Conserved Non-Pocket Interactions Drive the Diversity of Peptide Presentation by MHC Class I Molecules

Kyle Jackson

Follow this and additional works at: https://digitalcommons.library.tmc.edu/utgsbs_dissertations

Part of the Immunity Commons, Medicine and Health Sciences Commons, and the Other Immunology and Infectious Disease Commons

Recommended Citation
Jackson, Kyle, "Conserved Non-Pocket Interactions Drive the Diversity of Peptide Presentation by MHC Class I Molecules" (2021). The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access). 1135. https://digitalcommons.library.tmc.edu/utgsbs_dissertations/1135

This Dissertation (PhD) is brought to you for free and open access by the The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences at DigitalCommons@TMC. It has been accepted for inclusion in The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact digitalcommons@library.tmc.edu.
CONSERVED NON-POCKET INTERACTIONS DRIVE THE DIVERSITY OF PEPTIDE PRESENTATION BY MHC CLASS I MOLECULES

by
Kyle Ross Jackson, B.S.

APPROVED:

[Signatures]

Dean, The University of Texas
MD Anderson Cancer Center UTHHealth Graduate School of Biomedical Sciences
CONSERVED NON-POCKET INTERACTIONS DRIVE THE DIVERSITY OF
PEPTIDE PRESENTATION BY MHC CLASS I MOLECULES

A

Dissertation

Presented to the Faculty of

The University of Texas

MD Anderson Cancer Center UTHealth

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

by

Kyle Ross Jackson, B.S.

Houston, Texas

August, 2021
ACKNOWLEDGEMENTS

First I would like to thank my mentor, Dr. Greg Lizée, for everything he has done over the past seven years. I would not be the scientist I am today without his guidance.

Next I would like thank my wife, Jill Jackson, for her loving kindness and support through all of these years. Every time I experienced a setback she was there to help me and I would also not be where I am today without her.

I would also like to thank and acknowledge all the members of the Lizee lab, past and present, for their support and friendship. Brenda Melendez was the first to teach me about cloning, molecular biology and lentivirus, Sherille Bradley always answered my questions and put up with my messy desk and Heather Sonnemann always had a listening ear when things got frustrating. Amjad Talukder and Arjun Katalilha have been instrumental in helping with all aspects of my research work from cell lines to peptide elutions. Fenge Li and Yimo Sun have also been helpful in the short time we have been in the lab together. I would also like to thank and acknowledge all the many undergraduate, summer, graduate and medical students I had the pleasure to work with.

I would also like to acknowledge and thank all the collaborators I have had the pleasure of working with other the past several years. Dinler Antunes has become a good friend and I am thankful for his help, knowledge and prowess in computational simulations. Lydia Kavraki and her lab group at Rice University have also been supportive and a pleasure to work with. I would also like to thank Dr. David Hawke, who always took time to talk with me. I would also like to thank all the current and previous members of my advisory and exam committees for their support and guidance. Last, but certainly not least, I would like to thank my mom for her support from afar and my children, Emery and Eli, who remind me every day to find the joy in life and not take any moment for granted.
Conserved Non-Pocket Interactions Drive the Diversity of Peptide Presentation by MHC Class I Molecules

Kyle Ross Jackson, B.S.
Advisory Professor: Gregory Lizée, Ph.D.

Cytotoxic T-lymphocytes (CTL) can lyse infected or transformed cells through recognition of peptides presented on human leukocyte antigen (HLA) molecules. A thorough understanding of peptide-HLA interactions is needed for improvement of CTL-based immunotherapies. We observed that aspartic acid (D) and glutamic acid (E) at peptide position 4 are highly prevalent in HLA-I peptide ligands, and discovered that they interact with arginine (R) in position 65 and lysine (L) in position 66 of the α1 helix of the binding groove in HLA-A*0201 and HLA-A*2402. Since this interaction differed from well-characterized peptide-HLA anchor interactions mediated by peptide position 2 and the C-terminus, we investigated if the charged interactions between D/E in position 4 and R65 and K66 on the α1 helix of A*0201 and A*2402 are important for peptide antigen presentation. Mutations to R65 and K66 caused loss of HLA cell surface expression, reduced peptide repertoire diversity, and loss of charged interactions between D/E and p65-66 in both A*0201 and A*2402. In addition, mutating R65 and/or K66 altered the peptide motifs of both A*0201 and A*2402 specifically in peptide positions 1, 2 and the C-terminus. Peptide binding and stability assays revealed that D/E residues in position 4 contribute significantly to both peptide binding and stability, particularly in peptides with weaker N and C-terminal anchors. Several other HLA-I alleles and MHC molecules from different animal species also showed conservation of R65 and K66, in addition to D/E in bound peptide ligands, suggesting that charged interactions between D/E and R65/K66 represent a conserved, non-canonical pan-MHC-I anchor interaction that should be accounted for when identifying peptide targets for immunotherapies.
# TABLE OF CONTENTS

| APPROVAL SIGNATURES | ............................................................................. | i |
| TITLE PAGE | ........................................................................................................ | ii |
| ACKNOWLEDGEMENTS | ........................................................................................................ | iii |
| ABSTRACT | ................................................................................................................ | iv |
| TABLE OF CONTENTS | ........................................................................................................ | v |
| LIST OF FIGURES | ........................................................................................................ | ix |
| LIST OF TABLES | ........................................................................................................ | xiii |
| ABBREVIATIONS | ........................................................................................................ | xiv |

## CHAPTER I: INTRODUCTION AND BACKGROUND ........................................... 1

1.1 Introduction to immunity ........................................................................... 2
1.2 Human Leukocyte Antigens ...................................................................... 6
1.3 HLA-I molecules bind intracellular peptides via pockets and anchors ...... 11
1.4 Antigen presentation of intracellular peptides for CTL recognition ......... 15
1.5 Identifying HLA-I-restricted peptide targets for cancer immunotherapies ........................................................................................................ | 18 |
1.6 Mass spectrometry permits direct analysis of the HLA-I immunopeptidome ........................................................................................................ | 21 |
1.7 Molecular dynamics and HLA-I/peptide interactions ................................ | 25 |
1.8 Middle peptide residues interact with the HLA-I peptide binding groove | 26 |
1.9 Survey of eluted HLA-I peptides reveals a preference for negatively charged amino acids in peptide position 4 ............................................. 31
1.10 Interactions may occur between D/E residues in peptide position 4 and R65/K66 residues of A*0201 and A*2402 .......................................................... 38

1.11 Rationale, Hypothesis, Key Questions and Approach .............................................. 45

CHAPTER II: NON-POCKET INTERACTIONS BETWEEN HLA-A*0201 AND BOUND PEPTIDES ARE REQUIRED FOR OPTIMAL PEPTIDE BINDING STABILITY AND COMPLETE PEPTIDE REPERTOIRE DIVERSITY ............................................. 49

2.1 R65 and K66 interactions in WT HLA*0201 molecules ............................................. 50

2.2 Creation of mutated alleles of HLA-A*0201 .............................................................. 56

2.3 HLA*0201 mutants demonstrate loss of cell surface expression and peptide repertoire ............................................................................................................. 59

2.4 Immunopeptidome analysis reveals alterations in peptide position 4 frequencies of mutated HLA A*0201 ........................................................................................................................................................................... 65

2.5 Loss of R65 and/or K66 in HLA-A*0201 alters HLA-I/peptide interactions .......... 69

2.6 D/E4 peptide residues contribute to peptide binding and stability in A*0201 ...... 77

2.7 Mutation of R65 and/or K66 alters HLA-A*0201 peptide-binding motifs ........ 80

2.8 Increased frequencies of K/R1 residues in peptides eluted from A2-K66N and A2-DM mutants ............................................................................................................................................................................. 83

2.9 Peptide motif alterations in primary anchor positions of A*0201 mutants .... 90

2.10 Discussion ................................................................................................................... 93

CHAPTER III: NON-POCKET INTERACTIONS BETWEEN HLA-A*2402 AND BOUND PEPTIDES ARE REQUIRED FOR OPTIMAL PEPTIDE BINDING STABILITY AND COMPLETE PEPTIDE REPERTOIRE DIVERSITY ............................................... 97
3.1 Creation of mutated alleles of HLA-A*2402……………………………………………………98
3.2 Positively-charged amino acids in p65-66 of HLA-A*2402 facilitate cell surface
expression, peptide repertoire diversity, and D/E4 composition of bound peptide
ligands…………………………………………………………………………………………………104
3.3 Substituting p65 or p66 in A*2402 alters HLA-I/peptide interactions ..........110
3.4 D/E4 peptide residues contribute to peptide binding to HLA-A*2402………114
3.5 Mutating p65 or p66 residues alters HLA-A*2402 peptide-binding motifs……118
3.6 D/E4 stabilizes HLA-A*2402 peptides with weaker primary anchors………..126
3.7 Discussion……………………………………………………………………………………129

CHAPTER IV: NON-POCKET CHARGED INTERACTIONS SHOW SIMILARITIES
IN A*0201 AND A*2402, AND MAY BE CONSERVED IN MHC CLASS I
MOLECULES……………………………………………………………………………………132

Comparison of HLA-A*0201 and HLA-A*2402 results………………………………………133
4.1 Basic residues in p65/66 of HLA-A*0201 and A*2402 are required for optimal
antigen presentation and full peptide repertoire diversity…………………………….133
4.2 D/E4 peptide residues enhanced the stability of peptides containing fewer
dominant primary anchors……………………………………………………………………136
4.3 Discussion……………………………………………………………………………………143

Non-pocket interactions involving R65 and K66 may be conserved across the HLA-I
system and MHC-I molecules from other species …………………………………………144
4.4 R65 and K66 are conserved across HLA-I alleles and their presence correlate
with presentation of D/E4 peptide ligands………………………………………………144
4.5 D/E4 peptide preference is conserved in animal MHC-I molecules……….148
HLA-I molecules show a preference for positively charged amino acids in peptide position 1.

4.7 K/R1 peptides are prevalent in HLA-I molecules lacking K66, R62 or R163.

4.8 Discussion

CHAPTER V: OVERALL CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Overall Conclusions

5.2 Future directions

CHAPTER VI: MATERIALS AND METHODS

BIBLIOGRAPHY

VITA
LIST OF FIGURES

CHAPTER I: INTRODUCTION AND BACKGROUND

Figure 1.1 Activated cytotoxic T-lymphocytes lyse cells based on TCR recognition of HLA-I/peptide complexes .................................................................................................................. 7

Figure 1.2 Peptides bind HLA-I molecules via conserved residues at the N and C-termini .................................................................................................................................................... 13

Figure 1.3 Overview of the HLA-I antigen presentation pathway ........................................ 16

Figure 1.4 Mass spectrometry can identify HLA-I bound peptide sequences .......... 22

Figure 1.5 Middle peptide residues can form non-pocket HLA-I/peptide interactions .......................................................................................................................................................... 29

Figure 1.6 Overall survey of MS-eluted HLA-I peptides .................................................. 33

Figure 1.7 Aspartic acid (D) and glutamic acid (E) are highly preferred in HLA-A, B and C molecules .............................................................................................................................................. 36

Figure 1.8 Most HLA-A, B and C alleles prefer D or E residues in peptide position 4 ......................................................................................................................................................... 39

Figure 1.9 R65 and K66 may interact with negatively charged p4 amino acids in HLA-A*0201 and HLA-A*2402 ........................................................................................................................................ 43

CHAPTER II: NON-POCKET INTERACTIONS BETWEEN HLA-A*0201 AND BOUND PEPTIDES ARE REQUIRED FOR OPTIMAL PEPTIDE BINDING STABILITY AND COMPLETE PEPTIDE REPertoire DIVERSITY

Figure 2.1 In HLA-A*0201, peptide residues D4 and E4 both interact with R65 and K66 in multiple peptide backbones ............................................................................................................................................ 51
Figure 2.2 R65 and K66 interactions with GLEKGIPAL in the peptide binding groove.................................................................54

Fig 2.3 A*0201 and A*2402 peptide-binding motifs are distinct from those of endogenous HLA-I in H1975 cells.................................................................60

Figure 2.4 Loss of cell surface expression and peptide repertoire in mutated HLA-A*0201 molecules.................................................................63

Figure 2.5 Peptide position 4 amino acid frequencies are altered in the immunopeptidomes of HLA-A*0201 mutants.................................................................67

Figure 2.6 Charged interactions and salt bridges occur between E4 and p65-66 in A*0201 mutants.................................................................71

Fig 2.7 P65-66 substitutions in HLA-A*0201 alter HLA-I/peptide interactions in the binding groove..................................................75

Fig 2.8 D/E4 residues contribute to peptide binding and stability in HLA-A*0201....78

Fig 2.9 Mutations to positions 65/66 alter HLA-A*0201 peptide-binding motifs......81

Fig 2.10 K66N mutation changes amino acid composition of peptide position 1 in A*0201..............................................................................84

Fig 2.11 Interpeptide interactions between K1 and D/E4 occur in eluted HLA-A*0201 peptides........................................................................87

Fig 2.12 Mutation of R65 altered C-terminal anchors while mutation of K66 altered p2 peptide anchors.................................................................91

CHAPTER III: NON-POCKET INTERACTIONS BETWEEN HLA-A*2402 AND BOUND PEPTIDES ARE REQUIRED FOR OPTIMAL PEPTIDE BINDING STABILITY AND COMPLETE PEPTIDE REPertoire DIVERSITY
Figure 3.1 Network of K66 interactions observed in HLA-A*2402/QFKDNVILL complexes……………………………………………………………………………………..99
Figure 3.2 HLA-A*2402 cell surface expression and peptide repertoire size is dependent on the number of positively charged amino acids in p65-66……………….105
Figure 3.3 Composition of amino acids in peptide position 4 is influenced by the number of positively charged amino acids in p65-66 of HLA-A*2402……………….108
Figure 3.4 Mutations to p65-66 in HLA-A*2402 alters HLA-I/peptide interactions..112
Figure 3.5 D/E4 peptide residues enhance peptide binding to HLA-A*2402 in multiple peptide backbones……………………………………………………………………………………..116
Figure 3.6 Mutations to positions 65/66 alter HLA-A*2402 peptide-binding motifs………………………………………………………………………………………………………..119
Figure 3.7 Peptide K1 residues cause increased stability in HLA-A*2402………..121
Figure 3.8 Alterations in primary anchor positions of mutated HLA-A*2402 molecules…………………………………………………………………………………………………………….124
Figure 3.9 D4 residue is more critical for stability of HLA-A*2402 peptides with weaker primary anchors………………………………………………………………………………………………………………….127

CHAPTER IV: NON-POCKET CHARGED INTERACTIONS SHOW SIMILARITIES IN A*0201 AND A*2402, AND MAY BE CONSERVED IN MHC CLASS I MOLECULES
Figure 4.1 HLA molecules with more positively charged amino acids in p65-66 demonstrate higher cell surface expression and expanded peptide repertoires….134
Figure 4.2 Percentage of D/E4 eluted peptides correlates with number of positively charged amino acids in HLA-I p65-66……………………………………………..138
Figure 4.3 D/E4 peptide residues are more important for stability of peptides with fewer dominant primary anchors.................................................................141

Figure 4.4 Presence of R65 and/or K66 in HLA-I molecules correlates with frequency of bound D/E4 peptides.................................................................146

Figure 4.5 Animal MHC-I molecules prefer D/E4 peptides and conserve R65......149

Figure 4.6: HLA-I molecules lacking R62, R163 and K66 prefer K/R residues in peptide position 1.................................................................155
LIST OF TABLES

Table 1.1 Comparison of receptors in innate and adaptive immunity ..................4
Table 1.2 Worldwide phenotypic and allele frequencies of A*0201 and A*2402.....46
Table 2.1 HLA-A*0201 mutants and their altered p65-66 residues......................57
Table 3.1 HLA-A*2402 mutants and their p65-66 residue alterations......................102
ABBREVIATIONS

**MHC**: Major Histocompatibility Complex

**HLA-I**: Human Leukocyte Antigen Class I

**CTL**: Cytotoxic T-lymphocyte

**TCR**: T-cell receptor

**Asp / D**: Aspartic acid

**Glu / E**: Glutamic acid

**Lys / K**: Lysine

**Arg / R**: Arginine

**Leu / L**: Leucine

**Val / V**: Valine

**Trp / W**: Tryptophan

**Phe / F**: Phenylalanine

**p2**: Peptide position 2

**pΩ**: C-terminal peptide position

**pAPCs**: professional antigen-presenting cells
CHAPTER I: INTRODUCTION AND BACKGROUND
1.1 Introduction to immunity

Immunity can be broadly defined as the ability or capacity to ward off infectious diseases and lyse transformed cells. This is necessary as interaction with our environment causes exposure to pathogens, such as viruses and bacteria, or foreign substances that may alter cells, such as carcinogens. In response to these threats vertebrates, especially mammals, have evolved a complex immune system that utilizes several different danger signals and immune cells to detect and destroy foreign pathogens, as well as prevent re-infection and to provide surveillance against cellular transformation (1).

The immune system can be split into innate or adaptive immunity, though the entire system functions together as a harmonious whole (2). Innate immunity refers to front-line or immediate defenses that rapidly occur against any infectious organism through identification of multiple pathogen-derived signals (3). These defenses are static in that they do not change in response to any specific pathogen (1, 2). Activation of innate immunity leads to development of adaptive immunity, the second arm of the immune system. Adaptive immunity refers to the ability of specialized white blood cells, or lymphocytes, to respond to specific foreign substances or antigens for clearance of infection (4). Lymphocytes are also vital for the formation of immune memory, which is the ability to rapidly respond to re-infection by the same pathogen (2, 4).

Adaptive immunity can also be separated into humoral and cellular immunity. Humoral immunity is the capacity to create antibodies against specific targets or antigens that are freely circulating the blood and is mediated by B-cells (3, 4).
Cellular immunity involves responses against intracellular infection or cellular transformation and is mediated by T-cells (1). While several different types of T-cells exist, T-cells that have the ability to lyse infected or transformed cells are called cytotoxic T-lymphocytes (CTLs) (2, 4).

The ability to differentiate between self and foreign antigens is referred to as self-tolerance and is critical for the immune system to respond only against foreign antigens (1, 2). Several mechanisms are in place to allow both innate and adaptive immune responses to target antigens from outside the host only (Table 1.1). Innate immunity relies on multiple receptors (ex. Toll-like receptors), where each receptor recognizes a conserved trait of pathogens, such as flagellin from bacteria or double-stranded RNA in viruses (5). These receptors are invariant, meaning they cannot change or adapt, and are expressed by both normal cells and innate immune cells (3, 5).

Lymphocytes rely on two receptors for specific recognition of antigens: T-cell receptors (TCRs) for cellular immunity and B-cell receptors (BCRs) for humoral immunity (Table 1.1). Both types receptors are generated through genetic recombination events, resulting in a very high level of TCR and BCR diversity within individuals that are capable of recognizing millions of different antigens (4). Despite this collective diversity, each T or B lymphocyte bears only a single TCR or BCR receptor. T cells and B cells are selected for during early development to delete potentially dangerous cells with high-affinity receptors recognizing self-antigens, and several other mechanisms are also in place to preserve self-tolerance after maturation of lymphocytes (1, 2). Soluble forms of BCRs produced by B-cells are
### Receptors responsible for recognition of foreign substances

<table>
<thead>
<tr>
<th>Innate Immunity</th>
<th>Adaptive immunity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple receptors (ex. TLRs)</td>
<td>2 receptors (TCR and BCR)</td>
</tr>
<tr>
<td>Invariant: each receptor recognizes one trait</td>
<td>Adaptable: can recognize almost any antigen or pathogen</td>
</tr>
<tr>
<td>Are expressed on normal and immune cells</td>
<td>Expressed only by lymphocytes</td>
</tr>
<tr>
<td>One cell can express multiple receptors</td>
<td>Each lymphocyte bears TCR or BCR with same specificity</td>
</tr>
<tr>
<td>No selection process</td>
<td>Each receptor undergoes multiple selection processes</td>
</tr>
</tbody>
</table>

Table 1.1 Comparison of receptors in innate and adaptive immunity
Table 1.1 Comparison of receptors in innate and adaptive immunity:

Differences in the characteristics of receptors that recognize foreign substances or antigens in innate versus adaptive immunity.
called antibodies, and these can directly recognize and bind to antigens in the blood or other bodily fluids to activate innate and adaptive immunity and initiate pathogen clearance \((2, 4)\). By contrast, CTLs identify infected or transformed cells through interaction of their membrane-bound TCRs with human leukocyte antigen (HLA) molecules presenting antigenic peptides, which will be discussed in the next section \((7)\).

### 1.2 Human Leukocyte Antigens

Cell surface HLA molecules constitutively present peptide fragments derived from intracellular proteins to the immune system as a ‘snapshot’ of the current protein composition of the cell. The process of presenting peptide-HLA complexes to the immune system is called antigen presentation and occurs in all nucleated cells of the body \((6)\). Antigen presentation allows for the surveillance of intracellular activity by CTLs and for the detection of viral infection, since peptides from viral proteins are also presented by HLA molecules \((4, 6)\). Recognition of peptide-HLA complexes by activated CTLs leads to target cell lysis via perforin and Granzyme-B \((\text{Fig 1.1})\) \((7, 8)\).

The discovery of human HLA molecules dates back over 75 years and involves 5 Nobel laureates. Allogeneic tumor transplantation in mice in the early 1900s showed that blood group antigens were responsible for tumor growth in different mouse strains \((9)\). Studies in the 1940s in rabbits demonstrated that rejection of allogeneic grafts was caused by an immune response, and further examination in mice revealed that this was due to the presence of major histocompatibility antigens (MHC) \((9-11)\). In 1958, the first human MHC molecules, termed HLA, were
Cytotoxic T-lymphocytes recognize peptide-HLA complexes

Fig. 1.1 Activated cytotoxic T-lymphocytes lyse cells based on TCR recognition of HLA-I/peptide complexes
Fig 1.1: Activated cytotoxic T-lymphocytes lyse cells based on TCR recognition of HLA-I/peptide complexes: Activated cytotoxic T-lymphocytes (CTLs) lyse cells based on recognition of specific peptide-HLA complexes through release of perforin and granzyme B*. T-cell receptors (TCRs) recognize the combination of both peptide and HLA molecule through interactions on the surface of the binding groove⁺.


⁺ Courtesy of Dr. Dinler Antunes, University of Houston
discovered by Jean Dausset using sera that caused agglutination of white blood cells (12). This led to an explosion of research into MHC molecules which was aided by International Histocompatibility Workshops where researchers from around the world worked together to unravel the complexities of the MHC system (9, 13, 14).

However, the true biological function of MHC/HLA molecules remained a mystery for several years. The first discovery came in the 1960s where it was determined in mice that different MHC molecules controlled the ability of mice to make antibodies against synthetic antigens (9). In 1973, a strong association was reported between HLA-B27 and ankylosing spondylitis, an autoimmune condition, showing that HLA molecules were involved in disease (15, 16). Also in 1973, Zinkernagel and Doherty both demonstrated that CTLs from mice strains with 2 different MHC haplotypes were unable to kill infected cells from the opposing mice strain, showing that CTL killing was MHC-restricted (17, 18). This was later shown to be true in CTL-HLA interactions as well, but the mechanism for this restriction was not known until the late 1980s when the crystal structure of HLA-A*0201 was discovered (9). The crystal structure revealed the presence of a peptide-binding groove created by 2 parallel alpha helices and a Beta sheet floor which was bound to a peptide not derived from HLA-A*0201 (19, 20). Thus, HLA molecules restricted CTL recognition by presenting intracellular peptides to T-cells (9).

The human major histocompatibility complex is located on chromosome 6 and contains several loci which are split into three distinct groups. HLA class I (HLA-I) genes encode for molecules that present short peptides at the surface of every nucleated cell; these are the molecules specifically recognized by CD8+ T cells, or
CTLs (21). HLA class II molecules are expressed mainly in professional antigen-presenting cells (APCs) and present longer peptides derived largely from extracellular antigens to stimulate CD4+ helper T-cell responses (22). HLA class III gene products do not present antigens, but usually play other roles in the immune system.

HLA-I molecules can be split into several different classes. The HLA-A, B and C genes encode the ‘classical’ HLA molecules that present intracellular peptides to the immune system (4). Other ‘non-classical’ HLA molecules may also present peptides, such as HLA-E, HLA-F, and HLA-G, but can also play different roles in the immune system (23-25). Mature HLA-I molecules are tri-molecular complexes composed of the heavy alpha chain, Beta-2-microglobulin, and peptide that are assembled non-covalently in the endoplasmic reticulum prior to being transported to the cell surface for antigen presentation (6). The heavy chain contains a ~35 amino acid cytoplasmic domain, a transmembrane domain, and 3 extracellular immunoglobulin-like domains (alpha-1, -2, and -3). The alpha-1 and -2 domains together form a peptide-binding groove composed of beta-sheets for the floor and alpha helices that form the sides (19). The binding groove is closed at either end, limiting the size of peptide ligands that can bind to <13 amino acids, although 9- and 10-mers are the most common.

Every individual inherits one HLA-A, B and C allele from each parent for a total of six classical HLA-I alleles which are all co-expressed (21). Classical HLA-I are the most polymorphic genes in the human genome, with 21,353 alleles currently known for HLA-A, B and C (26). This allelic sequence variation has important implications
for peptide binding, antigen presentation, and cell-mediated immunity as will be described in the next section.

1.3 HLA-I molecules bind intracellular peptides via pockets and anchors

Once it was established that HLA molecules presented intracellular peptides, work began to understand the mechanisms that govern peptide-HLA interactions. It was quickly established that the binding groove contains several pockets in the beta sheet, designated A through F, and that the most hypervariable positions are in these pockets (20, 27). Comparative analyses of similar HLA molecules or mutation of specific pocket residues illustrated that these pockets govern overall HLA binding (28-31). Further studies determined that amino acids at the N- and C-termini are the most important anchor residues for peptide binding to HLA-I (32-35). This is illustrated with a peptide-binding assay measuring binding of the SYIGLKGLYF peptide to HLA-A*2402, where alanine substitutions were made in every peptide position. However, only substitution of the tyrosine at position 2 or the C-terminal phenylalanine abrogated binding (Fig 1.2A). Consistent with this, additional analyses showed that two dominant pockets bind the C-terminal amino acid (pΩ, pocket F) and peptide position 2 (p2, pocket B), and these constitute the primary anchors most important for stabilizing peptide binding (33, 36-38).

The biochemical characteristics, size and shape of pockets B and F dictate that certain amino acids will interact in the pocket more optimally than others. Thus, some amino acids are preferred in different peptide positions and this preference is specific for each HLA molecule. These preferences are revealed in peptide-binding
motifs compiled from peptides that bind to each HLA molecule, where the prevalence of each amino acid at different peptide positions is shown relative to each other. Examination of the binding motif for HLA-A*0201 shows that leucine is highly preferred in p2 whereas leucine, valine and isoleucine are all prevalent at pΩ of bound peptides (Fig 1.2B). In HLA-A*2402, tyrosine, a bulky and polar amino acid, is highly favored at p2 whereas phenylalanine, leucine and isoleucine are favored at pΩ (Fig 1.2C). Thus, each HLA molecule contains a unique peptide-binding motif based on peptide interactions with pocket residues (39, 40).

Binding assays of multiple peptides to HLA-I molecules reveals that some peptides bind more tightly than others do, even when they share primary anchor residues. These assays demonstrate that primary anchors are necessary but not always sufficient for optimal peptide binding and stability. Several studies using pooled sequencing and binding assays attempted to determine which non-anchor (i.e. non-p2 or pΩ) amino acid residues were most important for HLA-I binding (41-43). These studies showed that some amino acids were favorable or tolerated at certain peptide positions, while others were unfavorable and disrupted peptide binding. Thus, peptide-HLA binding consists of interactions of primary anchor residues in pockets B and F, in addition to secondary interactions that can occur throughout the peptide.
HLA-I molecules bind peptides by conserved end residues

Fig 1.2 Peptides bind HLA-I molecules via conserved residues at the N and C-termini
Fig 1.2 Peptides bind HLA-I molecules via conserved residues at the N and C-termini: (A) Competitive peptide binding assay of SYIGLKLGYF in HLA-A*2402 with alanine substitutions in each peptide position. Percent inhibition of fluorescence was calculated and compared to positive and negative control peptides. Alanine substitution of position 2 and C-terminus caused loss of peptide binding. (B) Peptide sequence logos of eluted peptides from both HLA-A*0201 and HLA-A*2402 reveal conserved amino acids at peptide ends (44).
1.4 Antigen presentation of intracellular peptides for CTL recognition

Intracellular peptides bind to HLA-I complexes in the endoplasmic reticulum (ER) (Fig 1.3). Proteins in the cytosol are degraded by the proteasome into shorter peptide fragments which are then transported into the ER through the transporter associated with antigen processing (TAP). Proteases in the ER trim the N-terminal peptide ends while chaperone proteins including tapasin, calreticulin, calnexin, and Erp57 assist in the loading peptides onto HLA-I molecules. Once stable peptide-HLA complexes are formed they are sent through the secretory pathway involving the Golgi and then to the cell surface. Multiple HLA-I peptide elutions by our lab and others has revealed that peptides from HLA-A molecules are the most prevalent, followed by HLA-B peptides and then HLA-C peptides, which mirrors the cell surface expression of these three gene products.

Peptide-binding motifs from the majority of HLA-I molecules show considerably less conservation of amino acids at middle peptide residues (i.e. p4 to p-1) compared to the positions at the N- and C- termini (see Fig 1.2B) (45). Crystal structures of peptide-HLA complexes reveal that the vast majority of middle residues are found at or near the top of the binding groove. In longer peptides, these middle residues bulge out and away from the binding groove, which is necessary to accommodate longer peptide lengths as the ends of the binding groove are closed. These residues in the middle of the peptide are most available for TCR contact upon TCR recognition of peptide-HLA complexes. The capacity of CTLs to target and kill cells expressing unique peptide-HLA targets has important implications for immunotherapeutic treatment of cancer, as will be described in the next section.
HLA-I Antigen presentation pathway

Fig 1.3 Overview of the HLA-I antigen presentation pathway
**Fig 1.3 Overview of the HLA-I antigen presentation pathway:** Peptides are processed by the proteosome and transported into the endoplasmic reticulum (ER) by the transporter associated with antigen processing (TAP) 1 and 2 and bind to nascent HLA molecules with the help of multiple chaperone proteins. Peptide-loaded HLA molecules are then processed through the secretory pathway and expressed on the cell surface*. 

*Created with BioRender.com
1.5 Identifying HLA-I-restricted peptide targets for cancer immunotherapies

Attempts to harness the power of the immune system to treat cancer have been ongoing for over 100 years. The earliest studies utilized mixtures of live and attenuated bacteria in an attempt to induce cancer remission, with mixed results (46, 47). In the late 1950s, Dr. Prehn showed that immunity developed against sarcoma in mice that later protected them from re-challenge (48, 49). This demonstrated that antigens from tumors can cause establishment of immune memory, potentially indicating the involvement of lymphocytes. In 1989, Wölfel et al showed that human melanoma cells could be lysed by CTLs due to interactions with peptide-HLA complexes (50). This and other discoveries led to the development of the first active immunotherapy using interleukin-2 by itself or in combination with tumor-infiltrating lymphocytes (TIL) to treat multiple cancer types, showing efficacy in reducing tumor burdens (51-53).

In recent years, the field of cancer immunotherapy has flourished with numerous approaches that utilize the ability of CTLs to lyse transformed cells. In the past decade, Jim Allison and his group pioneered the use of monoclonal antibodies to target specific T-cell receptors that inhibit the immune response, leading to significant tumor remissions and extended patient survival in melanoma and other cancer types (54-58). Significant advances have been made in other immunotherapeutic modalities, including infusions of genetically-engineered TIL or endogenous CTLs, and the potential of cancer vaccines to induce the production of tumor antigen-specific CTLs (58, 59).
Several cancer immunotherapies rely on identification of the peptide targets presented by HLA-I molecules. Once a peptide target has been identified there are two major routes for CTL-based immunotherapies. The first approach is to create a therapeutic cancer vaccine which can be administered to patients in multiple ways (60, 61). The objective of cancer vaccines is to activate the patient’s immune system to create CTLs that will lyse the transformed cells leading to tumor remission. Several different factors influence the success of cancer vaccines including the ability to identify personalized peptide targets, and the ability of the patient to mount a sufficient immune response (62, 63). Despite these challenges, successful generation of tumor antigen-specific immune responses and tumor regressions have been observed in some cancer vaccine studies (64, 65).

The other route that can be taken to generate CTL immunotherapies is to stimulate or modify CTLs ex vivo and then expand them for re-infusion back into the patient. Numerous approaches are under investigation including direct stimulation of CTLs derived from the patient or transducing TCRs reactive against specific peptide-HLA complexes directly into CTLs (66-68). Similar to cancer vaccines, challenges abound with this approach but successful tumor remissions have been demonstrated (69-72). However, both cancer vaccines and generation of CTLs require rapid and accurate identification of peptide targets presented by HLA-I.

The majority of tumor-associated peptides presented on HLA complexes are derived from self-proteins and therefore are not safe to target with CTLs. Some tumor-associated proteins are only expressed in tissues that can be more safely targeted such as melanocyte differentiation antigens and cancer-testis antigens in
melanoma (73-75). Recently, a new group of cancer-placenta antigens were discovered that are highly expressed in the placenta and have been identified as potential targets in multiple cancer types (76, 77). Other peptide targets are preferentially expressed in tumor cells and include proteins that necessary for transformation such as mutated or overexpressed oncogenes. However, one of the most intriguing categories of peptide targets are those derived from mutations in the DNA of tumor cells; as such, they are highly tumor-specific and are known as neo-antigens (78-80).

Identification and prediction of neo-antigens is currently one of the hottest topics in cancer immunotherapy. Since these peptides are derived from DNA mutations, they are not only tumor-specific but also highly personalized, due largely to the high HLA polymorphism within the human population combined with the private nature of most cancer mutations. Several approaches have been developed to identify and target neo-antigens for cancer immunotherapies (63). One unique approach involves transduction of several different mini-genes into a patient’s antigen-presenting cells and allowing the patient’s immune response identify and select the best neo-antigens to target (81-83). However, the mainstream approach has been to utilize genomic data to generate lists of potential neo-antigens and predict which peptides are most likely to bind to the patients expressed HLA-I molecules (84). This has led to an abundance of online HLA-I prediction programs utilizing sophisticated machine-learning techniques to identify similarities in HLA peptide training datasets that can then be used to predict whether neoantigen or other tumor-associated
peptides will bind to a given HLA-I molecule (85-88). In some instances, the top candidates are then validated by peptide-binding assays.

While highly promising, this approach has led to identification of several neoantigen peptides that are predicted to bind HLA-I, but are unable to generate effective antitumor immunity in vitro or in vivo (89-91). There are several potential reasons for this. One is that the binding prediction algorithms may be inaccurate for less well-studied HLA-I alleles. Multiple groups are addressing these concerns by updating or creating new peptide-binding prediction programs (92). Another reason is that the antigen-processing steps that occur before a peptide is presented cannot be accurately predicted, particularly in tumor cells where antigen processing is often dysfunctional to enable immune escape (63). A third possible reason for inaccurate binding predictions is that secondary peptide-HLA interactions may not be properly accounted for.

1.6 Mass spectrometry permits direct analysis of the HLA-I immunopeptidome

A major disadvantage of prediction-based tumor antigen discovery is the lack of information from the antigen processing and presentation process. This can be overcome through analyzing peptides identified directly from tumor-derived HLA molecules. Mass spectrometry (MS) is a proteomics method initially developed for protein identification but has been adapted for use in identifying HLA peptide sequences (93-95) (Fig 1.4). HLA molecules from any source, patient sample or cell
Mass spectrometry for immunopeptidome identification

Fig 1.4 Mass spectrometry can identify HLA-I bound peptide sequences
Fig 1.4 Mass spectrometry can identify HLA-I bound peptide sequences: HLA molecules are immunoprecipitated using a pan-HLA antibody and peptides are then eluted using a mild acid. Eluted peptides are then separated by hydrophobicity in liquid chromatography (LC) and then analyzed using tandem mass spectrometry (MS/MS). Each peptide generates a unique spectrum which is then compared against the human proteome to determine the peptide sequence*

lines, can be immunoprecipitated from lysates using HLA-specific antibodies (94, 96). Peptides are eluted from HLA molecules at low pH and then separated using high-pressure liquid chromatography based on hydrophobicity. Separated peptides are then injected into the mass spectrometer where they are ionized and fragmented (94, 97). Each peptide fragmentation generates a spectrum which can then be searched against peptide databases of the human proteome to determine the original peptide sequence (98, 99). Using this technique, thousands of HLA peptides can be identified from a tumor cell line in a short amount of time (95, 96).

The peptide repertoire generated from HLA-I molecules by MS analysis is referred to as the immunopeptidome. The ability to identify thousands of peptide sequences from individual patient tumor specimens or cell lines has the potential to revolutionize cancer immunotherapy in multiple ways. The immunopeptidomes of several different cancer types and HLA restrictions has now been annotated, allowing for direct identification of numerous shared tumor-associated antigens in multiple types of cancer (76). This windfall of validated MS-eluted peptides that have gone through antigen processing and presentation has been utilized to improve peptide-binding motifs for >100 common HLA alleles (100, 101). Most if not all online prediction programs now routinely include eluted peptides in their algorithms to improve prediction of unknown peptide antigens (45, 102-104). A number of studies have used MS to directly identify personalized neo-antigen peptide targets from patient-derived cell lines (105-108).

However, MS also has some significant limitations. The sensitivity level of MS is generally insufficient to detect peptides of lower abundance, which includes the vast
majority of neo-antigens and many shared tumor-associated peptide targets (63, 109). Further, MS analysis currently requires a large amount of tumor tissue (>200 mg), which far exceeds the amount typically provided by core needle biopsies (63, 110). Patient-derived tumor specimens also contain significant amounts of non-tumor cells (for example, infiltrating immune cells), providing a further challenge. Finally, deconvolution of peptide sequences from specific HLA molecules can be difficult and currently relies mainly on peptide binding prediction programs. Despite these limitations, MS remains a powerful tool for global analysis of immunopeptidomes and this information can facilitate a number of different CTL-based immunotherapies, as described above.

1.7 Molecular dynamics and HLA-I/peptide interactions

Another reason for the inability to accurately predict for binding or immunogenicity is a lack of understanding about the real-time interactions that occur in the peptide-binding groove. Currently, the majority of information available about potential interactions between peptides and HLA molecules is derived from crystal structures of a panel of different peptide-HLA complexes (111-113). Crystal structures are a valuable source of structural information that provides insights into a specific configuration of a peptide-HLA complex that occurs at a given moment in time (114). However, peptide-HLA interactions are naturally dynamic and changing over time, which cannot be accounted for from analyzing static crystal structures (115, 116).
Molecular dynamics is a computational technique that utilizes sophisticated programs and algorithms to simulate interactions between molecules including proteins and ligands over a specified period of time (117, 118). The 3-dimensional structure of a peptide-HLA complex derived from a crystal structure is used as a baseline and simulations are performed within specific restraints to calculate the specific interactions between the peptide and HLA binding groove under simulated physiological conditions over time (119-121).

Molecular dynamic simulations of peptide-HLA complexes have led to several discoveries that shape our understanding of peptide-HLA interactions. Some studies have focused on how a single HLA or peptide residue can alter the dynamics of the entire peptide-HLA complex (121, 122). For example, Abualrous and colleagues utilized truncated peptides to determine that occupation of the F pocket by the C-terminus is more important for long-term stability than the B pocket (123). Multiple groups have also used computational simulations to improve prediction of peptide binding or immunogenicity (124-126). Molecular dynamics is a powerful tool that can be utilized to study specific molecular interactions within the peptide-binding groove to gain insights into antigen presentation by different HLA alleles.

1.8 Middle peptide residues interact with the HLA-I peptide binding groove

Despite the advances in peptide binding prediction programs, MS analysis, and molecular dynamics, quick and reliable identification of immunogenic HLA peptide targets remains challenging. One potential reason involves the difference between peptide binding and stability, which are related but not synonymous. Peptide binding
refers to the ability of the peptide to initially interact with the HLA molecule and is a necessary step in antigen presentation as described above. Of equal importance is peptide stability, which is the capacity of the peptide to remain bound to the HLA molecule over time under physiological conditions. Some peptides which bind tightly to HLA molecules may not be stable enough in the binding groove to be effectively recognized by CTLs (127, 128).

Peptides with more or stronger secondary interactions are more stable in the peptide-binding groove since they establish multiple bonding connections with the HLA molecule. Crystal structures of peptide-HLA molecules show that the middle of the peptide is the least stable region, which can lead to multiple crystallized conformations or an inability to localize the middle peptide residues with certainty (115, 129, 130). Joseph and colleagues demonstrated that instability of middle peptide residues led to weaker binding of the C-terminus (131). Therefore, enhanced interactions of middle peptide residues with the HLA molecules should increase overall peptide binding and stability.

Several studies have indirectly and directly shown that middle peptide residues can interact with residues in the binding groove. Early studies on secondary peptide interactions found that certain amino acids in middle peptide positions are favorable and may line up with HLA pockets (42, 43, 132). Kondo et al. discovered that aromatic amino acids and proline were both favored in positions 4 and 5 of HLA-A*0101, causing the formation of different sub-motifs (133). Similarly, crystal structures of both murine H2-Kb and human HLA-A*1101 revealed that amino acids in positions 5, 6 or 7 pointed down and interacted with secondary pockets in the
binding groove (134, 135). Narzi et al. similarly found that arginine in peptide position 5 interacted with Asparagine-116 at the bottom of the binding groove of HLA-B*2705 (136). Thus, in some alleles, middle residues can increase stability in specific peptides by interacting with the beta-sheet residues forming the floor of the binding groove.

In the majority of HLA-I molecules, middle peptide residues do not point downwards toward the floor of the groove but are oriented upwards. This has supported the notion that middle peptide residues are primarily TCR contact residues. However, these middle residues may also have the potential to interact with the alpha helices forming the sides of the binding groove (Fig 1.5). Schlueter et al. discovered that altering position 5 in peptide QL9 affected both TCR recognition and peptide binding to H2-Ld, indicating the amino acid was pointing upwards towards the TCR yet still impacted binding (137). Substitutions in positions 5 to 7 of a peptide bound to HLA-A*0201 illustrated that some mutations caused increased stability when amino acid residues were pointing upwards (131). Mohammed et al. discovered that phospho-serine in peptide position 4 of HLA-A*0201 peptides acted as a solvent-exposed anchor through interactions with positively-charged residues found on the HLA-I alpha-1 helix (138). These studies, especially the phospho-peptide study, demonstrate that middle residues can interact with the alpha helices and impact peptide binding. However, the importance of non-pocket peptide-HLA interactions in natural antigen processing and presentation has not been well-studied. Strong peptide-HLA interactions such as the dominant anchor positions in the B and F pockets illustrate that these interactions cause a strong preference for
HLA-I/peptide interactions can occur inside or outside of binding pockets

Fig 1.5 Middle peptide residues can form non-pocket HLA-I/peptide interactions
**Fig 1.5 Middle peptide residues can form non-pocket HLA-I/peptide interactions:** Crystal structure of HIV peptide SLYNTVATL in HLA-A*0201 illustrates that amino acids at peptide ends bind down into pockets while middle amino acids assume an upright position and are in a viable position to interact with HLA residues in the alpha helices*.

*Courtesy of Dr. Dinler Antunes, University of Houston*
amino acids with specific biochemical characteristics at peptide positions p2 and pΩ, respectively (Fig. 1.2). However, no strong preferences for amino acids in middle peptide positions have been described for individual HLA-I molecules. Since non-pocket peptide-HLA interactions are generally weaker than pocket-based interactions, this might result in less conservation of specific amino acids at middle positions. In the next section, we analyzed datasets of eluted peptides from several HLA-I molecules to determine if there is any conservation of particular amino acids with specific biochemical characteristics in middle peptide positions, as would be expected if they were mediating interactions that may be important for peptide binding and/or stability.

1.9 Survey of eluted HLA-I peptides reveals a preference for negatively charged amino acids in peptide position 4

Due to advances in MS-based immunopeptidomics, there are now hundreds of thousands of peptide ligands eluted from different HLA allotypes publicly available. Multiple databases exist which compile these peptides for easier analysis. The largest is the Immune Epitope Database (IEDB), which is a compendium of T-cell and B-cell epitopes from published studies and includes peptides eluted by MS (139). IEDB currently contains 1,043,909 peptide epitopes from over 22,000 studies, though only about half of those epitopes are peptides eluted from HLA-I molecules (139). SysteMHC Atlas is another database of eluted peptides that has been utilized by several online programs to improve peptide-binding predictions such as MHCFlurry (88, 140). Also, Keskin et al. created nearly 100 monoallelic cell lines
engineered to express individual HLA-I allotypes and have eluted thousands of peptides from them (141).

We used these HLA-I eluted peptide databases to investigate amino acid frequencies of middle peptide residues. Collectively, these databases cover >90 of the most prevalent HLA-I allotypes, with thousands of peptides represented for each of them (88, 139, 141). We hypothesized that, similar to how specific $p_2$ and $p_\Omega$ amino acids are preferred in pockets B and F of different HLA molecules, any significant peptide-HLA interaction in middle peptide residues would also be revealed by preferences for specific amino acids. We grouped amino acids according to their biochemical properties into negative, positive, polar, hydrophobic or aromatic. The percentage of eluted peptides that contained amino acids from each category in each peptide position was determined and compared against known amino acid frequencies in vertebrate proteins (142). The most prevalent position was $p_\Omega$ where over 30% of eluted peptides preferred an aromatic amino acid, much higher than the background frequency of 8.6% of aromatic amino acids found in nature (Fig 1.6). This is consistent with several HLA-I binding motifs that showed preferences for one or more aromatic amino acids at $p_\Omega$, including HLA-A*2402 (45).

Middle peptide positions 4, 5, -3, -2, and -1 were also examined for amino acid preferences. There was a slight preference for hydrophobic amino acids in the majority of these positions, but this was also observed at the N- and C-terminal positions (Fig 1.6). No preference for polar, positive or aromatic amino acids was
Negatively charged amino acids are conserved in position 4

Figure 1.6 Overall survey of MS-eluted HLA-I peptides
Figure 1.6 Overall survey of MS-eluted HLA-I peptides: Three different eluted peptide databases were queried for the percentage of amino acids with specified biochemical characteristics in each peptide position. Positions with positive numerals are counted from the N-terminus while positions with negative numerals are counted from the C-terminus, with the C-terminal amino acid being 0. The black line in each graph represents the combined expected frequency of amino acids based on the frequencies of each amino acid from proteins in vertebrates.
found in middle peptide positions. However, HLA-I ligands showed a clear and striking preference for negatively-charged amino acids at position 4 (p4). HLA-I peptide ligands from all three databases contained glutamic acid (D) or aspartic acid (E) in approximately one-fourth of eluted peptides, significantly higher than the expected combined background D/E frequency of 11.7 percent in vertebrates (Fig 1.6). To our knowledge, this has not previously been reported in any similar fashion in the scientific literature.

We next examined all the amino acids found in peptide position 4 by determining their average frequency from the three HLA-I peptide databases then subtracting the known background frequency. We separated these analyses into HLA-A, B and C allotypes to determine whether any specific HLA-I class contained more D/E4 peptides than another. Both D4 and E4 showed elevated frequencies in all three HLA-I classes, with the following D/E frequency hierarchy: HLA-A > HLA-B > HLA-C (Figure 1.7). This analysis revealed little or no position 4 preference for any aromatic, polar, positive or hydrophobic amino acid except for proline, which was found to have a prevalence about 8-12 percent higher than background in all HLA-I classes. Proline did not stand out in our initial survey since it was combined with the other hydrophobic amino acid residues.
Amino acids D, E and P are preferred in peptide position 4

Figure 1.7 Aspartic acid (D) and glutamic acid (E) are highly preferred in HLA-A, B and C molecules
Figure 1.7 Aspartic acid (D) and glutamic acid (E) are highly preferred in HLA-A, B and C molecules: Preferred amino acids in position 4 were found by subtracting the percent of eluted peptides containing each amino acid with the known background frequency of the same amino acid. A value of zero indicates eluted and background frequencies were the same, a negative value indicates higher background frequencies and a positive value indicates higher eluted frequencies. Single letter amino acid codes are used with amino acids colored by biochemical characteristic (red- negatively charged; blue- positively charged; orange- polar; green- hydrophobic; purple- aromatic).
We next compared the prevalence of D/E4 peptides in different HLA-I alleles. The percent of peptides containing D/E4 was determined in several HLA-A, B and C molecules from all three datasets as shown in Figure 1.8. With the exception of only 2 HLA-B allotypes, peptides eluted from every HLA-A, B or C molecule examined contained D/E4 at higher than the background frequency (Fig 1.8). This was somewhat unexpected since HLA allotypes are known for their diverse and distinct peptide-binding motifs. Although peptides eluted from HLA-B molecules had the lowest prevalence of D/E4, multiple HLA-B alleles demonstrated > 25% D/E4 content. The majority of HLA-C alleles bound 24-30% D/E4 peptides. HLA-A molecules exhibited the largest range of D/E4 peptide frequencies, ranging from 20% to 40%, depending on the individual HLA-A allele.

1.10 Interactions may occur between D/E residues in peptide position 4 and R65/K66 residues of A*0201 and A*2402

Peptide ligands eluted from the HLA-A2 superfamily of allotypes demonstrated the highest frequencies of D/E4 peptides (40% or higher). We hypothesized that such a strong preference for D/E4 peptides might be an indication of strong interaction with other residues in the HLA peptide-binding groove. As discussed above, position 4 residues are oriented at the top of the binding groove and are therefore unlikely to interact with conventional binding pockets near the bottom of the groove. Therefore, we examined crystal structures of a number of HLA-A*0201/peptide complexes to look for HLA-I alpha helix residues that may interact with negatively-charged D/E4 residues of bound peptides. We found that HLA-
D/E4 are preferred in multiple HLA-A, B and C molecules

Figure 1.8 Most HLA-A, B and C alleles prefer D or E residues in peptide position 4
Figure 1.8 Most HLA-A, B and C alleles prefer D or E residues in peptide position 4: The percent of D4 and E4 peptides in HLA-A, B and C molecules from all three eluted peptide databases. Red lines in each graph indicate the background frequency of D and E combined.
A*0201 contains two adjacent positively-charged residues on the α1 helix, arginine (R) in position 65 (R65) and lysine (K) in position 66 (K66), that are found in close proximity to peptide position 4 residues (Fig 1.9A). A crystal structure of A*0201 bound to peptide GLKEGIPAL showed that E4 is located on top of K66 and very close to R65 (Fig 1.9B) (143). Therefore, we investigated whether negatively-charged amino acids in peptide position 4 may interact directly with R65 or K66 in A*0201. In collaboration with Dr. Dinler Antunes at the University of Houston, molecular dynamics simulations were performed for A*0201/GLKEGIPAL. The location of E4 allows it to move around freely near the top of the binding groove, so we first calculated the average distance between E4, R65, K66 and other potential interacting HLA residues. We observed that the closest HLA-I residues throughout the simulation were R65 and K66 (Fig 1.9C). Calculation of bonds revealed charged interactions with both R65 and K66, and hydrogen bonds with R65 resulting in the formation of a salt bridge between E4 and R65 (Fig 1.9D). No salt bridges were observed between E4 and K66 in this simulation. Thus, the E4 residue from the GLKEGIPAL peptide appeared to be capable of interacting with both R65 and K66 of HLA-A*0201.

Like A*0201, HLA-A*2402 is another highly prevalent HLA-I allele within the human population that also demonstrates a preference for D/E4 peptide ligands (~24% frequency, see Fig 1.8). Interestingly, A*2402 also has a lysine residue at p66 but contains a glycine instead of an arginine at position 65 (G65) (Fig 1.9E). We next examined if the D/E4 residues of bound peptides could potentially interact with K66 in A*2402. Analysis of the crystal structure of A*2402 bound to QFKDNVILL
revealed that D4 was found in close proximity to K66 (Fig 1.9F) (144).

Computational simulations on the crystal structure from A*2402/QFKDNVILL were also performed, which revealed that K66 was the closest HLA-A*2402 residue to D4 on either alpha helix throughout the simulation (Fig 1.9G). Calculation of different interactions revealed very strong charged interactions and hydrogen bonds between D4 and K66, leading to the formation of salt bridges between these residues (Fig 1.9H). Thus, negatively charged residues in peptide position 4 can potentially make meaningful interactions with K66 of A*2402.
R65 and K66 may interact with D/E4 in A*0201 and A*2402

Figure 1.9 R65 and K66 may interact with negatively charged p4 amino acids in HLA-A*0201 and HLA-A*2402
Figure 1.9 R65 and K66 may interact with negatively charged p4 amino acids in HLA-A*0201 and HLA-A*2402: (A) Amino acid sequence of the α1 helix from position 50-73 of A*0201. Consecutive positively charged amino acids R65 and K66 are highlighted in blue. (B) Crystal structure of HLA-A*0201 bound to GLKEGIPAL focusing on peptide position 4 and HLA positions 65-66. (C) Graph showing distances in angstroms throughout the simulation between E4 and HLA residues on both alpha helices*. The blue line indicates distance cutoffs for hydrogen bonds, the green line indicates distance cutoffs for charged interactions and the purple line indicates distance cutoffs for salt bridges. (D) Graphs showing charged or coulombic interactions between E4 and R65/K66 and hydrogen bonds and salt bridges between peptide and HLA residues. Negative values in coulombic graphs indicate attractive interactions while positive values indicate repulsive interactions. (E) Amino acid sequence of the α1 helix from position 50-73 of A*2402. Positively charged amino acids in p65-66 are highlighted in blue and E62 is highlighted in red as it is different than p62 in A*0201. (F) Crystal structure of HLA-A*2402 bound to QFKDNVILL focusing on peptide position 4 and HLA p65-66. (G) Graph showing distances in angstroms throughout the simulation between D4 and HLA residues on both alpha helices*. Line cutoffs are similar to C. (H) Coulombic interactions, hydrogen bonds and salt bridges in QFKDNVILL and HLA-A*2402.

*Computational simulations were run in collaboration with Dr. Dinler Antunes, University of Houston
1.11 Rationale, Hypothesis, Key Questions and Approach

Rationale

Most HLA-I molecules demonstrate a preference for negatively charged amino acids (D and E) at peptide position 4 of bound peptide ligands. The orientation of p4 amino acids does not indicate interactions with conventional pockets deep within the peptide-binding groove. Computational simulations in HLA-A*0201 and HLA-A*2402 showed charged interactions and salt bridges occurring between D/E4 residues in the peptide and R65/K66 residues located near the top of the binding groove. However, the contribution of these residues to peptide binding and stability and overall impact on the composition of the immunopeptidome is unknown.

Hypothesis:

R65/K66 in HLA-I and negatively charged amino acids glutamic acid and aspartic acid in peptide position 4 (D/E4) enhance peptide binding and stability and shape the composition of the immunopeptidome.

Key Questions:

- Is position 4 of HLA-binding peptides acting as a non-pocket anchor position?
- How much does the loss of positively charged residues in HLA-A p65/66 alter the immunopeptidome?

Approach:

Model and experimentally test potential interactions between D/E4 residues of peptide ligands and R65/K66 of HLA-A*0201 and HLA-A*2402 (two HLA-I alleles with the highest worldwide prevalence) (Table 1.2).
### Phenotype and allele frequencies for A*0201

<table>
<thead>
<tr>
<th>Country/Region</th>
<th>Ethnicity (USA)</th>
<th>Phenotype frequency</th>
<th>Allele frequency</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russia/Karelia</td>
<td></td>
<td>0.505</td>
<td>0.2963</td>
<td>1,075</td>
</tr>
<tr>
<td>USA</td>
<td>Caucasian</td>
<td>0.476</td>
<td>0.2755</td>
<td>1,242,890</td>
</tr>
<tr>
<td>Ireland/Northern</td>
<td></td>
<td>0.473</td>
<td>0.274</td>
<td>1,000</td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td>0.462</td>
<td>0.2667</td>
<td>39,689</td>
</tr>
<tr>
<td>Spain</td>
<td></td>
<td>0.418</td>
<td>0.2368</td>
<td>4,335</td>
</tr>
<tr>
<td>USA</td>
<td>Hispanic</td>
<td>0.375</td>
<td>0.2095</td>
<td>146,714</td>
</tr>
<tr>
<td>China/Hong Kong</td>
<td></td>
<td>0.343</td>
<td>0.1891</td>
<td>7,595</td>
</tr>
<tr>
<td>Columbia</td>
<td></td>
<td>0.296</td>
<td>0.1613</td>
<td>1,463</td>
</tr>
<tr>
<td>South Korea</td>
<td></td>
<td>0.291</td>
<td>0.158</td>
<td>4,128</td>
</tr>
<tr>
<td>China/Jiangsu Han</td>
<td></td>
<td>0.243</td>
<td>0.13</td>
<td>3,238</td>
</tr>
<tr>
<td>USA</td>
<td>African-American</td>
<td>0.231</td>
<td>0.1235</td>
<td>416,581</td>
</tr>
<tr>
<td>Japan</td>
<td></td>
<td>0.218</td>
<td>0.1162</td>
<td>18,604</td>
</tr>
</tbody>
</table>

### Phenotype and allele frequencies for A*2402

<table>
<thead>
<tr>
<th>Country/Region</th>
<th>Ethnicity (USA)</th>
<th>Phenotype frequency</th>
<th>Allele percent</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan</td>
<td></td>
<td>0.597</td>
<td>0.3648</td>
<td>18,604</td>
</tr>
<tr>
<td>USA</td>
<td>Filipino</td>
<td>0.423</td>
<td>0.2399</td>
<td>50,614</td>
</tr>
<tr>
<td>USA</td>
<td>Pacific Islander</td>
<td>0.421</td>
<td>0.2393</td>
<td>11,499</td>
</tr>
<tr>
<td>South Korea</td>
<td></td>
<td>0.393</td>
<td>0.2209</td>
<td>4,128</td>
</tr>
<tr>
<td>Malaysia</td>
<td></td>
<td>0.357</td>
<td>0.1977</td>
<td>951</td>
</tr>
<tr>
<td>India/Tamil Nadu</td>
<td></td>
<td>0.300</td>
<td>0.1629</td>
<td>2,492</td>
</tr>
<tr>
<td>China/Hubei Han</td>
<td></td>
<td>0.292</td>
<td>0.159</td>
<td>3,732</td>
</tr>
<tr>
<td>China/Hong Kong</td>
<td></td>
<td>0.273</td>
<td>0.1472</td>
<td>7,595</td>
</tr>
<tr>
<td>USA</td>
<td>Hispanic</td>
<td>0.246</td>
<td>0.1316</td>
<td>146,714</td>
</tr>
<tr>
<td>USA</td>
<td>Middle Eastern</td>
<td>0.221</td>
<td>0.1174</td>
<td>70,890</td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td>0.181</td>
<td>0.0954</td>
<td>39,689</td>
</tr>
<tr>
<td>USA</td>
<td>Caucasian</td>
<td>0.163</td>
<td>0.0846</td>
<td>1,242,890</td>
</tr>
</tbody>
</table>

Table 1.2 Worldwide phenotypic and allele frequencies of A*0201 and A*2402
Table 1.2 Worldwide phenotypic and allele frequencies of A*0201 and A*2402:

Phenotypic and allele frequencies of A*0201 and A*2402 in different countries and ethnicities. A*0201 is highly prevalent in multiple ethnicities and countries while A*2402 is more prevalent in Asian and Pacific Islander populations.
Aim 1: Analyze interactions between D/E4 and R65/K66 of HLA-A*0201

- Perform site-directed mutagenesis of R65 and/or K66 of HLA-A*0201
- Perform in-depth examination of the immunopeptidome
- Analyze impact of changes in peptide p4 to A*0201 binding and stability
- Collaborate with Dr. Dinler Antunes of University of Houston to perform and analyze computational simulations of WT and mutated A*0201 molecules

Aim 2: Analyze interactions between D/E4 and G65/K66 of A*2402

- Perform site-directed mutagenesis of G65 and/or K66 of A*2402
- Perform in-depth examination of the immunopeptidome
- Analyze impact of changes in peptide p4 to A*2402 binding and stability
- Collaborate with Dr. Dinler Antunes of University of Houston to perform and analyze computational simulations of WT and mutated A*2402 molecules

Relevance:

- Perform direct comparisons between WT and mutant A*0201 and A*2402 and examine if similar non-pocket interaction may occur in other HLA-I molecules
- HLA-I/peptide complexes are highly relevant immune targets for several different human diseases; therefore, understanding the molecular interactions that drive peptide binding to HLA-I molecules are of primary importance.
CHAPTER II: NON-POCKET INTERACTIONS BETWEEN HLA-A*0201 AND BOUND PEPTIDES ARE REQUIRED FOR OPTIMAL PEPTIDE BINDING STABILITY AND COMPLETE PEPTIDE REPERTOIRE DIVERSITY
2.1 R65 and K66 interactions in WT HLA*0201 molecules

Based on the analyses outlined in Chapter 1, our first priority was to examine the role of arginine-65 (R65) and lysine-66 (K66) in A*0201 in more detail. E4 in GLKEGIPAL (GLKE) prefers charged interactions and salt bridges with R65 instead of K66, which could be due to the longer length of the glutamic acid (E) side chain compared to that of aspartic acid (D), as well as the longer length of R compared to K. It was also unclear from simulations with A*0201/GLKE whether K66 interacted with position 4, since this residue has been noted previously for its interaction with peptide position 2 residues (143, 145, 146). We also wanted to test whether D4 behaved similarly to E4 with respect to A*0201 binding.

To answer these questions, we analyzed two additional crystal structures of A*0201 bound to peptides which contained D4 or E4 to validate findings in GLKE (147, 148). Simulations were performed on structures containing VLHDDLLEA (VLHD) or FLKEPGHGV (FLKE) peptides complexed to A*0201, which allowed examination of D4 and E4 in more detail. All computational simulations were run by our collaborator Dr. Dinler Antunes at the University of Houston.

First, we examined the charged interactions of D/E4 peptide residues with R65 and K66 as well as formation of salt bridges. Interestingly, D4 in VLHD formed strong attractive interactions with K66 and much weaker interactions with R65 (Fig 2.1A). Examination of hydrogen bonds showed that salt bridges were forming with D4 between both K66 and R65, but that the salt bridges with K66 were much more prevalent, consistent with observed coulombic interactions (Fig 2.1A).
Peptide residues D4 and E4 can interact with R65 and K66 of A*0201.

Figure 2.1 In HLA-A*0201, peptide residues D4 and E4 both interact with R65 and K66 in multiple peptide backbones.
Figure 2.1 In HLA-A*0201, peptide residues D4 and E4 both interact with R65 and K66 in multiple peptide backbones: (A) Coulombic interactions and alluvial plots showing interactions between peptide and HLA in simulations of A*0201 and VLHDDLEA.* (B) Coulombic interactions and alluvial plots from simulations showing interactions between A*0201 and FLKEPGHGV.*

*Simulations were run in collaboration with Dr. Dinler Antunes, University of Houston
In A*0201/FLKE complexes, E4 made coulombic interactions with both R65 and K66, but the strongest interactions occurred with K66 (Fig 2.1B). This was in contrast to GLKE, in which the E4 residue interacted primarily with R65. For FLKE, analysis of hydrogen bond interactions revealed highly prevalent salt bridges between E4 and K66, but no salt bridges occurring with R65 (Fig 2.1B). These simulations illustrate the plasticity and complexity of the peptide-binding groove, in which same amino acid position can form interactions with different HLA backbone residues. In the case of A*0201 binding peptides, it appears that D/E4 residues may interact preferentially with either R65 or K66, depending on the individual peptide.

The most common structures from the A*0201/GLKE simulation were overlaid on top of each other which allowed for three-dimensional examination of the simulation in more detail. Adjacent HLA-I residues were included to provide a sense of the ‘neighborhood’ of interactions around R65 and K66. These structures showed that R65 appears to be the “odd man out”, as it sits on top of the alpha helix of the binding groove and interacts only with peptide position 4 (Fig 2.2). By contrast, K66 is found on the inner side of the groove and is surrounded by a network of potential interactions. The most prominent peptide interaction of K66 in these structures are hydrogen bonds with the backbone of leucine in position 2. A smaller number of structures reveal interactions between K66 and E4 when this p4 residue is oriented downwards. K66 also interacts with several other A*0201 residues, forming prevalent salt bridges with E63 as well as hydrogen bonds with G62 and H70 (Fig 2.2). Thus, K66, much more so than R65, seems to be play a central role in the network of interactions occurring on this portion of the α1 helix.
K66 can interact with several surrounding HLA and peptide residues

A*0201 / GLKEGIPAL

Figure 2.2 R65 and K66 interactions with GLEKGIPAL in the peptide binding groove
Figure 2.2 R65 and K66 interactions with GLKEGIPAL in the peptide binding groove: Prevalent structures pulled from simulations of A*0201 and GLKEGIPAL and overlaid to illustrate the movement of residues in the peptide binding groove*. Yellow dashes indicate hydrogen bonds while red dashes indicate salt bridges.

*Simulations were run in collaboration with Dr. Dinler Antunes, University of Houston
2.2 Creation of mutated alleles of HLA-A*0201

To more fully examine the role of R65 and K66 in antigen presentation, we constructed three A*0201 mutant alleles. The majority of mutations made in proteins change the mutated position to alanine, which is a small, hydrophobic amino acid that does not form many interactions (149). However, we were concerned that alanine mutations would cause unwanted side effects in peptide and HLA interactions or potentially disrupt the alpha helix. Therefore, we created A*0201 mutant alleles that used substituted residues naturally found in other HLA-A molecules. Since similar experiments and mutations were also planned for A*2402, reciprocal mutations were made to make A*0201 more A*2402-like, and vice-versa (see Chapter 3).

In order to study the charged interactions suggested by the computational modeling, three mutant alleles were constructed to reduce the number of positively-charged amino acids in HLA-A*0201 in p65-66. Two of the mutants retained one positively charged amino acid at either p65 or p66, while the third mutant contained no positively charged amino acids in either position. Since A*2402 conserves K66 but substitutes glycine (G) for arginine at position 65, the first A*0201 mutant allele changed R65 to G65 (hereafter referred to as A2-R65G) (Table 2.1). In addition to A*2402, this combination of G65/K66 is also present in the naturally occurring allele HLA-A*02:70.

We examined other HLA-A alleles to find another amino acid to substitute for K66, since both A*0201 and A*2402 contained this K66 residue. We discovered that every other prevalent HLA-A allele contained asparagine (N) in position 66 (Table
## HLA-A*0201 mutants constructed

<table>
<thead>
<tr>
<th>A*0201 construct</th>
<th>Amino acid sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-A cons</td>
<td>--QETRNVKAH--</td>
</tr>
<tr>
<td>A2-WT</td>
<td>--GETRKVKAH--</td>
</tr>
<tr>
<td>A2-K66N</td>
<td>--GETRNVKAH--</td>
</tr>
<tr>
<td>A2-R65</td>
<td>--GETGKVKAH--</td>
</tr>
<tr>
<td>A2-DM</td>
<td>--GETGNVKAH--</td>
</tr>
</tbody>
</table>

Table 2.1 HLA-A*0201 mutants and their altered p65-66 residues
Table 2.1 HLA-A*0201 mutants and their altered p65-66 residues: Amino acid sequences of positions 62-70 of the HLA-A consensus sequence, A2-WT and all three A2 mutant molecules. Positively charged amino acids in p65-66 are in blue and other p65-66 residues are in green.
2.1) A polar amino acid, asparagine is smaller than K but able to form hydrogen bonds in multiple positions. This was an important consideration, since K66 is involved in multiple hydrogen bond interactions, including hydrogen bonds with the backbone of peptide position 2. Therefore, we mutated K66 in A*0201 to N66 (hereafter called A2-K66N), which is also identical to the naturally occurring allele HLA-A*02:20 (R65/N66) (Table 2.1).

As both R65 and K66 are positively charged and both potentially interact with peptide position 4, we hypothesized that they may compensate for each other. Thus, the third A*0201 mutant allele removed both positively charged amino acids in positions 65 and 66 by combining the previous two mutations, creating the G65/N66 double mutant (called A2-DM) (Table 2.1). While this pairing of G65/N66 has not been reported to occur in the HLA-A2 family of allotypes, it does occur naturally in the A24 family allotype HLA-A*24:145.

2.3 HLA*0201 mutants demonstrate loss of cell surface expression and peptide repertoire

All three A*0201 (A2) mutants were created by site-mutagenesis, validated by Sanger sequencing and then subcloned into a lentiviral expression vector to produce lentivirus. A*0201 (A2-WT) and all three A2 mutants were transduced into H1975, a lung cancer cell line that expresses HLA-A*0101, HLA-B*4101 and HLA-C*1701 (150). Peptide motifs from endogenous HLA-I molecules are different in both p2 and pΩ anchors compared to HLA-A*0201, allowing for separation of peptides from endogenous and transduced HLA molecules (Fig 2.3).
A*0201 and A*2402 peptide-binding motifs are distinct from those of endogenous HLA-I in H1975 cells

Fig 2.3 A*0201 and A*2402 peptide-binding motifs are distinct from those of endogenous HLA-I in H1975 cells
Fig 2.3 A*0201 and A*2402 peptide-binding motifs are distinct from those of endogenous HLA-I in H1975 cells: H1975 is homozygous for A*0101, B*4101 and C*1701. Peptide binding motifs of all three HLA-I molecules are distinct from the peptide binding motifs of A*0201 and A*2402 as A*0101 peptides prefer T/S2, D/E3 and YΩ, B*4101 prefers peptides with E2 and C*1701 prefers A2-LΩ peptides.
After transduction of H1975 with A2-WT and all three A2 mutants, A*0201 cell surface expression was examined by flow cytometry, which revealed differing levels of expression depending on the number of positively charged amino acids in p65/66 (Fig 2.4A). A2-WT, which contains two positively charged amino acids, had the highest expression level while A2-DM, which contains no positively charged amino acids, had the lowest expression. A2-K66N and A2-R65G, each containing one positively charged amino acid, showed intermediate expression between that of A2-WT and A2-DM. A2-K66N also exhibited a higher level of surface expression compared with A2-R65G (Fig 2.4A).

To ensure that different cell surface expression levels was not due to differing dosages of A*0201 transduction, all A2 constructs were created with green fluorescent protein (GFP) attached to the C-terminus of A*0201. These GFP-tagged A2-WT and A2 mutant molecules were transduced into H1975 and overall cellular expression of A*0201 was determined by analyzing GFP expression by flow cytometry (Fig 2.4B). This experiment revealed comparable overall cellular A*0201 expression levels for both A2-WT and all three A2 mutants, while cell surface A*0201 expression showed the same hierarchy: A2-WT > A2-K66N > A2-R65G > A2-DM.

As discussed previously, the alpha-chain of HLA molecules is unstable as a monomer, and will not be transported to the cell surface unless bound to both beta-2-microglobulin and a peptide in the binding groove. Since both mutations were made to the peptide-binding groove, we hypothesized that cell surface expression decreased due to loss of peptide binding and complex stabilization. To analyze this
Loss of A*0201 cell surface expression and peptide repertoire

Fig 2.4 Loss of cell surface expression and peptide repertoire in mutated HLA-A*0201 molecules
Fig 2.4 Loss of cell surface expression and peptide repertoire in mutated HLA-A*0201 molecules: (A) Cell surface expression of A*0201 in H1975 cells transduced with A2-WT or A2 mutant alleles. (B) Overall cellular expression of A*0201 in H1975 transduced with A2-WT or A2 mutant alleles with GFP attached to the end of the intracellular domain. (C) Number of A*0201 peptides eluted from A2-WT and all three A2 mutated molecules in H1975 cells. (D) Number of peptides from endogenously expressed HLA molecules in H1975 in A2-WT and all three A2 mutants.
possibility further, HLA-I bound peptides were eluted from H1975 cells transduced to express A2-WT, A2-K66N, A2-R65G, or A2-DM proteins, and tandem MS/MS was performed on these samples. Using the pan-HLA-I mAb, we identified between 1,500-2,500 peptide sequences from each cell line, though some of these peptides were derived from endogenously-expressed HLA-A*0101, -B*4101, and -C*1701 molecules. We were able to separate peptides that were eluted from the A*0201 variants based on differences in binding motifs for these endogenous HLA-I alleles as discussed previously (see Fig 2.3).

First, we determined the total number of unique A*0201 peptides that were eluted from A2-WT and the A2 mutants. Comparison of A2-WT to A2 mutants showed that loss of positively charged residues caused loss of diversity within the A*0201 peptide repertoire (Fig 2.4C). 1945 unique A*0201 peptide sequences were eluted from A2-WT, compared with 555 unique peptides identified from A2-DM. A2-R65G and A2-K66N both contained more unique peptides than A2-DM, with A2-K66N presenting more unique peptides (1319) than A2-R65G (864). Analysis of the peptides eluted from endogenous HLA-I molecules revealed 850 or more unique peptide sequences from all cell lines, indicating functional antigen presentation machinery within both A2-WT and A2 mutant-transduced H1975 cells (Fig 2.4D).

2.4 Immunepeptidome analysis reveals alterations in peptide position 4 frequencies of mutated HLA-A*0201

Since we had previously computationally characterized the charged interactions between A*0201 WT R65/K66 and D/E in peptide position 4, we suspected that the
loss of these interactions was related to the observed loss of peptide repertoire diversity and cell surface expression. To determine this, the percentage of eluted D/E4 peptides was compared between A2-WT and the A2 mutants. We observed a significant decrease in the frequency of D4 (p <0.001) to 13% in all three A2 mutants compared with 20% D4 peptides eluted from A2-WT (Fig 2.5A). A significant decrease in E4 peptide frequencies was also observed for all A2 mutants (p<0.001): while A2-R65G and A2-DM showed 15% E4 peptides, A2-K66N showed E4 frequency of 21%; this compares to an E4 frequency of 26% in A2-WT.(Fig 2.5A). This data provides evidence in support of the notion that R65 and K66 can interact with D/E4 residues to stabilize peptide binding.

Decreases in D/E4 peptide frequencies in the A2 mutants meant that increased frequencies occurred for other amino acids in peptide position 4. The difference in percent for each p4 amino acid in all three A2-mutants compared to A2-WT were calculated and compared (Fig 2.5B). Several hydrophobic and polar amino acids showed slight increases in the A2 mutants compared to A2-WT, though only proline, alanine and serine demonstrated increases in all three mutants. The increase in proline was interesting as it was the only other prevalent amino acid in peptide position 4 of HLA-A alleles (see Fig 1.7). Proline in position 4 showed increased frequencies in all three A2 mutants, though only the increase in A2-DM was statistically significant (p<0.01) (Fig 2.5C). This suggested that proline binds independently of R65 and K66, and that additional stability is required in position 4 when positively charged amino acids are lost at HLA p65/66.
Loss of D/E4 in peptide position 4 A*0201

Figure 2.5 Peptide position 4 amino acid frequencies are altered in the immunopeptidomes of HLA-A*0201 mutants
Figure 2.5 Peptide position 4 amino acid frequencies are altered in the immunopeptidomes of HLA-A*0201 mutants: (A) Percent of eluted A*0201 peptides with D or E in peptide position 4 in A2-WT and all three A2 mutant molecules. *** indicates p<0.001 using 2-sample test for equality of proportions without continuity correction. (B) Heat map of differences in percent of amino acids in position 4 of all three A2 mutants compared to A2-WT. Negative values indicate higher percent of amino acids in A2-WT while positive values indicate higher percent of amino acids in A2-mutants. (C) Percent of eluted peptides in A2-WT and A2 mutant molecules with proline in peptide position 4. ** indicates p< 0.01 using 2-sample test for equality of proportions without continuity correction.
2.5 Loss of R65 and/or K66 in HLA-A*0201 alters HLA-I/peptide interactions

To further characterize the charged interactions between peptide position 4 and A*0201 p65-66, computational simulations were run of the GLKEGIPAL peptide bound to A2 mutant molecules and compared to that of the A2-WT simulation (Fig 2.6). In modeled A2-R65G/GLKE complexes, we observed loss of the charged interactions and salt bridges between E4 and p65, and stronger coulombic interactions with K66 leading to a higher prevalence of salt bridges with K66 than is observed in A2-WT (Fig 2.6A,B). However, the E4-K66 interactions in A2-R65G are not as strong as those seen in A2-WT, indicating E4 prefers to interact with R65. A2-K66N/GLKE showed charged interactions between E4 and R65, although they were not as strong as E4-R65 interactions in A2-WT (Fig 2.6C). Furthermore, the salt bridges between E4 and R65 in A2-K66N are less prevalent than those in A2-WT, indicating a loss of interaction even though both amino acids are still present (Fig 2.6C). A2-DM forms no charged interactions or salt bridges with G65, similar to that observed for A2-R65G (Fig 2.6D).

Coulombic plots for A2-K66N and A2-DM identified both negative and positive interactions between E4 and N66 in the GLKE peptide complexes (Fig 2.6C,D). Further investigation revealed that the amide group in asparagine confers partial negative and positive charges on the side chain and thus, depending on its conformational arrangement, can either attract or repel E4. N66 also forms hydrogen bonds with E4 similar in stability to hydrogen bonds involved in salt bridges in A2-WT which led us to explore whether N66 was also capable of forming salt bridges with E4 (Fig 2.6C,D). Re-examining the computational simulations showed that salt
E4 can interact with R65, K66 and N66 in A*0201 mutants
E4 can interact with R65, K66 and N66 in A*0201 mutants

**GLKEGIPAL / HLA-A*0201**

**C**

- **K66N**
  - E4 ↔ R65
  - E4 ↔ N66

**D**

- **DM**
  - E4 ↔ G65
  - E4 ↔ N66

**Fig 2.6** Charged interactions and salt bridges occur between E4 and p65-66 in A*0201 mutants
Fig 2.6 Charged interactions and salt bridges occur between E4 and p65-66 in A*0201 mutants: Coulombic interactions and alluvial plots for computational simulations with GLKEGIPAL bound to A2-WT and all three A2 mutants.* In alluvial plots, red lines indicate salt bridges and blue lines indicate hydrogen bonds.  
(A) A2-WT  (B) A2-R65G (C) A2-K66N (D) A2-DM

*Simulations were run in collaboration with Dr. Dinler Antunes, University of Houston
bridges can occur between N66 and E4 in A2-K66N and A2-DM which were more stable than the E4-K66 salt bridges in A2-WT but less stable than the E4-R65 salt bridges in A2-WT (Fig 2.6 A,C,D). Therefore, the HLA-A consensus residue N66 appears to be capable of forming hydrogen bonds and charged interactions with peptide E4 residues.

Examination of prevalent structures from molecular dynamics simulations of A2 mutant molecules revealed that N66 replicated some of the network of interactions that K66 was involved in (Fig 2.7). Loss of R65 caused E4 to move downward to form more prevalent salt bridges with K66 than in A2-WT (Fig 2.7A). In A2-K66N, E4 and R65 form salt bridges over the top of N66 while N66 does not interact much in these structures, though we do see hydrogen bonds, salt bridges and charged interactions between E4 and N66 in Figure 2.6 (Fig 2.7B). N66 in A2-DM does show prevalent hydrogen bonds with the backbone of peptide position 2, similar to K66. However, the vast majority of structures did not show N66 interacting with E63, although hydrogen bonds still occurred with G62 and H70. While analysis of the entire simulation showed some interactions between N66 and E4, these were not observed in a survey of the most prevalent conformations. Instead, the backbone of E4 interacted with the side chain of K3, which is the closest charged residue (Fig 2.7).
P65-66 substitutions alter HLA-I/peptide interactions in A*0201/GLKEGIPAL complexes
P65-66 substitutions alter HLA-I/peptide interactions in A*0201/GLKEGIPAL complexes

Fig 2.7 P65-66 substitutions in HLA-A*0201 alter HLA-I/peptide interactions in the binding groove
Fig 2.7 P65-66 substitutions in HLA-A*0201 alter HLA-I/peptide interactions in the binding groove: Interactions between HLA and peptide residues surrounding p65-66 in (A) A2-R65G (B) A2-K66N and (C) A2-DM.* Yellow lines indicate hydrogen bonds while red bonds indicate salt bridges.

*Simulations were run in collaboration with Dr. Dinler Antunes, University of Houston
2.6 D/E4 peptide residues contribute to peptide binding and stability in A*0201

Previous analyses focused on examining the impact of alterations to A*0201 on antigen presentation, the immunopeptidome and molecular interactions with bound peptides. Next, we altered amino acids in position 4 of A*0201 peptides to study differences in peptide binding and stability. *In vitro* peptide binding and stability assays were performed on peptide sequences derived from crystal structures and sequences from peptide elutions. In these assays, lower nanomolar (nM) affinity values indicate tighter peptide binding while higher percent stability compared to a reference peptide indicates better stability. All four peptides analyzed below contained D, E or a phospho-serine (pS) at position 4. Negative charges were added by changing position 4 to phospho-serine (pS), or removed by changing D/E4 residues to hydrophobic amino acids.

In the VLHDDLLEA peptide, position 4 was altered to pS, E or P and binding and stability assays were performed (*Fig 2.8A*). Both VLHD and VLHpS exhibited very tight binding (about 2nM) but VLHpS was more stable at 84% compared to 68% in VLHD. VLHE showed increased stability (80%) compared to VLHD but decreased binding (20nM). Changing position 4 to proline in VLHP led to decreases in both binding (45 nM) and stability (40%) (*Fig 2.8A*).

Similar peptide binding and stability assays were performed on three other peptide sequences. RQApSLISV exhibited tight binding (2nM) and high stability (78%) but increasing loss of negative charges at position 4 led to loss of both binding and stability (*Fig 2.8B*). The VQIEEVRQV peptide (eluted from A2-WT transduced H1975 cells) showed loss of both binding and stability when position 4
D/E4 residues contribute to peptide binding and stability

Fig 2.8 D/E4 residues contribute to peptide binding and stability in HLA-A*0201
Fig 2.8 D/E4 residues contribute to peptide binding and stability in HLA-A*0201: Comparison of binding and stability values in multiple peptide backbones with substitutions in peptide position 4. (A) VLHDDLLEA (B) RQApSLSISV (C) VQIEEVRQV (D) LTKELLHSV.
was altered to glycine (Fig 2.8C). LTKELLHSV is another peptide eluted from A2-WT transduced H1975 that contains threonine, a less prevalent position 2 anchor, and demonstrated weak stability (24%) and intermediate binding (585nM). Changing position 4 to pS caused a large increase in both binding (61nM) and stability (57%), whereas altering position 4 to alanine caused further loss of binding (3942nM) and stability (21%). (Fig 2.8D). In these peptides, there was more loss of stability than binding with fewer negative charges in position 4, indicating that D/E4 may be more important for peptide-HLA complex stability than for initial peptide binding.

2.7 Mutation of R65 and/or K66 alters HLA-A*0201 peptide-binding motifs

While loss of negatively charged amino acids in peptide position 4 was significant and affected peptide stability, it was insufficient to explain the overall loss of cell surface expression and peptide repertoire observed in mutated A*0201 molecules. Therefore, other peptide positions were examined and differences in the percent of amino acids in each position compared to A2-WT were calculated (Fig 2.9).

We hypothesized that A2-R65G and A2-K66N would demonstrate similar trends of losses or gains in amino acid frequencies, while A2-DM would also show similar trends as the single A2 mutants, but more exaggerated. This occurred in peptide position 4, as shown in the heat map and described above. However, at other peptide positions, A2-K66N and A2-R65G trends diverged from each other. For example, in peptide position 3 A2-R65G showed loss of leucine, phenylalanine and tryptophan while A2-K66N demonstrated increased frequencies of these same
Amino acid alterations in peptides eluted from A*0201 mutants

Fig 2.9 Mutations to positions 65/66 alter HLA-A*0201 peptide-binding motifs
Fig 2.9 Mutations to positions 65/66 alter HLA-A*0201 peptide-binding motifs:

Heat map of amino acid percent differences between A2-WT and all three A2 mutant molecules in all peptide positions. Negative values indicate higher percentage of amino acids in A2-WT while positive values indicate higher percentage of amino acids in A2-mutants.
amino acids (Fig 2.9). Similar opposing trends between A2-K66N and A2-R65G occurred in position 1, position 2 and the C-terminus as well. Interestingly, A2-DM followed similar trends to A2-K66N at peptide positions 1, 2, and 3, but at the C-terminus A2-DM was more similar to A2-R65G. These differences suggested that the K66N mutation impacts N-terminal positions whereas R65G alters the C-terminus; furthermore, these alterations were distinct from each other, and A2-DM demonstrated both sets of changes.

Overall analysis of the A*0201 mutants revealed that the most dramatic and unexpected differences occurred at the peptide termini. For example, there was a dramatic change in amino acid frequencies in position 1 of peptides eluted from A2-K66N and A2-DM compared to A2-WT (Fig 2.9). Furthermore, we also observed significant changes in anchor positions p2 and pΩ in different A2 mutants, differences that will be discussed in detail below.

2.8 Increased frequencies of K/R1 residues in peptides eluted from A2-K66N and A2-DM mutants

The most dramatic peptide-binding motif shift in the A*0201 mutants occurred in residues in peptide position 1 (p1). Approximately 25% of peptides bound to A2-WT contain either lysine (K) or arginine (R) at p1, which designates them as preferred residues in light of the combined background K/R frequency of 15% in vertebrate proteins (Fig 2.9). Stability assays comparing VLHDDLLEA to KLHDDLLEA revealed that KLHD showed better stability than its wild-type counterpart (Fig
K66N substitution allows for increased K/R1 peptide frequencies

Fig 2.10 K66N mutation changes amino acid composition of peptide position 1 in A*0201
Fig 2.10 K66N mutation changes amino acid composition of peptide position 1 in A*0201: (A) Peptide stability curves with VLHDDLLEA and KLHDDLLEA in A*0201 WT compared to positive control peptide LLFGYPVYV. (B) Logos of 9mer peptides eluted from A2-WT, A2-R65G, A2-K66N and A2-DM showing increase in K/R in position 1 in A2-K66N and A2-DM. (C) Fold change loss of amino acids with different biochemical characteristics in peptide position 1 of A2-K66N and A2-DM.
2.10A). This indicates lysine and arginine in peptide position 1 may improve peptide stability when bound to A*0201.

In peptides eluted from the A2 mutants, large differences were observed in the percentage of positively charged amino acids in position 1. A2-R65G showed a slight decrease in both K1 and R1 frequencies, while both A2-K66N and A2-DM showed a drastic increase in positively charged amino acids: ~70% of eluted peptides contained either K1 or R1 in these two A2 mutants (Fig 2.10B). K1 was the dominant amino acid, found in ~50% of all eluted peptides, while R1 was present in ~20% of all peptides from A2-K66N and A2-DM. The fact that A2-DM demonstrated the same trends as A2-K66N indicated that R65 is not likely to be involved in this interaction. This increase in K/R1 caused substantial decreases in all other amino acids in position 1 (Fig 2.10C). The largest fold decrease was observed in aromatic amino acids, specifically phenylalanine and tyrosine, but substantial frequency decreases also occurred in several polar and hydrophobic amino acids (Fig 2.9).

Since K1 and K66 are both positively charged, it was initially thought that the reduction in repulsion when mutating position 66 lysine to asparagine may have been responsible for the dramatic increase in K/R1 peptides. However, while coulombic interactions did show repulsion between K1 and K66 throughout the simulation, the repulsion was not strong (Fig 2.11A). Furthermore, the only amino acids which increased in frequency were lysine and arginine. Since this indicated an interaction that favored positive charges in position 1, we next examined other nearby residues for potential interactions.
K1 shows increased interactions with E4 in K66N A*0201 mutants

Fig 2.11 Interpeptide interactions between K1 and D/E4 occur in eluted HLA-A*0201 peptides
**Fig 2.11 Interpeptide interactions between K1 and D/E4 occur in eluted HLA-A*0201 peptides:** (A) Coulombic interactions between K1 and K66 in simulations of A2-WT bound to KLKEGIPAL.* (B&C) Alluvial plots of hydrogen bonds and salt bridges between peptide positions 1-4 of KLKEGIPAL in (B) A2-WT and (C) A2-K66N and A2-DM simulations. * Red lines indicate salt bridges and blue lines indicate hydrogen bonds. (D&E) Overlays of prominent crystal structures pulled from computational simulations of KLKEGIPAL bound to (D) A2-WT and (E) A2-DM.* (F&G) Analysis of co-occurrences between K/R1 and D/E4 in peptides eluted from A2-WT and A2 mutant molecules.

*Simulations were run in collaboration with Dr. Dinler Antunes, University of Houston
Interpeptide interactions were previously observed between E4 and K3 in the A2-WT/GLKEGIPAL complex, so we wondered whether K1 might also form interpeptide bonds. Interestingly, simulations of A2-WT/KLKEGIPAL showed that ~2% of the frames contained salt bridges between K1 and E4, which was less frequent than the salt bridges observed between K3 and E4 (Fig 2.11B). However, repeating the same simulations with A2-K66N and A2-DM mutants revealed that salt bridges between K1 and E4 increased to ~24% of all frames, thus providing an explanation for preference for positively charged amino acids in position 1 (Fig 2.11C). Examination of prevalent structures pulled from the simulation of A2-WT/KLKE and A2-DM/KLKE showed that the closed nature of the binding groove forces the K1 side-chain upwards, but the repulsion from K66 forced K1 away from E4 (Fig 2.11D). However, when K66 was replaced with N66, K1 was able to bend over top of peptide position 2 and form salt bridges with E4 (Fig 2.11E).

A*0201 peptides eluted from the H1975 cell line panel were analyzed to determine how often K/R1 co-occurred with D/E4 in the same peptide. The results showed that D/E4 peptides contained a higher percent of K/R1 compared with all other position 4 amino acids, and this proportion was more pronounced in A2-K66N and A2-DM (Fig 2.11F). In addition, K/R1 peptides contained a higher percent of D/E4 compared with all other p4 amino acids, and this proportion was also more pronounced in A2-K66N and A2-DM (Fig 2.11G). Therefore, the dramatic increase in K/R1 peptides seemed to be due to increased interaction with D/E4 when K66 is mutated to N66.
2.9 Peptide motif alterations in primary anchor positions of A*0201 mutants

Somewhat surprisingly, amino acid frequencies at eluted peptide C-termini showed alterations in A2-R65G and A2-DM compared to A2-WT and A2-K66N. In peptides eluted from A2-WT, valine at the C-terminus (VΩ) was present in ~50% of eluted peptides, while leucine (LΩ) was present in ~33% of peptides (see Fig 2.9). However, in A2-R65G and A2-DM VΩ frequencies increased by ~10 percent while LΩ decreased by ~10 percent (Fig 2.12A). Since p65/66 do not directly interact with the C-terminus of peptides, we wondered whether these changes were due to the fact that LΩ might represent a less stable anchor than VΩ. Stability assays in of A*0201 bound to VLHDDLLEA where the C-terminus was altered to VΩ or LΩ confirmed that VLHD-V was more stable than VHLD-L (Fig 2.12B). Furthermore, altering D4 to A4 destabilized VLHD-L but had a negligible impact on stability of the VLHD-V peptide (Fig 2.12B). These data suggest that charged non-pocket peptide-HLA interactions in A*0201 may be more important for enhancing the stability of peptides with weaker primary anchors.

We also examined position p2 in peptides eluted from the A2 variants. Peptides eluted from A2-WT showed that ~70% of peptides contained leucine in p2 while isoleucine, methionine and valine were all present in ~8% of eluted peptides (Fig 2.9). However, peptides eluted from both A2-K66N and A2-DM show a ~15% decrease in the frequency of L2 and a ~10% increase in V2 compared to A2-WT (Fig 2.12C). Although this was unexpected, previous computational modeling with K66 and N66 showed direct interactions with peptide position 2 (see Fig 2.6). In addition, stability assays with position 2 of RQApSLSISV substituted for V2 indicated
A2-K66N alters position 2 and A2-R65G alters the peptide C-terminus

Fig 2.12 Mutation of R65 altered C-terminal anchors while mutation of K66 altered p2 peptide anchors
Fig 2.12 Mutation of R65 altered C-terminal anchors while mutation of K66 altered p2 peptide anchors: (A) Differences in percent of amino acids in the C-terminus compared to wild-type of peptides eluted from all three A2 mutant molecules. **** indicates p<0.0001 using 2-sample test for equality of proportions without continuity correction (B) Peptide stability curves of VLHDDLLEL and VLHDDLLEV with D4 and A4 compared to positive control peptide LLFGYPVYV. (C) Differences in percent of amino acids in peptide position 2 compared to wild-type of peptides eluted from all three A2 mutant molecules. (D) Peptide stability curves of RQApSLSISV with position 2 changed to valine compared to positive control peptide LLFGYPVYV. ** indicates p<0.01, *** indicates p<0.001, **** indicates p<0.0001 using 2-sample test for equality of proportions without continuity correction
that valine is a stable p2 anchor (Fig 2.12D). Thus, the K66N mutation caused the substitution of a dominant p2 anchor (L2) with a less dominant but equally stable anchor (V2).

2.10 Discussion

Analysis of interactions between A*0201-bound peptides and R65 and K66 demonstrate that both residues are important for antigen presentation, but affect peptide binding in different ways. Loss of either R65 or K66 decreased A*0201 cell surface expression and peptide repertoire diversity, but substitution of both residues showed an additive effect. Since overall cellular protein expression was similar in both A2-WT and all of the A2 mutants, it indicates that these mutations may hinder binding of some peptides in the endoplasmic reticulum (ER), and thus some HLA molecules are unable to continue moving through the secretory pathway to the cell surface. This is further substantiated by loss of the peptide repertoire in both A2-K66N and A2-R65G compared to A2-WT. The fact that A2-DM shows a larger loss of peptide repertoire and cell surface expression than either single A2 mutant suggests that R65 and K66 are able to compensate somewhat for the loss of the other.

However, A2-K66N retained a higher cell surface expression and a more diverse peptide repertoire than A2-R65G, which may have occurred for a number of reasons. Though both single point mutants lose D/E4 peptides compared to A2-WT, A2-K66N retains more E4 peptides than either A2-R65G or A2-DM. The side chain of E4 is extended by one carbon atom compared to D4 and thus may be able to interact more easily with R65, located at the top of the binding groove. Molecular
dynamic simulations of A2-WT/GLKEGIPAL complexes reveal prevalent salt bridges between E4 and R65, and examination of other A*0201/peptide crystal structures showed clear interactions between E4 and R65 as well, indicating this interaction is prevalent in multiple A*0201-binding peptides (143, 148). Another potential reason for the increased peptide repertoire diversity is that N66, unlike G65, is polar and able to form hydrogen bonds with other peptide positions and HLA residues as shown in the simulations here. In fact, we observed that N66 was still able to form weak salt bridges with negatively charged amino acids in peptide position 4 due to its polar nature. These additional interactions supplement the intact p4-R65 interaction to give A2-K66N a peptide-binding advantage over A2-R65G.

A2-K66N also differed from A2-R65G in that it dramatically altered the binding motif, particularly in peptide position 1. The K66N mutation impacted the frequency of every amino acid typically found in peptide position 1 except for histidine. We determined that changing K66 to N66 caused more interactions between K/R1 and D/E4, but it is quite possible that other interactions may be occurring in different peptides that would be worthy of further investigation. A previous study analyzing HLA-A*0220, which also contains a K66 to N66 alteration, also reported increased frequencies of K/R1 compared to A*0201 WT and also postulated that the K66N change was crucial (151). This corroborating evidence demonstrates that data from our A2 mutants reflects actual motif changes in naturally-occurring HLA-A alleles.

Another difference between A2-R65G and A2-K66N occurred at the peptide C-terminus, where peptides bound to A2-R65G and A2-DM showed increases in the dominant anchor \( \Psi \Omega \), and decreases in less dominant C-terminal anchors. Since
R65 does not directly interact with the peptide C-terminus, this is likely another indication that middle non-pocket interactions affect overall peptide stability, as the more dominant C-terminal anchor was preferred. Thus, it makes sense that more stability in the middle of the peptide would be preferable for peptides with weaker primary anchor residues. Crystal structures with increased mobility in middle peptide residues showed decreased electron density or multiple conformations (115). Kuhns et al. showed that instability in the middle of a peptide derived from HER-2/neu resulted in poor HLA-I binding (130). This corroborates our stability analysis of the VLHDDLLEV and VLHDDLLEL peptides, where VLHD-L was shown to be less stable than VLHD-V, and therefore more stability loss was observed in VLHD-L with a D4A substitution compared to the same substitution in VLHD-V.

A2-DM demonstrated both peptide motif alterations: the impact on p1 seen in A2-K66N peptides, and the changes in pΩ observed in A2-R65G peptides. This indicated that R65 and K66 may have different modes of action related to peptide binding and stability. K66 is closer to the peptide and the binding groove, and is thus able to directly interact with multiple peptide positions (1, 2 and 4) and A*0201 backbone residues, as shown above. The role of R65 is simpler to interpret since it is located farther away from the binding groove and current simulations show interactions only with peptide position 4. However, we also observed a larger loss in peptide repertoire diversity, A*0201 surface expression, and loss of weaker C-terminal anchors with A2-R65G. One potential interpretation to explain this is when a peptide begins to become unstable and detach from the binding groove, R65 is in a prime position to interact with the D/E4 residues in the middle of the peptide and
stabilize interactions long enough for strong primary anchor attachments to be restored. The differences observed in the pΩ residue frequencies in A2-R65G peptides may indicate that detachment occurs more in this position compared with position 2. Further computational studies and mutational analysis in A*0201 or other HLA-A alleles with R65 and will be necessary to determine the accuracy of this potential model.
CHAPTER III: NON-POCKET INTERACTIONS BETWEEN HLA-A*2402 AND BOUND PEPTIDES ARE REQUIRED FOR OPTIMAL PEPTIDE BINDING STABILITY AND COMPLETE PEPTIDE REPertoire DIVERSITY
3.1 Creation of mutated alleles of HLA-A*2402

Like HLA-A*0201, A*2402 is the one of the most prevalent HLA allotypes in the human population. However, unlike A*0201, which is distributed worldwide, A*2402 demonstrates a particularly high prevalence in Asian populations (Table 1.2). While the frequency of D/E4 peptide ligands in A*0201 is ~40%, D/E4 is found in ~25% of A*2402-bound peptides (see Fig 1.9A). A*2402 contains K66 (similar to A*0201), but due to an arginine to glycine substitution (G65), it lacks a positively-charged residue at p65. Therefore, one of our hypotheses was that D/E4 peptides occurred at a lower frequency in A*2402 due to a lack of charged interactions between D/E4 and G65. Despite the high prevalence of A*2402, relatively few HLA-peptide complexes have been crystallized, and only one of the peptides, QFKDNVILL (QFKD), contained a negatively charged amino acid in p4 (144). Since molecular dynamic simulations are more accurate when based directly on crystal structures, we performed computational simulations on the A*2402/QFKDNVILL complex. As shown in Chapter 1, D4 showed strong charged interactions and formed highly prevalent salt bridges with K66 (see Fig 1.9H).

Closer inspection of the most predominant structures from A*2402/QFKD revealed that K66 is involved in several interactions with surrounding amino acid residues. Unlike A*0201, A*2402 has a glutamic acid position 62 (E62) in addition to the E63 residue that is common to both molecules. K66 interacts with both E62 and E63, though more structures showed interactions with E62 than E63 (Fig 3.1). K66 also moved around the alpha helix to interact with H70 and formed hydrogen bonds.
Wild-type HLA-A*2402 interactions with peptide QFKDNVILL

A24-WT / QFKDNVILL

Figure 3.1 Network of K66 interactions observed in HLA-A*2402/QFKDNVILL complexes
Figure 3.1 Network of K66 interactions observed in HLA-A*2402/QFKDNVILL complexes: Overlay of prevalent structures from computer simulations of A*2402 bound to QFKDNVILL. Yellow lines indicate hydrogen bonds while red lines indicate salt bridges*.

*Simulations were run in collaboration with Dr. Dinler Antunes, University of Houston
with the backbone of the phenylalanine in peptide position 2 (F2) (Fig 3.1). Thus, in this region of the peptide binding groove, K66 is at the center of a network of interactions involving both HLA and peptide residues.

In order to test our hypothesis, two mutant A*2402 alleles were created: the first mutant substituted an arginine for glycine in position 65 (A24-G65R), thus adding a positively-charged amino acid; the second mutant substituted the positively-charged lysine in position 66 with asparagine (A24-K66N), the consensus amino acid found in most HLA-A allotypes (Table 3.1). R65/K66 is naturally found in the HLA-A*24:29 allele, and represents the same configuration of amino acids found in positions 65-66 of WT A*0201; thus, A24-G65R is more “A*0201-like”. The corresponding A*0201 mutant, A2-R65G, is more “A*2402-like”, and was described in Chapter 2. The A24-K66N mutant has the configuration G65/N66 (like the A2-DM mutant), and also has an identical sequence to the naturally-occurring A24 allele HLA-A*24:145.
### HLA-A*2402 mutant constructs

<table>
<thead>
<tr>
<th>A*2402 construct</th>
<th>Amino acid sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-A cons</td>
<td>--QETRNVKAH--</td>
</tr>
<tr>
<td>A24-WT</td>
<td>--GETGKVKAH--</td>
</tr>
<tr>
<td>A24-G65R</td>
<td>--GETRKVKAH--</td>
</tr>
<tr>
<td>A24-K66N</td>
<td>--GETGNVKAH--</td>
</tr>
</tbody>
</table>

Table 3.1 HLA-A*2402 mutants and their p65-66 residue alterations
Table 3.1 HLA-A*2402 mutants and their p65-66 residue alterations: Amino acid sequences of positions 62-70 of HLA-A*2402 WT and mutated residues as well as the HLA-A consensus sequence. Positively charged amino acids in p65-66 are in blue and other amino acids in p65-66 are in green.
3.2 Positively-charged amino acids in p65-66 of HLA-A*2402 facilitate cell surface expression, peptide repertoire diversity, and D/E4 composition of bound peptide ligands

Lentivirus expressing A24-WT, A24-G65R, or A24-K66N were used to transduce H1975 cells, which do not endogenously express A*2402. Flow cytometric analysis using an A*2402-specific mAb showed a similar trend to that observed with the A*0201 mutants: A24-G65R demonstrated the highest cell surface expression, A24-K66N showed the lowest expression, and A24-WT had intermediate levels of expression (Fig 3.2A). Similar to the case with the A*0201 mutants, the levels of A*2402 surface expression were directly correlated with the number of positively charged amino acids in HLA p65/66.

Tandem mass spectrometry was also performed to identify peptides eluted from immunoprecipitated A24-WT and A24 mutant molecules. Using the pan-HLA-I antibody W6/32, >2,000 unique peptide sequences were identified from A24-G65R and A24-WT cells, while ~1,000 total peptide sequences were identified in A24-K66N cells. Since this represented a mixture of A*2402 peptides and peptide ligands bound to endogenous HLA-A, B, and C molecules, we identified the A*2402-binding peptides using known peptide motifs identified from large datasets of A*2402 peptides, which were sufficiently distinct from the peptide motifs of the endogenous HLAs (see Fig 2.3). This analysis revealed that we eluted 1265 unique peptides from A24-WT. Interestingly, the number of unique peptide ligands increased to 1891 in A2-G65R, while only 200 unique peptides were eluted from A24-K66N (Fig 3.2B). Between 700-800 unique peptides from HLA alleles endogenous to H1975 cells
Cell surface expression and peptide repertoire of A*2402 mutants

Figure 3.2 HLA-A*2402 cell surface expression and peptide repertoire size is dependent on the number of positively charged amino acids in p65-66.
Figure 3.2 HLA-A*2402 cell surface expression and peptide repertoire size is dependent on the number of positively charged amino acids in p65-66: (A) Cell surface expression of A*2402 in H1975 cells transduced with A24-WT, A24-G65R and A24-K66N. (B) Number of eluted A*2402 peptides in H1975 cells transduced with A24-WT or A24 mutant molecules. (C) Number of eluted peptides from HLA-I molecules endogenous to H1975 in H1975 cells transduced with A24-WT or A24 mutant molecules.
were also eluted from H1975 cells expressing A24-WT, A24-G65R, and A24-K66N, indicating that antigen presentation machinery was functional in all cells (Fig 3.2C). Thus, in A*2402 the K66 residue is required for complete cell surface expression and full diversity of peptide presentation, while the addition of R65 enhanced cell surface expression and expanded the repertoire of presented peptides.

We next analyzed peptide position 4 of peptides eluted from A24-WT and both A24 mutants. A24-G65R peptides demonstrated a slightly higher percentage of D4 peptides compared to A2-WT, but bound to a significantly higher percentage of E4 peptides (p<0.01) (Fig 3.3A). By contrast, A24-K66N showed a significant decrease of both D4 and E4 peptides compared to A24-WT (p<0.01) (Fig 3.3A). Further examination of the A24 mutants also showed a slight decrease of proline residues in p4 from A24-G65R compared to A24-WT (Fig 3.3B). A24-K66N peptides showed an increased frequency of the most prevalent p4 amino acids in A24-WT, including proline, serine, threonine and asparagine, indicating that this mutation induced a significant alteration of the p4 amino acid composition of bound peptides (Fig 3.3B).
R65 and K66 influence peptide position 4 composition in A*2402

Figure 3.3 Composition of amino acids in peptide position 4 is influenced by the number of positively charged amino acids in p65-66 of HLA-A*2402
Figure 3.3 Composition of amino acids in peptide position 4 is influenced by the number of positively charged amino acids in p65-66 of HLA-A*2402: (A)

Percent of eluted peptides with D or E in position 4 in A24-WT, A24-G65R and A24-K66N molecules. ** indicates p<0.01 using 2-sample test for equality of proportions without continuity correction. (B) Heat map of percent difference of amino acids in position 4 of both A24 mutants compared to A24-WT. Negative values indicate higher percentage of amino acids in A2-WT while positive values indicate higher percentage of amino acids in A2-mutants.
3.3 Substituting p65 or p66 in A*2402 alters HLA-I/peptide interactions

Molecular dynamic (MD) simulations of A*2402/QFKDNVILL complexes were run for A24-WT and both A24 mutants. Examination of cumbic interactions in A24-G65R showed weaker but more frequent interactions between D4 and K66, as well as some weak interactions with R65 (Fig 3.4A). A24-K66N exhibited attractive and repulsive interactions between D4 and N66, although attractive interactions were more frequent (Fig 3.4A). Analysis of hydrogen bonds and salt bridges showed that salt bridges between D4 and K66 in A24-G65R were similar to wild-type (Fig 3.4B). Addition of the R65 residue in A24-G65R did not induce the formation of salt bridges with D4 in QFKDNVILL. D4 did form salt bridges with N66 in the A24-K66N mutant, but they were less prevalent than salt bridges observed between D4 and K66 in A24-WT and A24-G65R (Fig 3.4B).

The most prevalent MD structures for both A24-G65R and A24-K66N were identified to illustrate changes in peptide-HLA interactions. In A24-G65R, the K66 residue shifted downward and formed predominant salt bridges with E63 instead of E62 as observed in A24-WT (Fig 3.4C). This downward shift also caused an increase in hydrogen bonds with the backbone of the peptide F2 residue (Fig 3.4C). While not observed in these structures, alluvial plots summarizing the entire MD simulation showed that K66 likely does interact with D4, though the downward shift of K66 positioning may explain why the interactions were weaker in this mutant. We expected that R65 would interact with D4, but the MD structures showed that R65 at the top of the alpha helix was attracted by the E62 HLA residue instead. While no salt bridges were formed, the residues were close enough for charged interactions to
QFKDNVILL peptide D4 residue interacts with R65 and K66 of A*2402
QFKDNVILL peptide D4 residue interacts with R65 and K66 of A*2402

Figure 3.4 Mutations to p65-66 in HLA-A*2402 alters HLA-I/peptide interactions
Figure 3.4 Mutations to p65-66 in HLA-A*2402 alters HLA-I/peptide interactions: (A) Coulombic interactions between D4 and positions 65-66 in computer simulations of QFKDNVILL in A24-WT, A24-G65R and A24-K66N.* (B) Alluvial plots of hydrogen bonds and salt bridges in peptide positions 1-4 and HLA residues in computer simulations of QFKDNVILL bound to A24-WT and mutant molecules.* Red lines indicate salt bridges and blue lines indicate hydrogen bonds. (C&D) Overlaps of prevalent crystal structures pulled from computer simulations of (C) A24-G65R and (D) A24-K66N bound to QFKDNVILL.*

*Simulations were run in collaboration with Dr. Dinler Antunes, University of Houston
occur. This was not observed in the A*0201 MD analyses, since A*0201 contains a glycine at position 62. This may explain why few interactions were seen between D4 and R65 throughout the simulation of the A*2402 peptide/HLA complex (Fig 3.4C). Thus, in A*2402 R65 may shift the positioning of K66 to cause altered interactions with the bound peptide.

Examination of the most prevalent MD structures of A2-K66N/QFKD showed that N66 formed strong hydrogen bonds with the backbone of F2 in addition to the lysine in peptide position 3 (K3), which was not observed in A24-WT simulations (Fig 3.4D). N66 also formed a few hydrogen bonds with HLA residues E62 and E63. D4 did form a few hydrogen bonds with N66, though no salt bridges between N66 and D4 were seen in the most prevalent structures of this simulation. Thus, in A*2402/QFKDNVILL complexes, N66 appears to interact in a network with several surrounding amino acids, but in a pattern that is distinct from K66.

3.4 D/E4 peptide residues contribute to peptide binding to HLA-A*2402

We next examined the role of D/E4 in the binding in multiple peptide backbones to wild-type A*2402. A competitive binding assay was utilized in which a fluorescently-labeled, known A*2402-binding peptide was co-incubated with increasing concentrations of test peptide in the presence of H1975 A24-WT cells. Test peptides that bind tightly outcompete the fluorescent peptide, while weaker binding peptides are less able to compete, especially at lower peptide concentrations (152).
This assay was performed on the QFKDNVILL peptide backbone altered to E or A in position 4 to determine the effect of negatively charged amino acids in p4 on A*2402 binding. We observed that QFKDNVILL and QFKENVILL both bound to A*2402, but not as strongly as the control peptide; furthermore, QFKDNVILL bound with slightly higher affinity (2- to 3-fold) compared to QFKENVILL (Fig 3.5A). Changing position 4 to alanine decreased peptide binding at least 10-fold compared to QFKD (to a level similar to the negative control peptide) particularly at lower test peptide concentrations, indicating that the presence of D/E4 improved binding of this peptide (Fig 3.5A). Peptide binding assays employing the VFGFVRACL or VFIDKQTNL peptide backbones substituted at p4 also confirmed improved binding to A*2402 when D/E4 were present and reduced binding when they were not (Fig 3.5B,C). These results support the notion that D/E4 peptide residues make a significant contribution to peptide binding to wild-type A*2402.
D/E4 peptide residues enhance peptide binding to HLA-A*2402

Figure 3.5 D/E4 peptide residues enhance peptide binding to HLA-A*2402 in multiple peptide backbones
Figure 3.5 D/E4 peptide residues enhance peptide binding to HLA-A*2402 in multiple peptide backbones: Competitive fluorescent binding assays of multiple A*2402 peptides with positive control peptide AYLEAIHNF and negative control peptide DVAEPYKVV. (A) QFKDNVILL with position 4 altered to E and A. (B) VFGFVRACl with position 4 altered to D4. (C) VFIDKQTNL with position 4 altered to A.
3.5 Mutating p65 or p66 residues alters HLA-A*2402 peptide-binding motifs

Alterations in peptide-binding motifs were examined in all peptide positions for A24-G65R and A24-K66N (Fig 3.6). Aside from position 4, A24-G65R amino acid frequencies were very similar to A24-WT from positions p3 through to p-1. By contrast, A24-K66N peptides showed altered frequencies in most of these middle peptide positions, which could indicate compensation for the loss of D/E4.

However, the largest differences observed in peptides from both A24-K66N and A24-G65R were in p1, p2 and pΩ. A24-WT peptide frequencies demonstrated a preference for positively charged amino acids in peptide position 1 above the expected background frequency, indicating a potential secondary interaction (Fig 3.6). Consistent with this, competitive binding assays demonstrated increased binding to A24-WT when position 1 was changed to lysine (Fig 3.7A).

Analysis of position 1 in peptides eluted from A24-G65R mutants showed a decrease in both K1 and R1 in A24-G65R, with a combined 7.25% reduction in K/R1 peptides compared to A24-WT frequency of ~25% (Fig 3.7B). By contrast, A24-K66N exhibited a 5% increase in peptides with K1 and no significant increase in peptides with R1 (Fig 3.7B). This demonstrated a phenotype quite distinct from that of K66N mutations in A*0201, and may be partially due to the dramatic loss of binding peptides shown in A24-K66N, or for other reasons discussed below.
Amino acid alterations in peptides eluted from A*2402 mutants

<table>
<thead>
<tr>
<th>AA</th>
<th>Position 1</th>
<th>Position 2</th>
<th>Position 3</th>
<th>Position 4</th>
<th>Position 5</th>
<th>Position -2</th>
<th>Position -1</th>
<th>C-terminus</th>
<th>Difference in amino acid percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>1.0</td>
<td>0.0</td>
<td>3.6</td>
<td>16.3</td>
<td>5.8</td>
<td>2.2</td>
<td>3.3</td>
<td>1.7</td>
<td>0.0</td>
</tr>
<tr>
<td>E</td>
<td>4.2</td>
<td>0.0</td>
<td>1.6</td>
<td>13.4</td>
<td>5.9</td>
<td>4.0</td>
<td>8.9</td>
<td>6.8</td>
<td>0.0</td>
</tr>
<tr>
<td>H</td>
<td>3.4</td>
<td>0.0</td>
<td>1.4</td>
<td>2.1</td>
<td>2.9</td>
<td>3.2</td>
<td>6.7</td>
<td>6.8</td>
<td>0.0</td>
</tr>
<tr>
<td>K</td>
<td>15.0</td>
<td>0.0</td>
<td>3.0</td>
<td>5.3</td>
<td>8.1</td>
<td>2.9</td>
<td>4.6</td>
<td>13.6</td>
<td>0.0</td>
</tr>
<tr>
<td>R</td>
<td>11.4</td>
<td>0.0</td>
<td>0.7</td>
<td>1.7</td>
<td>3.0</td>
<td>1.8</td>
<td>3.4</td>
<td>10.9</td>
<td>0.0</td>
</tr>
<tr>
<td>P</td>
<td>3.4</td>
<td>0.0</td>
<td>8.6</td>
<td>15.2</td>
<td>5.5</td>
<td>9.7</td>
<td>5.6</td>
<td>0.7</td>
<td>0.0</td>
</tr>
<tr>
<td>A</td>
<td>6.4</td>
<td>0.0</td>
<td>7.0</td>
<td>4.6</td>
<td>4.9</td>
<td>5.6</td>
<td>6.9</td>
<td>4.1</td>
<td>0.0</td>
</tr>
<tr>
<td>G</td>
<td>1.2</td>
<td>0.0</td>
<td>3.9</td>
<td>4.8</td>
<td>5.3</td>
<td>5.7</td>
<td>1.4</td>
<td>2.8</td>
<td>0.0</td>
</tr>
<tr>
<td>I</td>
<td>7.1</td>
<td>0.6</td>
<td>10.6</td>
<td>3.4</td>
<td>9.0</td>
<td>9.6</td>
<td>2.6</td>
<td>1.5</td>
<td>16.2</td>
</tr>
<tr>
<td>L</td>
<td>6.6</td>
<td>2.6</td>
<td>19.6</td>
<td>3.7</td>
<td>6.5</td>
<td>10.8</td>
<td>8.3</td>
<td>10.6</td>
<td>31.3</td>
</tr>
<tr>
<td>M</td>
<td>1.2</td>
<td>1.1</td>
<td>3.7</td>
<td>0.2</td>
<td>0.9</td>
<td>1.8</td>
<td>1.2</td>
<td>1.1</td>
<td>1.2</td>
</tr>
<tr>
<td>V</td>
<td>12.6</td>
<td>0.0</td>
<td>5.8</td>
<td>3.4</td>
<td>11.3</td>
<td>9.5</td>
<td>7.1</td>
<td>4.7</td>
<td>0.0</td>
</tr>
<tr>
<td>C</td>
<td>1.5</td>
<td>0.0</td>
<td>1.3</td>
<td>0.1</td>
<td>0.7</td>
<td>0.6</td>
<td>0.8</td>
<td>1.1</td>
<td>0.0</td>
</tr>
<tr>
<td>N</td>
<td>3.7</td>
<td>0.1</td>
<td>5.1</td>
<td>5.3</td>
<td>5.6</td>
<td>2.9</td>
<td>8.4</td>
<td>4.4</td>
<td>0.0</td>
</tr>
<tr>
<td>Q</td>
<td>2.8</td>
<td>0.4</td>
<td>4.5</td>
<td>3.7</td>
<td>4.2</td>
<td>4.0</td>
<td>9.0</td>
<td>5.1</td>
<td>0.0</td>
</tr>
<tr>
<td>S</td>
<td>6.7</td>
<td>0.0</td>
<td>8.2</td>
<td>7.5</td>
<td>7.0</td>
<td>5.9</td>
<td>6.8</td>
<td>7.6</td>
<td>0.0</td>
</tr>
<tr>
<td>T</td>
<td>6.2</td>
<td>0.0</td>
<td>5.9</td>
<td>6.0</td>
<td>5.3</td>
<td>6.7</td>
<td>4.5</td>
<td>10.8</td>
<td>0.0</td>
</tr>
<tr>
<td>F</td>
<td>1.5</td>
<td>9.2</td>
<td>3.8</td>
<td>1.6</td>
<td>3.3</td>
<td>9.5</td>
<td>4.9</td>
<td>2.3</td>
<td>44.9</td>
</tr>
<tr>
<td>W</td>
<td>0.0</td>
<td>2.4</td>
<td>0.2</td>
<td>0.6</td>
<td>1.0</td>
<td>1.8</td>
<td>0.8</td>
<td>0.2</td>
<td>6.4</td>
</tr>
<tr>
<td>Y</td>
<td>4.1</td>
<td>83.7</td>
<td>1.4</td>
<td>1.1</td>
<td>3.8</td>
<td>1.7</td>
<td>4.6</td>
<td>3.2</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Figure 3.6 Mutations to positions 65/66 alter HLA-A*2402 peptide-binding motifs
Figure 3.6 Mutations to positions 65/66 alter HLA-A*2402 peptide-binding motifs: Heat map showing percent differences between A24 mutants and wild-type in amino acids in all peptide positions of peptides eluted from A24 mutant molecules. Negative values indicate a higher percentage of amino acids in A24-WT while positive values indicate a higher percentage of amino acids in A24-mutants.
Peptide K1 residues increase stability in HLA-A*2402

Figure 3.7 Peptide K1 residues cause increased stability in HLA-A*2402
Figure 3.7 Peptide K1 residues cause increased stability in HLA-A*2402: (A)
Competitive fluorescent binding assay of QFKDNVILL with position 1 changed to lysine compared to positive control peptide AYLEAIHNF and negative control peptide DVAEPYKVY. (B) Percent of eluted peptides from A24-WT and both A24 mutants with lysine (K) or arginine (R) in peptide position 1. ** indicates p<0.01 using 2-sample test for equality of proportions without continuity correction.
Based on the known peptide-binding motif for WT A*2402, tyrosine (Y) is the predominant p2 anchor, with approximately 84% of all peptides containing this amino acid (see Fig 3.6). The next most prevalent p2 anchor is phenylalanine (F2, at 9%), while leucine and tryptophan are present at p2 frequencies of 2.5 to 3% of A*2402-bound peptides (see Fig 3.6). The vast majority of peptides eluted from A24-G65R and A24-K66N also contained Y2, though subtle frequency differences were noted. Peptides from A24-G65R showed a ~4% decrease in Y2 frequency and a concurrent increase in the less prominent p2 anchors leucine, phenylalanine and tryptophan (Fig 3.8A). This indicated that A24-G65R might allow for binding of some peptides with weaker anchor residues. Interestingly, A24-K66N showed the opposite trend, demonstrating a 2% increase in Y2 frequency and slight decreases in F2 and W2, indicating a slight disfavoring of less dominant anchors (Fig 3.8A).

The preferred C-terminal anchors of A24-WT peptides are phenylalanine (FΩ, at ~45%), and leucine (LΩ, at ~31%), while isoleucine and tryptophan are also present at lower frequencies (see Fig 3.6). Unlike p2, amino acid frequencies at this pΩ did not differ between A24-WT and the A24-G65R mutant (Fig 3.8B). However, peptides eluted from A24-K66N demonstrated a 14% increase in the frequency of FΩ, with concurrent decreases in LΩ and WΩ, indicating that K66 loss results in a favoring of peptides containing the more dominant C-terminal anchor (Fig 3.8B).
Dominant primary anchors show increased frequencies in peptides eluted from A24-K66N

Fig 3.8 Alterations in primary anchor positions of mutated HLA-A*2402 molecules
Fig 3.8 Alterations in primary anchor positions of mutated HLA-A*2402 molecules: (A&B) Differences in percent of amino acid anchors in both A24 mutants in compared to A24-WT in (A) peptide position 2 (Y2 is the dominant anchor) and (B) the C-terminus (FΩ is the dominant anchor). * indicates p<0.05, *** indicates p<0.001, **** indicates p<0.0001 using 2-sample test for equality of proportions without continuity correction.
3.6 D/E4 stabilizes HLA-A*2402 peptides with weaker primary anchors

To investigate the importance of charged interactions between peptide position 4 and HLA-p65-66 on stability of bound peptides, assays were performed on several different substituted variants of QFKDNVILL. Although QFKDNVILL has been crystallized and studied for its unique position 3 to position 5 interaction, stability assays indicated that this peptide was not very stable in the peptide-binding groove under physiological conditions (Fig 3.9A). Since neither 
F2 nor L\fect are dominant primary anchors, we decided to alter both amino acids to examine the change in stability with different anchors. Substituting F\fect for L\fect in QFKDNVILL (F2-F\fect) improved stability by 15%, while substituting Y2 for F2 (Y2-L\fect) caused stability to improve by 38%, indicating that both Y2 and F\fect constitute more stable primary anchors (Fig 3.9A). Substituting in both dominant anchors (Y2-F\fect) improved stability by 71%, demonstrating that p2 and p\fect anchors work in combination with each other to determine stability of HLA/peptide binding (Fig 3.9A).

The impact of position 4 on peptide stability was also examined by exchanging D4 for A4 in each peptide above and measuring peptide stability. For the highly stable QYKDNVILF peptide, the D4 to A4 change reduced stability by \(~15\%)\), which was measurable although the peptide binding was still very stable (Fig 3.9B). Changing D4 to A4 in QYKDNVILL resulted in a similar drop in stability (\(~18\%)\) (Fig 3.9C). Interestingly, in QFKDNVILF, the least stable of the three primary anchor-mutated peptides, substituting A4 for D4 completely reduced the stability of this peptide to 0\%, indicating that the D4 residue is crucial for stability of this peptide backbone in the A*2402 groove (Fig 3.9D). Collectively, these results demonstrate
Peptides with weaker primary anchors utilize D4 residue for increased stability

Figure 3.9 D4 residue is more critical for stability of HLA-A*2402 peptides with weaker primary anchors
Figure 3.9 D4 residue is more critical for stability of HLA-A*2402 peptides with weaker primary anchors: (A) Peptide stability curves of QFKDNVILL and peptides with the following substitutions: F0, Y2, Y2-F0 compared to positive control peptide TYTQDFNKF. (B-D) Peptide stability curves of (B) QFKDNVILF, (C) QYKDNVILL and (D) QFKDNVILF with position 4 altered to alanine.
that D4 contributes to peptide stability, particularly for peptides with weaker primary anchors.

3.7 Discussion

A*2402 is a highly prevalent HLA-A allele that binds to a significant proportion of D/E4 peptides (~25%), but not as many as A*0201 (~40%). Since A*2402 shares K66 but not R65 with A*0201, it presented an ideal opportunity to assess the importance of these residues for D/E4 peptide interactions and for antigen presentation by making reciprocal mutants of these two HLA-I allotypes. Based on our previous work in A*0201, we hypothesized that adding R65 to A*2402 would enhance D/E4 peptide binding and improve antigen presentation, while removing K66 would be deleterious to both of these parameters. As predicted, A24-G65R demonstrated increased cell surface expression, increased peptide repertoire diversity, and an increased proportion of eluted D/E4 peptides. By contrast, A24-K66N showed decreased cell surface expression, a dramatically reduced peptide repertoire, and the few peptides that were presented showed a reduced percentage of D/E4 peptides.

The increase in cell surface expression and peptide repertoire observed in A24-G65R compared to A24-WT indicated that A24-G65R might possess an expanded binding motif compared to A24-WT. However, when we analyzed this further, we discovered high amino acid similarity in all peptide positions except for positions 2 and 4. In position 2, there was a slight increase in the number of peptides with subdominant primary anchors (i.e. not Y2) but overall, the binding motifs were very
similar. This led us to the conclusion that the G65R change enhanced the binding and/or stability of peptides containing similar binding motifs to A24-WT, resulting in the observed increase in surface expression and peptide repertoire diversity.

Further analysis of p4 in eluted A24-G65R peptides showed that the frequency of D4 peptides increased only slightly, while E4 peptide frequencies increased significantly. This indicates E4 and R65 are more inclined to interact favorably with each other, which is similar to what we observed in simulations of GLKEGIPAL/A*0201 complexes. However, we also observed that R65 in A24-G65R may interact with E62, which is not present in A*0201. This may explain why there was not a larger impact of the G65R mutation on the A*2402 peptide-binding motif. In addition, D4 peptides may not have increased as much as E4 peptides because D4 has a smaller side-chain than E4 and may have been too short to reach and interact with R65, located at the top of the binding groove alpha-helix.

Unlike in A*0201, altering K66 to N66 in A*2402 had a huge negative impact on antigen presentation, resulting in a catastrophic loss of peptide repertoire diversity (~80% loss). This indicates that the K66 residue is heavily involved in peptide binding and/or stability and loss of this residue is not easily tolerated in A*2402. Interestingly, N66 can still make several hydrogen bonds with surrounding residues, including the backbone of peptide position 2 and even weak salt bridges with position 4 in QFKDNVILL. However, we only simulated one A*2402/peptide complex due to a lack of available A*2402 crystal structures; therefore, it is possible that N66 does not interact well with the predominant Y2 anchor of A*2402-binding peptides,
or that the peptides can bind but N66 is insufficient to stabilize the peptides, and they dissociate more readily from the binding groove.

Unlike A24-G65R, we observed changes in the peptide motifs from A24-K66N at multiple peptide positions. The increase in lysine in position 1 was not nearly as dramatic as we saw for A2-K66N, but stability studies with A24-WT indicated that K1 does offer some additional binding stability that may enhance binding to A24-K66N. Both p2 and pΩ demonstrated an increase of dominant primary anchors in this mutant, though pΩ showed the larger increase. Since p66 cannot directly interact with the peptide C-terminus, this likely indicates a loss of overall peptide stability resulting from the K66N substitution.

This was directly tested using synthetic peptides containing different primary anchor pairs of the peptide backbone QFKDNVILL in A*2402 WT, which showed that peptides with only one or zero dominant primary anchors relied more on the stability added by a negatively charged amino acid in position 4. This makes sense, since high stability at both peptide ends would anchor the peptide more strongly and allow for more instability in the middle of the peptide. However, a weakening of primary pocket interactions at p2 or pΩ would make stable non-pocket interactions between HLA residues and peptide p4 residues more advantageous.
CHAPTER IV: NON-POCKET CHARGED INTERACTIONS SHOW SIMILARITIES IN A*0201 AND A*2402, AND MAY BE CONSERVED IN MHC CLASS I MOLECULES
Comparison of A*0201 and A*2402 results

4.1 Basic residues in p65/66 of HLA-A*0201 and A*2402 are required for optimal antigen presentation and full peptide repertoire diversity

Analysis of HLA-A*0201 and HLA-A*2402 mutants were considered separately in Chapters 2 and 3, respectively. However, both A2 and A24 mutant alleles were created with reciprocal mutations that changed HLA positions 65-66 to make them more similar to the other allele, or to other HLA-A molecules. For example, both A2-WT and A24-G65R contain R65/K66, while both A24-WT and A2-R65G have G65/K66. Likewise, A2-K66N and the most other HLA-A alleles contain the consensus configuration R65/N66. Therefore, it was of interest to compare the wild-type and mutant A2 and A24 molecules to each other to characterize the similarities and differences in detail.

The first comparison focused on changes in cell surface expression of WT and mutant HLA molecules. The percentage change in cell surface expression between mutant and WT HLA molecules was calculated for each allotype and then compared against the other mutants. A24-G65R showed an increase in cell surface expression of >150% compared to wild-type A24, indicating a substantial change in antigen presentation. (Fig 4.1A). All of other A2 and A24 mutants showed decreased cell surface expression compared to WT. A2-DM showed the highest decrease in surface expression (>80%) compared to A2-WT; interestingly, A24-K66N showed a less dramatic decrease in surface expression compared to A24-WT (~50%), even though both A2-DM and A24-K66N contained the G65/N66 configuration (Fig 4.1A).
Loss of R65 and/or K66 causes loss of antigen presentation

Figure 4.1 HLA molecules with more positively charged amino acids in p65-66 demonstrate higher cell surface expression and expanded peptide repertoires
Figure 4.1 HLA molecules with more positively charged amino acids in p65-66 demonstrate higher cell surface expression and expanded peptide repertoires:

(A) Percent change of cell surface expression in mutant HLA-A molecules compared to their wild-type counterparts. Positive percent change indicates increased cell surface expression while negative percent change indicates decreased cell surface expression compared to wild-type. (B) Percent change of peptide repertoire in HLA-A mutants compared to their wild-type counterparts. Positive percent change indicates increase peptide repertoire while negative percent change indicates decreased repertoire compared to wild-type.
Changes in the number of unique peptides eluted from A*0201 and A*2402 WT and mutant molecules were then compared. Percent changes were calculated compared to the WT allotypes to ensure a fair comparison of the mutants. A24-G65R was the only mutant to show an increased peptide repertoire compared to the corresponding wild-type HLA-I (Fig 4.1B). A24-G65R exhibited a ~50% increase in unique peptides compared to A24-WT, which was less than the observed increase in cell surface expression, suggesting that some of the increased expression might due to an overall increased density of each HLA-I/peptide complex. A2 mutants containing mutations to either R65 or K66 showed a loss of 30-60% of unique peptide repertoire compared to WT, with A2-R65G exhibiting a larger loss than A2-K66N (Fig 4.1B). HLA mutants lacking both R65 and K66 lost 70-85% of their peptide repertoire compared to their wild-type counterparts. A24-K66N exhibited the largest loss of peptide repertoire of all the mutants, indicating that A*2402-binding peptides rely more on the single positively charged residue K66 to stabilize peptide binding. By contrast, A*0201 contains two positively charged residues at R65/K66, which can each apparently compensate somewhat for the loss of the other; the peptide repertoire only takes a dramatic hit when both residues are lost, as in A2-DM.

4.2 D/E4 peptide residues enhanced the stability of peptides containing fewer dominant primary anchors

We next compared the percentage of D/E4 peptides eluted from WT A*0201, WT A*2402, and the five mutants. Supporting our initial hypothesis, HLA molecules
containing both R65 and K66 showed the highest percentage of D/E4 peptides, with A2-WT containing ~45% D/E4 peptides compared to ~35% D/E4 peptides from A24-G65R (Fig 4.2). The majority of the differences were due to E4, as both A2-WT and A24-G65R showed similar proportions of D4 peptides (18-20%). This indicated that E4 peptides may be interacting somewhat differently with A*0201 compared to A*2402.

Since several WT and mutant HLA-I molecules contained one positively charged residue in p65-66, we next compared HLA-I molecules that contained either G65/K66 or R65/N66. Both alleles with the G65/K66 configuration (A2-R65G and A24-WT) exhibited similar overall levels of D/E4 peptides of ~30% (Fig 4.2). However, A2-R65G bound to more E4 peptides while A24-WT bound to slightly more D4 peptides, indicating some differences between A*0201 and A*2402. No mutants analogous to A2-K66N were made in A*2402 (i.e. containing R65/N66), but several other HLA-A allotypes do possess this R65/N66 configuration. Therefore, percentage of D/E4 peptides eluted from these other HLA-A molecules was averaged and compared to that of A2-K66N. A2-K66N bound ~30% D/E4 peptides, while other HLA-A allotypes bound a mean of ~24% D/E peptides. Both showed similar D4 levels (~13%), but A2-K66N bound to a significantly higher proportion of E4 peptides (~80% higher) compared to other HLA-A alleles (Fig 4.2). HLA mutants with the G65/N66 configuration demonstrated the lowest D/E4 percentage of all alleles, though A2-DM bound to more E4 peptides than A24-K66N. In all cases, peptides eluted from A*0201 molecules exhibited higher levels of E4 peptides.
HLA-I molecules with more positively charged residues in p65-66 bind to a higher proportion of D/E4 peptides.

Figure 4.2 Percentage of D/E4 eluted peptides correlates with number of positively charged amino acids in HLA-I p65-66.
Figure 4.2 Percentage of D/E4 eluted peptides correlates with number of positively charged amino acids in HLA-I p65-66: Comparison of the percent of eluted peptides with D4 and E4 in A*0201 and A*2402 WT and mutated molecules by amino acids in positions 65-66.
compared to A*2402 or other HLA-A allotypes possessing similar p65/66 configurations. This indicates that A*0201 may have a potential advantage for binding E4 peptides that may occur independently of p65-66.

Stability assays testing both A*0201 and A*2402 peptides demonstrated that the reliance on D/E4 residues for peptide stability was dependent on the primary anchors in peptide position 2 and at the C-terminus. To quantitate this, we separated the A*0201- and A*2402-restricted peptides into two groups, depending on the number of dominant anchors found in p2 and pΩ (L2-VΩ for A*0201, Y2-FΩ for A*2402), and then calculated the change in stability that occurred when position 4 was changed to D or E. Peptides with one dominant anchor in either p2 or pΩ showed an average stability increase of 12%, though the differences with individual peptides ranged from a +3% to +25% increase (Fig 4.3). By contrast, peptides containing dominant anchors at both p2 and pΩ showed significantly lower increases in stability (mean 3.5%, range -2.5% to +10% range) with addition of D/E4 (Fig 4.3). Thus, both A*0201- and A*2402-restricted peptides with two dominant primary anchors are less reliant on D/E4 for stability compared to peptides with only one dominant anchor.
Peptides with weaker primary anchors are more dependent on D/E4 residues for stability

Figure 4.3 D/E4 peptide residues are more important for stability of peptides with fewer dominant primary anchors
**Figure 4.3 D/E4 peptide residues are more important for stability of peptides with fewer dominant primary anchors:** A*0201 and A*2402 peptides were split by the number of dominant anchors in position 2 and the C-terminus, the dominant anchors being Y2-F0 for A*2402 peptides and L2-V0 for A*0201 peptides. The stability of peptide backbones from each HLA-A molecule with and without D/E4 was determined and the increase in stability when a peptide contained D/E4 was calculated and plotted. ** indicates p < 0.01 using unpaired t-tests with Welch’s correction.
4.3 DISCUSSION

It is clear that similar trends were observed in A*0201 and A*2402 with respect to preference for D/E4 peptides and the importance of D/E4 for peptide stability. Both HLA allotypes showed increased preference for D/E4 peptides when there were more positively charged amino acids in p65-66, indicating a significant correlation between these HLA-I and peptide residues. Furthermore, in both allotypes the contribution of these charged, non-pocket interactions to the stabilization of peptide-HLA complexes appears to be more important in peptides with one dominant anchor as opposed to two dominant primary anchors.

Some key differences were also observed between A*0201 and A*2402. For example, it was clear that A24-G65R bound significantly fewer E4 peptides than A2-WT. One potential explanation for this is that simulations of A24-G65R/QFKDNVILL revealed that R65 spends the majority of the simulation time interacting over the top of the binding groove with E62 (see Fig 3.3C). E62 is situated in a similar position to R65 at the top of the alpha helix and is not present in A*0201, which contains a glycine at this position. We previously showed that E4 and R65 in A*0201 prefer to interact with each other; therefore, it is plausible that R65 in A*2402 is less available to interact with E4 than R65 in A*0201 due to the competing interaction between R65 and E62.

However, E4 was also more prevalent in peptides eluted from all of the A2 mutants compared to A*2402 or to other HLA-A allotypes. In the case of the other HLA-A molecules, there could be a similar phenomenon where R65 interacts preferentially with other HLA residues instead of E4. However, there were also more
E4 peptides eluted from in A2 mutants that lacked R65. The E4 peptides we modeled predominately interacted with the alpha-1 helix, but also showed some interactions with the residues of the opposite alpha-2 helix. Perhaps the longer reach of the extended side chain of E4 allows for interactions with other HLA residues that are not available to D4. We also observed intra-peptide interactions between E4 and positively charged residues in peptide positions 1 and 3, so it is likely that peptide-peptide interactions occur in other backbones as well. Thus, E4 residues may possess a unique plasticity of binding that seems to be preferred by A*0201.

**Non-pocket interactions involving R65 and K66 may be conserved across the HLA-I system and MHC-I molecules from other species**

4.4 R65 and K66 are conserved across HLA-I alleles and their presence correlate with presentation of D/E4 peptide ligands

Our initial survey of HLA-I peptides showed elevated frequencies of D/E4 peptides across almost all HLA-I molecules examined (Fig 1.6). Based on our studies with A*0201 and A*2402, we analyzed how prevalent R65 and K66 were across all classical HLA-A, -B, and -C molecules. Examination of HLA-A allotypes showed that ~60% of HLA-A have R65/N66, ~10% have G65/K66 (unique to the HLA-A24 family of alleles), and the remaining ~30% contain both R65 and K66, which are predominately from the HLA-A2 family (Fig 4.4A). Thus, HLA-A molecules all show a strong preference for D/E4 peptides, and also all contain at least one positively charged amino acid in either position 65 or position 66.
In HLA-B, no allotypes expressed both R65 and K66. Furthermore, only ~15% of HLA-B allotypes contained either R65 or K66, and of those the majority expressed R65 (Fig 4.4B). The other ~85% of HLA-B molecules possess a consensus glutamine (Q) in position 65 and an isoleucine (I) in position 66. Q is a polar amino acid similar in structure to glutamic acid (E) but contains a side chain amide group (-NH₂) instead of a hydroxyl group (-OH). All HLA-C allotypes contained Q65, similar to HLA-B; ~90% of HLA-C allotypes contained K66, while the other ~10% contained N66 (Fig 4.4C).

We then analyzed whether HLA-I molecules with more positively charged p65/66 amino acids naturally bound to a higher proportion of peptides with D/E in position 4. HLA-A, -B, and -C allotypes were separated based on their p65/66 configurations and the percentage of bound D/E4 peptides was determined from public peptide ligand databases. As expected, HLA-I molecules with both R65 and K66 showed a significantly higher percentage of D/E4 peptides compared to HLA-I allotypes with one or neither of these two amino acids (Fig 4.4D). Furthermore, HLA-I allotypes with either R65 or K66 bind to a significantly higher proportion of D/E4 peptides than those allotypes lacking both R65 and K66. This analysis demonstrates that the positively charged amino acid content in HLA-I p65-66 correlates with presentation of D/E4 peptides throughout the HLA-A, B, and C system.
Most HLA-A, B, and C molecules conserve R65 and/or K66 and show correlation with D/E4 peptide frequencies.

Figure 4.4 Presence of R65 and/or K66 in HLA-I molecules correlates with frequency of bound D/E4 peptides.
Figure 4.4 Presence of R65 and/or K66 in HLA-I molecules correlates with frequency of bound D/E4 peptides: (A-C) Examination of positions 65-66 in (A) HLA-A, (B) HLA-B and (C) HLA-C molecules previously examined for prevalence of D/E4 peptides. (D) HLA-A, B and C molecules were separated according to expression of positively charged amino acids in p65-66 and percent of D/E4 peptides was determined. Brown-Forsythe ANOVA test showed significant difference between means (p<0.0001) and multiple comparisons using Dunnett’s T3 test were performed to identify significant differences between specific groups. *** indicates p<0.001.
4.5 D/E4 peptide preference is conserved in animal MHC-I molecules

As discussed in the Introduction section, HLA-I homologues are found in other species (referred collectively here as MHC-I molecules) and are present in all jawed vertebrates due to the necessity of having an adaptive immune response capable of responding to intracellular pathogens. Since we observed a correlation between D/E4 and R65/K66 across the HLA-I system, we wondered whether this interaction was also found in animal MHC-I molecules. Interest in animal MHC-I molecules has intensified in the past several years, with multiple studies emerging that analyze MHC-I molecules in several different animal species (153-158). The NetMHCpan4.1 algorithm allows peptide-binding predictions for peptides from a wide variety of MHC-I alleles derived from different species (45). Publicly available datasets of eluted MHC-I peptides from animal species are still limited; however, we found sufficient peptides eluted from MHC-I molecules in cattle, dog, rhesus monkey and mice to perform an analysis (45, 88). Examination of the percentage of D/E4 peptides in these animal species showed that they also occurred above the expected background frequency, particularly for dog and cattle MHC-I peptides (Fig 4.5A).

Some of the most prevalent MHC-I allotypes from each species were then aligned to analyze the amino acids found at positions 65 and 66. Interestingly, we found that there is a high conservation of R65 in the majority of examined MHC-I molecules from all four species, while a small handful of MHC-I molecules have lysine or arginine in position 66 (Fig 4.5B). These data illustrate that D/E4 and R65 are prevalent in multiple animal species as well. Further studies will be required to
Figure 4.5 Animal MHC-I molecules prefer D/E4 peptides and conserve R65

A

![Graph showing peptide preference and conservation](image)

B

<table>
<thead>
<tr>
<th>Species</th>
<th>MHC-I Allele</th>
<th>MHC-I Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>60  65  70  75</td>
</tr>
<tr>
<td>Dog</td>
<td>DLA-88*501:01</td>
<td>EYWDPQT RTIKETARTFR</td>
</tr>
<tr>
<td></td>
<td>DLA-88*34:01</td>
<td>EYWDGET RKVQTAQLYR</td>
</tr>
<tr>
<td>Mouse</td>
<td>H-2-Kb</td>
<td>EYWERETQ KAKGNEQSF R</td>
</tr>
<tr>
<td></td>
<td>H-2-Db</td>
<td>EYWERETQ KAGGEQWF R</td>
</tr>
<tr>
<td></td>
<td>H-2-Kd</td>
<td>EYWEEQTQ RAKSDEQWF R</td>
</tr>
<tr>
<td></td>
<td>H-2-Kk</td>
<td>EYWERNTQI KAKGNEQIF R</td>
</tr>
<tr>
<td></td>
<td>H-2-Ld</td>
<td>EYWERITQIAKQQEQWF R</td>
</tr>
<tr>
<td></td>
<td>H-2-Dd</td>
<td>EYWERTQ KAKGNEQSF R</td>
</tr>
<tr>
<td>Cattle</td>
<td>BoLA-1*019:01</td>
<td>EYWDRET RISKENTLYVR</td>
</tr>
<tr>
<td></td>
<td>BoLA-2*012:01</td>
<td>EYWDRET RNFKD TAQTFR</td>
</tr>
<tr>
<td></td>
<td>BoLA-2*016:01</td>
<td>EYWDRTQ IRYKDTAQTF R</td>
</tr>
<tr>
<td></td>
<td>BoLA-2*018:01</td>
<td>EYWDRTQ IRYKDTAQTF R</td>
</tr>
<tr>
<td></td>
<td>BoLA-3*002:01</td>
<td>EYWDRTQ RKAGNAQFFR</td>
</tr>
<tr>
<td></td>
<td>BoLA-3*017:01</td>
<td>EYWDRTQ RAKGNAQIF R</td>
</tr>
<tr>
<td></td>
<td>BoLA-4*024:01</td>
<td>EYWDEQTR IKVDTAQTFR</td>
</tr>
<tr>
<td></td>
<td>BoLA-6*013:01</td>
<td>EYWDRETQ ISKENALWYR</td>
</tr>
<tr>
<td></td>
<td>BoLA-6*014:02</td>
<td>EYWDRETQ ISKENALYR</td>
</tr>
<tr>
<td></td>
<td>Mamu-A1*001:01</td>
<td>EYWDRET RMKTETQ NAP</td>
</tr>
<tr>
<td>Rhesus monkey</td>
<td>Mamu-A1*002:01</td>
<td>EYWDRET RMKAETQ NAP</td>
</tr>
<tr>
<td></td>
<td>Mamu-A1*011:01</td>
<td>EYWDRET RSKANTQTYR</td>
</tr>
<tr>
<td></td>
<td>Mamu-B*001:01</td>
<td>EYWDRET RKAQNAQTDR</td>
</tr>
<tr>
<td></td>
<td>Mamu-B*003:01</td>
<td>EYWEEET RAKGHAQTD R</td>
</tr>
<tr>
<td></td>
<td>Mamu-B*017:01</td>
<td>EYWEEKT RRAKEAAQTHR</td>
</tr>
<tr>
<td></td>
<td>Mamu-B*039:01</td>
<td>EYWEEET RRAKETAQ TF R</td>
</tr>
<tr>
<td></td>
<td>Mamu-B*083:01</td>
<td>EYWEEET RNSKARAQTD R</td>
</tr>
</tbody>
</table>
Figure 4.5 Animal MHC-I molecules prefer D/E4 peptides and conserve R65:

(A) Eluted peptides from cattle, mice, dog and rhesus monkey were examined for percent of D/E in all peptide positions. (B) Sequence alignment of multiple MHC molecules from dog, mouse, cattle and rhesus monkey. Positively charged amino acids in positions 65-66 are in blue.
determine whether they interact in an analogous fashion as they do in human HLA-A molecules.

4.6 Discussion

Since D/E4 peptides are prevalent in multiple HLA-I molecules, we were very curious to explore whether the non-pocket interactions identified between D/E4 and p65-66 of A*0201 and A*2402 might have relevance throughout the HLA-I system. The evidence presented here demonstrates that the presence of D/E4 and R65/K66 are correlated and most likely interact with each other, though more studies in other HLA-I allotypes are needed to formally demonstrate this.

The results presented here raise the possibility that the interaction between D/E4 residues of bound peptides and R65/K66 residues of the HLA-I backbone constitutes a hitherto undescribed, subdominant pan-HLA-I anchor. Extensive searches of the published literature show no reports of conserved peptide-HLA interactions that span multiple HLA-I allotypes. It is well-known that allotypes from the same HLA superfamilies share very similar peptide-binding motifs, which has been the basis of several HLA classification systems (159, 160). Sarkizova et al. performed extensive MS-based analyses to identify peptides from >90 different HLA molecules, and identified several sub-motifs that were characteristic of peptides derived from individually distinct clusters of HLA molecules (141). However, these sub-motif characterizations have generally focused on the dominant primary anchors found at p2 and pΩ, and the conventional HLA-I binding pockets found deep within the binding groove. The lack of attention to subdominant, non-pocket HLA-I/peptide interactions may be one major reason that such a p4-specific pan-HLA anchor has
not been previously described. Here, we showed that such non-pocket interactions involving HLA residues p65-66 play a pivotal role in driving diversity of peptide antigen presentation, and may be conserved across HLA-A, -B, and C molecules.

The average frequency of D/E4 peptides in HLA-I molecules without either R65 or K66 is ~20%, which is still significantly higher than the background D/E frequency of ~11% (see Fig 4.4D). This may indicate that negatively charged p4 amino acids are preferred over others due to other interactions independent of R65 and K66. The location of position 4 at the top of the binding groove in most HLA-I/peptide complexes allows D and E side chains to “roam” over large regions of the peptide binding groove, potentially making interactions with both alpha helices and with other peptide residues. In multiple MD simulations, E4 formed hydrogen bonds and salt bridges with N66, a polar amino acid. HLA-B conserves a different polar amino acid, Q65, which could also conceivably interact with D4 or E4. Most HLA-B alleles also contain an arginine at position 62, another residue which could potentially interact well with D/E4. Other polar amino acids on the opposite alpha-2 helix are within range for interactions with D4 and E4 side chains, and these residues are very highly conserved across HLA-A, B, and C molecules. While further studies will be needed to demonstrate the contributions of these alternate interactions to peptide binding and stability, it is clear that peptides containing D/E4 have opportunities for interactions with multiple surrounding residues in the binding groove.

The finding that several animal MHC-I molecules also showed a preference for peptides with D/E4 and contain R65 (and sometimes K66) is quite fascinating, and suggests that non-pocket interactions between D/E4 and R65/K66 might be
It is plausible that as MHC-I molecules changed and evolved across species that those molecules which conserved the ability to form stable interactions in the middle of the peptide retained an evolutionary advantage over MHC-I molecules that could not. This would potentially allow such alleles to outcompete other alleles and remain within populations over the long term.

**HLA-I molecules show a preference for positively charged amino acids in peptide position 1**

**4.7 K/R1 peptides are prevalent in HLA-I molecules lacking K66, R62 or R163**

Perhaps the most surprising discovery of this thesis project was the substantial changes observed in p1 of peptides eluted from the A*0201 mutants bearing the K66N mutation. Lysine, arginine and histidine were highly favored at this position, along with a concurrent decrease in all other p1 amino acids (see position 1 of Fig 2.6). While A24-K66N did not exhibit nearly as dramatic a phenotype as A2-K66N and A2-DM, there was an observed increase in lysine at p1 compared to A24-WT (see Fig 3.5). We next explored if other HLA-I allotypes showed a preference for positively charged amino acids at peptide p1. Analysis of eluted peptide ligands from multiple HLA-I databases revealed that K/R1 are present in ~22% of peptides overall, which is higher than the expected background frequency of ~14% and higher than the percentage of K/R at all other peptide positions (see Fig 1.6).
Based on our results, we hypothesized that HLA-I molecules lacking K66 would bind to more K/R1 peptides than HLA-I molecules with K66. Initial analysis showed that some HLA-I molecules lacking K66 showed an increased percentage of K/R1 peptides, but the majority of them did not. Closer examination of the peptide binding groove revealed that arginine in positions 62 and 163 (R62 and R163) are located in close proximity to peptide position 1 and thus could be repulsing positively charged amino acids in position 1. To assess this hypothesis, HLA-A allotypes were sorted based on whether they contained R or K in positions 62, 66 and/or 163 and the frequency of bound K/R1 peptides was compared between groups. Peptides from HLA-A molecules that lacked either arginine or lysine in p62, 66 and 163 showed a mean frequency of ~50% K/R1 peptides, with one HLA-A allotype being as high as 75%, similar to the percentage of K/R1 peptides eluted from A2-K66N and A2-DM (Fig 4.6A). HLA-A molecules containing only K66 averaged ~20% K/R1 peptides, which is close to what was observed in wild-type A*2402 and A*0201. Peptides from HLA molecules containing only R163 showed a further decrease in the frequency of K/R1 peptides to ~10%, while peptides eluted from HLAs containing R62 with or without K66 or R163 decreased even further to 2% to 6% K/R1 peptides (Fig 4.6A).

Consistent with this, in HLA-B only 4 allotypes lack R62 and they all show elevated frequencies of ~50% K/R1 peptides (Fig 4.6B). Conversely, the majority of HLA-B molecules with R62 showed significantly less K/R1 peptides (mean ~18%). Interestingly, HLA-B molecules containing both R62 and E163 showed ~35% K/R1 peptides, potentially due to favorable interactions with E163. All but two
K/R1 peptides are preferred in HLA-I molecules without R62, K66 and R163

Figure 4.6: HLA-I molecules lacking R62, R163 and K66 prefer K/R residues in peptide position 1
Figure 4.6: HLA-I molecules lacking R62, R163 and K66 prefer K/R residues in peptide position 1: (A-C) HLA-I molecules were separated according to expression of amino acids in positions 62, 66 and 163 and the percent of eluted peptides containing K/R in peptide position 1 were determined for (A) HLA-A, (B) HLA-B and (C) HLA-C and HLA-G. Brown-Forsythe ANOVA test showed significant difference between means ((A): p<0.0001; (B) p<0.001; (C) p<0.001) and multiple comparisons using Dunnett’s T3 test were performed to identify significant differences between specific groups.* indicates p<0.05, ** indicates p<0.01, *** indicates p<0.001, **** indicates p<0.0001. (D) All possible conformations for R62, R163, K66 and K1 in A*0201 were simulated and overlayed on top of each other to determine the overlap between K in peptide position 1 and the three HLA residues.* (E) The percent of D/E4 in peptides with K/R or other amino acids in position 1 was determined as well as the percent of K/R1 in peptides with D/E or other amino acids in position 4 was calculated to determine whether K/R1 co-occurs with D/E4. **** indicates p<0.0001 using unpaired t-tests with Welch’s correction.

*Computer simulations were run in collaboration with Dr. Dinler Antunes of the University of Houston
HLA-C allotypes contain both R62 and K66, and these molecules bind K/R1 peptides at frequencies ranging from ~3% to ~15% or less K/R1 peptides (Fig 4.6C). Remarkably, both HLA-C allotypes containing a K66N substitution bound K/R1 peptides at a frequency of ~45%. Comparison of HLA-C*0701 and C*0702 was particularly revealing in this regard, since these allotypes differ only at HLA-I residues 66 and 99. C*0702 (K66) bound ~15% K/R1 peptides, while C*0701 (N66) bound ~50% K/R1 peptides. The S99/Y99 difference is located at the floor of the binding groove and is not expected to affect large, basic p1 residues like K or R, which extend upwards to near the top of the binding groove. HLA-G is a non-classical allele that presents a limited number of peptides to the immune system. None of the 3 HLA-G allotypes in the peptide database contains R62, K66 or R163, and peptides eluted from these molecules bind to ~70% K/R1 peptide on average (Fig 4.6C). Collectively, this analysis reveals a strong negative correlation between the occurrence of positively charged residues in peptide p1 and HLA-I residues R62, K66, and R163 across all HLA-A, B, C, and G allotypes.

To examine why R62, K66 and R163 have such a large impact on the frequency of K/R1 peptides, we used MD simulations to examine the peptide binding groove in more detail. We analyzed all possible movements of K1, R62, R163 and K66 side chains and overlaid them onto the same A*0201 backbone structure. From this analysis, it was readily apparent that all three HLA-I residues can partially overlap in the same space as K1, with R62 and R163 overlapping with K1 the most (Fig 4.6D). This is consistent with our observations from multiple HLA-I molecules where the lowest frequencies of K/R1 peptides were found in allotypes with R62 and R163.
Our analysis of eluted A*0201 peptides showed that K/R1 and D/E4 were more likely to be present within the same peptide in wild-type and especially in mutated A2 HLA molecules. We next explored if similar co-occurrences existed in peptides eluted from other HLA-I molecules. Thus, the percentage of D/E4 residues was analyzed in peptides with or without K/R1, which revealed that K/R1 peptides are more likely to contain D/E in position 4 than other peptides (Fig 4.6E). Reciprocal analysis of the percentage of K/R1 in peptides with or without D/E4 also showed that D/E4 peptides are more likely to contain K/R1. These data indicate a significant co-occurrence between K/R1 and D/E4 in multiple HLA-I molecules, and supports our structural analyses showing the potential for charged interactions between these intra-peptide residues.

4.8 Discussion

HLA-I molecules contain closed binding grooves, meaning that peptide amino acids cannot extend from the end of the binding groove. This physical barrier dictates that peptide position 1 amino acids with longer side chains, such as lysine and arginine, must position themselves in an upright orientation towards the top of the binding groove. This is why K1 and R1 residues, although three positions away from D/E4, can directly interact with p4 residues and also be affected by interactions with HLA-I position 66 residues. The high prevalence of K/R1 residues in peptides eluted from multiple HLA molecules suggests a non-pocket interaction that may represent another subdominant pan-HLA-I anchor. While the occurrence of K/R1 is more restricted than D/E4 (likely due to interference from R62, R163 and K66 in
many HLA-I allotypes), it also occurs at a much higher frequency in certain HLA-I allotypes (>70%) compared to the maximum observed D/E4 frequencies (~45%).

The fact that multiple HLA residues could directly affect the frequency of K/R1 peptides was intriguing. The impact of specific HLA-I residues in pocket interactions has been well studied in multiple HLA-I molecules, and demonstrate that a slight change in size, biochemical properties or space can radically alter the binding motifs of different pockets (161-164). However, our studies showed that multiple HLA-I residues could have a significant impact on a single peptide position. The presence of multiple positive charges in close proximity to each other is not conducive to stable peptide-HLA interactions and explains why K/R1 peptides are less likely to bind to HLA-I allotypes containing R62 or R163. Conversely, removal of the spatial restrictions in the space above peptide position 1 allows for a dramatic increase of peptides with K/R1, indicating that these amino acids must have a significant competitive advantage over other residues. Although we observed some interactions with D/E4 in A*0201 MD simulations, it is likely that K/R1 also forms stable interactions with other surrounding residues.

The fact that K/R1 and D/E4 frequently co-occur together across the HLA-I system was also a unique observation. All co-occurrences identified to date are HLA allotype-specific or peptide-specific, as they either involve specific interactions within pockets or interactions between non-conserved peptide residues (41, 133, 134, 144). However, since K/R1 and D/E4 interact across the top of the peptide binding groove, pocket differences are likely to be irrelevant. Taken together, these studies
suggest not only that pan-HLA co-occurrences are possible, but may occur independently of specific, conventional pocket interactions.
CHAPTER V: OVERALL CONCLUSIONS AND FUTURE DIRECTIONS
5.1 Overall Conclusions

Cytotoxic T-lymphocytes (CTLs) are the branch of the adaptive immune system tasked with eliminating intracellular pathogens and transformed cells (3). In the thymus, each T-cell receptor (TCR) is positively selected for recognition of peptide-HLA-I molecules or deleted if they react too strongly or too weakly with self peptide-HLA-I complexes. Naïve CTLs then circulate through lymph nodes and other immune organs, waiting for signals from the innate immune system and professional antigen-presenting cells (pAPCs) to be activated. Once activated by HLA-I peptide complexes of pAPCs, CTLs proliferate rapidly and disseminate throughout the body, scanning the surfaces of somatic cells to find those that express the specific peptide-HLA-I complex that their TCR recognizes. Once found, the CTL can lyse the cell using perforin and granzyme B (8). Thus, the ability of nucleated cells to bind to and present peptides from viral or other foreign proteins in the context of HLA-I to CTLs is of immense importance for immune responses and overall survival.

HLA-I/peptide molecular interactions are a vital component of antigen presentation, as demonstrated by the mutations made to HLA p65-66 residues in A*0201 and A*2402. Mutations that removed the conserved, positively charged residues R65 and/or K66 of A*0201 and A*2402 negatively impacted antigen presentation in both allotypes by restricting the number of peptides capable of interacting with the mutated HLA molecules, with accompanying concurrent decreases in cell surface expression. Conversely, replacing G65 with R65 in A*2402 improved antigen presentation by increasing cell surface expression and the number
of unique peptides able to bind to the A24-G56R molecule. Thus, R65 and K66 both play a significant role in antigen presentation in both HLA-A*0201 and HLA-A*2402.

Changing residues in the HLA peptide binding groove has long been known to alter peptide binding (161-164). However, previous studies focused on the hypervariable residues found in pockets on the floor of the binding groove or at the base of the alpha helices. Our study was unique in focusing on R65 and K66, which are not part of any pocket as they located at the top and inner side of the alpha-1 helix, respectively. However, we demonstrated that these non-pocket residues also played an important role in peptide binding and stability.

The orientation of D/E4 and HLA p65-66 are such that they are in a position to be recognized by TCRs. Investigation of TCR interactions with peptide-HLA complexes revealed that TCRs can directly interface with peptide p4 residues and these interactions may drive the immune response in some instances (165-167). Li et al. noted that peptide position 4 is the most important peptide position for predicting the immunogenicity of peptides involving both peptide-HLA and peptide-TCR interactions (168). Based on our studies, we propose that negatively charged amino acids in position 4 may play a dual role, being not only visible to TCRs but also crucial for facilitating HLA-I/peptide interactions.

Several studies have demonstrated that R65 and K66 can also be recognized by TCRs (146, 169, 170). Wang et al. showed that mutation of R65 or K66 disrupted TCR recognition of a specific peptide-HLA-A*0201 complex (171). Another study observed that K66 interacts with bound peptides and investigated whether the change in TCR recognition after mutating K66 to A66 was due to altered peptide
binding or abrogated TCR recognition (146). They determined that both were simultaneously occurring and posited that K66 may constitute a hybrid TCR-anchor residue in A*0201 (146). Our studies corroborate the importance of K66 in peptide binding and antigen presentation and revealed the importance of R65 as well. Thus, we propose that R65 and K66 both constitute hybrid anchor-TCR residues in A*0201 and A*2402.

In Chapter 4, the possibility that D/E4 and R65/K66 are evolutionarily conserved across multiple species was discussed. The fact that both D/E4 and R65/K66 play distinct roles in TCR recognition suggests some potential reasons for the conservation of both the peptide and MHC-I residues. Negative and positive charges on the surface of the peptide binding groove are readily visible to TCRs and likely play a key role in sustaining interactions between these receptors (166, 167, 169). Thus, having adjacent positive and negative charges at HLA p65-66 and peptide p4 may be advantageous for TCR recognition of MHC-I/peptide complexes, which would imply that animals expressing these HLA alleles might be more likely to survive encounters with infectious pathogens and propagate. While highly speculative, it is plausible that the importance of peptide and HLA residues for both antigen presentation and TCR binding is the reason these amino acids are evolutionarily conserved.

Improved understanding of interactions between antigenic peptides and HLA-I molecules can help design more effective immunotherapies, including facilitating the identification of peptide targets for cancer therapies (172). The implications of this study for identifying peptide targets are simple: prioritize peptides that contain
aspartic or glutamic acid residues at p4 in the creation of peptide vaccines or development of CTL-based therapies for A*0201 and A*2402 patients. D/E4 peptides with K/R1 may be even more beneficial as peptides with these amino acids are more likely to be stable in the peptide binding groove and thus may be more likely to be immunogenic than other peptides with similar predicted binding affinities. Furthermore, for a specific subset of HLA-I molecules peptides with K1 or R1 are also likely to be better binders or more stable in the binding groove, regardless of position 4. The insights provided by these studies will also potentially be very useful to guide peptide prediction and selection for patient HLA-I molecules that have not been well-studied, of which there are several thousands of alleles in the human population. For example, knowing the identity of the amino acids found at p62, p63, p65, p66, and p163 of uncharacterized HLA-I allotypes will now allow for improved prediction of peptide-binding preferences, and hopefully improved efficacy of personalized CTL-based immunotherapies.

5.2 Future directions

Since interactions with D/E in peptide position 4 and HLA-I R65/K66 were suggested to be relevant for other HLA-A, B and C peptide-HLA complexes, future studies could be performed to confirm that this interaction is important functionally. In HLA-A molecules, D/E4 would likely interact preferably with R65 while in HLA-C molecules, K66 would likely make the primary interaction with D/E4. Thus, it would be informative to create similar or complementary mutations in other HLA-I molecules: for example, N66K and R65G mutants of HLA-A allotypes, or Q65R and K66N for HLA-C molecules. These mutants would contain zero, one or two positively
charged amino acids in p65/66 that could then be analyzed as A*0201 and A*2402 were in this dissertation.

HLA-B molecules also show a preference for D/E4 peptides even though most allotypes lack both R65 and K66. It may be that D/E4 residues interact with Q65 or possibly other residues in the HLA-B binding groove. Mutant HLA-B molecules could be made at position 65 (Q65R or Q65G) and position 66 (I66K, I66N) with mutants containing full positive charges, partial charges with polar amino acids or no charges to examine these interactions. Alzipar et al. found that four HLA-B allotypes demonstrated a preference for binding peptides phosphorylated at position 4, and a crystal structure of HLA-B*4001 showed an interaction between the negatively charged phosphate moiety and an arginine at position 62 (173). HLA-B p62 mutants could be made by substituting amino acids found in p62 of other HLA-I molecules (R62Q, R62G) to examine the loss of charge on peptide stability and the immunopeptidome. Computational simulations in A*0201 and A*2402 also showed hints of interactions with HLA residues on the opposite alpha helix including glutamine in position 155, whose role could be computationally investigated in further studies since it is conserved in almost all HLA-A, -B, and -C molecules.

Further examination of the importance of HLA-I p65/66 and D/E4 for immunogenicity of peptide targets is highly warranted. This can be examined in several different ways. Known immunogenic peptides in A*0201 and other HLA-I molecules can be analyzed for the percentage of peptides containing D/E in position 4, especially with known dominant or weaker p2 and pΩ anchors. Computational models and simulations can be performed on crystal structures of TCRs complexed
with A*0201 molecules bound to D/E4 peptides to analyze whether D/E4 and HLA p65-66 interactions are still occurring and whether or how these residues are recognized by the TCR. *In vitro* tests of immunogenicity can be performed using predicted or MS-identified peptide sequences from a known tumor or viral antigen, separated into D/E4 and non-D/E4 peptides, pulsed onto target cells (K562 transduced with HLA-I molecules or cancer cells with appropriate HLA-I molecules) and incubated with CTLs known to target this tumor or viral antigen. We would expect to see increased killing of target cells pulsed with D/E4 peptides as opposed to non-D/E4 peptides. These peptide-pulsed K562 cells could also be used to stimulate PBMC in order to assess the relative immunogenicity of the p4-substituted peptides. This could also be done with mutated HLA-I molecules to examine their impact of specific positions on immunogenicity. Multiple other assays could be done such as ELISPOT to quantitate the extent of T-cell activation, TCR-sequencing of T-cells primed by D/E4 peptides to assess the breadth and strength of elicited TCRs, and assessment of the ability of D/E4 peptides to generate T-cells for use in endogenous T-cell therapy.

Since some murine MHC-I allotypes also showed an increased prevalence of D/E4 peptides, *in vivo* models could be utilized to examine D/E4 and MHC p65/66 interactions. Mice bearing tumors with specific antigens could be vaccinated using peptide targets with or without D/E4 and another immune therapy such as checkpoint blockade to determine whether D/E4 peptides caused increased tumor clearance. Tumor cells could be transduced with wild-type or mutated MHC molecules, implanted into mice and then challenged with different immune therapies
to examine how mutated MHC molecules affect tumor clearance, although allo-specific immune responses generated against the MHC-I mutations could create challenges in interpreting the data. To avoid these issues, transgenic models of mice expressing mutated MHC molecules could be developed to examine the general impact of these mutations on T-cell mediated immunity and the ability to clear tumors or viral infections.

Analysis of the immunopeptidome of A*2402 G65R indicated a substantial increase in peptide repertoire diversity, but few alterations in peptide motifs. This could possibly be exploited to enable increased sensitivity of identification of peptide sequences for specific tumor or viral antigens in A*2402. The tumor or viral antigen of interest would be transduced along with A*2402 G65R in K562 or an A*2402-negative cancer cell line, and cells would be grown up to large numbers for MS analysis. We would expect to see increased tumor or viral antigen sequences as well as increased A*2402 peptides which could enable enhanced identification of new therapeutic peptide targets. Since the K66N mutation induced major alterations of peptide motifs in A*0201, it would not be advisable to use in this fashion. However, it is possible that HLA-B and -C molecules, the majority of which do not contain R65, could be augmented in a similar fashion. This should increase identification of peptide targets from these allotypes, which represents an unmet need since fewer HLA-B and C peptides are identified by MS compared to HLA-A peptides. However, peptide-binding assays would need to be performed to ensure no major differences in binding of identified peptides was observed between the R65 mutant compared to the wild-type HLA-I.
It is very likely that other significant non-pocket interactions occur throughout the peptide binding groove in multiple HLA-I molecules. We observed some slight preferences for aromatic amino acids in position 3 and positively charged amino acids in position -1 that could potentially be investigated in future studies. Other non-pocket peptide-HLA interactions may be more prevalent in specific amino acids or different HLA-I families. The prevalence of proline in peptide position 4 was not addressed experimentally in these studies but is likely to be important for multiple HLA-I allotypes. Studies could also be done on specific HLA-I families such as A1, A3, B7 or others to identify more specific non-pocket interactions. The fact that we discovered novel interactions in HLA-A*0201, the first identified and most studied HLA-I molecule, indicates that there are numerous other discoveries yet to be made within the greater HLA-I system. Any amino acid found above expected background frequencies could be analyzed as described in this dissertation.

It is worth mentioning that unraveling the interactions between D/E in peptide position 4 and R65/K66 in A*0201 and A*2402 would not have been possible without computational simulations, *in vitro* experiments and proteomics analysis working in tandem with each other. All three approaches were needed to provide different facets of the project data, and all were crucial for understanding these interactions: making direct mutations of the HLA-I molecules were the critical first steps in determining that R65 and K66 impacted antigen presentation. MS-based immunopeptidome analysis was necessary to characterize peptide-binding motif changes of mutated HLA-I molecules. *In vitro* assays were needed for quantitative determination of the relevance of D/E4 and K/R1 residues in peptide-HLA binding.
and stability. Computational MD simulations were crucial for understanding the molecular interactions that drove the observations made in the wet-bench experiments. The strengths and weaknesses of each technique complemented each other and drove the development of this project.

This powerful pipeline, incorporating global immunopeptidome investigation, quantitative immune assays and detailed molecular analysis could be used to investigate several other basic or translational questions surrounding HLA-I/peptide interactions. For example, to identify TCR cross-reactivity of different peptide-HLA complexes a database of peptides eluted from highly expressed genes could be created and a program generated that would compare the HLA-I/peptide surface of the peptide target of interest against peptide-HLA surfaces from same HLA-I molecule in the database of eluted peptides. Top matches could be validated experimentally for peptide-HLA binding and stability as well as TCR recognition. To investigate alterations in peptide presentation of cancer cells, immunopeptidomes from tumors and matched normal tissue could be compared against each other. Any observed mutations in HLA-I molecules or other proteins involved in the antigen presentation pathway could be computationally modeled and compared with observed immunopeptidome alterations. These and several other questions about HLA-I/peptide interactions can be answered using this unique combinatorial experimental approach.
CHAPTER VI: MATERIALS AND METHODS
Cell lines and generation of HLA-A*0201 and HLA-A*2402 mutant alleles

H1975 cells were purchased from ATCC (CRL-5908). Site-directed mutagenesis using the QuikChange II site-directed mutagenesis kit (Agilent, 200521) was performed on HLA-A*0201 and HLA-A*2402 sub-cloned into pDONR221 (ThermoFisher Scientific, 12536017). Creation of A*0201 mutants (R65G, K66N and R65G/K66N) and A*2402 mutants (G65R, K66N) were confirmed using Sanger sequencing (MDACC DNA Analysis Core Facility). A*0201 and A*2402 wild-type and mutant alleles were sub-cloned into destination vector pLV460 using the LR reaction (ThermoFisher Scientific, 11791020).

Lentiviral production and transduction

293METR packaging cell line was a kind gift from Dr. Amer Najjar (Assistant Professor, Department of Pediatrics, UT-MDACC, Houston, TX). 293METR cells were plated in 10cm plates (Corning, 430167) and transfected with 10 µg pCMVdelta 8.91, 1µg VSV-g, 10 µg pLV460 with A*0201 or A*2402 and 50ul lipofectamine 2000 (ThermoFisher Scientific, 11668019). Viral supernatant was collected after 2 and 3 days, centrifuged and concentrated using Lenti-X (Takara BioSciences, 631232). H1975 was transduced with wild-type or mutated HLA-A*0201 and HLA-A*2402 lentivirus in 4 ug/ml polybrene (Millipore Sigma, TR-1003-G) and expression was analyzed using HLA-A*2402 (17A10, MBL International, K0208-3) or HLA-A*0201 (BB7.2, BioLegend, 343308) antibodies by flow cytometry (BD FACS Canto II).
Eluting peptides from HLA molecules

Cell lines were eluted as described previously (76). In brief, H1975 wild-type and transduced cell lines were expanded in 100 cm plates (Corning, 430599), lysed using Triton X-100 (Sigma Aldrich, T9284), and incubated overnight at 4 °C under gentle agitation with 50 µg of HLA-I-specific antibody W6/32 from hybridoma cell line (ATCC HB-95) for every 10mg of protein. HLA molecules were immunoprecipitated by using Protein A/G UltraLink resin beads (ThermoFisher Scientific, 53133) and then directly eluted on columns (ThermoFisher Scientific, 89897) using 0.1 N acetic acid (Fisher Scientific, BP2401). HLA-I purification was confirmed by Western blot and elutes were pooled for continued analysis.

LC-MS/MS of eluted peptides

Analysis of peptides eluted from HLA-I molecules was done as described previously (76). In brief, peptide eluted were performed with high sensitivity LC-MS/MS which used a Nano-cap LC system hooked up to tribid MS. Chromatography was performed in a 15cm column packed with 2.6 micron XB-C18 particles using 0.1% formic acid in water as solvent A and 0.1% formic acid in acetonitrile as solvent B. MS/MS was performed using the linear ion trap and ESI where the full MS1 range was 265-1500 m/z for singly charged precursors, 400-750 m/z for doubly charged precursors and 270-500 m/z for triply charged precursors. Spectra acquired by LC-MS/MS was analyzed using the MASCOT algorithm to search spectra against the SWISSPROT human proteome database. Peptides eluted from HLA-I molecules on H1975 cells were split in half with one half analyzed by the Proteomics Facility at
U.T. MD Anderson Cancer Center and the other half were analyzed by the ThermoFisher facility in La Jolla, CA.

Multiple parameters were utilized for creation of a ‘reality score’ for each peptide sequence. At the proteomics level, Mascot ion score, the difference between observed and calculated peptide mass at MS1 (delta mass) and number of matched ions were the main parameters used for quality control. Differences between observed HPLC retention time and predicted retention time were also utilized. Predicted binding to transduced and endogenous HLA-I molecules via NetMHCpan4.1 was another parameter. All of these factors were weighted differently to give a reality score with 100 being the highest possible score. Peptide sequences with the highest reality score and that had a Mascot ion score above 25 were used for immunopeptidome analysis.

**Separation of A*0201 and A*2402 peptides from endogenous alleles**

To separate A*0201 and A*2402 peptides from other peptides from endogenous HLA alleles, dual anchor clusters were determined from p1-3 of the N-terminus and pΩ of the C-terminus from other HLA-I peptide databases containing A*0201, A*2402 and endogenous alleles A*0101, B*4101 and C*1701. Peptides containing dual anchor clusters from endogenous alleles were removed.

**A*0201 Peptide synthesis and peptide binding assays**

All A*0201 peptides used in the study were purchased from Genscript at purity of 70% with quality control by reverse-phase HPLC and mass spectrometry (SC1208). Recombinant HLA-A*0201 and HLA-A*2402 molecules and were made and re-
folded as described elsewhere (174). Conformation-dependent antibody W6/32 antibody used in these assays were produced in vitro from hybridoma cell lines (Integra, Chur, Switzerland). All A*0201 binding assays were done as described previously (174). In brief, streptavidin donor beads (PerkinElmer, 6760002) and acceptor beads (PerkinElmer, 6762001) conjugated to W6/32 were incubated with A*0201 that had been pre-incubated with varying peptide concentrations. Plates were read by PerkinElmer EnVision and data was analyzed using GraphPad Prism.

A*2402 Peptide synthesis and A*2402 peptide binding assays

Peptides used for A*2402 binding assays were synthesized using MultiPep RSi synthesizer. After synthesis, peptides were cleaved from resin using trifluoracetic acid and then desalted using C18 columns. Peptide quality was checked using 4700 proteomics analyzer mass spectrometer using MALDI-TOF. Peptides were then dried and reconstituted for use in competitive fluorescent peptide assays.

Competitive fluorescent peptide assays were performed as described previously (152). In brief, H1975 cells expressing high levels of A*2402 were used along with a known A*2402 binder with FITC attached to position 5 (RYLKXQQLL where X is the fluorescent residue). Test peptides were incubated with fluorescent peptides at increasing concentrations overnight and fluorescence was measured using flow cytometry. Percent inhibition was calculated and used as data points in GraphPad prism.

A*0201 and A*2402 stability assays
All peptides used for stability assays were purchased from Genscript at purity of 70% with quality control by reverse-phase HPLC and mass spectrometry (SC1208). Peptide stability assays were performed as described previously (175). In brief, biotinylated A*0201 and A*2402 molecules were refolded with excess peptide then incubated with 0, 2, 4 and 6 M urea. Non-denatured peptides were detected by W6/32 using ELISA as described previously (176). Stability was calculated as a percentage relative to the reference peptide with a stability index of 100%. LLFGYPVYV (HTLV Tax peptide) was used as a reference peptide for both A*0201 assays and TYTQDFNKF was used as a reference peptide in both A*2402 assays (177-179).

**Statistical analysis**

Differences between wild-type and mutated frequencies in the immunopeptidome were analyzed using the sample test for equality of proportions without continuity correction. Comparisons of two means with unequal variances were analyzed using unpaired t-tests with Welch’s correction. Comparisons of multiple means with unequal variances were performed using Brown-Forsythe ANOVA and multiple pairwise comparisons were then performed using Dunnett’s T3 test.
Bibliography


145. R. C. Hillig, P. G. Coulie, V. Stroobant, W. Saenger, A. Ziegler, M. Hülsmeyer, High-resolution structure of HLA-A*0201 in complex with a tumour-specific


156. J. Xiao, W. Xiang, Y. Chai, J. Haywood, J. Qi, L. Ba, P. Qi, M. Wang, J. Liu, G. F. Gao, Diversified Anchoring Features the Peptide Presentation of DLA-


Kyle Ross Jackson was born in Moscow, Idaho, the son of Brian Jackson and Daphne Jackson. He earned his high school diploma from Ledyard High School in 2006 in Ledyard, Connecticut. He then attended Brigham Young University in Provo, Utah and earned a Bachelor of Science with a major in microbiology. In August of 2013 he joined the Immunology Program in The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences.

**Permanent Address:**

112 E Cornell Dr

Meridian, ID 83646