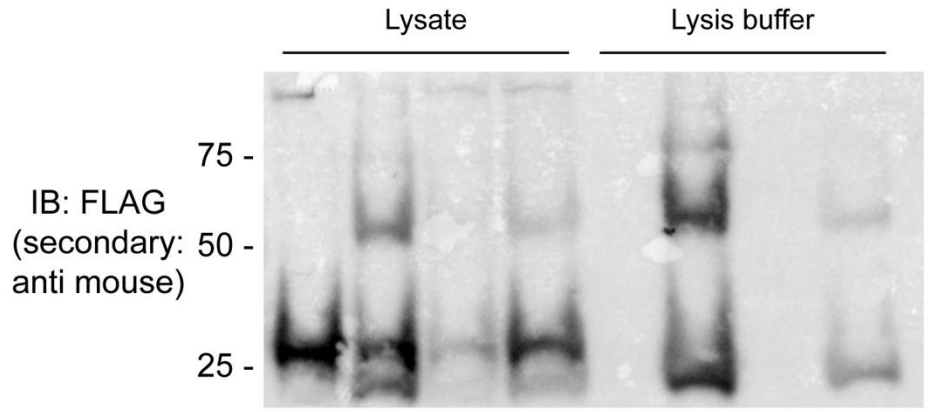
Figure 7 Affinity purification protocol of HLA-DRA-FLAG/HLA-DRB should be optimized for higher yield.



Lanes: 1. High Five cell lysate, 2. High Five cell culture supernatant, 3. “flow through” (solution after binding to beads for 2 hours), 4. wash (2nd out of 10 CV), 5. wash (5th out of 10 CV), 6. eluate, 7. acid regeneration (2nd out of 3 CV), 8. neutralization (1st mL), 9. concentrated affinity purified-protein after MWCO (diluted 12X).

A trouble-shooting experiment was done to check whether the anti-FLAG beads bind to FLAG-tagged proteins as advertised. In this experiment, purification was done in 1.5 mL tubes, and the bead to protein ratio was increased from previous purifications. The Sadegh-Nasseri laboratory reported 1 mg/L as good yield from High Five cells (personal communication). For HLA-DM and -DR affinity purification, 500 µL of slurry, corresponding to binding capacity of at least 0.3 mg, was used for 80 mL of supernatant (maximum yield of ~0.08 mg). This ratio was increased three-fold for trouble-shooting, as the proportional volume of slurry to purify 1.5 mL of supernatant was too small, and the only objective was to observe whether the beads bind to their intended target. The resulting western blot (Figure 8) showed that most DRA-FLAG were found bound to the beads as opposed to being left in the flow-through.

Figure 8 Trouble-shooting batch binding experiment confirms the functionality of anti-FLAG affinity beads.



Lanes: 1. lysate, 2. boiled anti-FLAG beads, 3. “flow-through” of lysate after incubation/binding, 4. leftover beads in incubation tube, 5. lysis buffer, 6. boiled beads that incubated with lysis buffer, 7. “flow-through” of lysis buffer after incubation, 8. leftover beads in incubation tube.

2.4 Discussion

Immunoprecipitation is a useful biochemical technique to extract and study MHC epitopes, however it is dependent on the availability of appropriate antibodies and alleles. On the other hand, an *in vitro* toolkit can also be used to identify antigenic peptides, with the added flexibility to examine any allele or combinations of alleles. The development of this technology will further our understanding on the role of polymorphism in MHCs and its impact on the adaptive immune system. Indeed, a method on the reconstitution of the class I antigen presentation system was described recently[25]. Interestingly, the same insect cell transfection system was employed in the study, indicating its robustness and effectiveness for expression of secreted subunit protein complexes. In their approach, however, Wearsch and Cresswell purified components of the antigen presentation complex from cell lysate[25]. This major difference is due to the observation of multiple bands from probing HLA-DR-expressing cell lysates with anti-FLAG antibody, in contrast to a single band seen in the

supernatant[59]. This trend was observed in the current study as well (Figure 7), and is attributed to most likely incomplete signal sequence cleavage and partial glycosylation[59].

Typically, purification of culture supernatant is done after proteins are concentrated[60]; a general concern for this method is the presence of serum proteins accompanying the culture medium. However, given that High Five cells are maintained in serum-free medium, this hurdle does not have to be overcome in this project. Regardless, purifying proteins from culture supernatant itself is a challenge, as larger volume leads to lower effective concentrations of protein to be purified. To improve protein yield, the supernatant was passed through anti-FLAG affinity beads multiple times, and a batch binding experiment was conducted. However neither approaches yielded satisfactory results, as the FLAG-tagged MHCs did not bind to the beads well, and HLA-DR could not be found in any of the fractions after size exclusion purification. It is also worth nothing that affinity purification of HLA-DM and DR was done at room temperature, whereas the Sadegh-Nasseri laboratory purified them at 4°C. Usually, protein purification procedures are performed at 4°C to protect its stability, but because the pilot purification of DM yielded positive results, despite the inefficiency, I decided to purify DR at room temperature as well. Strategies to moving forward include increasing the bead to supernatant ratio further in affinity purification, scaling up protein production, and comparing yield under different purification temperatures.

2.5 Conclusion

HLA-DM and HLA-DR were successfully cloned and transfected into Sf9 cells. The viral particles were subsequently used to infect High Five cells, and soluble DM, but not DR, was purified from the culture supernatant. As such, the original intent to develop an *in vitro* reconstitution assay for class II MHC antigen presenting system was not realized. However, the goal of this project is still very valuable, and optimizations on the protein purification procedures are strongly recommended to continue exploring this unique aspect of antigen presentation.

3. Analysis of acid eluted cell surface peptides by mass spectrometry-based peptidomics

3.1 Introduction

One of the ultimate goals of this thesis is to develop techniques that are capable of characterizing personalized MHC peptide repertoires from patients for subunit vaccine discovery/production. While the *in vitro* approach is an invaluable tool for deciphering allele-specific antigenic peptides, an *in vivo* counterpart should also be used as it more accurately represents the human immune system and validates the result from *in vitro* experiments. Two major ways to extract MHC peptides *in vivo* are immunoprecipitation (IP) and acid elution. In IP, cells are lysed and MHC-peptide complexes are purified using anti-MHC antibodies. The peptides are then eluted from MHCs in a separate step. This procedure has the advantage that almost all peptides ionized into the mass spectrometer are of MHC origin, however it requires at least 10^9 cells to perform[6]. In contrast, acid elution requires only 10^8 cells or less[32]. However, one has to first validate the data and show that the peptides are derived from MHC. With the long term application in mind, we chose the acid elution approach to examine MHC peptides, as one cannot culture primary cells to the amount required for an IP experiment and generate enough peptides for mass spectrometry. The acid elution method has actually been used previously[32], but this was back when the biological mass spectrometry field was in its infancy, and the technology was simply not there for the method to be applied to study MHC peptides.

Studying MHC peptide repertoires in an *in vivo* setting is not a trivial task, as MHC genes are highly polymorphic. Given that our current knowledge on anchor residues of class II MHCs is lacking[61], eluting a random assortment of primary cells without having a reference will lead to meaningless data. Therefore, for this project, consanguineous cell lines from a family of 7 were purchased and examined, as this ensures individual

alleles to be present in at least two cell lines, while maximizing potential combinations of alleles to observe the variability in antigen presentation.

3.2 Experimental methods

3.2.1 Materials

The materials used in this experiment can be found below:

- B lymphocyte cell lines GM02705, GM02707, GM02709, GM02711, GM02713, GM02728, GM03027 (Coriell Institute for Medical Research; Camden, NJ)
- Rosewell Park Memorial Institute 1640 medium, fetal bovine serum (FBS), penicillin and streptomycin antibiotics, and L-glutamine (Life Technologies)
- Lentiviral short hairpin RNAs, negative control (MISSION® pLKO.1-puro non-mammalian shRNA control plasmid DNA; Sigma-Aldrich) and targeting HLA-A (GE healthcare/Dharmacon clone ID TRCN0000057238; Ottawa, ON)
- Puromycin (Sigma-Aldrich) and Polybrene (EMD Millipore)
- HEK293T/17 cells (American Type Culture Collection; Manassas, VA)
- Trypsin/EDTA (0.5%) and Glutamax (Life Technologies)
- DMEM (Caisson labs; North Logan, UT)
- Lentiviral packaging mix (Sigma-Aldrich)
- Opti-MEM (Life Technologies)
- FuGENE® HD transfection reagent (Promega; San Luis Obispo, CA)
- HALT protease/phosphatase inhibitor cocktail (Thermo Scientific/Pierce; Rockford, IL)
- Rabbit monoclonal anti-DRA antibody (Abcam 52922)
- Rabbit polyclonal anti-calnexin antibody (Enzo Life Sciences ADI-SPA-865-D; Farmingdale, NY)
- Chemicals to make up common buffers such as phosphate-buffered saline, saline solution, and 2% acetic acid (Sigma-Aldrich)

- C18 disks (3M; London, ON)
- Aqua 5U C18 200A bulk packing material (Phenomenex)

3.2.2 Cell lines

B lymphocytes were maintained in RPMI-1640 with 15% or 20% fetal bovine serum (FBS), 1% pen/strep, and 2 mM L-glutamine, as instructed by Coriell's culture protocols[62]. HEK293T/17 cells were maintained in DMEM with high glucose, sodium pyruvate (1 mM), Glutamax (1X), and 10% FBS, as described by ATCC[63].

3.2.3 Lentiviral short hairpin RNA knockdown

Both the negative control and HLA-A knockdown shRNAs came within the pLKO.1 plasmid. The primer 5'- CAA GGC TGT TAG AGA GAT AAT TGG A -3' was used to confirm the identities of these sequences. Plasmids were extracted with Life Technologies' miniprep kit.

Prior to knocking down HLA-A, a titration curve experiment was performed to determine the concentration of polybrene and puromycin to use during transduction. Cells were seeded into 2 6-well plates at $5X10^5$ cells/mL. For the first plate, polybrene was added at concentrations 0, 6, 8, 10, 12 µg/mL. For the second plate, puromycin was added at 0, 0.5, 1, 2, 3, 4 µg/mL. Optimal concentration for polybrene was determined by an intermediate concentration that did not kill cells in 3 days (e.g., if at 3 days 6 and 8 µg/mL treated cells are alive, but at 5 days the cells with 8 µg/mL looked sick, 6 µg/mL was used for transduction), and optimal concentration for puromycin was dictated by the minimum concentration that killed cells in 3 days. These concentrations were used as a starting point; it is possible that they kill the cells during transduction. In the case that the cells died at these concentrations, lower concentration would be used.

HEK293T/17 cells were used for viral amplification, and seeded at $1X10^6$ cells on a 6 cm plate one day before transfection. Transfection was done by mixing 1 µg of plasmid containing shRNA with 10 µL lentiviral packaging mix, 100 µL Opti-MEM, and 6 µL FuGENE transfection reagent. The mixture was gently vortexed, then left incubating at room temperature for 20 min. Finally, the mixture was applied to HEK293T/17 cells drop by drop to cover the whole plate with the transfection mix.

Lentiviral particles carrying either non-mammalian targeting shRNA or HLA-A shRNA were harvested in the supernatant 48 hr post-transfection by centrifugation at 330 rcf for 5 min. 4 mL of supernatant was first passed through a 0.45 µm filter, then added into 4 mL B cells at cell density $5X10^5$ cells/mL. Polybrene was added to a final concentration of 8 µg/mL to aid the transduction process. The resulting mixture was incubated at 37°C, 5% CO_2 for 48 hr. Cells were then pelleted and replaced with fresh medium and continued to be cultured for 48 hr, at which point 1 µg/mL puromycin was added to select for cells containing lentiviral particles. Knockdown efficiency was determined by lysing recovered cells with lysis buffer (20 mM Tris, pH 7.5, 150 mM NaCl, 1% NP-40, 10 mM sodium pyrophosphate, 50 mM sodium fluoride) and HALT protease/phosphatase inhibitor cocktail (1X), then performing a western blot on individual cell lines.

3.2.4 Cell surface peptide extraction by acid elution

2×10^8 cells were harvested, washed three times with phosphate buffered saline (PBS: 9 g/L sodium chloride, 1.44 g/L potassium phosphate monobasic, 7.95 g/L sodium phosphate dibasic, pH 7.4), then washed with saline (PBS without the phosphate components). Cell surface peptides were extracted by mixing 2% acetic acid in saline with cells for 30 sec, followed by centrifugation at 200 rcf for 5 min to pellet the cells. An additional centrifugation step was performed on the supernatant at 2000 rcf for 15 min to remove leftover cell debris. 3 replicates were performed for each of the B cell lines,

GM02709 transduced with negative control shRNA, GM02709 transduced with HLA-A knockdown shRNA, and THP-1 cells.

3.2.5 Sample cleanup

The supernatant, containing all surface peptides, was desalted with two stop-and-go extraction (STAGE) tips[35]. STAGE tips are made by fitting C18 disks to the end of p200/p1000 pipette tips. The disks act as a physical barrier as well as a chemical filter when the supernatant is pushed through the tip with a syringe. In particular, peptides are retained by hydrophobic interactions while salts are passed through.

Peptides were washed with 0.1% acetic acid in water (Buffer A), then eluted with 50% acetonitrile, 0.1% acetic acid in water (Buffer B), vacuum dried, and reconstituted in Buffer A for analysis by LC-MS/MS.

3.2.6 Liquid chromatography tandem mass spectrometry

The analysis of peptides eluted from each cell line was performed using a Q-Exactive mass spectrometer equipped with the EASY-nanoLC system (Thermo Scientific; San Jose, CA). Peptides were eluted with a gradient from 100% Buffer A to 40% Buffer B (80% ACN, 0.5% acetic acid) over 142 min at a constant flow of 250 nL/min. The instrument was operated using Xcalibur v2.2 (Thermo Scientific) in data-dependent acquisition mode, with fragmentation of the 5 most abundant ions per scan and dynamic exclusion of 30 seconds enabled. MS resolution was set to 70,000 with an automatic gain control (AGC) target of 3×10^6, maximum fill time of 20 msec and a mass window of 300 to 2000 *m/z*. Higher collision dissociation (normalized collision energy 26 with 20% stepping, done in accordance to previous findings to obtain optimal spectra[64]) was performed with an AGC target of 1×10^6, maximum fill time of 120 msec, mass resolution of 35,000, and charge exclusion set to unassigned.

3.2.7 Data analysis

Raw data was searched using MaxQuant (v 1.4.1.2). Default parameter values were selected with these exceptions: Match Between Runs checked, unspecific digestion mode (i.e., no enzyme), no fixed modifications, N-terminal protein acetylation and methionine oxidation as variable modifications, revert decoy mode, human.first.search.fasta as the separate fasta file for first search, default contaminant database used, PSM and protein FDR of 0.1. The search was conducted against a protein databases containing UniProtKB/TrEMBL human sequences (88844 sequences, retrieved July 2014), cow sequences (2151 proteins, retrieved July 2014) that were identified in a separate search against cow sequences alone, and 11 likely viral contaminants including Epstein-Barr virus, adenovirus, and bovine diarrhea virus.

The resulting peptides were re-matched against the database and all non-exact matches were removed. The remaining peptides were sorted by species. Those that matched both bovine and human databases were categorized as human, as MHCs present endocytosed proteins in addition to cellular proteins. Peptides matching the forward or decoy versions of the extremely large (~34000 residues) protein titin tended to provide either false matches or were erroneously processed by Excel due to its large size, so they were removed.

3.3 Results

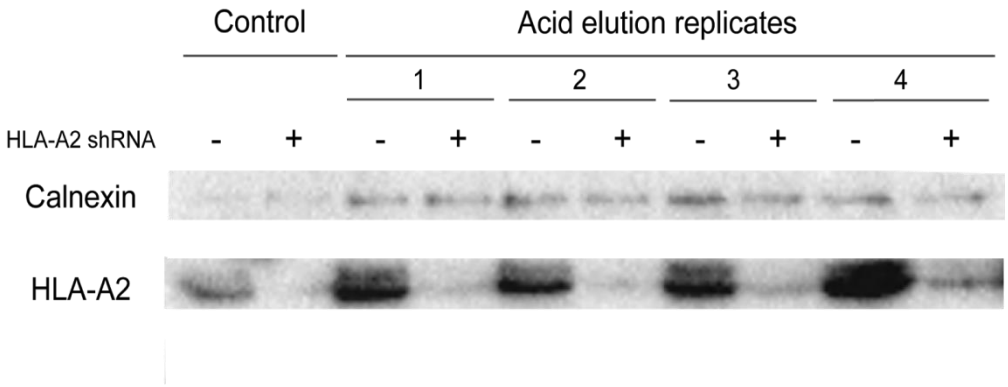
3.3.1 Evaluation of HLA-A knockdown efficiency

As acid elution was picked as the method of choice to obtain MHC peptides, a validation experiment was performed to ensure that most peptides did come from MHC. This was done by knocking down a subtype of MHC. In this experiment, HLA-A was chosen

because it is homozygous (HLA-A2) in two of the cell lines and its well-studied anchor residues allow for proper analysis and interpretation of the data.

To analyze the efficiency of HLA-A2 knockdown, a western blot was done on cell pellet of transduced GM02709 cells (non-targeting and HLA-A*02 knockdown shRNA). Puromycin-selected cells recovered 11 days post-transduction, and acid elution was performed 21, 42, and 52 days post-transduction (Figure 9, replicates 1, 3, and 4). In addition to probing for HLA-A, anti-calnexin antibody was used as a normalization control. The intensities of the bands were determined by Adobe Photoshop[65], and normalized HLA-A2 intensities were calculated by dividing the intensity of the HLA-A2 band in each sample by the intensity of calnexin. HLA-A2 knockdown (KD) efficiency could then be calculated by subtracting the intensity ratio of normalized HLA-A*02 intensity in the KD construct over the non-targeting construct (neg. control) by one. For example, if the normalized HLA-A*02 band is 25% as intense in KD when compared to neg. control, then the knockdown efficiency would be 75%. As seen in Figure 9, HLA-A2 knockdown efficiency was at 90%, 75% and 73% respectively on the three acid elution replicates, with an average knockdown efficiency of 79.4%.

Figure 9 HLA-A2 knockdown efficiency analysis via western blot for 4 acid elution biological replicates.



HLA-A2 intensities were first normalized against calnexin, then the knockdown efficiency was calculated by taking the ratio of knockdown to negative control.
Lanes 1 and 2 – recovered B lymphocytes transduced with non-mammalian targeting shRNA (negative control) and HLA-A2 shRNA, respectively, 11 days post-transduction.
Lanes 3 to 10 – protein expression in B lymphocytes 21, 31, 42, and 52 days post-transduction, corresponding to the days on which acid elution was performed.

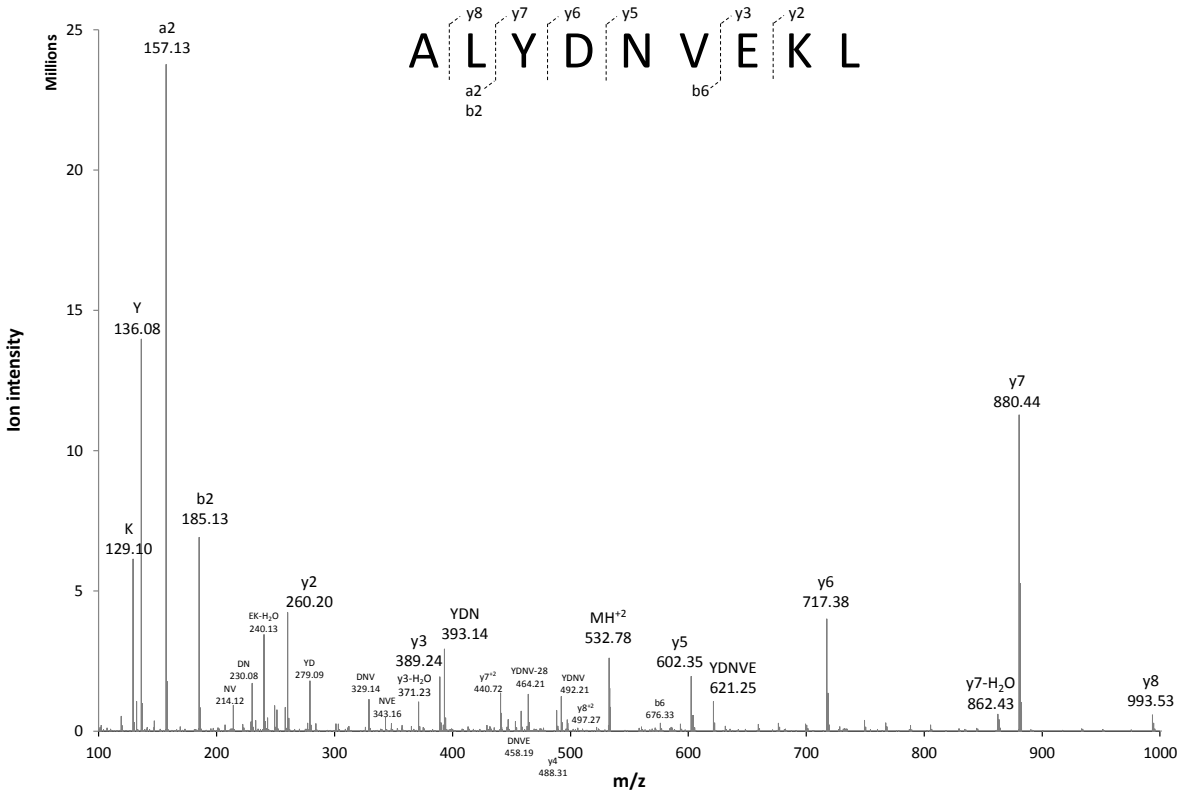
3.3.2 Mass spectrometry data processing

Surface-derived peptides from acid elution were analyzed on the Q Exactive mass spectrometer, and the raw data were searched using MaxQuant[66]. Results by MaxQuant showed matches to 40413 peptides, of which about 3.7% (1495 peptides) were removed from analysis because they were not exact matches to the protein database (817 peptides), had no associated precursor intensity value (677 peptides), or matched both human and virus proteins (1 peptide). After further excluding all matches to the reverse database (also including palindromic matches to the forward database), there are a total of 38814 peptide hits with a 1.8% peptide false-discovery rate. A majority of these (94%) were self (human) peptides; about a half of them (19222) also matched sequences the bovine database, but they were highly likely to be derived from lysed cells in the culture. Strictly bovine peptides, coming from fetal bovine serum added to the media, made up 5.3% (2048 peptides) of the total hits. A minute 0.2% (76 peptides) were virus-derived. Finally, it is important to note that for the purpose of this

project, cow epitopes are just as useful as human epitopes as long as they are correctly identified; cow proteins can just as feasibly be phagocytosed and processed as other cellular debris inside a B cell. Examples of mass spectra of true positive MHC peptides are shown in Figure 10.

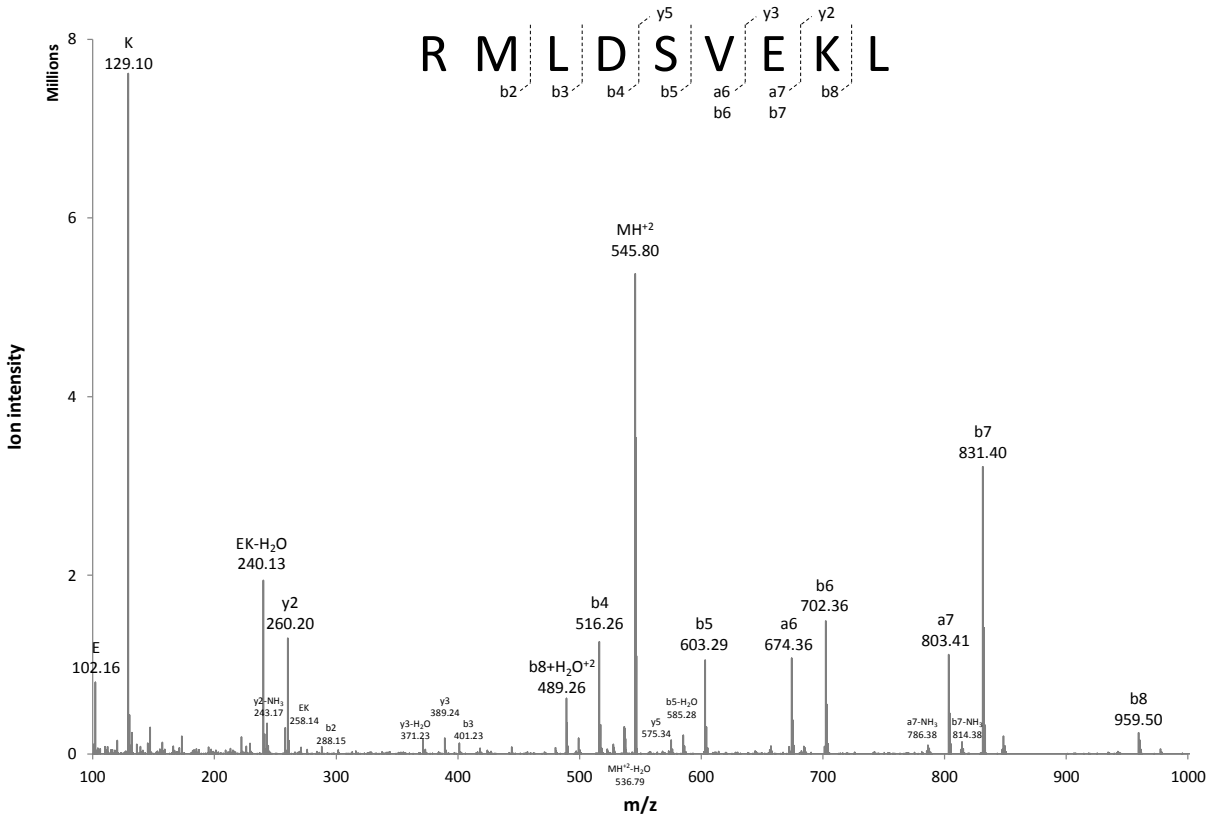
Figure 10 Sample mass spectra of acid-eluted MHC peptides.

Figure 10a Mass spectrum of ALYDNVEKL.



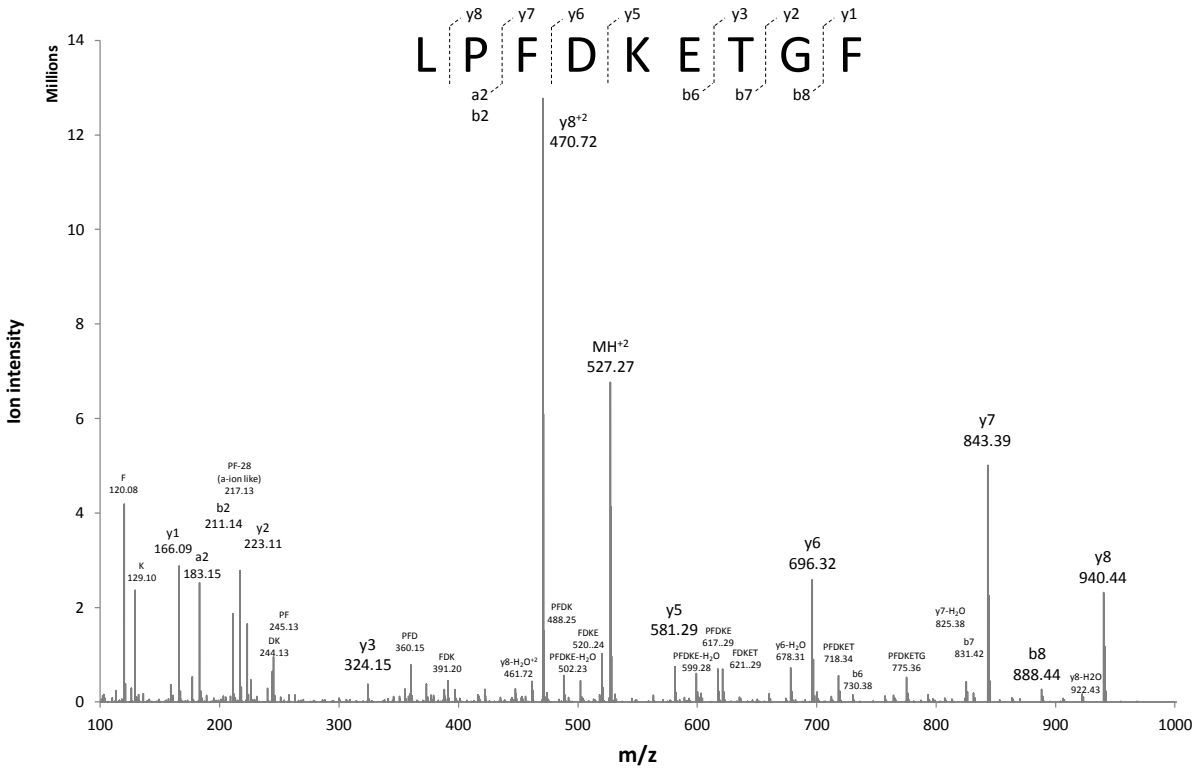
ALYDNVEKL, a likely HLA-A2 peptide identified in all three replicates of HLA-A2 shRNA knockdown GM02709 cells.

Figure 10b Mass spectrum of RMLDSVEKL.



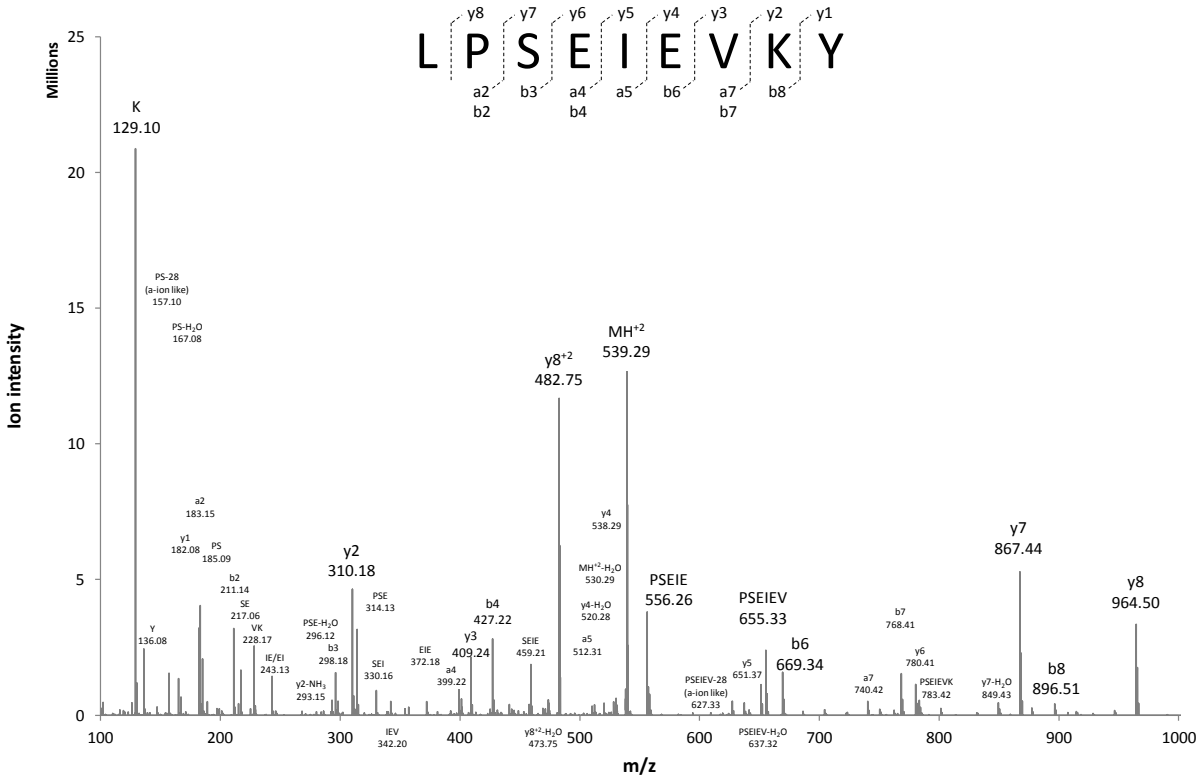
RMLDSVEKL, a likely HLA-A2 peptide identified in all three replicates of HLA-A2 shRNA knockdown GM02709 cells.

Figure 10c Mass spectrum of LPFEKETGF.



LPFDKETGF, a likely HLA-B peptide identified in all three replicates of scramble sequence shRNA-transduced (negative control) GM02709 cells.

Figure 10d Mass spectrum of LPSEIEVKY.



LPSEIEVKY, a likely HLA-B peptide identified in all three replicates of scramble sequence shRNA-transduced (negative control) GM02709 cells.

For analysis of MHC class I peptides, all human peptides, including those that matched to bovine database, were used. In order to compare the amount of a specific peptide being presented across cell lines, a normalization step was performed by dividing the ion intensities of individual peptides by the 10^{th}-percentile intensity of all peptides found in the cell line. Normalized values were then averaged across replicates should the peptide be found in more than one replicate. The end result is a list of peptides found in at least one cell line, with their associated normalized ion intensity values.

The next step was to correlate surface peptide intensities from GM02709 cells to knockdown efficiency. Each peptide can be found in one of three scenarios: found in KD cells only, found in neg. control cells only, and found in both (shared). Shared peptides can be separated within the group based on their KD/neg. control intensity ratios. Peptides with intensity ratio less than “1 minus KD efficiency” were assigned into a subgroup, as they can be inferred to be a group that became greatly diminished in amount as a direct result of HLA-A2 knockdown. For example, the average KD efficiency among three replicates (90%, 75%, 73%) is 79.4%, therefore, peptides associated with HLA-A2 knockdown should be found at 20.6% intensity in KD cell line compared to neg. control on average. The reciprocal value (4.85) of the above threshold (0.206) was also applied to check how many peptides fall beyond this value. Peptides found in only KD cells were arbitrarily assigned a KD/neg. control ratio of 50, whereas those found in only neg. control cells were assigned a ratio of 0.02. These intensity ratios were used to elucidate biological implications of the HLA-A2 knockdown experiment.

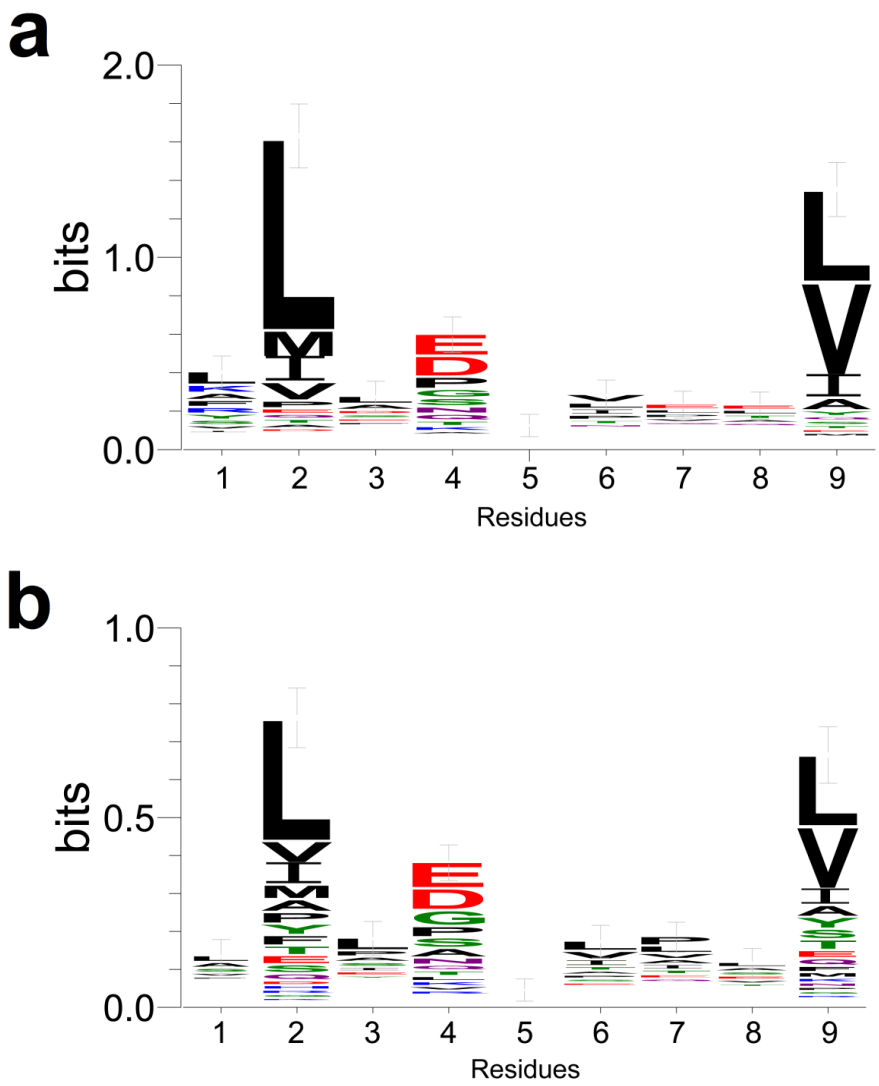
3.3.3 MHC bias assessment in eluted surface peptides

By knocking down HLA-A2 in GM02709 cell line, two biases should be logically present if most peptides are of MHC origin. The first bias is that peptides below KD/neg. control intensity ratio 0.206 should show a bias toward having HLA-A2 characteristics. In other words, if the peptides displaying intensities proportional to the amount that HLA-A2 expression is suppressed in the knockdown cell line are found to possess HLA-A2 characteristics, then the result strongly suggests that these peptides originate from MHC. The second bias is that the percentage of class I peptides should be highest in the group of peptides found only in the KD cell line, as it should contain the most HLA-A peptides relative to the shared group and neg. control cell line. Following the same logic, % class I peptides should be the lowest in the group of peptides found only in the neg. control cell line.

The binding pocket of MHCs is known to bind to 9 amino acid residues in both classes[61, 67]. To confirm the HLA-A2 characteristic bias, one needs to first align the MHC class I peptides, which vary between 8-11 amino acids in length. Clustal X2 was used for this purpose, aligning amino acids by their chemical properties[68]. The gap opening parameter under multiple alignment parameters was changed to 100 in order to prohibit gaps among aligned sequences. Finally, Weblogo was used to visualize the data, displaying amino acid residues in different heights and widths depending on their frequency of occurrence within a specific position[69].

To address the first bias, peptides in each group (i.e., neg. control only, shared < 0.206, shared between 0.206 and 4.85, shared > 4.85, KD only) were subjected to alignment and plotted on Weblogo. Peptides with KD/neg. control intensity ratios < 0.206 display a bias toward HLA-A2 (Figure 11), whereas peptides above the 0.206 threshold do not show any particular bias toward either HLA-A2 or HLA-B7/35 (Figure A1). The same intensity ratio/Weblogo analysis was also conducted between neg. control cells and healthy GM02709 cells, and no bias could be observed in any of the groups (Figure A2).

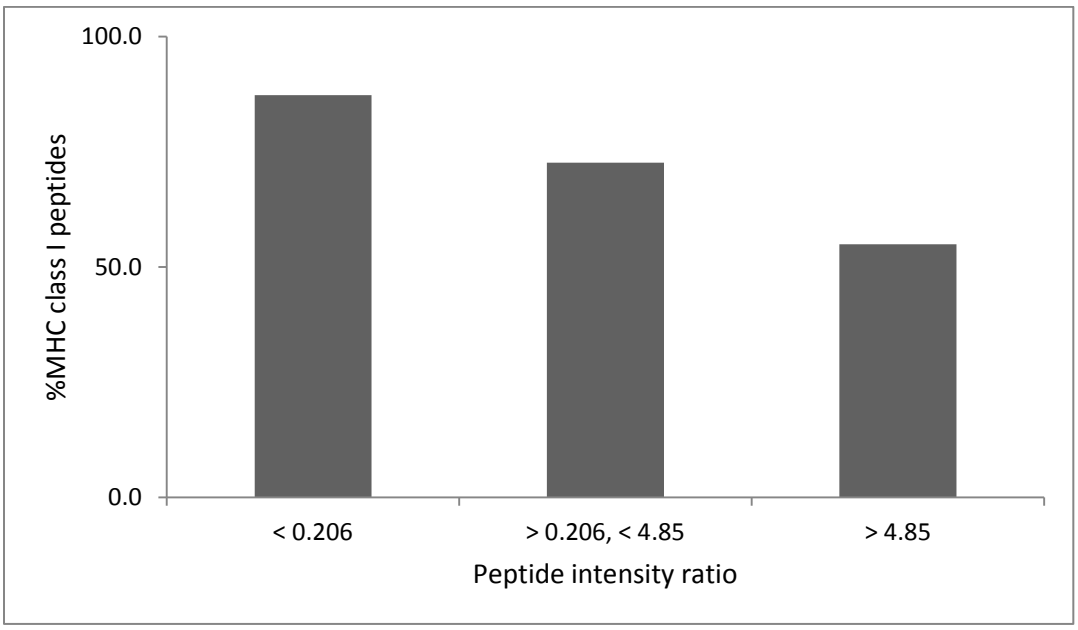
Figure 11 Weblogos of multiple sequence aligned peptides derived from surface of HLA-A2 knockdown and negative control shRNA GM02709 cells.



Logos are compiled of peptides found in **(a)** both HLA-A2 knockdown and negative control cell lines, with knockdown-to-negative control peptide intensity ratios less than 0.206, and **(b)** only the negative control cell line. [Anchor] or preferred residues, extracted from SYFPEITHI, for relevant alleles are as follows for positions 2, 4, 6, and 9: HLA-A2 [LM]-E-V-[VL], HLA-B7 [P]-D-L-[LF], HLA-B35 [P]-DEG-VL-[YFMLI].

To check the second bias, the ratio of the number class I peptides over the total number of peptides identified was determined. Peptides were separated into three groups using 0.206 and 4.85 as thresholds, and class I peptides were defined by those with 8 to 11 amino acid residues. Peptides with KD/neg. control intensity ratios < 0.206 have the highest %class I among the three groups (87.3%), followed by the group with intensity ratios between 0.206 and 4.85 (72.6%), and lastly those with intensity ratios > 4.85 (54.9%) (Figure 12).

Figure 12 Class I MHC peptide percentage bias.



Peptides found in each replicate were normalized to the 10th percentile intensity of all identified peptides, and normalized intensities were averaged across replicates. The combined peptide list was separated into three groups: 1. peptides with knockdown-to-negative control intensity ratios less than 0.206 (87.3% class I MHC), 2. peptides with intensity ratios between 0.206 and 4.85 (72.6% class I MHC), and 3. peptides with knockdown-to-negative control intensity ratios greater than 4.85 (54.9% class I MHC).

3.4 Discussion

To understand the relationship between MHC polymorphism and the immunogenic peptides presented on the cell surface in an *in vivo* setting, we modified an old peptide elution method to examine these peptides using LC-MS/MS[32]. While immunoprecipitation would yield almost exclusively MHC peptides, the technique is not compatible with primary B cells and therefore cannot be applied to clinical samples. Choosing to elute all surface peptides, however, require a validation experiment to show that most peptides are of MHC origin. To that end, a lentiviral shRNA knockdown experiment was conducted. Label-free quantitation was selected over metabolic labeling because label signals for epitopes would be diluted across peptides and not every MHC peptide is guaranteed to contain the heavy isotope amino acids used in labeling.

The HLA-A shRNA knockdown experiment validated surface acid elution as a means to extract MHC peptides, as all results match the anticipated biases and trends. Peptides with low knockdown-to-negative control intensity ratios (< 0.206), mirroring the HLA-A2 expression level in the knockdown cell line, displayed a bias toward having HLA-A2 anchor residues. However, peptides with intensity ratios > 4.85 (inverse of 0.206) along with peptides with intensity ratios between 0.206 and 4.85 showed no particular bias toward any type of MHC class I. No difference was observed when all three categories of peptides were multiple sequence aligned and displayed with Weblogos between healthy and non-targeting shRNA lenvirus-infected GM02709 cells. Similarly, no biases could be found between healthy GM02709 and GM02711 cells, which share the same HLA-A allele.

If acid elution of cell surface produces MHC peptides, there should be also a percentage class I bias between the three groups of peptides mentioned above, and the data validated this hypothesis. The percentage of class I peptides was highest in the group of peptides with KD/neg. control intensity ratios < 0.206 (87.3%). This group is thought to include many HLA-A peptides that were not presented or diminished greatly in amount in the KD cells, as well as some HLA-B and HLA-C peptides. The second highest in class I peptide percentage was the group containing peptides with intensity

ratios between 0.206 and 4.85 (72.6%), which should contain most HLA-B and HLA-C peptides and a small portion of HLA-A peptides that were still presented by KD cells. The group with the lowest class I peptide percentage consists of peptides with intensity ratios > 4.85 (54.9%), with the source of peptides mostly from HLA-B and HLA-C, as most HLA-A peptides would be found in neg. control cell lines only and a small amount found shared between the two cell lines at moderate intensity ratios. In other words, a ratio of 4.85 or greater for an HLA-A peptide would be quite an outlier as this requires the peptide intensity to be much greater in KD than neg. control cell line.

Ultimately, the goal of this project is two-fold. First, to develop a technique that enables personalized MHC peptide profiling in a clinical setting. With the continuous speed and cost improvements of genome sequencing, it is quite realistic for us to be able to gather personalized genomic and immunopeptidomic data in the near future. This not only enables more accurate diagnostics but also help the immunology research field by gathering “real data”, giving us the chance to really understand how genotype translates to phenotype in human MHC proteins. Second, to develop a bioinformatics prediction tool to more accurately define the anchor residues in binding pockets of MHC alleles. The tool would be particularly useful for class II, which trails behind current knowledge on class I because it is much harder to describe and define an epitope (9 residues) when the peptide themselves can be 15-25 amino acids in length.

It is important to note that although the project’s aim is to devise an innovative system to look at MHC class II peptides, knockdown of a class I allele was chosen for two reasons. Firstly, to ensure maximal knockdown efficiency, a homozygous MHC gene was selected as candidate for knockdown. This enabled the introduction of just one shRNA construct to knockdown both copies of the HLA-A2 gene in cell line GM02709. Secondly, biases are only meaningful if the sequence aligned data can be searched against a robust database. Between the two classes, current knowledge on anchor residues of class I peptides far exceeds what is known on class II peptides. With validation completed, we can now move onto developing tools which will aid our understanding on the immunopeptidome.

3.5 Conclusion

A cell surface acid elution method was used to develop an epitope prediction tool for MHC peptides, with the goals of being able to deconvolute and assign peptides to individual alleles and the method being applicable for clinical samples. Lentiviral shRNA knockdown was performed in cell lines with homozygous HLA-A2 as validation, and the resulting biases confirmed acid elution as a good way to extract MHC peptides. Current work revolves around the bioinformatics of deciphering class II MHC epitopes for peptides extracted from cell lines of a family of seven.

4. Conclusion

MHC plays a prominent role in the adaptive immune system. With infectious diseases being the number one cause of premature mortality worldwide[70] and the rise in number of antibiotic-resistant pathogens, there is great interest in shifting the paradigm for treating diseases from a reactive stance to a proactive one (i.e., vaccines). Various aspects of antigen presentation has been studied in great detail in recent years, such as the peptide loading mechanism[2, 3, 71, 72], transcriptional[73] and translational control[74], anchor residues for individual alleles[75, 76], and peptide binding prediction tools[77, 78]. To date, we have a greater understanding on the anchor residues of class I MHC alleles than their class II counterparts, mainly due to the difference in length of peptides between the two classes; the binding cleft for both class I and II can accommodate around 8-10 amino acids, however the length of the entire peptide that is bound onto MHC is much longer for class II (15-25 amino acids) than class I (8-11 amino acids). This makes it difficult to decipher the interacting residues for class II peptides via any elution method. Crystallography is a potential option, but it can only be done on a per-allele basis, and the time investment would be tremendous given the amount of alleles and the peptides they bind to. Another obstacle to overcome is that it is very difficult to assign peptides to individual alleles in a cell-based system. This can be resolved by using the recently developed *in vitro* systems for expressing class I and II MHCs, which has been mostly used for mechanistic studies but not high-throughput identifications as of yet[25, 26].

Mass spectrometry has been the go-to technology in recent years to reliably identify and quantify peptides. While identifying non-tryptic peptides significantly increases the search space and adds difficulty in obtaining confident results, we have developed a protocol that generates a list of MHC peptides at a false discovery rate of below 1% in-house (manuscript in preparation for submission). With this technical hurdle gone, it is clear that MS is the method of choice to address the genotype-phenotype relationship in the antigen presenting process.

4.1 Addressing the project aims and hypothesis

In this thesis, two complementary yet independent approaches were taken to investigate the relationship between polymorphism in MHC and the peptide repertoire each allele gives rise to. First, an *in vitro* reconstitution system was set up for class II MHCs in chapter 2 with the goal of purifying different alleles and identifying high-affinity antigenic peptides from various allelic combinations. Soluble domains of HLA-DR and HLA-DM were cloned and transfected into Sf9 cells using the baculovirus transfection system. To obtain maximal yield of secreted proteins, viral particles were harvested and used to infect High Five cells and the supernatant was purified using affinity chromatography followed by size exclusion chromatography. While the purification of DM was successful, DR was not observed in any of the fractions recovered from size exclusion separation. Further trouble-shooting experiments revealed that optimization of affinity purification protocol is required. One such optimization experiment was performed on DR via batch binding instead of gravitational flow. However the setup used proved to be even less efficient than purification of DM.

Chapter 3 covered an acid elution protocol as a means to examine MHC peptides in an *in vivo* setting. There are two predominant approaches to extract MHC peptides from the cell surface – immunoprecipitation and acid elution. The latter technique was selected because 1. it contrasts but also complements the first project in that it provides a pan-view of the peptide repertoire and can serve as validation for the first project, and 2. it aligns with our long-term goal for this project, which is to apply the technique on clinical samples such as primary B cells. Using the acid elution method means a validation experiment is required, however, given there are likely other peptides found on the cell surface. Therefore, an shRNA knockdown experiment was performed, with the hypothesis that if most of the eluted peptides originated from MHC, then knocking down the expression of a subtype of MHC (e.g., HLA-A2) would result in marked under-representation of anchor residues from that MHC (e.g., less L/M residues at second position) from the subpopulation of peptides whose abundance drop proportionately to the knockdown efficiency. After aligning the peptides and displaying the data in

Weblogo, the result showed a strong bias toward HLA-A2 in the group of peptides whose intensities are reduced proportionately to HLA-A2 expression level after knockdown, confirming the initial hypothesis. With the acid elution technique verified, the focus now shifts to developing ways to extract epitope information from seven consanguineous B cell lines; specifically, the goal is to produce a universal pipeline that result in visual display of allele-specific anchor residues.

4.2 Future directions

4.2.1 Optimization of anti-FLAG protein purification protocols

All of the purification experiments thus far showed binding inefficiency between FLAG-tagged proteins and anti-FLAG beads. A small-scale batch binding experiment was performed using DR with higher bead to protein ratio, and considerable improvement in binding efficiency was observed. This suggests that even higher amount of beads should be used, despite the fact that each experiment was carried out using at least 3 times the advertised binding capacity. Another way to enhance binding efficiency is by concentrating the supernatant prior to purification. Lastly, all purification experiments have been done on a small scale thus far. Scaling up the cell culture to 500 mL or 1 L will likely yield visible bands in fractions with the correct molecular weight even if the binding efficiency stays the same. Given that a 100 mL culture can afford approximately 3 *in vitro* presentation experiments (personal communication with Sadegh-Nasseri laboratory), it is recommended that optimizations via higher amount of beads and reducing the volume of supernatant be done first before moving onto scaling up.

4.2.2 Combining components of *in vitro* produced MHCs to determine allelic preferences within the immunopeptidome

As the *in vitro* reconstitution technique was established by the Sadegh-Nasseri laboratory, we have signed a material transfer agreement and therefore will be brief on the protocol for peptide presentation by various MHC alleles. As mentioned before, DM will always be included in any presentation experiment, along with a peptide-presenting class II MHC. We will first test a list of known binder and non-binder peptides (from Dr. Wilfred Jefferies, Michael Smith Laboratories) to ensure the assay works as intended. DM and DR (or DP/DQ) will be incubated with peptides for 3 hr, followed by an affinity purification to separate MHC-peptide complexes from the unbound substrates. Peptides can be eluted off by 2% acetic acid, and separated from MHCs by applying the eluate through a 10 kDa MWCO filter. Peptides obtained in the filtrate should reflect the known binders in MS analysis.

To step up in complexity, one can try incubating the MHCs with antigens with a known strong binder, such as hemagglutinin from influenza virus[2]. With antigen instead of peptides as substrate for MHC binding, cathepsins are required. Therefore, the protocol will be modified such that the antigen is allowed to bind with MHCs first, with cathepsins added later. Following incubation, the same peptide extraction scheme is carried out. Once we are able to observe known strong binders from a few antigens, we will move onto antigens with less well known epitopes.

4.2.3 Determination and assignment of anchor residues to MHC alleles

There are three goals for the acid elution project: to develop a method to extract and identify MHC peptides, to define epitopes from these peptides, and to determine the anchor residues for these epitopes and assign them to a specific allele. The bioinformatics procedures to derive epitope information are being worked on by coworker Q. Chan and several parameters are still being constantly improved.

Therefore preliminary results will be summarized here, along with comments on future directions of this project.

As mentioned previously, the binding cleft for both MHC classes can accommodate around 9 amino acid residues[61, 67], but the length of MHC peptides is often greater than this number, particularly for MHC class II. Recognizing that different identified peptides may represent the same epitope, peptides from the same region of a protein were aligned to define epitopes. In fact, this exercise of “nesting” the peptides can be used to further ensure that the analysis only include potential immunogenic peptides. To determine how many observed peptides are needed to ensure confident identification of a real epitope, we asked how frequently identified reversed peptides formed nested sets and chose two parameters to tinker with: number of overlapping peptides and final length of the epitope. For example, a sequence with 3 or more overlapping peptides and spans 12-30 residues may be a class II MHC epitope. By applying these 2 rules, 1645 potential epitopes fall into this category with a technical false discovery rate of 0.06% (i.e., search epitope sequences against the reverse database). A similar set of rules can also be applied for class I MHC epitopes, but the stringency for requiring overlapping peptides is relaxed, given the short length in individual class I peptides. Therefore, in its current iteration we define a class I MHC epitope by requiring 2 or more overlapping peptides that form an epitope spanning 9-11 residues or singletons that are found in at least one-third of the samples. Under these criteria, a total of 4799 potential epitopes fall into the class I MHC category with 1.9% of these being reverse hits.

Currently, our work involves optimizing the aforementioned parameters, as well as defining the term “anchor residues”. This will take form in a threshold score, where overrepresentation of specific amino acids will be calculated at each position. As with the knockdown experiment, our knowledge on class I anchor residues will be of tremendous value here as we can compare the epitopes we generated against known, established anchor residues. Once we are able to confidently assign epitopes to class I alleles, then we will apply the same logic and metrics to assign epitopes to class II alleles, revealing the anchor residues in the process.

4.2.4 Subunit vaccine discovery

In the age where multi-drug resistance pathogens are on the rise, vaccines are one of the most cost-effective medicines available. Vaccination is also the only conceivable way of completely eradicating a disease[79, 80]. Currently, “reverse vaccinology” is the method of choice for discovering subunit vaccines[81, 82]. The basic idea behind reverse vaccinology is to sequence the genome of a pathogen, use bioinformatics to predict genes as potential antigens, and produce and test them one by one to check for efficacy *in vivo*. While reverse vaccinology generates lots of potential antigens compare to traditional vaccinology (e.g., knock out of a gene to see if a pathogen has the same virulence without it), realistically it is a form of controlled serendipity[82] that produce hundreds or thousands of potential candidates, making it a time-consuming and expensive bottleneck still in vaccine discovery. With the tools available in modern proteomics and ideas introduced in this thesis, more filters can be applied to improving the frequency of getting immunogenic epitopes, and ultimately lower the cost while accelerate the rate of vaccine development . For the acid elution project, our laboratory has already purchased more cell lines to apply the technique to a wider range of HLA alleles, allowing us to gather more personalized antigen repertoires to improve on the allele assignment and anchor residue prediction tool. Meanwhile, the *in vitro* reconstitution project can act as validation for these predictions.

4.3 Closing

For its role in the adaptive immune system, major histocompatibility complexes have always been one of the focal points in the field of immunology research. While there have been plenty of discoveries, important questions, such as the effect of MHC polymorphism on peptide repertoire, remain unanswered. With the advances in the field of biological mass spectrometry, it is now possible to set up high throughput experiments to explore the immunopeptidome. In this thesis, two approaches have been

devised to examine the role of allelic differences in MHC on the peptide repertoire presented on the cell surface. The *in vitro* approach requires further optimization, but will provide great insight on the effect of individual alleles on peptide presentation once established. The *in vivo* approach to extract MHC peptides has been validated, and it not only paves the way to decoding more anchor residues, particularly for class II MHC alleles, but also serves as confirmation on the results obtained from the *in vitro* approach. Taken together, the technologies explored by this thesis will ultimately contribute to our fundamental knowledge on the adaptive immune system and host-pathogen interaction as a whole, while offering potential clinical applications down the road.

References

1. Kimball, J.W. Kimball's biology pages. http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/ (accessed December 2014).

2. Narayan, K. *et al*. HLA-DM targets the hydrogen bond between the histidine at position β81 and peptide to dissociate the HLA-DR—peptide complexes. *Nat. Immunol.* **8**, 92-100 (2007).

3. Guce, A. I. *et al*. HLA-DO acts as a substrate mimic to inhibit HLA-DM by a competitive mechanism. *Nat. Struct. Mol. Biol.* **20**, 90-98 (2013).

4. Fortier, M. *et al*. The MHC class I peptide repertoire is molded by the transcriptome. *J. Exp. Med.* **205**, 595-610 (2008).

5. Granados, D. P. *et al*. Impact of genomic polymorphisms on the repertoires of human MHC class I-associated peptides. *Nat. Commun.* **5** (2014).

6. Mommen, G., P.M. *et al*. Expanding the detectable HLA peptide repertoire using electron-transfer/higher energy collsion dissociation (EThcD). *Proc. Natl. Acad. Sci.* **111**, 4507-4512 (2014).

7. Yu, C. Y., Yang, Z., Blanchong, C. A. & Miller, W. The human and mouse MHC class III region: a parade of 21 genes at the centromeric segment. *Immunol. Today* **21**, 320-328 (2000).

8. Schubert, U. *et al*. Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature* **404**, 770-774 (2000).

9. Princiotta, M. F. *et al*. Quantitating protein synthesis, degradation, and endogenous antigen processing. *Immunity* **18**, 343-354 (2003).

10. Abele, R. & Tampe, R. Function of the transport complex TAP in cellular immune recognition. *Biochim. Biophys. Acta* **1461**, 405-419 (1999).

11. Antoniou, A. N., Powis, S. J. & Elliott, T. Assembly and export of MHC class I peptide ligands. *Curr. Opin. Immunol.* **15**, 75-81 (2003).

12. Giroux, M., Schmidt, M. & Descoteaux, A. IFN-gamma-induced MHC class II expression: transactivation of class II transactivator promoter IV by IFN regulatory factor-1 is regulated by protein kinase C-alpha. *J. Immunol.* **171**, 4187-4194 (2003).

13. Berger, A. C. & Roche, P. A. MHC class II transport at a glance. *J. Cell Sci.* **122**, 1-4 (2009).

14. Kropshofer, H. *et al*. Editing of the HLA-DR-peptide repertoire by HLA-DM. *EMBO J* **15**, 6144-6154 (1996).

15. Kropshofer, H., Hämmerling, G. J. & Vogt, A. B. How HLA-DM edits the MHC class II peptide repertoire: survival of the fittest? *Immunol. Today* **18**, 77-82 (1997).

16. Bryant, P. & Ploegh, H. Class II MHC peptide loading by the professionals. *Curr. Opin. Immunol.* **16**, 96-102 (2004).

17. Busch, R., Doebele, R. C., Patil, N. S., Pashine, A. & Mellins, E. D. Accessory molecules for MHC class II peptide loading. *Curr. Opin. Immunol.* **12**, 99-106 (2000).

18. Brocke, P., Garbi, N., Momburg, F. & Hammerling, G. J. HLA-DM, HLA-DO and tapasin: functional similarities and differences. *Curr. Opin. Immunol.* **14**, 22-29 (2002).

19. Poluektov, Y. O., Kim, A., Hartman, I. Z. & Sadegh-Nasseri, S. HLA-DO as the optimizer of epitope selection for MHC class II antigen presentation. *PLoS One* **8** (2013).

20. Miller, P. J. & Collins, E. J. in *Immunodoimnance - the choice of the immune system* (ed Frelinger, J. A.) 3 (Wiley-VCH, Weinheim, Germany, 2006).

21. Blum, J. S., Ma, C. & Kovats, S. Antigen-presenting cells and the selection of immunodominant epitopes. *Crit. Rev. mmunol.* **17**, 411-417 (1997).

22. Fairchild, P. J. Reverseal of immunodominance among autoantigenic T-cell epitopes. *Autoimmunity* **30**, 209-221 (1999).

23. Deng, H., Fosdick, L. & Sercarz, E. The involvement of antigen processing in determinant selection by class II MHC and its relationship to immunodominance. *APMIS* **101**, 655-662 (1993).

24. Meyer, D. & Thomson, G. How selection shapes variation of the human major histocompatibility complex: a review. *Ann. Hum. Genet.* **65**, 1-26 (2001).

25. Wearsch, P., A. & Cresswell, P. In vitro reconstitution of the MHC class I peptide-loading complex. *Methods Mol. Biol.* **960**, 67-79 (2013).

26. Hartman, I. Z. *et al*. A reductionist cell-free major histocompatibility complex class II antigen processing system identifies immunodominant epitopes. *Nat. Med.* **16**, 1333-1340 (2010).

27. Wearsch, P., A., Peaper, D., R. & Cresswell, P. Essential glycan-dependent interactions optimize MHC class I peptide loading. *Proc. Natl. Acad. Sci.* **108**, 4950-4955 (2011).

28. Lehner, P. J., Surman, M. J. & Cresswell, P. Soluble tapasin restores MHC class I expression and function in the tapasin-negative cell line.220. *Immunity* **8**, 221-231 (1998).

29. Leonhardt, R. M., Keusekotten, K., Bekpen, C. & Knittler, M. R. Critical role for the tapasin- docking site of TAP2 in the functional integrity of the MHC class I-peptide-loading complex. *J. Immunol.* **175**, 5104-5114 (2005).

30. Vigneron, N., Peaper, D., R., Leonhardt, R. M. & Cresswell, P. Functional signi fi cance of tapasin membrane association and disul fi de linkage to ERp57 in MHC class I presentation. *Euro. J. Immunol.* **39**, 2371-2376 (2009).

31. Sadegh-Nasseri, S. *et al*. Multiple paths to immunodominance: dominant epitopes are captured by MHC II prior, or post antigen proteolysis, and are enriched by HLA-DM (APP2P.103). *J. Immunol.* **192**, 43:4 (2014).

32. Storkus, W. J., Zeh, H. J. 3., Salter, R. D. & Lotze, M. T. Identification of T-cell epitopes: rapid isolation of class I-presented peptides from viable cells by mild acid elution. *J. Immunother. Emphasis Tumour Immunol.* **14**, 94-103 (1993).

33. Yates III, J. R., Ruse, C. I. & Nakorchevsky, A. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu. Rev. Biomed. Eng.* **11**, 49-79 (2009).

34. Shevchenko, A., Tomas, H., Havli, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1**, 2856-2860 (2007).

35. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663-670 (2003).

36. Hao, Z. *et al*. A quadrupole-orbitrap hybrid mass spectrometer offers highest benchtop performance for in-depth analysis of complex proteomes. *Thermo Scientific application note* **552** (2012).

37. Yates III, J. R. Mass spectral analysis in proteomics.
Annu. Rev. Biophys. Biomol. Struct. **33**, 297-316 (2004).

38. Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science* **312**, 212-217 (2006).

39. Karas, M., Bachmann, D. & Hillenkamp, F. Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules.
Anal. Chem. **57**, 2935-2939 (1985).

40. Tanaka, K. *et al*. Protein and polymer analyses up to *m/z* 100 000 by laser-ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2**, 151-153 (1988).

41. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C., M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71 (1989).

42. Knochenmuss, R. Ion formation mechanisms in UV-MALDI. *Analyst* **131**, 966-986 (2006).

43. Walther, T., C. & Mann, M. Mass spectrometry-based proteomics in cell biology. *J. Cell Biol.* **190**, 491-500 (2010).

44. Sadygov, R. G., Cociorva, D. & Yates III, J. R. Large-scale database searching using tandem mass spectra: Looking up the answer in the back of the book. *Nat. Methods* **1**, 195-202 (2004).

45. Eng, J. K., McCormack, A. L. & Yates III, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976-989 (1994).

46. Cox, J. *et al*. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794-1805 (2011).

47. Reiter, L. *et al*. Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry. *Mol. Cell. Proteomics* **8**, 2405-2417 (2009).

48. Gygi, S. P. *et al*. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994-999 (1999).

49. Thompson, A. *et al*. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895-1904 (2003).

50. Krey, J. F. *et al*. Accurate label-free protein quantitation with high- and low-resolution mass spectrometers. *J. Proteome Res.* **13**, 1034-1044 (2014).

51. Cox, J. *et al*. MaxLFQ allows accurate proteome-wide label-free quantitation by delayed normalization and maximal peptide ratio extraction. *Mol. Cell. Proteomics* **13**, 2513-2526 (2014).

52. Weinzierl, A. *et al*. Features of TAP-independent MHC class II ligands revealed by quantitative mass spectrometry. *Eur. J. Immunol.* **38**, 1503-1510 (2008).

53. Weinschenk, T. *et al*. Integrated functional genomics approach for the design of patient-individual antitumor vaccines. *Cancer Res.* **62**, 5818-5827 (2002).

54. Flad, T. *et al*. Direct identification of major histocompatibility complex class I-bound tumor-associated peptide antigens of a renal carcinoma cell line by a novel mass spectrometric method. *Cancer Res.* **58**, 5803-5811 (1998).

55. Invitrogen. Bac-to-Bac baculorivus expression system - an efficient site-specific transposition system to generate baculovirus for high-level expression of recombinant proteins. **Version F** (2010).

56. Invitrogen. Growth and maintenance of insect cell lines. (2010).

57. Mislick, K. A. & Baldeschwieler, J. D. Evidence for the role of proteoglycans in cation-mediated gene transfer. *Proc. Natl. Acad. Sci. USA* **93**, 12349-12354 (1996).

58. Davis, T. R., Trotter, K. M., Granados, R. R. & Wood, H. A. Baculovirus expression of alkaline phosphatase as a reporter gene for evaluation of production, glycosylation and secretion. *Nat. Biotechnol.* **10**, 1148-1150 (1992).

59. Stern, L., J. & Wiley, D. C. The human class II MHC protein HLA-DR1 assembles as empty αβ heterodimers in the absence of antigenic peptide. *Cell* **68**, 465-477 (1992).

60. Menge, U., Fraune, E., Lehmann, J. & Kula, M. R. Purification of proteins from cell culture supernatants. *Dev. Biol. Stand.* **66**, 391-401 (1987).

61. Rammensee, H., Bachmann, J., Emmerich, N. P., Bachor, O. A. & Stevanović, S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* **50**, 213-219 (1999).

62. Coriell Institute for Medical Research. GM02709 cell line information. https://catalog.coriell.org/0/sections/Search/Sample_Detail.aspx?Ref=GM02709&product=CC (accessed March 2014).

63. American Type Cell Culture.293T/17 culture method. http://www.atcc.org/products/all/CRL-11268.aspx#culturemethod (accessed April 2014).

64. Diedrich, J. K., Pinto, A. F. & Yates III, J. R. Energy dependence of HCD on peptide fragmentation: stepped collision energy finds the sweet spot. *J. Am. Soc. Mass Spectrom.* **24**, 1690-1699 (2013).

65. Valanne, S. *et al*. Genome-wide RNA interference in *Drosophila* cells identified G protein-coupled receptor kinase 2 as a conserved regulator of NF-κB signaling. *J. Immunol.* **184**, 6188-6198 (2010).

66. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367-1372 (2008).

67. Lundegaard, C. *et al*. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11**
. *Nucleic Acid Res.* **36, W509-W512 (2008).

68. Larkin, M. A. *et al*. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948 (2007).

69. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188-1190 (2004).

70. Institute for Health Metrics and Evaluation. Global burden of disease – years of life lost. http://viz.healthmetricsandevaluation.org/gbd-compare/ (accessed May 2014).

71. Tector, M. & Salter, R. D. Calnexin influences folding of human class I histocompatibility proteins but not their assembly with beta 2-microglobulin. *J. Biol. Chem.* **270**, 19638-19642 (1995).

72. Oliver, J. D., Roderick, H. L., Llewellyn, D. H. & High, S. ERp57 functions as a subunit of specific complexes formed with the ER lectins calreticulin and calnexin. *Mol. Biol. Cell* **10**, 2573-2582 (1999).

73. Ting, J. P. & Trowsdale, J. Genetic control of MHC class II expression. *Cell* **109**, S21-S33 (2002).

74. Gonalons, E., Barrachina, M., Garcia-Sanz, J. A. & Celada, A. Translational control of MHC class II I-A molecules by IFN-gamma. *J. Immunol.* **161**, 1837-1843 (1996).

75. Falk, K., Rotzschke, O., Stevanović, S., Jung, G. & Rammensee, H. G. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* **351**, 290-296 (1991).

76. Stern, L. J. *et al*. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* **368**, 215-221 (1994).

77. Lin, H. H., Ray, S., Tongchusak, S., Reinherz, E. L. & Brusic, V. Evaluation of MHC class I peptide binding prediction servers: Applications for vaccine research. *BMC Immunol.* **9** (2008).

78. Wang, P. *et al*. A Systematic Assessment of MHC Class II Peptide Binding Predictions and Evaluation of a Consensus Approach. *PLOS Comput. Biol.* **4**, e1000048 (2008).

79. Fenner, F., Henderson, D. A., Arita, I., Jezek, Z. & Ladnyi, I. D. in *Smallpox and its eradication* (World Health Organization, Geneva, 1988).

80. Food and Agriculture Organization. The global Rinderpest eradication programme. (2011).

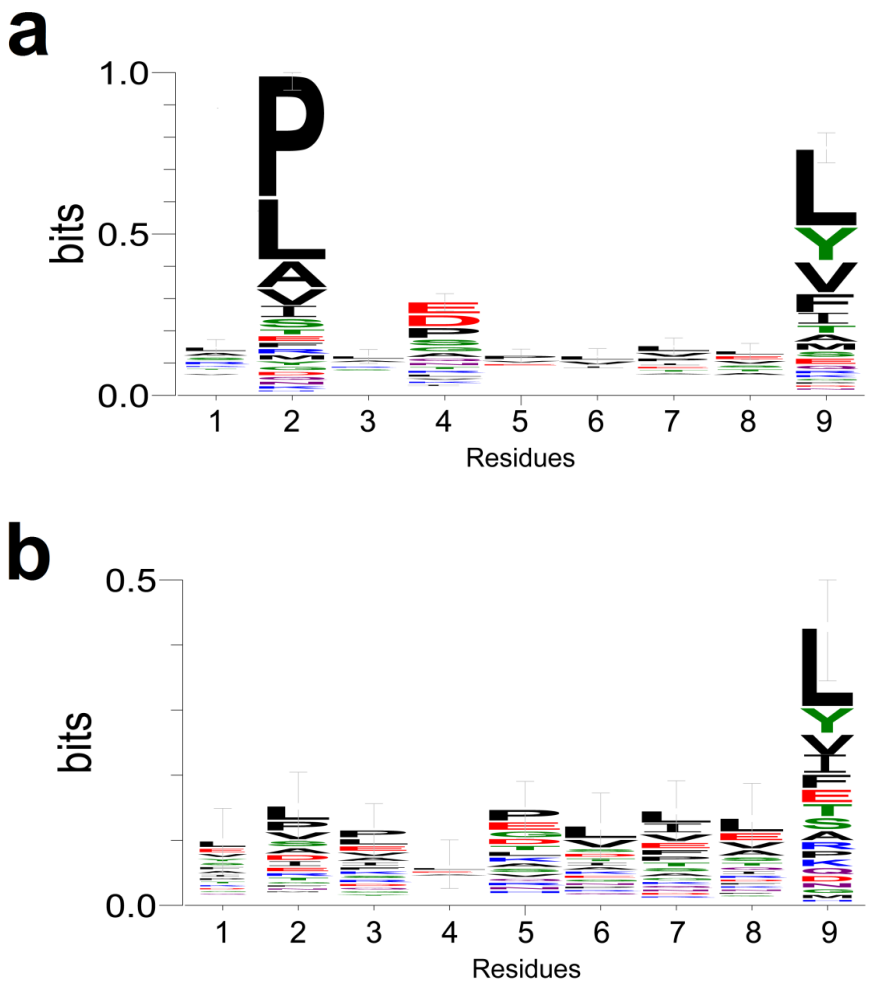
81. Masignani, V., Rappuoli, R. & Pizza, M. Reverse vaccinology: a genome-based approach for vaccine development. *Expert Opin. Biol. Ther.* **2**, 895-905 (2002).

82. Bowman, B. N. *et al*. Improving reverse vaccinology with a machine learning approach. *Vaccine* **29**, 8156-8164 (2011).

Appendix

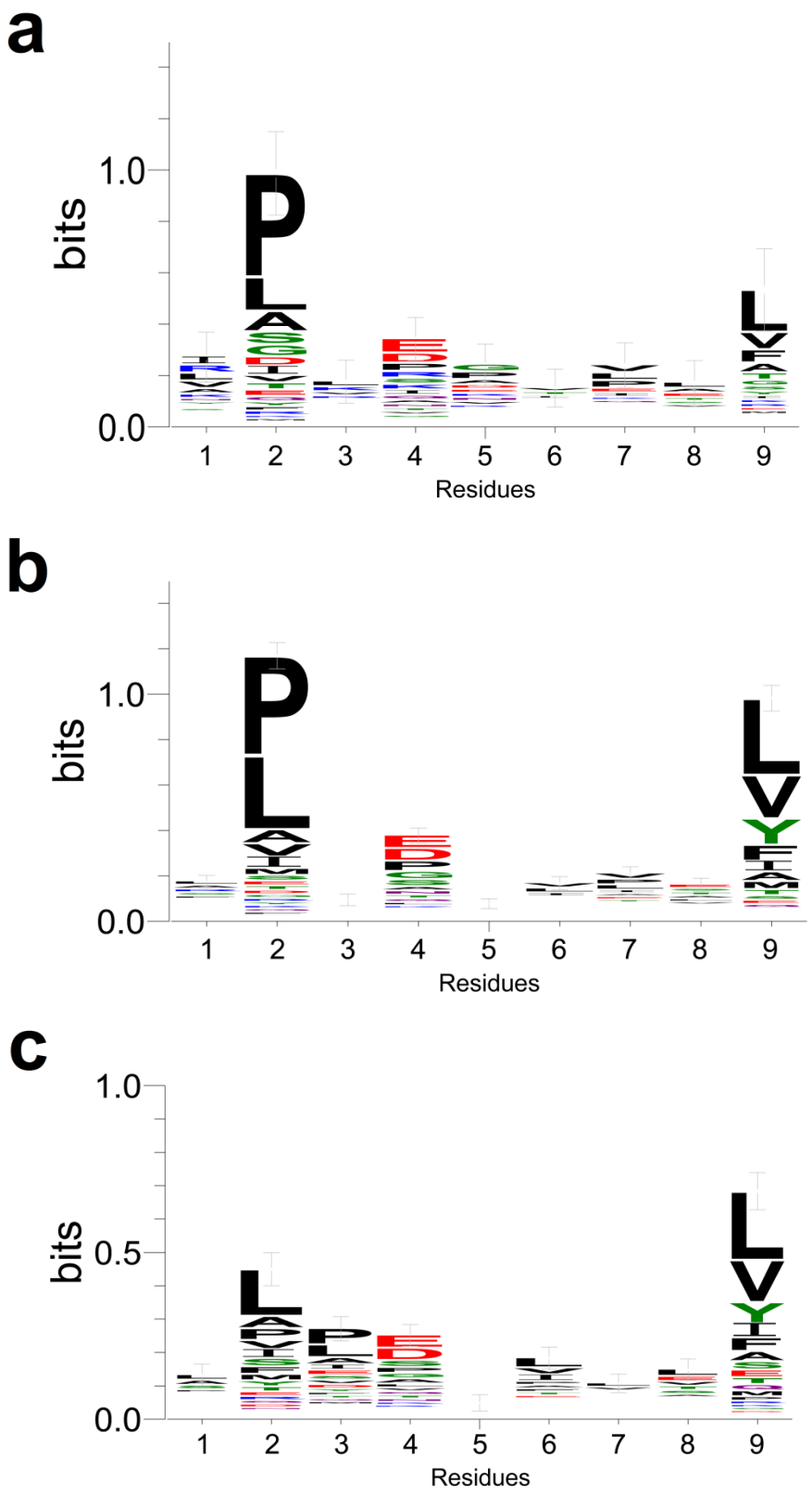
Appendix 1. Additional Weblogos for surface peptide bias assessment

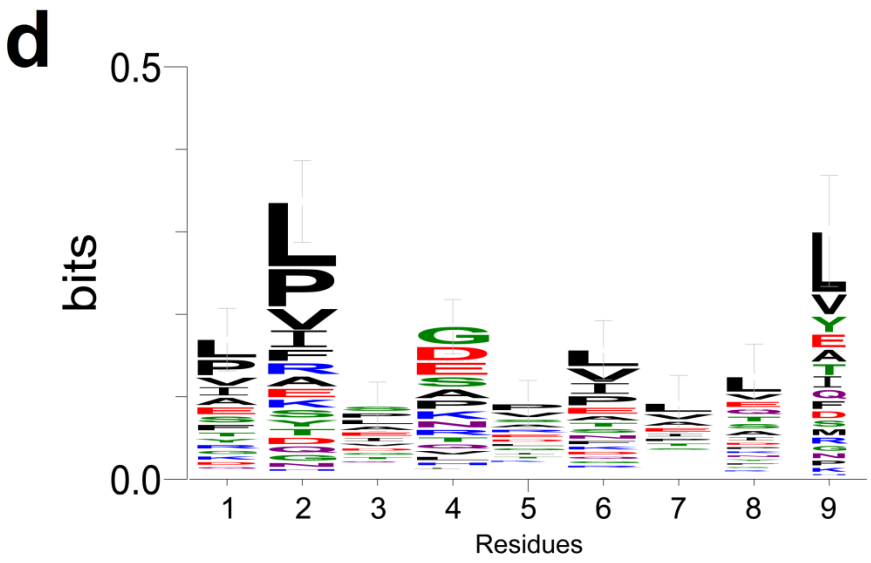
Figure 13 Additional Weblogos of multiple sequence aligned peptides derived from surface of HLA-A2 knockdown and negative control shRNA GM02709 cells.



Logos are compiled of peptides found in **(a)** both HLA-A2 knockdown and negative control cell lines, with knockdown-to-negative control peptide intensity ratios of greater than 0.206, and **(b)** only the knockdown cell line. [Anchor] or preferred residues, extracted from SYFPEITHI, for relevant alleles are as follows for positions 2, 4, 6, and 9: HLA-A2 [LM]-E-V-[VL], HLA-B7 [P]-D-L-[LF], HLA-B35 [P]-DEG-VL-[YFMLI]

Figure 14 Weblogos of multiple sequence aligned peptides derived from surface of negative control shRNA and healthy GM02709 cells.





Logos are compiled of peptides found in **(a)** both negative control shRNA transduced and healthy cell lines, with negative control-to-healthy peptide intensity ratios of less than 0.206, **(b)** both negative control shRNA and healthy cell lines, with negative control-to-healthy peptide intensity ratios of greater than 0.206, **(c)** only the negative control shRNA cell line, and **(d)** only the healthy cell line. [Anchor] or preferred residues, extracted from SYFPEITHI, for relevant alleles are as follows for positions 2, 4, 6, and 9: HLA-A2 [LM]-E-V-[VL], HLA-B7 [P]-D-L-[LF], HLA-B35 [P]-DEG-VL-[YFMLI]