Molecular mechanisms underlying pMHC-II recognition

A thesis submitted to Cardiff University

in candidature for the degree of

Doctor of Philosophy

Andrea Schauenburg

October 2016

DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed ……………………………………………. (candidate)   Date ……..………

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD

Signed ……………………………………….……… (candidate)   Date ……………

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated, and the thesis has not been edited by a third party beyond what is permitted by Cardiff University's Policy on the Use of Third Party Editors by Research Degree Students. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed ……………………………………….…..… (candidate)   Date ……………

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ………………………………………..…….. (candidate)   Date ……………

Summary

The immune system is a complex network of cells and molecules working together with the purpose of fending off potentially harmful pathogens. CD4$^+$ T cells take key roles within this network by orchestrating a multitude of its players. They recognise pathogen or self-derived peptides (p) bound to molecules of the major histocompatibility class II (MHC-II) through their T cell receptor (TCR). Cytokines secreted in response to recognition aid antibody production and cytotoxic T cell activity, both critical for anti-viral immunity. In this thesis, TCR/pMHC-II interactions were investigated using a range of functional and molecular approaches in order to gain valuable insight into the mechanisms underlying successful antigen recognition.

To aid these investigations, a versatile, insect cell based expression system for HLA-DR1 was successfully implemented to generate soluble protein for use in multimer stainings and biophysical assays. Two HLA-DR1 restricted peptides encoded within influenza heamagglutinin (HA) were confirmed as being highly conserved making them ideal targets for vaccine development and allowing identification of influenza specific CD4$^+$ T cells.

Furthermore, the various roles of peptide flanking residues (PFR) were investigated using two experimental models. In a HA derived peptide, C-terminal PFR proved essential for peptide binding stability to HLA-DR1 and in consequence, CD4$^+$ T cell activation. Clonotyping of CD4$^+$ T cells grown against peptides of varying PFR lengths showed that TCR gene selection was heavily influenced by PFR. A HIV gag24 derived peptide displaying an unusual secondary structure within its N-terminal PFR gave further insight into how seemingly small modifications to PFR can have wide reaching impact on CD4$^+$ T cell activation. Both studies highlighted the need for more in depths investigations into this emerging field and the wide reaching impacts of this inherent feature of MHC-II restricted peptides.

Overall, the results in this thesis added novel insight into the mechanisms underlying TCR/pMHC-II interactions.

Acknowledgments

First and foremost, I would like to thank Dr. David Cole for taking me on as a Research Assistant and for making it possible for me to pursue a PhD alongside (including taking two months off for writing). Without his support and advice this PhD would not have happened.

I would also like to thank my co-supervisor Prof. Andy Sewell for his support and input throughout my thesis. I am grateful to Prof. Andy Godkin for his supervision on the epitope mapping project. Further thanks go to Pierre Rizkallah for sharing his knowledge on crystallography and to Dr. Meriem Attaf who guided me through the clonotyping. A special thanks goes to Dr. Garry Dolton for his help and advice on T cell culture and his good humour (and for gerbil-sitting!). Further thanks go to Anna Fuller who was always there to help and Andreas for his advice on protein expression in insect cells. I would also like to thank Prof. John Miles who encouraged me in my decision to pursue a PhD.

I would like to thank every past and present member of the T-cell modulation group who made my time in the lab unforgettable. Thanks go to Bruce, Alex, Katie T (thanks for proof reading!), Angharad L, Debbie and everyone else I have shared an office or TC with over the last seven years. Special thanks goes to my friends inside and outside the lab who always had my back and kept me sane, in person and/or via the internet: Katharina, Joh, Val, Aileen and Anna Rita.

I am deeply grateful to my parents for their continuing support, encouraging me in my choices and always asking how "my littles ones" were doing. Thanks also go to the Clemitson family for being my family away from my own. I particularly want to thank Andrew, for his love and support, especially over the last few months. You're the best travel partner I could wish for.

Publications based on data presented in this thesis

Identifying conserved HLA-DR1 restricted epitopes within internal and external H3N2 influenza A proteins.
Schauenburg AJ, Greenshields Watson A, Dolton G, MacLachlan B, Sewell AK, Godkin AJ, Cole DK
Under preparation

Peptide flanking regions influence CD4+ T cell activation, TCR binding and TCR gene selection
Schauenburg AJ, Dolton G, Greenshields Watson A, Attaf M, Sewell AK, Godkin AJ, Cole DK
Under preparation

Poster and oral Presentations based on data presented in this thesis:

Oral presentation at the weekly **Institute Seminar** April 2014
Oral presentation at the **South West Structural Biology Consortium meeting** 2013

Poster presentation at the **International Congress for Immunology** 2013
Poster presentations at **BSI congresses** 2011, 2012 & 2013
Poster presentations at the **South West Structural Biology Consortium meetings** 2011, 2012, 2013 & 2014
Poster presentations at the **Annual Institute meetings** 2011, 2012, 2013 & 2014

Other publications:

Using X-ray crystallography, biophysics and functional assays to determine the mechanisms governing T cell receptor recognition of cancer antigens
*MacLachlan B, Greenshields Watson A, **Schauenburg AJ**, Mason G, Bianchi V, Fuller A, Rizkallah PJ, Sewell AK, Cole DK*
JoVE 2016, in press

Hotspot autoimmune T cell receptor binding underlies pathogen and insulin peptide cross-reactivity.
*Cole DK, Bulek AM, Dolton G, **Schauenberg AJ**, Szomolay B, Rittase W, Trimby A, Jothikumar P, Fuller A, Skowera A, Rossjohn J, Zhu C, Miles JJ, Peakman M, Wooldridge L, Rizkallah PJ, Sewell AK.*
J Clin Invest. 2016 Jun

Identification of human viral protein-derived ligands recognized by individual MHCI-restricted T-cell receptors.
*Szomolay B, Liu J, Brown PE, Miles JJ, Clement M, Llewellyn-Lacey S, Dolton G, Ekeruche-Makinde J, Lissina A, **Schauenburg AJ**, Sewell AK, Burrows SR, Roederer M, Price DA, Wooldridge L, van den Berg HA.*
Immunol Cell Biol. 2016 Jul

Enhanced Detection of Antigen-Specific CD4$^+$ T Cells Using Altered Peptide Flanking Residue Peptide-MHC Class II Multimers.
*Holland CJ, Dolton G, Scurr M, Ladell K, **Schauenburg AJ**, Miners K, Madura F, Sewell AK, Price DA, Cole DK, Godkin AJ.*
J Immunol. 2015 Dec 15

A molecular switch in immunodominant HIV-1-specific CD8 T-cell epitopes shapes differential HLA-restricted escape.
*Kløverpris HN, Cole DK, Fuller A, Carlson J, Beck K, **Schauenburg AJ**, Rizkallah PJ, Buus S, Sewell AK, Goulder P.*
Retrovirology. 2015 Feb

Naive CD8$^+$ T-cell precursors display structured TCR repertoires and composite antigen-driven selection dynamics.
*Neller MA, Ladell K, McLaren JE, Matthews KK, Gostick E, Pentier JM, Dolton G, **Schauenburg AJ**, Koning D, Fontaine Costa AI, Watkins TS, Venturi V, Smith C, Khanna R, Miners K, Clement M, Wooldridge L, Cole DK, van Baarle D, Sewell AK, Burrows SR, Price DA, Miles JJ.*
Immunol Cell Biol. 2015 Aug

Structural basis for ineffective T-cell responses to MHC anchor residue-improved "heteroclitic" peptides.
*Madura F, Rizkallah PJ, Holland CJ, Fuller A, Bulek A, Godkin AJ, **Schauenburg AJ**, Cole DK, Sewell AK.*
Eur J Immunol. 2015 Feb

Antibody stabilization of peptide-MHC multimers reveals functional T cells bearing extremely low-affinity TCRs.
*Tungatt K, Bianchi V, Crowther MD, Powell WE, **Schauenburg AJ**, Trimby A, Donia M, Miles JJ, Holland CJ, Cole DK, Godkin AJ, Peakman M, Straten PT, Svane IM, Sewell AK, Dolton G.*
J Immunol. 2015 Jan

T-cell receptor (TCR)-peptide specificity overrides affinity-enhancing TCR-major histocompatibility complex interactions.
*Cole DK, Miles KM, Madura F, Holland CJ, **Schauenburg AJ**, Godkin AJ, Bulek AM, Fuller A, Akpovwa HJ, Pymm PG, Liddy N, Sami M, Li Y, Rizkallah PJ, Jakobsen BK, Sewell AK.*
J Biol Chem. 2014 Jan

T-cell receptor specificity maintained by altered thermodynamics.
*Madura F, Rizkallah PJ, Miles KM, Holland CJ, Bulek AM, Fuller A, **Schauenburg AJ**, Miles JJ, Liddy N, Sami M, Li Y, Hossain M, Baker BM, Jakobsen BK, Sewell AK, Cole DK.*
J Biol Chem. 2013 Jun

Peptide length determines the outcome of TCR/peptide-MHCI engagement.
Ekeruche-Makinde J, Miles JJ, van den Berg HA, Skowera A, Cole DK, Dolton G,
Schauenburg AJ*, Tan MP, Pentier JM, Llewellyn-Lacey S, Miles KM, Bulek AM,*
Clement M, Williams T, Trimby A, Bailey M, Rizkallah P, Rossjohn J, Peakman M,
Price DA, Burrows SR, Sewell AK, Wooldridge L.
Blood. 2013 Feb

TCR/pMHC Optimized Protein crystallization Screen.
Bulek AM, Madura F, Fuller A, Holland CJ, ***Schauenburg AJ****, Sewell AK, Rizkallah*
PJ, Cole DK.
J Immunol Methods. 2012 Aug

**Anti-CD8 antibodies can trigger CD8[+] T cell effector function in the absence of
TCR engagement and improve peptide-MHCI tetramer staining.**
Clement M, Ladell K, Ekeruche-Makinde J, Miles JJ, Edwards ES, Dolton G,
Williams T, ***Schauenburg AJ****, Cole DK, Lauder SN, Gallimore AM, Godkin AJ,*
Burrows SR, Price DA, Sewell AK, Wooldridge L.
J Immunol. 2011 Jul

**Genetic and structural basis for selection of a ubiquitous T cell receptor
deployed in Epstein-Barr virus infection.**
Miles JJ, Bulek AM, Cole DK, Gostick E, ***Schauenburg AJ****, Dolton G, Venturi V,*
Davenport MP, Tan MP, Burrows SR, Wooldridge L, Price DA, Rizkallah PJ, Sewell
AK.
PLoS Pathog. 2010 Nov

Contents

Table of Figures

Abbreviations used in this thesis

(I)U	(International) Unit
Aa	Amino acid
APC	Antigen presenting cell
APC fluorochrome	Allophycocyanin
BCR	B-cell receptor
B-LCL	B-cell Lymphoblastoid cell line
C	Constant TCR gene fragment
CBS	Central Biotechnology Services
CD	Cluster of differentiation
cDNA	Complementary deoxyribonucleic acid
CDR-loops	Complementarity determining region-loops
CLIP	Class II-associated invariant-chain peptide
C-terminus	Carboxyterminus
CTLs	Cytotoxic T lymphocytes
D	Diversity TCR gene fragment
DC	Dentritic cells
DN	Double negative thymocytes (CD4$^-$ CD8$^-$)
DP	Double positive thymocytes (CD4$^+$ CD8$^+$)
DTT	Dithiothreitol
E.coli	*Escherichia coli*
EDC	N-(3-dimethylaminopropyl)-N^3-ethylcarbodiimide
ELISA	Enyzme-Linked-ImmunoSorbet-Assay
ELISPot	Enzyme-Linked-ImmunoSpot
ER	Endoplasmic reticulum
FCS	Foetal calf serum
FMO	Fluorescence minus one
g	Gram
GAD	Glutamate decarboxylase
GF	Gel Filtration
HA	Haemagglutinin
H-Bond / HB	Hydrogen bond
HEL	Hen egg lysosome
HIV	Human Immunodeficiency Virus

HLA	Human leukocyte antigen
IBs	Inclusion bodies
IE	Ion exchange
IFN-γ	Interferon-γ
Ig	Immunoglobulin
I_i	Invariant chain
IPTG	Isopropyl β-D-thio-galactoside
J	Joining TCR gene fragment
K_D	Equilibrium dissociation constant
MFI	Mean fluorescence intensity
mg	Milligram
MHC-I	Major histocompatibility complex class I
MHC-II	Major histocompatibility complex class II
ng	Nanogram
NK	Natural Killer cells
N-terminus	Amino-terminus
OD	Optical density
PAMPs	Pathogen associated molecular patterns
PB	Peptide bond
PB fluorochrome	Pacific blue
PBS	Phosphate buffer solution
PCR	Poly chain reaction
PDB	Protein data bank
PE	Phycoerythrin
PerCP	Peridinin Chlorophyll Protein
PFR	Peptide flanking residue / region
PHA	Phytohaemagglutinin
PKI	Protein Kinase Inhibitor
pMHC	peptide- major histocompatibility complex
pTα	pre-T-cell α chain
RACE	Rapid Amplification of cDNA Ends
RAG	Recombination activation genes
R-factor	Residual factor
R-factor	reliability factor
RNA	Ribonucleic acid

RU	Response units
SDS-PAGE	Sodium dodecyl sulfate – polyacrylamide gel electrophoresis
*Sf*9	*Spodoptera frugiperda* 9 cells
SMART^{TM}	Switching Mechanism At the 5' end of RNA Template
SP	Single positive thymocyte (CD4$^+$ or CD8$^+$)
SPR	Surface plasmon resonance
TCR	T-cell receptor
T$_{FH}$	T-follicular helper cells
T$_H$	T helper cells
TRAJ	T-cell receptor alpha joining gene
TRAV	T-cell receptor alpha variable gene
TRBJ	T-cell receptor beta joining gene
TRBV	T-cell receptor beta variable gene
T$_{reg}$	Regulatory T-cells
V	Variable TCR gene fragment
vdW	Van der Waals
β2m	β2 microglobulin
μg	Microgram

1. Introduction

1.1. Overview of the immune system

1.1.1. *The two arms of the immune system*

The immune system is a highly complex system of cells and molecules which all have their roles to play in order to protect us from pathogens, as well as eliminate cancerous cells, while avoiding autoimmune reactions. The immune system can be divided into the innate immune system, which recognises common patterns on pathogens and affected cells, and the adaptive immune system which is able to mount a highly specific response to any given pathogen. Players of both arms communicate and work together in order to fight off pathogens. Traditionally, innate and adaptive immunity have their own set of cells and molecules, however, more recent research shows that the boundaries between the two systems are more blurred than initially believed (Lanier 2013). For my introduction I will concentrate on the adaptive immune system.

1.1.2. *Adaptive immune system*

The adaptive immune system consists of two major cell types, B lymphocytes (or B cells) and T lymphocytes or (T cells). During embryonic development, precursors of B lymphocytes establish themselves within the bone marrow where they develop into mature B cells which then will enter the blood stream. Precursor cells within the bone marrow replenish the B cell pool whenever needed (LeBien et al. 2008). B cells are characterised by the expression of the membrane bound B cell receptor (BCR). It consists of two heavy chains and two shorter light chains that assemble in a distinctive Y-shape.

The distal ends of the two arms are characterised through antigen binding sites that bind to the surface of pathogens e.g. viral capsid proteins. These antigen-binding sites contain so called complementarity determining regions, or CDR loops, which vary greatly between individual B cell clones and determine the three-dimensional surface epitope the BCR will bind to. The pathogen will then be taken up by the B cell via phagocytosis and degraded within lysosomes (Drake et al. 1997). B cells also produce a soluble form of the BCR called immunoglobulin (Ig) or antibody. Once secreted these Igs bind to the surface of the pathogens primarily via the CDR loops. Depending on the class of antibody produced by the B cell, Igs differ in their constant region that confers them specific functions. For example, IgG can bind to surface proteins of virus and neutralise them by inhibiting their ability to infect new host cells. During infection, B cells specific to the pathogen in question will develop into antibody producing plasma cells and expand rapidly in order to generate the amount of antibody needed to clear the pathogen. Following clearance, some B cells will remain as memory cells allowing the body to quickly secrete specific antibodies in case of a second infection with the same or highly similar pathogen. This process is part of how vaccinations confer protection against infectious disease (Pulendran and Ahmed 2011).

T lymphocytes or T cells also express a surface antigen receptor called the T cell receptor or TCR. It resembles the short arm of the Y-shaped antibody with the highly variable CDR loops located at the distal end. In contrast to BCRs, TCRs only exist in membrane bound form. While BCRs and Igs bind to 3D epitopes on the surface of pathogens, TCRs bind to membrane ligands on other cells. The nature of these ligands vary depending of the subset of T cells.

1.2. T cell biology

1.2.1. Overview of T cell subsets

The TCR is a heterodimer consisting of either an α- and a β-chain (αβTCR) or a γ- and a δ-chain (γδTCR). While the majority of T lymphocytes in the peripheral blood express the αβTCR, γδT-cells are found in large proportions amongst tissue resident T cells (J. Zheng et al. 2013). Ligands for so called "conventional" αβT-cells are peptides bound to molecules of the major histocompatibility complex (MHC) class I or class II expressed at the surface of target cells. Unconventional T cells on the other hand include αβ- and γδT-cells which recognise a plethora of other ligands such as CD1, HLA-E and other members of the non-classical MHC family (Godfrey et al. 2015).

Conventional αβT-cells can be divided into CD4$^+$ T-cells and CD8$^+$ T cells depending on the expression of the corresponding co-receptor on their cell surface. As a general rule, CD8$^+$ T cells recognise peptides 8 to 13 amino acid (aa) in length bound to MHC class I (MHC-I). Their main function is to recognise infected cells presenting pathogen derived peptides on their surface and induce cell death through the secretion of cytotoxic molecules thereby eliminating the pathogens within the target cell. CD4$^+$ T cells recognise longer peptides of at least 9 aa in length bound to MHC-II (see**Figure** 1**.**1). Their main function is the coordination of the immune response directly (through cell-cell interactions with antibody producing B-cells) and indirectly (through secretion of cytokines) and are therefore also called helper T cells. For my thesis I studied the interaction of CD4$^+$ αβT-cells with pMHC-II and will therefore concentrate on these in this introduction.
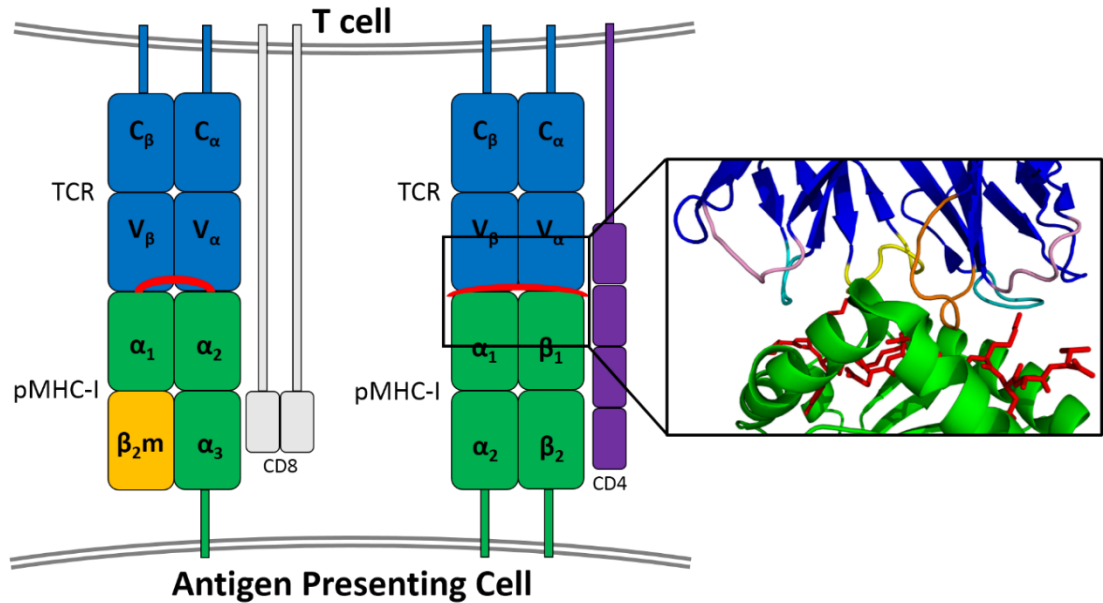
Figure 1.1: TCRs bind pMHC via their CDR3 loops. MHC (shown in green and yellow) present peptides at the cell surface. T cells recognise pMHC through their TCR (shown in blue) which binds both MHC and peptide via 6 flexible CDR loops (coloured regions in cut-out). Co-receptors CD8 and CD4 expressed by T cells bind to MHC-I and MHC-II, respectively and strengthen the interaction (PDB: 1FYT).

1.2.2. The T cell receptor

Each chain of the TCR consists of several segments: constant (C), joining (J) and variable (V) region and, in the case of the β-chain, an additional diversity (D) diversity segment (See **Figure** 1**.**2). During T cell development these regions are assembled into their final form in a process called somatic recombination or V(D)J recombination. There are several gene segments encoding each of these regions: (C_α: 1, J_α: 61, V_α: 70-80, C_β: 2, J_β: 13, D_β: 2 and V_β: 52. During V(D)J recombination one segment of each region is randomly selected and spliced together by a pair of enzymes called RAG1 and RAG2 (**Figure** 1**.**2) (Schatz and Ji 2011). This mix-and-match process leads to a vast number of possible TCR sequences particularly since each potential VαJα recombination can be paired with virtually any VβDβJβ recombination. However, this diversity is further increased by the

addition of random nucleotides (and as a result random non-germline amino acids in the final protein) at the junction between the V- and the J-region of the α-chain and the V-, D- and J-region of the β-chain (Bassing, Swat, and Alt 2002). It is estimated that 10^{15}-10^{20} different TCRs can be generated through this process however, only ~2×10^7 unique TCR sequences can be found in humans at any time (Miles, Brennan, and Burrows 2009)

Crystallographic studies show that the correctly refolded TCR expressed at the cell surface can be divided into a constant and a variable domain consisting of the Cα-and Cβ-regions and the VαJα- and VβDβJβ-regions, respectively (see **Figure** 1**.**2). Adjacent to the constant region is the short transmembrane tail that also constitutes the N-termini of both chains. The TCR is flanked by the CD3 complex consisting of CD3εδ, CD3εγ and C3ζζ, three transmembrane proteins playing an important role in TCR signalling (Alarcon et al. 2003). The variable domain constitutes the distal end of the TCR and contains six flexible CDR loops. Each chain encodes three of these: CDR1α, CDR2α, CDR3α, CDR1β, CDR2β and CDR3β. While the CDR1 and CDR2 loops of each chain are encoded entirely by germline DNA, the CDR3 loops are located at the junction of the Vα- and Jα-regions and Vβ-, Dβ- and Jβ-regions, respectively **Figure** 1**.**2). Due to the hypervariable nature of these junctions, CDR3 loops can vary greatly even between TCRs otherwise bearing the same VDJ recombination for both chains. Contacts between TCR and pMHC are mainly mediated through residues within or adjacent to the CDR loops of both chains (M G Rudolph, Stanfield, and Wilson 2006).
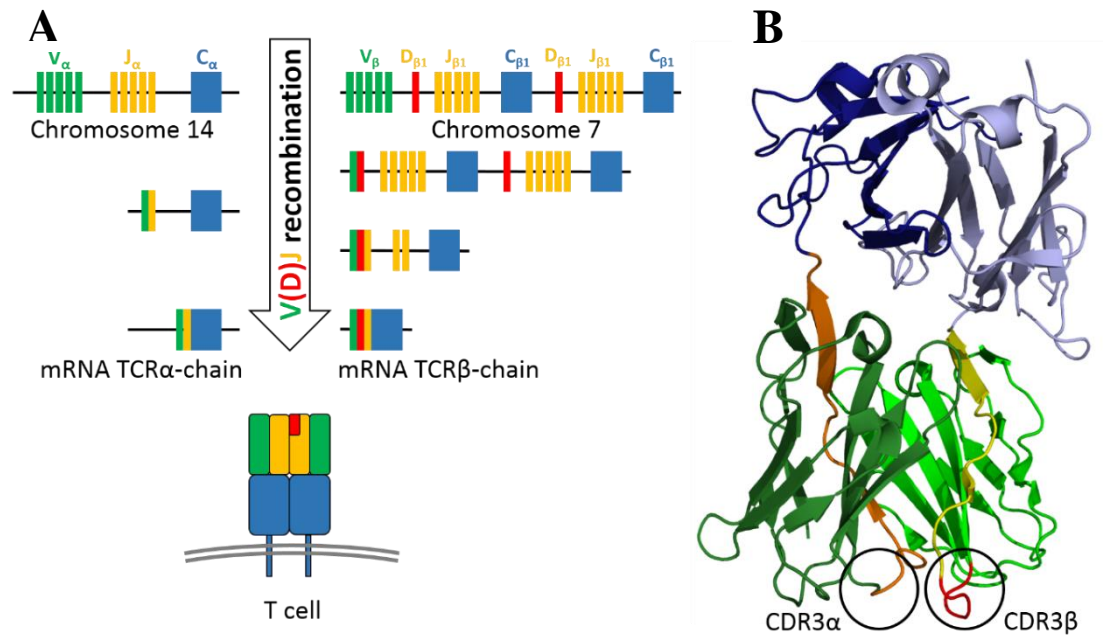
Figure 1.2: TCR gene segments are assembled through V(D)J recombination. (A) During T cell development, the V, J, D and C segments of each chain are assembled one after the other. Since one allele of each segment is chosen at random this creates an immense diversity of possible TCRs. In addition, random nucleotides are added at the junction of V_α and J_α sections and V_β, D_β and J_β, respectively further increasing diversity. Rearranged DNA will be transcribed into mRNA and the translated and refolded protein transported to the cell surface. (B) Cartoon model the HA1.7 TCR (PDB code: 1FYT). CDR3 loops are highlighted by circles and segments are colour coded as follows: V_α: dark green; J_α: orange; C_α: dark blue; V_β: light green, D_β: red (part of CDR3β loop); J_β: yellow; C_β: light blue.

1.2.3. *MHC-I present peptides of short length*

The MHC-I heterodimer consists of an α-chain and a shorter β_2m-chain. Together, β_2m and the α3 form the constant part of the MHC-I while the α1- and α2-subunits form the peptide-binding groove. In contrast to MHC-II, it is closed at both ends inhibiting peptides to protrude from the groove and limiting their maximal length to ~14aa. Peptides bind through anchor residues located at or close to the first and last peptide position and a conserved network of interactions between MHC and peptide N- and C-terminus,

respectively. With increasing length, peptides will bulge out of the groove (Holland, Cole, and Godkin 2013).

1.2.4. *The MHC-II accommodates peptides of varying length*

The MHC-II heterodimer also consists of an α- and a β-chain which together form a constant and a variable region. Peptides bind within a dedicated groove in the variable region. This groove is made of a floor of β-sheets flanked by an α-helix on each side. In the case of MHC-II, the ends of the groove are open allowing the peptide to protrude from the groove. Pockets in the β-sheet floor accommodate residue side chains at certain positions of the peptide called anchor residues. The nature of the amino acids determines the properties (e.g. charge and size) of potential amino acids. Different MHC-II alleles mainly differ in composition of these pockets and therefore require different sets of anchor residues for peptides to bind. MHC II heterodimers can be divided into HLA-DP, HLA-DQ and HLA-DR (see Figure 1.3). The HLA-DP and HLADQ loci both encode one α- and on β-chain each. Depending on which alleles have been inherited on the maternal and paternal chromosomes, each person can express up to four different versions of HLA-DP and HLA-DQ. The HLA-DR locus on the other hand encodes one α-chain and up to four β-chains, depending on the individual's haplotype. The three HLA-DRA genes encode the same peptide binding pockets and are therefore regarded as one. Thus, any one person can express up to four different HLA-DR molecules. In total, a person can express up to twelve different MHC-II heterodimers in addition to up to six possible MHC-I heterodimers (Shiina et al. 2009). This ensures that a wide variety of peptides can be presented to CD4[+] T cells is therefore one of the key mechanisms of a successful immune response.

Various HLA alleles have been associated with a predisposition for certain diseases. Some examples include: Autoimmune thyroid disease (HLA-DR3), Behcet disease (HLA-B51), Celiac disease (HLA-DQ8), Rheumatoid arthritis (HLA-DR3) and Type 1 diabetes (HLA-DR4, HLA-DR3) (Mackie et al. 2012; Shiina et al. 2009; C. Nguyen et al. 2013). These disease associations underline the importance of studying TCR/pMHC interactions.
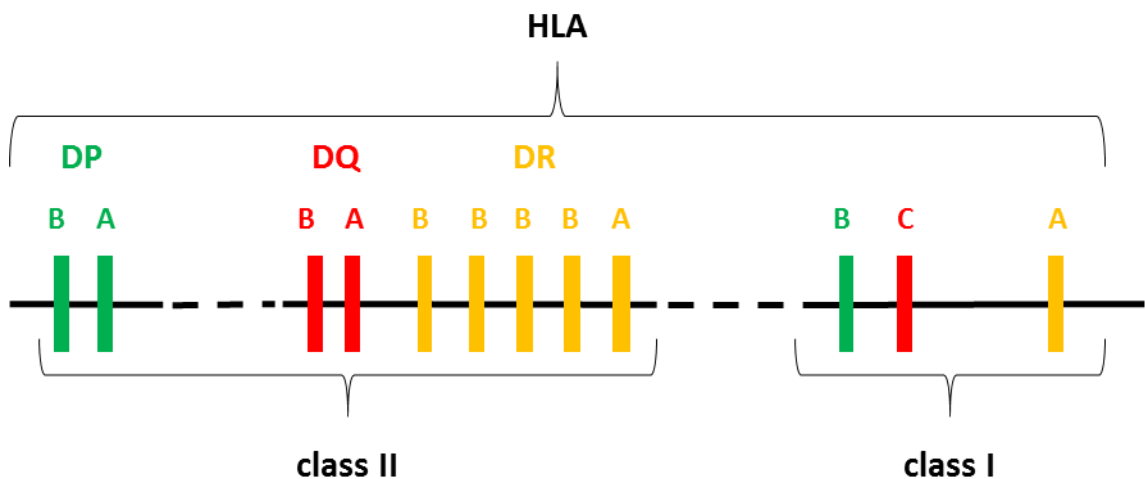
Figure 1.3: Schematic overview of the HLA locus.

Peptides bind to MHC-II in a flat and extended fashion (see **Figure** 1**.**4). In addition to the anchor residues, the MHC-II forms a series of H-bonds along the backbone of the peptide. Peptide residues located between the first and last anchor residue constitute the core of the peptide and are numbered from 1 to 9. Residues flanking this core on either side are called peptide flanking residues or region (PFR). N-terminal PFR are numbered in reverse order starting from P-1 while C-terminal PFR continue the numbering from 10 onwards. Due to the open conformation of the peptide-binding groove PFR and therefore

peptides binding to MHC-II can vary greatly in length (more detail in Chapter 4). Therefore, the topology presented to the TCR differs considerable between pMHC-I (short, bulging peptide) and pMHC-II (long, extended peptide with variable flanks).
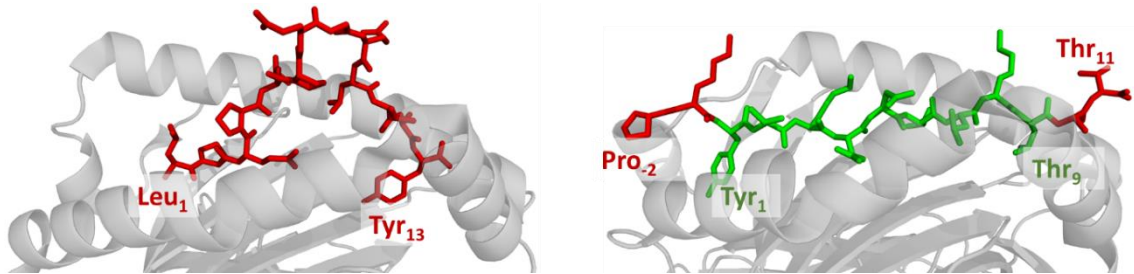
Figure 1.4: MHC-II present peptides of different lengths in an extended fashion. (A) MHC-I binding grooves are closed on both sides causing longer peptides to bulge out of the peptide binding groove such as the EBV derived 13mer LPEP peptide presented by HLA-B3508 (PDB code: 1ZHL). (B) The open peptide binding groove of MHC-II allows peptides of the same length (13aa) to bind in a flat conformation such as the influenza derived PKY peptide presented by HLA-DR1. The 9mer core is highlighted in green and PFRs in red. (PDB code: 1FYT).

1.2.5. General rules of TCR/pMHC-II interactions

An ever-growing database of crystal structures of TCRs in complex with pMHC-II backed by additional biophysical data allows drawing of some general conclusions about TCR/pMHC-II interactions. These have been reviewed extensively by Rossjohn and colleagues and Rudolph and colleagues (Rossjohn et al. 2015; Rudolph, Stanfield, and Wilson 2006).

Upon binding, the TCR is oriented in a roughly diagonal angle (varying between 37° and 90°) above the peptide binding groove with the TCR α- and β-chains located above the N- and C-terminus of the peptide, respectively (Rossjohn et al. 2015). As a general rule,

CDR1 and CDR2 loops are responsible for most of the interactions between TCR and pMHC. Several investigations found that these interactions show a degree of conservation with certain residues within the MHC-II being contacted in most crystal structures solved to date. Similarly, certain positions within the CDR1 and CDR2 loops are more likely to contact the MHC and also show some degree of conservation. Since CDR1 and CDR2 loops are entirely encoded by germline DNA, this bias is a sign of co-evolution between genes encoding TCR and MHC, respectively (Scott-Browne et al. 2011). The majority of contacts with the peptide is made by CDR3 loops. The inherent hyper variability of CDR3 loops ensures that virtually every possible pMHC-II combination will be recognised by at least one TCR present in the body. Furthermore, comparing the structures of TCRs in their unbound and bound states, respectively, shows a certain degree of flexibility within all six CDR loops in general and the CDR3 loops in particular (Rossjohn et al. 2015).

As their name indicates, CD4$^+$ T cells also express the CD4 co-receptor which binds to pMHC-II on the target cell within its constant region and aids TCR/pMHC-II interactions. The crystal structure of free CD4 show that it consists of four domains. Until recently no structure of CD4 binding a TCR-pMHC complex has been solved most likely due to low affinity interactions between co-receptor and MHC. Affinity maturation of murine CD4 however made it possible to generate a higher affinity mutation located within the distal D4 domain of CD4 (Wang et al. 2011). This made it possible to solve the structure of a TCR-pMHC-II-CD4 ternary complex showing that CD4 and TCR are building an arched conformation with the CD4 D4 domain interacting with the MHC-II α1 and β1 domains (Yin et al. 2011). Most interestingly, all MHC II residues contacted by CD4 are conserved across all human and most murine MHC-II alleles. Moreover, the space between CD4 and MHC-II at the cell membrane (~70Å) is wide enough for all three CD3 molecules to fit into. Engagement of CD3 is crucial for TCR signalling and T cell activation. Thus, this

structure provides evidence supporting the hypothesis that the conserved docking mode is due to positioning of co-receptors although this has yet to be proven to be the case for CD8.

Until recently, all structures solved support the binding mechanisms described above. However, Beringer and colleagues published the structures of two TCRs binding pMHC-II in an orientation that was 180° reversed to what has been seen in any other structure so far (Beringer et al. 2015). This shows that TCR/pMHC interactions are most likely not restricted to what has been defined as the consensus. It also demonstrates the need for more ternary structures in order to gain better insight into the TCR/pMHC interactions.

1.2.6. T cell development

Once pre-cursor T cells leave the bone marrow they migrate to the thymus to undergo further development. At this stage they don't express either co-receptor (CD4 or CD8) or TCR and are called double negative (TN) (Singer, Adoro, and Park 2008). First, the TCR β-gene locus undergoes recombination as described above. Once the β-chain is successfully assembled it associates with a pre-TCRα-chain forming a preTCR expressed at the cell surface where it co-locates with the CD3 complex. Expression of the preTCR induces the expression of both CD4 and CD8 co-receptors leading to double positive (DP) lymphocytes (Spits 2002). Signalling through the preTCR leads to recombination of the α-chain gene locus and generation of the final TCR (Michie and Zúñiga-Pflücker 2002).

At this stage, T cells undergo a selection process during which they are presented with MHC class I and class II loaded with self-peptides. In order to represent the majority (but not all) possible self-pMHC T cells might encounter in the periphery, proteins from all tissues are temporarily expressed by epithelial cells in thymic medulla (Klein et al. 2009).

T cells failing to recognise any of these self-pMHC undergo apoptosis (death by neglect) while T cells recognising self-pMHC with high sensitivity are also deleted (negative selection). This process results in a T cell population expressing TCRs that bind self-pMHC with moderate affinity (positive selection). They are therefore unlikely to cause autoimmune responses while still being able to recognise pathogenic peptides in the context of the HLA alleles expressed by the individual. During the selection process, one of the co-receptors (CD4 or CD8) will be downregulated mirroring the TCR's preference for binding MHC-I or MHC-II and become single positive (SP) naïve T cells (Rothenberg, Moore, and Yui 2008). While the majority of our T-cells mature during childhood, the thymus remains active throughout life replenishing the pool of self-tolerant T cells as necessary.

1.2.7. *CD4$^+$ T cell subsets and their function*

Upon leaving the thymus, naïve T cells will circulate within secondary lymphoid organs such as lymph nodes and the spleen. When professional antigen presenting cells (APCs) such as dendritic cells (DCs) or macrophages encounter and engulf pathogens they will present pathogen derived peptides on their MHC (1.2.4). These APC then migrate to lymph nodes where they present these pMHC to T cells for inspections. TCR recognition of any of these pMHC will lead to activation of the T cell in presence of co-stimulatory signals through molecules such as the CD28 receptor expressed on T cells. In addition, APCs and other cells activated during the primary immune response secrete stimulatory cytokines. The composition of these cytokines will not only aid T cell activation but also determine their differentiation into different subsets, each with their own markers and functions (Zhu and Paul 2010). T$_{reg}$ are responsible for down-regulating immune response both in autoimmune setting as well as following pathogen clearance. There are several

different TH subsets (see Table 1.1). Although each have their own function, there is a certain degree of overlap. T_{FH} reside in germinal centres where they aid B cell maturations and antibody production (X. Zhang et al. 2013). T_H17 are characterised through the expression of IL-17 and are implicated in inflammatory responses as well as clearance of bacterial infections (Korn et al. 2009). T_H1 and T_H2 were the first CD4$^+$ T cell subsets to be identified. Both stimulate antibody production in B cells. T_H2 cells have been shown to interact with cells of the innate immune system such as eosinophils and macrophages and are generally involved in defence against extracellular pathogens (Zhu, Yamane, and Paul 2010). T_H1 cells on the other hand are involved in the clearance of intracellular pathogens such as virus by stimulating cytotoxic CD8$^+$ lymphocytes (CTLs). Their signature cytokine INF-γ leads to upregulation of MHC-II on non-professional antigen cells such as infected epithelial cells and is crucial in antiviral immune response (Szabo et al. 2003). Effective immune responses require a fine balance between these subsets and when disturbed can lead to failure of defence against pathogens.

Subset	Function	Signature cytokines
T_H1	Stimulate CD8$^+$ T cells Defence against intracellular pathogens	IFNγ, TNFα
T_H2	Interact with innate immune system Defence against extracellular pathogens	IL-2
T_H17	Involved in inflammatory response Defence against bacterial infections	IL-17
T_{FH}	Stimulate B cells to secrete antibodies	IL-21

Table 1.1: Function and signature cytokines of some T$_H$ subsets.

1.3. Antigen processing and pMHC generation

1.3.1. MHC-I present peptides of endogenous origin

As a general rule, MHC-I present peptides derived from endogenous proteins. As part of the continuous turnover of intracellular material, proteins will be degraded into peptides by the proteasome. Upon infection or following stress signals, additional proteases are expressed and added to form the immunoproteasome. Peptides are transported from the cytosol into the ER by the Transporter Associated with antigen Processing (TAP) complex. Nascent and empty pMHC-I are generated in the ER as well where they are stabilised by chaperon proteins. Upon binding of a suitable peptide, these dissociate and pMHC-I are transported to the cell surface ready for inspection by T cells (Pamer and Cresswell 1998).

1.3.2. MHC-II mostly present peptides of exogenous origin

As a general rule, CD4$^+$ T cells recognise peptides from extracellular material. Once taken up by phagocytosis, proteins are broken down into peptides inside lysosomes. Nascent MHC-II are assembled within the endoplasmatic reticulum (ER) where they associate with a chaperon protein called invariant chain (Ii) (see **Figure** 1.**5**). Upon transfer into the Golgi apparatus they co-locate in so called MHC-II compartments (MIIC) where Ii is cleaved leaving only the class II-associated invariant-chain peptide (CLIP) in the MHC-II binding groove (Riberdy et al. 1992). CLIP serves as placeholder peptide stabilising MHC-II. They are further stabilised by HLA-DM which binds MHC-II and facilitates the exchange of CLIP for pathogen derived peptides (Denzin and Cresswell 1995). This occurs at low pH (~pH 5) following fusing of MIIC with late endosomes. HLA-DM plays a crucial role in deciding which peptides will bind: only peptides with a strong enough

affinity will successfully displace CLIP in a process called peptide editing. Deficiencies in HLA-DM lead to a greatly altered peptidome presented by MHC-II (Kropshofer et al. 1996).

Proteases known to be involved in MHC-II antigen processing include cathepsin (Cat) B (cysteine peptidase with both endo- and exopeptidase activity), D (aspartic endopeptidase), E (aspartic endopeptidase), L (cysteine endopeptidase) and S (cysteine endopeptidase) as well as asparaginyl endopeptidase (AEP) (Lippolis et al. 2002; Delamarre et al. 2005; Mitrović et al. 2016; Benes, Vetvicka, and Fusek 2008; Chlabicz et al. 2012; Turk et al. 2012; Kirschke and Wiederanders 1994). While exopeptidase cleave residues from one of the termini, endopeptidases cleave non-terminal peptide bonds. Their exact roles in antigen presentation are still under investigation. Inhibition of Cat B leads to altered immune responses including a shift from T_H2 responses to a T_H1 phenotype while Cat D and L were found to be crucial in digestion of the MHC-II associated Invariant chain (Ii; see below) (T. Zhang et al. 2000; Driessen, Lennon-Duménil, and Ploegh 2001). Inhibition of Cat S has been found to reduce the amount of pMHC-II expressed at the cell surface and is upregulated in inflammatory conditions (Wiendl et al. 2003; Riese et al. 1996). Cat S has also been shown to be involved in Ii chain degradation (Riese et al. 1996). Another cathepsin involved in the processing of antigens such as the tetanus toxin is cathepsin E (Hewitt et al. 1997). Protein degradation by endoproteases is a well-regulated process that, when disturbed can lead to disease often associated with faulty MHC-II antigen presentation.

Nature Reviews | Immunology
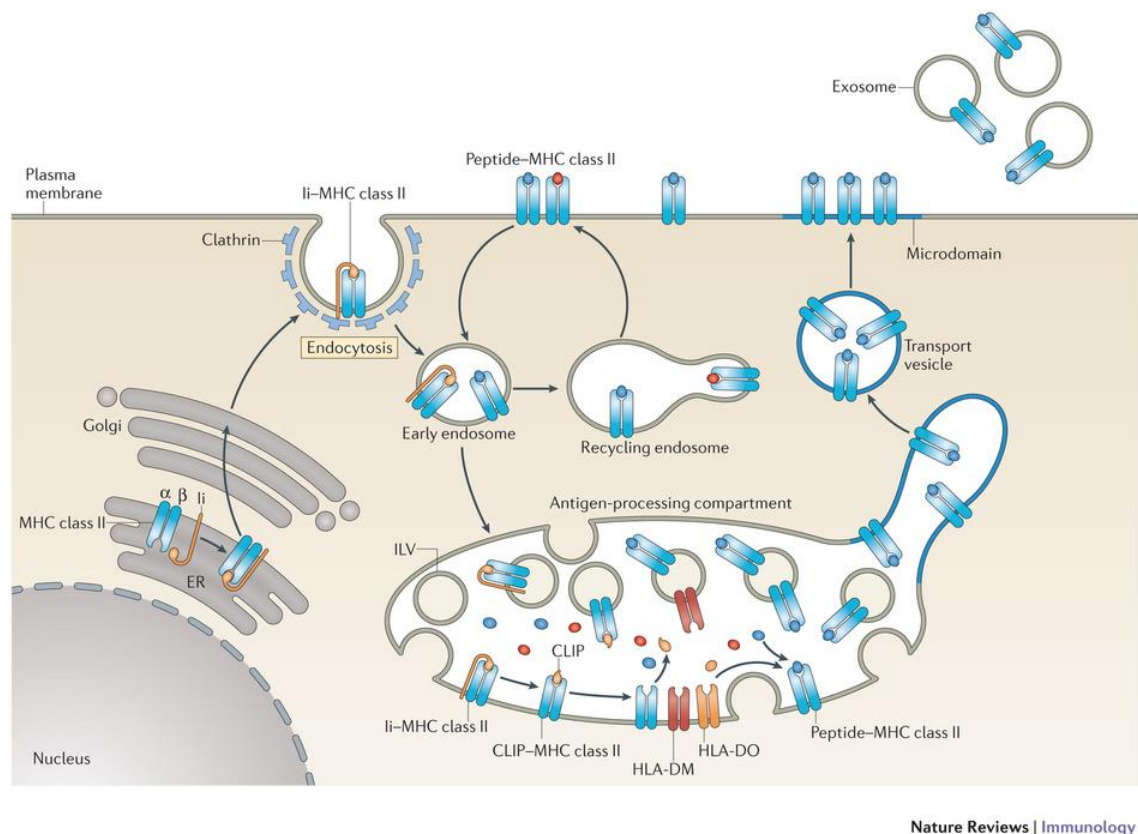
Figure 1.5: MHC-II antigen processing pathway. As a general rule, MHC-II present peptides from exogenous proteins degraded in phagosomes. Peptides are degraded by cathepsins before fusing with MHC-II containing compartments. Taken from Roche & Furuta 2015.

1.3.3. The MHC-II antigen processing machinery produces nested sets of peptides

Naturally processed peptides presented by MHC-II can be eluted from the surface of healthy cells and analysed by mass spectrometry (MS). The majority of eluted peptides are derived from self-proteins as expected in a healthy cell. More surprisingly, variants of the same peptide can be found in different lengths varying from 12-32aa (Chicz et al. 1992). These groups of peptides are called nested sets and are characterised by an overlapping, shared sequence as shown in **Table** 1**.**2 **HLA-DR4 bound peptides eluted from the cell surface.** MHC-II antigen processing machinery produces nested sets of

peptides of varying length as shown by peptides eluted from MHC-II (adapted from Chicz et al., 1992). Shared core sequence underlined. (Lippolis et al. 2002). This shared sequence is several residues longer than the 9mer core, as defined by the anchor residues, and therefore includes PFR. Without solving the crystal structures of these peptides, it is difficult to predict which residues consititure the anchor residues and therefore where the 9mer peptide core ends and PFR starts. For short peptides, the known anchor residue preferences of well-studied HLA alleles can aid the making of an educated guess of how the peptide binds backed up by online peptide binding algorithms such as NetPanMHCII (Andreatta et al. 2015). It is, however, possible that peptides contain more than one possible binding register. Most studies on T cells use peptides of set length for simplicity reasons. The impact of varying lengths peptides has not yet been investigated fully.

Source protein	Residues	Sequence	Length
HLA-A2	28-50	VDD<u>TQFVRFDSDAAS</u>QRMEPRAP	23
	28-48	VDD<u>TQFVRFDSDAAS</u>QRMEPR	21
	28-47	VDD<u>TQFVRFDSDAAS</u>QRMEPP	20
	28-46	VDD<u>TQFVRFDSDAAS</u>QRME	19
	30-48	D<u>TQFVRFDSDAAS</u>QRMEPR	19
	31-49	<u>TQFVRFDSDAAS</u>QRMEPRA	19
	28-44	VDD<u>TQFVRFDSDAAS</u>QR	17
	31-47	<u>TQFVRFDSDAAS</u>QRMEP	17
	31-47	<u>TQFVRFDSDAAS</u>QRM	15
	31-42	<u>TQFVRFDSDAAS</u>	12

Table 1.2 HLA-DR4 bound peptides eluted from the cell surface. MHC-II antigen processing machinery produces nested sets of peptides of varying length as shown by peptides eluted from MHC-II (adapted from Chicz et al., 1992). Shared core sequence underlined.

1.4. T cells in viral infections

1.4.1. CD4$^+$ T cells in influenza A infections

CD4$^+$ T cell play a major in protection against viral infections in general and influenza infections in particular. Thus, influenza derived, MHC-II restricted peptides provide an ideal experimental system for this thesis. The Influenza A virus is a RNA virus from the family of *Orthomyxoviridae* infecting a wide range of hosts including humans, birds, pigs and bats. Virus strains normally infecting animal hosts (mostly pigs and birds) occasional also infect humans in zoonotic infections. Humans can be infected by strains of the Influenza A, B and C genera. Influenza A strains are the most prevalent amongst humans where they infect organs of the respiratory tract such as the lungs and nasal epithelium causing coughing, sneezing and fever. Although most people will become infected with Influenza at some point in their life, it is estimated that only 75% will show symptoms (Hayward et al. 2014). In the remaining two thirds of cases, the immune system successfully has fought off the virus before symptoms manifest. Influenza viruses are in circulation all year round, although the majority of influenza associated illnesses occur in the colder months. Occasionally, particularly virulent strains can cause epidemics such the Spanish Flu in 1918 which is estimated to have killed 40 million people ("WHO | Influenza" 2004). Each year the WHO releases a list of the most prevalent strains in circulation in the coming winter season, which are then included in various vaccines. Like most RNA virus, Influenza A has a high mutation rate and strains change over time which poses an enormous challenge to the immune system as mutations can alter both B and T cell epitopes (Nobusawa and Sato 2006).

The Influenza virus enters host cells by first attaching to sialic acid on glycoproteins expressed on the cell surface via its hemagglutinin (HA) receptor (**see** Figure 1.6)**.** Viral

particles are then taken up through phagocytosis. HA also mediates fusion of the virus and host cell membranes releasing the virus RNA into the cytosol. Once viral proteins have been translated by host ribosomes, new viral particles bud from the cell membrane under the help of viral neuraminidase (NA) (Samji 2009). Viral proteins will be degraded into peptides while inside endosomes and thereby enter the MHC-II antigen-processing pathway. CD4[+] T cells recognising these pMHC-II play an important role in defence against the virus. Influenza-specific CD4[+] T cells stimulate antibody production by plasma cells and are important in maintaining antibody titres in the long term, therefore contributing to a stable memory response (Sridhar et al. 2015). In line with this, CD4[+] T cells aid generation of tissue resident memory CD8[+] T cells further contributing to the memory response (Laidlaw et al. 2014). Studies in mice showed that depletion of either CD4[+] or CD8[+] T cells did not significantly decrease efficacy of the vaccines tested while only 50-60% of mice survived a virus challenge following depletion of both (Cox et al. 2015; Guo et al. 2011). In another study, depletion of CD4[+] T cells led to a delayed viral clearance and recruitment of memory CD8[+] T cells resulting in increased mortality in mice (Doherty, Riberdy, and Belz 2000). Overall, CD4[+] T cells play an important role in anti-influenza immunity.
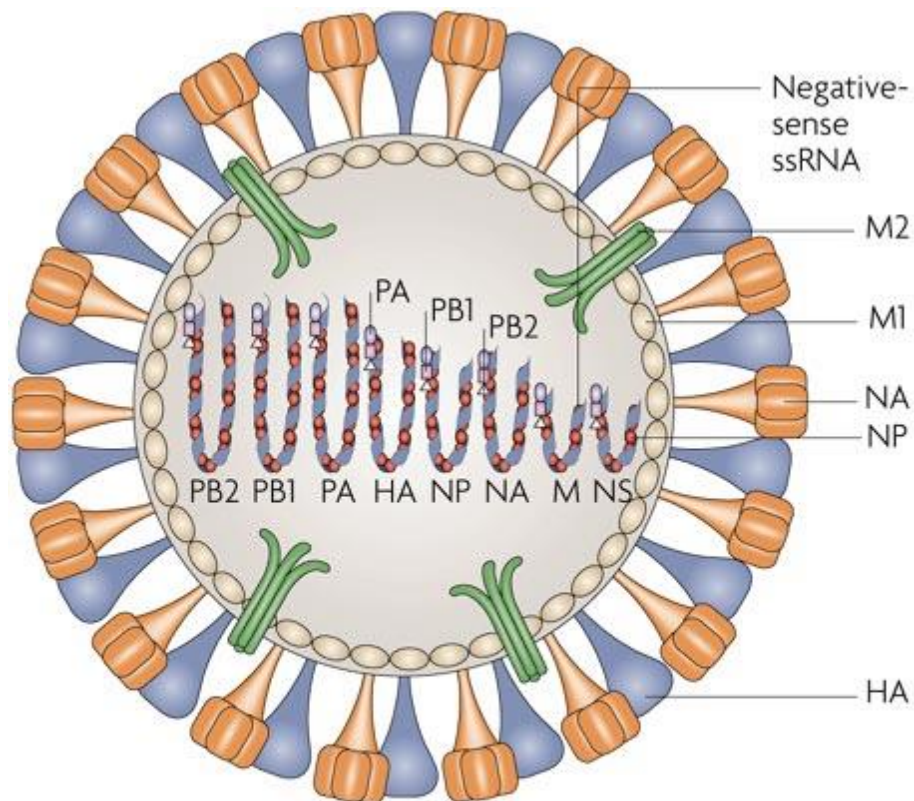
Figure 1.6: Schematic overview of the Influenza A virus. Taken from Nelson & Holmes, 2007.

1.4.2. CD4⁺ T cells in HIV infections

CD4$^+$ T cells also play an important role in HIV infections as they prime targets of the virus itself and have been shown to be crucial for controlling the infection. I also used an HIV derived peptide as an additional experimental system in this thesis. The Human immunodeficiency virus (HIV) is an RNA virus of the family of *Retroveridae* that causes acquired immunodeficiency syndrome (AIDS) (see **Figure** 1**.**7). In contrast to influenza which has been infecting humans for thousands of years (Suzuki and Nei 2002), the first

reported cases of AIDs occurred in the 1970's and the HI virus was only discovered in 1983 (F Barré-Sinoussi et al. 1983). The HIV-2 subtype is mostly confined to West Aftica while HIV-1 is the most prevalent and virulent subtype worldwide (Maartens et al. 2014). The HI virus is closely related to simian immunodeficiency virus (SIV) which is believed to be the origin of HIV following zoonotic infection (Keele et al. 2006). To date, 70 million individuals have been infected with HIV and about half of these have died of acquired immunodeficiency syndrome (AIDS) ("WHO | HIV/AIDS" 2016). Due to recent advances in combination antiretroviral therapy (cART), AIDS can be reasonably well managed even though the infections remains chronic (Françoise Barré-Sinoussi, Ross, and Delfraissy 2013). The HIV virus is transmitted through bodily fluids and infects CD4$^+$ T cells. Following attachment to the CD4 co-receptor and chemokine co-receptors CCR5 or CXCR4 on the cell surface, the virus membrane fuses with the host cell membrane releasing its contents into the cytoplasm (Maartens et al. 2014). Following reverse transcription of the viral RNA in double stranded DNA (dsDNA), viral proteins are transcribed by the host cell and assembled into fresh viral particles which will bud from the cell surface. The viral capsid (CA) protein, gag p24, plays a crucial role in the assembly of the virus (Freed 1998). Following the acute phase characterised though rapid viral replication and unspecific symptoms such as sore throat and fever the infection progresses into the asymptomatic latent phase. The latter can span years before progressing into AIDS which is characterised through opportunistic infections which will eventually lead to death if not treated with cART (Coffin, Hughes, and Varmus 1997). The major cause of AIDS and susceptibility to opportunistic infections is the gradual loss of CD4$^+$ T cells compromising the whole immune system. Interestingly, HIV specific CD4$^+$ T cells are preferentially infected by the virus, further compromising anti-HIV responses (Douek et al. 2002). Nevertheless, gag p24 specific CD4$^+$ T cells can be found

in all stages of infection and are associated with lower viral titres (Palmer, Boritz, and Wilson 2004; Christie et al. 1997; Rosenberg et al. 1997). Cytotoxic CD8[+] T cell responses are equally important in reducing viremia, particularly in early stages of infection mainly through elimination of infected CD4[+] T cells (Betts et al. 2001; Koup et al. 1994). The virus responds to pressure from the adaptive immune system by rapidly mutating and thereby generating so called escape variants where T cell epitopes are altered enough to escape recognition by T cells (Price et al. 1997). Although anti-HIV antibody responses can be detected in all stages of infection they seem to be unable to control viral replication, partly due to the loss of CD4[+] T cell help and exhaustion of B cells (Moir and Fauci 2009). Overall, HIV undermines the immune system by infecting and eliminating one of its key players thereby underlining their importance within the immune system.

Nature Reviews | Immunology
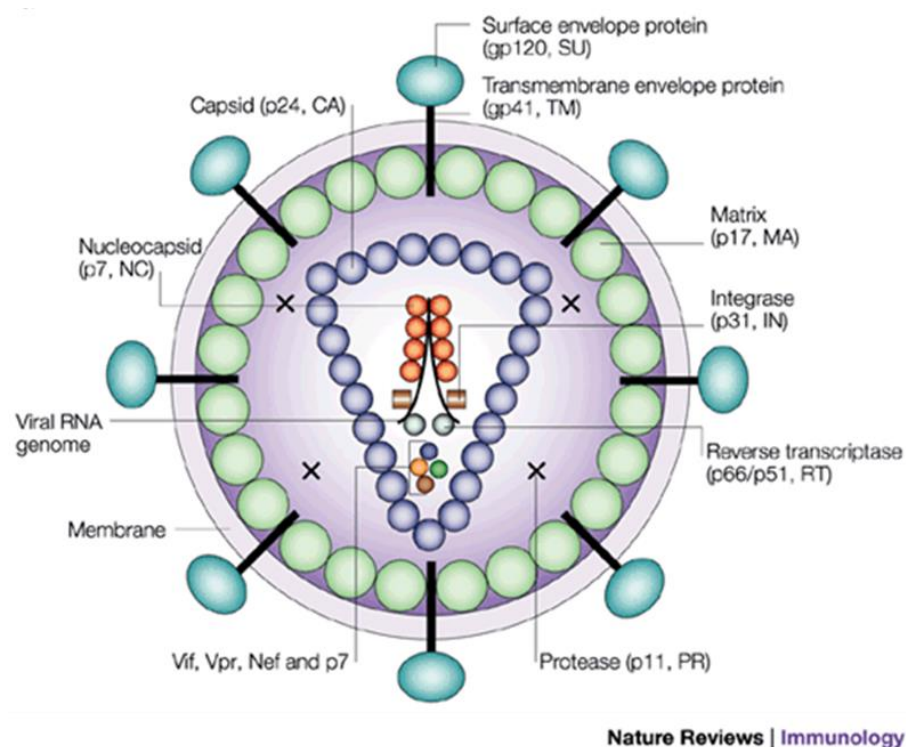
Figure 1.7: Schematic overview of the HI virus. Adapted from Robinson, 2002.

1.5. Aims of this thesis

CD4$^+$ T cells play an important role on many levels of the anti-viral immune response. The key ability of CD4$^+$ T cells to recognise intracellular infections through pathogen derived peptides presented on MHC-II is mediated through the TCR. Soluble pMHC and TCR are important tools for studying TCR/pMHC interactions. However, generating soluble MHC-II has proven to be more difficult compared to generation of their pMHC-I counterparts. Despite recent advances, research into CD4$^+$ T cell still lags behind research into CD8$^+$ T cells.

Therefore, the first aim of this thesis was to set up an insect cell based expression system for soluble MHC-II.

Peptides bound to pMHC-II are known to contain different lenghts PFRs. However, rather little is known about the role of PFR on T cell activation. Several studies have shown that PFRs indeed play a very important role in T cell activation. Previous investigations in our lab have shown that PFRs in an influenza and a HIV-1 derived peptide, respectively, can have substantial effects on TCR/pMHC-II interactions.

The second aim of this thesis was to investigate the impact of PFRs on T cell activation and peptide binding stability in these two experimental systems in more detail.

CD4$^+$ T cells play an important role in influenza infections. Monitoring of influenza specific populations can give valuable insight into CD4$^+$ T cell mediated defence and aids the development of effective vaccines. However, so far only one HA derived, MHC-II restricted peptide has been studied in great detail.

Thus, the third aim of this thesis. was to map conserved epitopes within the influenza heamagglutinin protein.

2. Methods and Materials

2.1. Expression of HLA-DR1 in Sf9 cells

2.1.1. Buffers, reagents and media

PBS: Phosphate buffered saline (Oxoid).

Elution buffer: 50 mM CAPS (Sigma-Aldrich®), pH 11.5.

Neutralising buffer: 300 mM Sodium phosphate (Sigma-Aldrich®), pH 6.

Storage buffer: PBS containing 0.02% sodium azide (Sigma-Aldrich®).

Luria-Bertani (LB) medium: 10 g/L tryptone (Fisher Scientific), 5 g/L yeast extract (Fisher Scientific) and 5 g/L NaCl (Fisher Scientific).

LB agar medium: 15 g/L agar bacteriological (Oxoid), 10 g/L tryptone (Fisher Scientific), 5g/L yeast extract (Fisher Scientific), 5 g/L NaCl (Fisher Scientific).

TYP media: 16g/L tryptone (Fisher Scientific), 16g/L yeast extract (Fisher Scientific) and 5g/L potassium phosphate dibasic (Acros Organics).

One Shot® (TOP10) *E. coli* (Lifetechnologies™) competent cells were used for DNA manipulation and amplification and grown in TYP medium.

Complete SFM: Sf-900™ II SFM (Gibco®), 100IU/ml penicillin (Gibco®), 100µg/ml streptomycin (Gibco®).

Insect cell freezing buffer: 60% Sf-900 II SFM (Gibco®), 30%FBS (Gibco®), 10%DMSO (Sigma-Aldrich®).

2.1.2. *Insect cell culture*

Sf9 cells for protein expression were kindly provided by Dr. Andreas Heil. *Sf9* cells were maintained in suspension culture in Erlenmeyer flasks complete SFM media at 25-27 °C and 80rpm on an orbital shaker. Cells were monitored daily and cell concentration was kept at 0.5-4x10^{6}cells/ ml in order to maintain cell growth in the logarithmic phase. *Sf9* cells were frozen at 10^{7}cells/ml in insect cell freezing buffer and stored for 1 h at -20° C before being transferred to -80 °C for long term storage. For initiation of fresh cultures, one vial of frozen *Sf9* cells was thawed at 37 °C and transferred into 10 ml of fresh complete SFM.

2.1.3. *Expression plasmids for expression of HLA-DR1 in Sf9 cells*

Sequences for the extracellular domains of HLA-DR1 (DRA*0101, residues 1-183 and DRB1*0101, residues 1-190) were obtained from Stern & Wiley 1992. The HLA-DR1α chain was further modified by adding a honey bee melitin leader sequence (MYIYADPSPA) to the N-terminus (Homa et al. 1995) and a fos leucine zipper sequence (Willcox et al. 1999) followed by a BirA biotinylation site to the C-terminus (O'Callaghan et al. 1999). The HLA-DR1β chain was further modified by adding a honey bee melitin leader sequence followed by the sequence of the CLIP peptide (MPVSKMRMATPLL) followed by a Thrombin cleavage site to the N-terminus and a jun leucine zipper sequence to the C-terminus (Willcox et al. 1999). Vectors encoding the modified HLA-DR1α chain and HLA-DR1β chains sequences were chemically synthesized by Genewiz and Geneart, respectively. Sequences were cloned and inserted

into BaculoDirect^{TM} (LifeTechnologies) expression vectors as described below allowing the sequences to be expressed in insect cells.

2.1.4. *Cloning into BacucloDirect^{TM} expression vectors*

First, the sequences were cut using an enzymatic digest (see 2.1.4.1) to separate the target sequences from the supplied vector. Following purification by gel electrophoresis, the target sequences were ligated into the pENTR11 entry vector (LifeTechnologies) (see 2.1.4.2). Second, the target sequences were inserted into the BaculoDirect^{TM} expression vectors using the site directed LR reaction (see 2.1.4.4).

2.1.4.1. *Enzymatic digestion*

Restriction endonucleases and buffers were purchased from Fermentas. 1000 ng of plasmid DNA were digested in 20 µl final volume containing no more than 10% of restriction enzyme. After 2 h incubation at 37 °C, digests were separated by gel electrophoresis on a 1% agarose gel.

2.1.4.2. *Agarose gel electrophoresis and ligation*

Restriction digests were prepared by adding 5 µl of loading dye (Bioline) and fragments were separated by gel electrophoresis for 50 minutes at 70 Volt on a 1% agarose (LifeTechnologies) gel in 100 ml of Tris-acetate-EDTA buffer containing 1x SYBR safe DNA gel stain (LifeTechnologies). Fragments were visualised using a transilluminator and sizes were estimated using a molecular weight marker (HyperLadder I^{TM}, Bioline) loaded simultaneously as the samples. Fragments of interest were extracted using a scalpel and purified using the GelWizard Kit (Promega) following the manufacturer's instructions. DNA concentrations were measures using Nanophotometer® by Implen.

T4 DNA ligase (New England Biolabs) was used to ligate inserts into 50 ng of the pENTR11 backbone at molecular ratios of backbone to insert of 1:1, 1:2 and 1:3 in DNA ligase (New England Biolabs) buffer in 10 µl final volume. Reactions were performed overnight at RT.

2.1.4.3. *Transforming competent bacteria cells and purifying plasmid DNA*

1 vial of 25 µl Top10 *E.coli* was thawed on ice for 5 min and 1 µl of the ligation mixture was added for 30min on ice. Cells were then heat shocked for 30 s at 42 °C before recovering 2 min on ice. 250 µl pre-warmed S.O.C. media (LifeTechnologies) was added and vials were incubated for 1h at 37 °C 220rpm in an orbital shaking incubator. 100 µl of culture was spread onto a pre-warmed LB agar plate containing 50 µg/ml kanamycin (Sigma-Aldrich®) and incubated overnight at 37 °C. Individual colonies were used to inoculate 12 ml of LB media containing 50 µg/ml kanamycin. Plasmid DNA was purified using the Zippy kit (Zymo Research) following manufacturer's instructions. DNA concentration and purity was analysed using Nanophotometer® by Implen. Sequences were confirmed by automated DNA sequencing by Central Biotechnology Services (CBS).

2.1.4.4. *Cloning into BaculoDirect^{TM} baculovirus expression vector*

Plasmid preparations that proved to contain the correct sequence were amplified again in Top10 *E.coli* as described above. Plasmid DNA was purified using the HiPure plasmid prep kit from LifeTechnologies in order to obtain ultra-pure DNA preparations to be used in subsequent steps. These pENTR11 entry clones were used to assemble recombinant baculovirus DNA using the BaculoDirect^{TM} Expression kit (Invitrogen). This system is based on the Lambda site-specific recombination (LR) reaction. Both the gene of interest

cloned into pENTR11 and a negative selection cassette located on the linearized baculovirus DNA are flanked by homologous DNA sequences. In presence of the LR clonase enzyme the gene of interest is inserted into the Baculovirus backbone through homologous recombination thereby replacing the negative selection cassette. 10 µl of linearized Baculovirus DNA were combined with 100-300 ng of the entry clone in a final volume of 16 µl TE buffer provided with the kit. As positive control 100 ng of the pENTR-Cat plasmid provided with the kit were used. LR clonase II was thawed on ice and vortexed twice for 2 s before adding 4 µl to the DNA mix. As negative control, no enzyme was added. All mixtures were incubated at 25 °C for at least 1 h or overnight and stored at 4 °C.

2.1.4.5. Verifying LR reaction

In order to verify successful insertion of the gene of interest into the Baculovirus vector, polymerase chain reaction (PCR) was conducted. 1 µl of the LR reaction from 2.1.4.4 was diluted 1:200 in water and 4 µl of this dilution were used in the PCR reaction. 4 µl sterile H_2O were used as a negative control. The following primers were used: Polyhedrin Forward Primer (5´-AAATGATAACCATCTCGC-3´) binding to the polyhedron promotor located upstream of the gene of interest and V5 Reverse Primer (5´-ACCGAGGAGAGGGTTAGGGAT-3´) binding to the V5 purification tag located downstream of the gene of interest. Primers were used at 200 nM final concentration. Deoxynucleoside triphosphates (LifeTechnologies) were used at 0.2 mM final concentration. 10 µl 5X Green GoTaq® Reaction Buffer (Promega) was added before adjusting final volume to 49.75 µl using sterile H2O. Finally, 0.25 µl GoTaq® DNA Polymerase (Promega) was added. PCR conditions recommended in the BaculoDirect^{TM}

Expression kit manual were used. The PCR products were verified by gel electrophoresis on a 1% agarose gel.

2.1.5. Transfection of Sf9 cells and generation of P1 viral stocks

0.8×10^6 *Sf9* cells were plated per well in 6-well tissue cultures plates and left to attach for 15 min at RT before aspirating medium entirely. 2.5 ml plating media (15% complete SFM + 85% antibiotic free Sf-900™ II SFM (Gibco®)) was added to each well. 8 µl of Cellfectin II (Invotrogen™) and 100 µl antibiotic free Sf-900™ II SFM (Gibco®) were combined in a 1.5 ml Eppendorf tube (mixture A). In a separate tube, 10 µl of each LR reaction including negative and positive controls were combined with 10 µl of antibiotic free Sf-900™ II SFM (Gibco®) (mixture B). For a mock transfection 10 µl of antibiotic free Sf-900™ II SFM (Gibco®) were used in place of the LR reaction. Mixtures A and B were combined, gently mixed by tapping and incubated 30 min at RT. Each transfection mixture was added dropwise to the corresponding well and incubated 3-5 h at 27 °C. Supernatants were aspirated entirely and 2 ml complete SFM containing 100µM ganciclovir (Source BioScience) were carefully added dropwise. Plates were placed in sealed plastic bags containing wet paper towels and incubated at 27 °C for 5 days. Supernatants from each well were collected and centrifuge at 4,000 rpm for 5min to remove cells before filtering supernatants. These P1 viral stocks were stored at 4 °C protected from light.

2.1.6. Generating high titre viral stocks

In order to amplify viral titres in a first step, *Sf9* cells were seeded at 1.5×10^6 cells/ml in 12 ml complete SFM containing 100 µM ganclicovir (Source BioScience) in Erlenmeyer flasks. 0.5 ml of P1 viral stocks were added to each flask and incubated at 27 °C and 80

rpm on an orbital shaker for 96 h. Cultures were centrifuged 5 min at 4000 rpm in a Hereaus Megefuge 16R with a Hereaus 75003629 rotor and supernatants were filtered into sterile tubes. P2 viral stocks were stored at 4° C protected from light. In order to obtain high titre viral stocks for protein expression *Sf9* cells were seeded at 1.5×10^6 cells/ml in 45 ml complete SFM containing no ganclicovir in Erlenmeyer flasks. 5 ml of P2 viral stocks were added and incubated at 27 °C and 80 rpm on and orbital shaker for 96 h hours. P3 viral stocks were harvested and stores as described above.

2.1.7. *Isolating viral DNA for verification by PCR*

Aliquots 750 μl of either P2 or P3 viral stocks were taken in order to extract viral DNA for verification of sequences using PCR. 750 μl cold 20% PEG 8000 (Molecular Dimensions) in 1M NaCl (Fisher Scientific) were added and gently mixed. After 30min incubation at RT, samples were centrifuged for 10 min at 12,000 to 13,000 at RT and medium was aspirated. Pellets were resuspended in 100 μl of PBS containing 0.1% Triton X-100 (Sigma-Aldrich®) before adding 10 μl Proteinase K (LifeTechnologies) at 5-10 mg/ml, mixing gently and incubating for 1 h at 50 °C. 110 μl of phenol:chloroform:isoamyl (25:24:1) (LifeTechnologies) were added and gently mixed. Samples were centrifuged for 5 min at 12,000 to 13,000 at RT and the aqueous phase was transferred to a fresh tube. The following reagents were added: 10 μl 3 M sodium acetate (Sigma-Aldrich®), 0.5 μl UltraPure Glycogen at 20 μg/μl (LifeTechnologies), 250 μl 100% ethanol (Fisher Scientifics). Samples were incubated at 20° C for 20 min before centrifuging for 15 min at 1300 rpm in a Hereaus Megefuge 16R with a Hereaus 75003629 rotor at 4 °C. Pellets were resuspended 500 μl 70% ethanol (Fisher Scientifics), centrifuged for 15 min at 13000 rpm in a Hereaus Megefuge 16R with a Hereaus

75003629 rotor at 4 °C and resuspended in 10 μl sterile H_2O. PCR were conducted as described in 2.1.4.5.

2.1.8. *Measuring of viral titre of P3 baculoviral stocks*

Viral titres of P3 stocks were measured using the BacPAK[TM] Baculovirus Rapid Titre Kit (Clontech). This kit is based on the expression of the baculoviral protein gp64 on the surface of infected insect cells. Briefly, *Sf9* cells were plated at 6.5×10^4 cells/well in a 96-well plate and let to attach for 1h at 27 °C. Medium was aspirated carefully and 25 μl of viral stocks at three different dilutions (10^{-3}, 10^{-4} and 10^{-5}) were added in triplicates or quadruplicates. 25 μl of complete SFM was added as negative control and plates were incubated 1 h at 27 °C. Inoculum was carefully replaced by 50 μl Methyl Cellulose Overlay and plates were incubated wrapped in a moist paper towel in a sealed plastic bag at 27 °C for 43-47 h. 150 μl of ice cold paraformaldehyde was added to each well and incubated 10 min at RT before washing with 200 μl PBS 0.05% Tween 20. 50 μl of Normal Goat Serum was added and incubated 5 min at RT before adding 25 μl Mouse gp64 Antibody per well and incubating at 37 °C for 25min. Following another wash step 50 μl of Goat Anti-mouse Antibody/HRP Conjugate was added and incubated 25 min at 37 °C. Following a last wash step, 50 μl Blue Peroxidase Substrate was added and incubated 3 h at RT. Numbers of blue foci were counted using a microscope and averages were used to calculate the viral titre in infectious units (IFU) per ml P3 viral stocks.

2.1.9. *Co-infection of Sf9 insect cells with high titre baculoviral stocks and harvesting of secreted HLA-DR1*

Sf9 cells were added to fresh shaker flasks and topped up with fresh SFM-II media to yield final cell concentration of 2×10^6 cells/ml. High titre viral stocks were added and

incubated 96 h at 27 °C and 80 rpm on an orbital shaker. Co-infected cells cultures were centrifuged for 10 min at 4000 rpm in a Hereaus Megefuge 16R with a Hereaus 75003629 rotor before filtering the supernatant at 0.33 µm into a sterile container. Harvested supernatants were concentrated using a viva flow 200 (polyethersulfone membrane) module to a volume of 50 ml. Further concentration to a volume of <1 ml was achieved using a 20 ml vivaspin column with a 10 kDa cut-off (Sartorious, France).

2.1.10. Purification of secreted HLA-DR1^{CLIP} from co-infected Sf9 insect cells

Correctly refolded HLA-DR1^{CLIP} was purified using an immunoaffnity column coated with the anti-HLA-DR antibody L243 (Brodsky 1984). Columns were equilibrated at RT for 1 h and washed with 5 ml PBS before applying samples. The collected flow through was collected and reapplied. This procedure was repeated a total of three times. 15 ml PBS was applied to wash the column until no more protein could be measured in the flow through. 10 ml of elution buffer was applied in order to elute the protein and immediately pH neutralised using an equal volume of neutralising buffer. The column was washed using 5 ml elution buffer before applying 5 ml storage buffer and storing the immunoaffinity column at 4° C. The eluate was concentrated using a vivaspin column and washed once using PBS. Protein purity was verified by Coomassie-stained SDS-PAGE and protein concentration measured using Nanophotometer® by Implen. Correctly refolded HLA-DR1 was further purified by size exclusion chromatography using a superdex 200HR column (GE Health care, UK) and an ÄKTA FPLC in conjunction with Unicorn software. Protein purity was verified by Coomassie-stained SDS-PAGE and protein concentration measured using Nanophotometer® by Implen.

2.1.11. Thrombin cleavage and peptide exchange.

Purified HLA-DR1^{CLIP} was cleaved using a Thrombin cleavage kit (Millipore) by adding 20 IU Thrombin/mg protein in 50 µl reaction buffer 10X and topping up with H$_2$O to a final volume of 500 µl. Reactions were incubated 90min at room temperature before adding the serine protease inhibitor 4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride (ABSF, Sigma-Aldrich®). 2 µl of peptide at 20 mg/ml was added and incubated overnight at room temperature. Excess unbound peptide was removed by size exclusion chromatography using a superdex 200HR column (GE Health care, UK) and an ÄKTA FPLC in conjunction with Unicorn software. Protein purity was verified by Coomassie-stained SDS-PAGE and protein concentration measured using Nanophotometer® by Implen.

2.2. Manufacturing of soluble pMHC-II and TCR in *E.coli*

2.2.1. Buffer, reagents and media

Luria-Bertani (LB) medium: 10 g/L tryptone (Fisher Scientific), 5 g/L yeast extract (Fisher Scientific) and 5 g/L NaCl (Fisher Scientific).

LB agar medium: 15 g/L agar bacteriological (Oxoid), 10 g/L tryptone (Fisher Scientific), 5 g/L yeast extract (Fisher Scientific), 5 g/L NaCl (Fisher Scientific).

TYP media: 16g/L tryptone (Fisher Scientific), 16g/L yeast extract (Fisher Scientific) and 5g/L potassium phosphate dibasic (Acros organics).

Lysis buffer: 10 mM Tris pH 8.1 (Fisher Scientific), 10 mM MgCl (Acros organics), 150 mM NaCl (Fisher Scientific), 10% glycerol (Sigma-Aldrich®).

Triton wash buffer: 5% Triton x100 (Fisher Scientific), 500 mM TRIS pH8.1, 1M NaCl (Fisher Scientific) and 10mM EDTA (Sigma-Aldrich®) pH 8.1.

Re-suspension buffer: 500 mM TRIS pH8.1and 10 mM EDTA (Sigma-Aldrich®) pH 8.1.

Urea buffer: 8 M Urea (Sigma-Aldrich®), 20 mM Tris pH 8.1 and 0.5 mM EDTA (Sigma-Aldrich®) pH 8.1.

MHC-II refold buffer: 25% glycerol (≥99.9) (Fisher Scientific), 20 mM Tris pH8.1, 1mM EDTA (Sigma-Aldrich®) pH 8.1, 0.74 g/L cysteamine, and 0.83 g/L cystamine.

Guanidine buffer: 6M guanidine, 50 mM TRIS pH8.1, 100 mM NaCl (Fisher Scientific), 2mM EDTA (Sigma-Aldrich®).

TCR refold buffer: 2.5M Urea, 50 mM TRIS (pH8), 2mM EDTA, 0.74g/L cysteamine, and 0.83g/L cystamine.

BIAcore Buffer: HBS-P+ buffer (10nM HEPES pH7.4, 150mM NaCl, 3mM EDTA (Sigma-Aldrich®) and 0.005% (v/v) Surfactant p20 by GE Healthcare).

Peptide binding assay blocking buffer: PBS containing 3% Bovine serum albumin (BSA, Sigma-Aldrich®).

Peptide binding buffer: 20 mM 2-(morpholino)ethane-sulfonic acid (MES, Sigma-Aldrich®), 300 mM 3-[(3-cholamidopropyl)dimethylammonio]-1-pro- panesulfonate (CHAPS, Sigma-Aldrich®), 140 mM NaCl (Fisher Scientific), pH5.

Neutralising buffer: 1 M Tris (pH 8), 10% BSA (Sigma-Aldrich®), 1% Tween 20 (Fisher scientific), 0.02% Sodium azide (Sigma-Aldrich®).

Triton wash buffer: 5% Triton x100 (Fisher Scientific), 500 mM TRIS pH8.1, 1 M NaCl (Fisher Scientific) and 10 mM EDTA (Sigma-Aldrich®) pH 8.1.

One Shot® (TOP10): *E. coli* (Invitrogen™) competent cells were used as host for DNA manipulation and grown in TYP medium.

Rosetta^{TM} 2(DE3) pLysS: *E. coli* (Novagen®) competent cells were used for protein expression and grown in TYP medium.

2.2.2. Generation of expression plasmids

The extracellular domains of HLA-DR1 (residues 1-118 of DRA*0101 and residues 1-190 of DRB*0101) were taken from published work (Stern and Wiley 1992). A biotinylation tag (GGGLNDIFEAQKIEWH (O'Callaghan et al. 1999) was added to the C-terminus of the HLA-DR1α chain. These sequences were previously inserted into the pGMT7 expression plasmid as described in 2.1.4.1 and 2.1.4.2. The pGMT7 plasmids allows the expression of genes of interest in *E.coli* as described in 2.2.3. It encodes an ampicillin resistance gene which allows transformed bacteria to grow in LB supplemented with 50 mg/L Carbenicillin (BiochemicalDIRECT^{TM}). Sequences coding for the extracellular domains of the following TCRs specific for the H3N2 derived epitope HA$_{306-318}$ were previously cloned into the pGMT7 expression plasmid as described in 2.1.4.1 and 2.1.4.2: 2C5 (TRAV 23/ TRAJ33 and TRBV 28/ TRBJ 1-2), 3A (TRAV4/ TRAJ30 and TRBV28/ TRBJ1-1), F11 (TRAV8-4/ TRAJ30 and TRBV24-1/ TRBJ 1-2), HA1.7 (TRAV 8-4/ TRAJ 48 and TRVB 28/ TRBJ 1-2). The sequence of the extracellular

domains of the HA$_{306-318}$ specific TCRs DC C8 (TRAV 8.4/ TRAJ9 and TRBV13 /TRBJ2.3) and DC D10 (TRAV9.2/ TRAJ 23.1 and TRBV5/ TRBJ2.3) were obtained by RNA-extraction and cloned into pGMT7 as described in 2.1.4.1 and 2.1.4.2. All TCR sequences were engineered to encode an artificial inter-chain disulphide bond near the C-terminus of each chain (Boulter et al. 2003). Sequences were confirmed by automated DNA sequencing by Central Biotechnology Services (CBS).

2.2.3. Expression of MHC-II α- and β-chains in E.coli

Transformation of DE3 *E.coli* "Rosetta" (LifeTechnologies^{TM}) was performed as described in 2.1.4.3 and grown over night on LB agar plates containing 50 μg/ ml Carbenicillin (BiochemicalDIRECT^{TM}). In order to confirm the expression of the transfected vectors, single colonies were cultured in TYP media containing 50 μg/ ml Carbenicillin (BiochemicalDIRECT^{TM}) at 37 °C and 220 rpm in an orbital shaking incubator to generate starter cutlures. Once optical density reached 0.4-0.6 expression of the transfected vector was induced in a 5-10 ml aliquot by addition of 0.5 mM Isopropyl *β*-D-thio-galactoside, (IPTG, Fisher Scientific) for 3 h. 10-20 μl of culture with and without 0.5 mM IPGT were analysed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie Blue staining using SimplyBlue™ SafeStain (Invitrogen™). The degree and purity of protein expression were determined by intensity of band observed. Remaining starter cultures were then added to 1L of TYP containing 50μg/ ml Carbenicillin (BiochemicalDIRECT^{TM}) and cultured at 37 °C and 220 rpm in an orbital shaking incubator until optical density reached 0.4-0.6 indicating the exponential growth phase of the DE3 *E.coli.* Expression of protein in the form of inclusion bodies was induced by the addition of 0.5 M IPTG for 3 h. Cells were harvested by centrifugation for 20 mins at 4000 rpm in a Sorvall Legend centrifuge with a Hereaus

6445 rotor. Pellets were resuspended in lysis buffer and sonicated for 30min on ice using a Sonopulse HD 2070 coupled to a MS73 probe (Bandelin, Germany) set to 60% power and using 2 s intervals before being incubated with 0.1g/L DNAse (Fisher Scientific) for 30 min at RT. The suspension was then added to 100 ml wash buffer and centrifuged for 20 min at 4 °C and 10000 rpm using an Evolution RC centrifuge in combination with a SLA-15000 rotor (Sorvall®). Pellets were resuspended in 100 ml wash buffer before centrifuging as before. Pellets were then treated with 100 ml re-suspension buffer and aliquot of 20 μl was taken before centrifuging as before. Pellets containing MHC-II inclusion bodies were resuspended in urea buffer. Inclusion body suspension were centrifuged for 30 min at 4000 rpm in a Sorvall Legend centrifuge with a Hereaus 6445 rotor and pellets were discarded. MHC α- and β-chain inclusion bodies were treated with TDM-8 mixed bed resin beads (Sigma-Aldrich) in order to neutralise any free urea radicals. Inclusion bodies were filtered and further purified by ion exchange using a 5 ml Hi-Trap column (GE Healthcare, UK) using a ÄKTA FPLC (GE Healthcare, UK) and the Unicorn software (GE Healthcare, UK). This purification step was not necessary for TCR α- and β-chain inclusion bodies. Protein concentrations in the supernatant were measured using Nanophotometer® by Implen and protein quality was verified by SDS-PAGE and Coomassie-staining as described above.

2.2.4. *Refolding, purification and biotinylation of pMHC-II complexes*

MHC-II refold buffer was prepared in advance and left to chill at 4 °C before adding 0.5 mg/l peptide. 3-5 mg/l DR1α-chain and 3 mg/l DR1β-chain inclusion bodies were incubated at 37° C for 15 mins and added simultaneously and dropwise. Refolds were stirred for 1 h and incubated 72 h to 2 weeks at 4 °C. Refolds were then filtered using a cellulose nitrate membrane filter with 0.45 μm pore size (Fisher Scientific) before being

concentrated to a volume of 30-50 ml using a viva flow 200 module with polyethersulfone membrane (Sartorius, France). Concentrated refolds were then washed with PBS and further concentrated to a volume 700 μl using a vivaspin column with a 10kDa cut-off (Sartorius, Fracne). pMHC-II with a biotinylation tag at the C-terminus of the α-chain were biotyinlated by adding 100μl Biomix A (0.5 M Bicine Buffer, pH8.3), 100 μl Biomix B (100mM ATP, 100 mM MGo(Ac)2), 100μld-Biotin and 2 μl BirA enzyme (all reagents from Avadin) and incubating at RT overnight. Biotinylated pMHC-II were washed with PBS and concentrated to a volume of ≤1 ml. Both biotinylated and non-biotinylated pMHC-II were further purified by size exclusion on a superdex 200HR column (GE Healthcare, UK) using a ÄKTA FPLC (GE Healthcare, UK) and the Unicorn software (GE Healthcare, UK). Protein concentrations in the supernatant were measured using Nanophotometer® by Implen and protein quality was verified by SDS-PAGE and Coomassie-staining as described above.

2.2.5. *Biophysical analysis of pMHC-II-TCR interaction using surface plasmon resonance*

SPR experiments investigating pMHC-II-TCR interactions were conducted on a BIAcore 3000 or T100® (GE Healthcare) using a CM5 sensor chip (GE Healthcare). In order to activate the sensor chip, its surface was exposed to an amine coupling solution (GE Healthcare) composed of 10 mM N-(3-dimethylaminopropyl)-N3-ethylcarbodiimide (EDC) and 400mM N-hydroxysuccinimide (NHS). Streptavidin (200 μg/ ml in 10 mM acetate pH 4.5 (Sigma-Aldrich®)) was covalently linked to the activated sensor chip surface of all four flow-cells until 5000 response units (RU) were achieved. In a last step, 1 M ethanolamine hydrochloride (GE Healthcare) was used to deactivate remaining active groups left on the chip surface. Streptavidin coated chips were loaded with

biotinylated pMHC molecules at a flow rate of 10 µl/min until 400-600 RUs were immobilised on each flow-cell. Anti-pan-HAL-DR L243 antibody injected at 30 µl/min. Results were analysed using BIAEvaluation 3.1.

2.2.6. *Peptide stability assay*

A 96-well microtiter plate was blocked with 250 µl blocking buffer for 3 h at 37 °C. Serial dilutions at three times the desired final concentration were prepared in peptide binding buffer. Half area EIA/RIA plates (Costar) were coated with 0.5 µg L243 per well in 50 µl peptide binding buffer. 150 µl of blocking buffer was transferred from the microtiter plates to L243 coated EIA/RIA plates. EIA/RIA plates were bashed dry and 0.05 µg HLA-DR1$^{\text{BT-CLIP}}$ was added per well in 40 µl peptide binding buffer before adding 20 µl of diluted was added to each well. Microtiter and EIA/RIA plates were covered in tinfoil and incubated 15-24 h at 37 °C. L243 coated plate was emptied and washed once with PBS. 10 µl neutralising buffer was added to each well on the microtiter plate, contents were transferred onto L243 coated plate and incubated 1 h at 37 °C. Plates were emptied and washed 3 times with wash buffer and 4 times with PBS. 50 µl streptavidin-HRP (R&D Systems) diluted 1:40 in PBS were added and plates were incubated 20-30 min at room temperature in the dark. Plates were washed 3 times in wash buffer and 50 µl of chromogen solutions A and B (mixed 1:1) were added. Plates were incubated in the dark until fully developed. The reaction was stopped by adding 25 µl of stop solution (2N H$_2$SO$_4$, R&D Systems) per well. Plates were read at 450 nm and 570 nm (background noise reference) using an iMARK$^{\text{TM}}$ Microplate reader (Biorad).

2.3. Mammalian cell culture

2.3.1. Buffer, reagents and media

R0: RPMI-1640 medium (Gibco®), 100IU/ ml penicillin (LifeTechnologies), 100IU/ ml streptomycin (LifeTechnologies) and 2 mM L-glutamine (LifeTechnologies).

R10: R0 medium supplemented with 10% heat inactivated foetal calf serum (FCS, Gibco®)

R5: R0 medium supplemented with 5% FCS (Gibco®)

R2: R0 medium supplemented with 2% FCS (Gibco®)

AB media: R0 medium supplemented with 5% filtered human AB serum (Welsh Blood Transfusion Services, WBTS)

T cell clone media: R0 supplemented with 10% FCS (Gibco®), 0.02M HEPES (Sigma), 1mM Non-essential amino acids (Gibco®), 1mM Sodium pyruvate (Gibco®), 20IU/ ml human recombinant IL-2 (Proleukin®, University Hospital of Wales (UHW) pharmacy)

Red blood cell (RBC) lysis buffer: 155 nM NH_4Cl (Sigma-Aldrich®), 10 nM $KHCO_3$ (Fisher Scientific), 0.5 M EDTA (Sigma-Aldrich®), pH 7.2-7.4.

Dextramer buffer: 0.05 M Tris-HCL, 15 mM NaN_3, 1% bovine serum albumin, pH 7.2

FACS buffer: PBS supplemented with 2% FCS (Gibco®).

2.3.2. *Immortalized cell lines and CD4$^+$ T cell clones*

An overview of the HLA-type of all cell lines used in this thesis is shown in **Table** 2**.**1.

HOM-2: The human, homozygous B-LCL HOM-2 is derived from a HLA-DR1 individual and was purchased from the European Collection of Cell Cultures (Public Health England (PHE)). It was maintained in R10 medium.

CD4$^+$ T cell clones specific for the HLA-DR1 restricted peptide HA306-318: CD4$^+$ T cell clones 2C5 was provided by Professor Andrew Godkin while DC C8 and DC D10 were provided by Garry Dolton. T cell clones were maintained in T cell clone medium and restimulated with allogeneic feeder cells every 2-4 weeks (see section 2.3.5).

Cell line /Donor	HLA-A	HLA-B	HLA-C	HLA-DR	HLA-DP	HLA-DQ
HOM-2	03/03	27/27	01/01	01/01	n/a	05/05
Donor 1	n/a	n/a	n/a	01/09	01/03	n/a
Donor 2	03/11	07/35	n/a	01/15	05/06	n/a
Donor 3	11/24	07/51	n/a	1/12	05/03	n/a

Table 2.1: HLA type of LCL used in this thesis. HLA types of donors 1, 2 and 3 were determined by PCR by the Welsh Blood Transfusion Service. The HLA type of HOM-2 was provided by the European Collection of Cell Cultures.

2.3.3. Blood donors

Buffy coats used to restimulated cell lines and T cell clones were obtained from the WBTS. Peripheral mononuclear cells (PBMC) used to generate CD4$^+$ T cell lines were obtained from healthy, HLA-DR1 positive, laboratory staff following informed consent.

2.3.4. Isolation of PBMCs from whole blood

PBMCs were isolated from 50 ml buffy coats (originating from 500 ml blood from healthy blood donors) purchased from the WBTS or from 50 ml fresh blood collected by venepuncture into a Falcon^{TM} tube (BD Biosciences) containing heparin (LEO Laboratories Ltd) at a concentration of 100IU/ ml of blood. Buffy coats were diluted 1:1 in PBS while fresh blood was processed neat. The blood was separated using Ficoll-Hypaque (Lymphoprep^{TM}, Axis-Shield) density gradient centrifugation for 20 min at 2000rpm in a Hereaus Megefuge 16R with a Hereaus 75003629 rotor) without breaks. The layer containing PBMCs was situated at the interface of the Ficoll-Hypaque (lower liquid layer) and the serum (upper liquid layer) and was aspirated carefully and transferred

into a fresh tube. Cells were then washed once in R0 before being treated with 5-10 ml of RBC lysis buffer. Cells were washed again in R0 before being resuspended in R0, R10 or AB media.

2.3.5. Expansion of CD4$^+$ T cell clones

CD4$^+$ T cell clones were restimulated every 2-4 weeks using allogeneic feeders. 10^6 CD4$^+$ T cell clones were resuspended in 10-15 ml T cell clone media supplemented with 1μg/ml PHA (Sigma) and 10-15x10^6 irradiated PBMCs from three different donors were added. Cells were incubated in T25 flasks for 7 days at 37 °C with half the media replaced on day 5. On day 7, CD4$^+$ T cell clones were plated on a 24 well plate at a cell concentration of 3-4x10^6 cells/ ml in 2 ml T cell clone media/well. Cells were then incubated for a further 7 days with half the media replaced every 3-4 days. After 14 days, CD4$^+$ T cell clones were ready to be used in *in vitro* assays.

2.3.6. Generation of short term CD4$^+$ T cell lines

PBMCs from healthy, HLA-DR1$^+$ volunteers were obtained as described in section 2.3.4. Cells were plated on round bottom 96 well plates at a concentration of 2x10^6 cells/ml in 100 μl AB media per well. Peptides (Peptide Protein Research Ltd) were added at a final concentration of 1-10 μg/ml. Where peptide pools were used, the final concentration of each individual peptide in the mixture was 10 μg/ ml. Where appropriate, PHA (Sigma) was added at a concentration of 1 μg/ml as a positive control. Outer wells were filled with sterile PBS to avoid evaporation of media and were incubated at 37 °C. On day three10 μl Cellkines (ck, Helvetica Healthcare) were added to each well. On day 6, 100 μl of AB media containing 40IU/ml IL-2 (Proleukin®, UHW pharmacy) were added per well. On day 9, half the media was replaced by fresh AB media containing 40 IU/ml IL-2

(Proleukin®, UHW pharmacy). This was repeated every 3-4 days when cells were cultured for more than 14 days. Cell lines were ready to be used in i*n vitro* assays from day 12 to 14 onwards.

2.3.7. *Peptide titration assay*

CD4$^+$ T cell clones were washed in R0 and rested in R2 overnight. APC were washed in R0 and 50x10^4 cells were added per well in a round bottom 96 well plate. Peptides were added at varying concentrations ranging from 10^{-9}-10^{-3}M and cells were incubated at 37 °C for 1-2 h. CD4$^+$ T cell clones were washed in R0 and 25x10^4 cells were added per well and plates were incubated overnight at 37 °C. 75 µl of the cell supernatant was collected and IFN-γ released in response to the peptide was measured in an ELISA as described in 2.3.9.

2.3.8. *INFγ-ELISpot*

Cells from short term CD4$^+$ T cell cultures were assayed for peptide responsive cells in an Interferon-γ Enzyme-Linked ImmunoSpot assay using the Human IFN-γELISpotBASIC (ALP) kit (Mabtech, Sweden) following manufacturer's instructions with slight modifications to the frequency of washing steps and incubation times. MultiScreen$_{HTS}$ IP Filter plates with a 0.45 µm pore size hydrophobic PVDF membrane (Merck Millipore) were pre-treated with 15 µl/well 70% ethanol (Fisher Scientifics) and washed twice with 150 µl/well. Capture antibody was diluted to 15 µg/ ml before coating plates with 50 µl/well and incubating over night at 4 °C. Wells were flicked empty and washed four times with 150 µl PBS and blocked by adding 100 µl AB media per well for 1h at 37 °C. Short term cell lines were washed and 150,000 cells were split between two wells before adding peptide or peptide pool at 1-10 µg /ml to one of the wells. In cases

where antigen presenting cells (APC) were used, these were washed in R0, pulsed with peptide or peptide pools at 1-10 μg/ ml and washed three times in R0. 50,000 APC were added per well prior to adding short term cell lines. Where appropriate, PHA (Sigma) was added at a concentration of 1μg/ ml as a positive control. Plates were incubated 4-18 h at 37 °C wrapped in tin foil. Plates were flicked empty and washed 5 times with 150 μl PBS per well. Detection antibody was diluted to 1 μg/ ml in PBS and 50 μl were added to each well. Plates were incubated 2 h at RT or 1 h at 37. Plates were flicked empty and washed 4 times with 150 μl PBS per well. Streptavidin-ALP was diluted 1:1000 in PBS and 50 μl were added per well. Plates were incubated at RT away from light for 30-60 min. AP Conjugate substrate (Biorad) was prepared according to manufacturer's instructions. Plates were flicked empty and washed 4 time with 150 μl PBS per well. 100 μl of substrate solution were added to each well and incubated at RT away from light for at least 10 min or until spots appeared. Plates were washed extensively in tap water and left to dry. Spot forming units (SFU) per 100,000 PBMCs were determined using an AID EliSpot reader in conjunction with the AID EliSpot software. Dry plates were stored at RT protected from light.

2.3.9. IFNγ-ELISA

Cell supernatants were assayed for Interferon-γ secretion using the IFN-γ Duoset Enzyme-linked Immunosorbent assay (ELISA) kit (R&D Systems) following the manufacturer's recommandations. All wash steps were performed using 300 μl wash buffer (0.05% Tween® 20 (Sigma Aldrich) in PBS) per well in conjunction with a MultiWash III plate washer (TriContinent). Briefly, 96 well half area EIA/RIA plates (Costar) were coated with 50 μl capture antibody per well at a concentration 4 μg/ml, sealed and incubated at 4 °C overnight. Following 3 wash steps the plates were blocked

with 1% BSA (Sigma-Aldrich®) in PBS, sealed and incubated 1 h at RT. Plates were washed three times and 50 μl of cell supernatant were added to each well. At the same time, IFN-γ standards provided with the kit were added at range of concentrations (1000 pg/ml, 500 pg/ml, 250 pg/ml, 125 pg/ml, 62.5 pg/ml, 31.2 pg/ml, and 15.6 pg/ml). Plates were sealed and incubated 75 min at RT. Plates were then washed three times and 50 μl of detection antibody was added per well at a concentration of 200 ng/ml. Plates were sealed and incubated 75 min at RT before being washed three times. 50 μl of a 1 in 40 dilution of streptavidin-HRP provided with the kit were added to each well. Plates were sealed and incubated 20 min at RT in the dark. Substrate solution was prepared by mixing colour reagent A (stabilised peroxide solution, R&D Systems) and colour reagent B (stabilised chromagen solution, R&D Systems) with the kit in a 1:1 ratio. Plates were washed and 50 μl of substrate solution were added to each well, keeping the plate away from direct light. Plates were sealed and incubated 10 min at RT in the dark or until no more colour change could be observed. The reaction was stopped by adding 25 μl of stop solution (2N H_2SO_4, R&D Systems) per well. Plates were read at 450 nm and 570 nm (background noise reference) using an iMARK™ Microplate reader (Biorad).

2.3.10. Multimerisation of pMCH-II

Tetramerisation of pMHC-II was performed as previously published (Altman et al. 1996). 1 μg of biotinylated pMHC-II were incubated with 1.5 μl streptavidin-PE (LifeTechnologies) for 20 min at 4 °C in the dark. This step was repeated 4 times in order to obtain a streptavidin-fluochrome ratio of 5:1. pMHC-II tetramers were then diluted in PBS to give a final concentration of 0.1 μg/μl and centrifuged for 1min at 16,000 rpm in VWR Microstar tabletop centrifuge in order to separate aggregates.

pMHC-II dextramers were assembled by incubating 1 µg biotinylated pMHC-II to 2 µl dextran-PE (Immudex) for 30 min on ice before topping up to a final volume of 10 µl with dextramer buffer. pMHC-II-dextramers were centrifuged for 1min at 16,000 rpm in VWR Microstar tabletop centrifuge in order to separate aggregates.

2.3.11. pMHC-II dextramer staining of CD4$^+$ T cell clones and CD4$^+$ T cell lines

Cells were pre-treated with PKI (protein kinase inhibitor Destatinib Axon Meddchem, Reston) at a final concentration of 50 nM for 30 min at 37 °C. 10 µl of pMHC-II tetramers or dextramers, respectively were added and incubated 20-30 min in the dark on ice. Cells were washed once in FACS buffer and once in PBS before adding LIVE/DEAD® Violet stain (LifeTechnologies) and incubated 5 min in the dark at room temperature. Appropriate surface stain antibodies (anti-CD14-PB (Biolegend), anti-CD19-PB (Biolegend), anti-CD3-PerCP (Miltenyi Biotec) and anti-CD4-APC (Miltenyi Biotec)) were added and incubated 20 min in the dark on ice. Cells were washed in FACS buffer prior to analysis by flow cytometry. Flow cytometric analysis was performed on a Canto BD FACSCanto™ II flow cytometer (BD Biosciences) and data was analysed using FlowJo (Tree Star Inc). Sorting of short term T cell lines was performed on a BD LSRFortessa™. Cells were sorted directly into lysis buffer (Qiagen) and data was analysed using FlowJo (Tree Star Inc).

2.3.12. Clonotyping of FAC sorted short term T cell lines

2.3.12.1.Generation of cDNA

Total RNA was sorted samples was extracted using the RNeasy kit (Qiagen) and cDNA was generated using the SMARTER™ kit (Clontech). Briefly, extracted mRNA was

incubated with Oligo-dT in a thermal cycler for 3 min at 72 °C followed by 2 min at 42 °C to anneal the oligo-dT primer to the oligo-A tail of the mRNA. 8 µl master mix containing 4 µl 5X First Strand buffer, 0.5 µl 100 mM DTT, 1 µl 20 mM dNTP, 0.5 µl RNAse Inhibitor (20 U) and 2 µl SMARTScribe RT (100 I) and 1 µl oligo-A primer II were added. Tubes were incubated 90 min at 42 °C followed by 10 min at 70 °C.

2.3.12.2. Amplification of cDNA

In a first PCR, the entire TCRVB region and part of the TCRCB region was amplified by mixing 2.5 µl of cDNA sample with 10 µl 5x Phusion® Green buffer, 0.5 µl 100 mM DMSO, 1 µl 20 mM dNTPs, 5 µl 10X Universal Primer A (forward primer), 1 µl Primer C_β-R1 (reverse primer), 0.25 µl Phusion® HF DNA polymerase and topped up to 50 µl final volume with H_2O. Samples were incubated at 94° C for the initial denaturation followed by 30 cycles of 30 s at 94, 30 s at 63 and 3 min at 72 °C before being incubated at 72 °C for the final extension.

In a second PCR, 2.5 µl of sample from the first PCR were mixed with the same reagents as stated above replacing the forward primer with 1 µl Primer A short and the reverse primer with Primer C_β-R2. Samples were incubated at 94 °C for the initial denaturation followed by 30 cycles of 30 s at 94 °C, 30 s at 66 and 3 min at 72 °C before being incubated at 72 °C for the final extension. Samples were analysed by electrophoresis on a 1% agarose gel. Amplified samples of the expected size were extracted and purified using the PCR Clean-up kit (Clontech).

2.3.12.3. Sequencing of TCRB cDNA

PCR products were cloned into a PCR-Blunt II-TOPO® vector using the Zero Blunt® TOPO® PCR cloning kit (LifeTechnologies) as described in. One Shot® TOP10 *E.coli*

(LifeTechnologies) were transformed as described in 2.1.4.3 and grown on LB-Agar plates containing 50 µg/ ml kanamycin. Individual colonies were screened by colony PCR by mixing each with 1 µl M13 forward primer, 1 µl M13 reverse primer and 23 µl DreamTaq® Green master mix and run on a thermocycler following manufacturer instructions. Positive colonies were sent off to be sequenced (Eurofins, Germany).

2.4. Bioinformatical and statistical analysis

2.4.1.1. Analysis of sequenced TCRB cDNA

Sequences were visualised using the BioEdit software (http://www.mbio.ncsu.edu/BioEdit/bioedit.htm) and V-, D-, J-segments were identified using IMGT/V-QUEST (http://www.imgt.org/IMGT_vquest). Frequency analysis was performed using Microsoft Office™ Excel.

2.4.1.2. Analysis of influenza HA derived epitopes

HA sequences of all identified H2N3 strains found on the Influenza Research Database (fludb.org) were aligned using the inbuilt multiple sequence aligned tool. Aligned sequences were visualised using either BioEdit or Jalview (www.jalview.org). Consensus sequence and Shannon entropies were calculated using BioEdit. Epitope conservation frequencies were calculated using the inbuilt tool on the Immune Epitope Databse (www.iedb.org) while conseration frequencies of individual residues were calculated using Microsoft Office™ Excel.

2.4.1.3. Figures and other data analysis

Figures were generated using Microsoft Office^{TM} Excel unless otherwise mentioned. Pearson's coefficients were calculated using Microsoft Office^{TM} Excel. EC_{50} and IC_{50} values were calculated using GraphPad Prism 5. Contact tables from crystal structures were generated using the CCP4 package. Figures visualising crystal structures were generated using the PyMOL software.

3. Manufacture of soluble MHC-II molecules in insect cells

3.1. Background

Investigating the nature of TCR recognition of pMHC-II is an important step towards understanding CD4$^+$ T cells and their role in disease and vaccination. Soluble pMHC-II molecules are essential in these studies as they allow for detection of antigen specific CD4$^+$ T cells by flow cytometry using fluorochrome-conjugated multimerized versions of pMHC-II. Furthermore, soluble pMHC-II is essential for studying biophysical properties of TCR-pMHC-II interactions using methods such as Surface Plasmon Resonance (SPR) and X-ray crystallography..

3.2. Introduction

3.2.1. *Limitations of bacterial and mammalian cell based expression systems*

Over-expression of soluble, functional proteins in cells is a well-established method of generating sufficient material for a wide range of applications. Generally, cellular expression systems can be divided into three categories: bacterial, mammalian cell and insect cell systems. All three have in common that the gene encoding the protein of interest has to be cloned into an expression vector which can then be transduced or transfected into the host cell prior to protein expression.

Bacterial expression systems are the most widespread. Expression vectors are relatively easy to assemble and thanks to the wide range of bacteria strains commercially available, many proteins can be over-expressed. Most commonly *E. coli* strains modified for the specific requirements (mammalian proteins, toxic proteins, etc.) are used. Culturing *E. coli* is simple, economic and requires only basic laboratory equipment. On the other hand, non-bacterial proteins expressed in bacteria are often produced as non-refolded insoluble inclusions bodies and require artificial refolding *in vitro*. MHC-II α- and β-chains fall into this category (Frayser et al. 1999). This is a labour intensive process that often goes hand in hand with low yields of protein. Furthermore, bacteria are incapable of post-translational modifications (PTMs) such as glycosylation which can be crucial for the function of certain proteins.

Mammalian proteins expressed in mammalian cells are usually soluble, properly refolded and bear natural PTMs. For uses where PTMs are more of a hindrance, such as crystallographic studies, they can be removed using specific enzymes. However, handling of mammalian cell cultures requires more expertise and equipment than bacterial cells.

Namely, the culturing of mammalian cells in a CO_2 infused incubator for protein expression on laboratory scale can be costly and challenging. Furthermore, yields can be small depending on the individual protein (Verma, Boleti, and George 1998).

Despite recent advances, expression of pMHC-II in *E. coli* remains a slow and laborious process while mammalian cell expression systems are costly. Insect cell expression systems, on the other hand, are becoming increasingly easy to implement. Thus, it was attractive to establish an insect cell based expression system for generating soluble pMHC-II using a Baculovirus Expression Vector system (BEVS).

3.2.2. *BaculoDirect^{TM} expression system*

Until recently, most BEVS relied on the generation of bacmids in competent *E. coli* under the control of a helper plasmid. This requires the selection of bacmids which have a successfully inserted the gene of interest (GOI). With the development of novel cloning techniques such as the Gateway® cloning technology, generation of Baculovirus expression vectors has been simplified. This cloning technology is based on the LR reaction used by phage Lambda to insert segments of its own genome into its bacterial host's genome (Liang et al. 2015). The donor (i.e. phage DNA segment) is flanked by so called *att*L sites which are homologous to *att*R sites on the acceptor (i.e. bacterial DNA strand). In the presence of the LR Clonase^{TM} enzyme mix, these regions align and DNA strands in between are interchanged between donor and acceptor. The BaculoDirect^{TM} expression system we used here to generate soluble pMHC utilises Gateway® technology to assemble the baculovirus expression vector in vitro (shown in Figure 3.1A). The so called entry clone serves as donor DNA and encodes the GOI flanked by *att*L sites. The entry clone is generated by inserting the GOI into the pENTR11 vector (available from LifeTechnologie^{TM}) using standard restriction digestion and ligation technology. The

Baculovirus backbone (i.e. the acceptor) used in this system is based on the *Autographa californica* multiple nuclear polyhedrosis bacoluvirus (AcMNPV). As shown in Figure 3.1B, modifications for heterologuous gene expression included replacing the viral polyhedrin gene with a Gateway® compatible negative selection cassette consisting of the Herpes simplex virus thymidine gene (HSV1 tk) under the control of an immediate-early promotor (P_{IE}) followed by a lacZ gene segment. This negative selection cassette is flanked by attR sites and replaced by the GOI during the LR reaction. During generation of high titre viral stocks, insect cells are cultured in the presence of gancliclor which allows the elimination of any insect cells transfected with vectors that have not undergone successful homologous recombination. This nucleoside analogue undergoes phosphorylation by HSVtk leading to its incorporation into DNA and inhibition of DNA replication (Godeau, Saucier, and Kourilsky 1992). Only cells carrying the desired expression vector replicate, generate new viral particles and express the GOI. As an additional negative selection mechanism, the BacoluDirect^{TM} expression vector is linearized by restriction digestion with Bsu36 I. During homologous recombination, the vector becomes circular again through insertion of the GOI. The negative selection cassette is followed by a V5 epitope and a His-tag which can be used for purification of the protein encoded in the GOI.

Following successful assembly of the bacmid and verification by PCR, it is transfected into insect cells using a lipid mediated method in order to generate a first generation of recombinant baculovirus called P1. This low titre virus is then used to generate higher titre stocks by successive transfection of insect cells. Once the titre of the desired generation of baculovirus (generally P3-P5) is determined it can be used to express the desired protein in insect cells.

Figure 3.1: The BaculoDirectTM expression system uses Gateway® technology. (A) The *att*L and *att*R sites on entry clone and BaculoDirect^{TM} backbone, respectively are homologues. In the presence of LR Clonase II enzyme mix, the DNA strands in between the LR sites are swapped over generating the desired expression vector and a by-product. (B) The negative selection cassette on the BaculoDirect^{TM} expression bacmid is flanked by *att*R sites. The cassette itself encodes the Herpex simplex virus thymidine kinase (HSV1 tk) and a lacZ gene. It also includes a Bsu36 I linearization site. Following the downstream *att*R site, a V5 epitope and His-tag are pre-encoded into the bacmid. The successfully inserted GOI is under the control of the strong baculoviral P_{PH} promotor. Figure adapted from BaculoDirect^{TM} Baculovirus Expression System (2010, LifeTechnologies).

3.3. MHC-II constructs used here

The HLA-DR1α and HLA-DR1β constructs used here are shown in

Figure **3.2.** In order to facilitate secretion of HLA-DR1 molecules into insect cell culture medium, a honey bee melittin leader sequence (MYIYADPSPA) was added to the 5' end of the DNA construct of both chains. Originating from another insect species (*Apis melifica*), this signal peptide targets the protein product to the secretory pathway following translation and has been shown to increase protein yield (Tessier et al. 1991). In order to facilitate assembly of both chains during expression in insect cells, one half of a leucine zipper was added to the α- and β-chain, respectively. This has been shown to increase stability of heterologous expressed proteins (Scott et al. 1996). In addition, the target sequence of the BirA enzyme (GLNDIFEAQKIEWHE) was added to the C-

terminus of the α-chain to enable biotinylation of pMHC II monomers in order to use them in a variety of techniques such as surface plasmon resonance (SPR) and multimer staining of CD4[+] T cells. While MHC II molecules have been shown to refold without a peptide in the binding groove, the addition of such increases their stability (H Kozono et al. 1994). Therefore, a CLIP sequence was added to the β-chain. This helps the generation of correctly refolded pMHC II molecules with a placeholder peptide in the binding groove. The CLIP peptide was followed by a Thrombin cleavage site (LVPRAGS (Waugh 2011)) allowing its removal and exchange with any peptide of interest binding to HLA-DR. The β-chain sequence was followed by the other half of the leucine zipper. Two separate β-chain constructs were designed encoding the HLA-DRB1*0101 and HLA-DRB1*0401 genes, respectively. Our laboratory routinely utilises immunoprecipitation to purify HLA-DR1, therefore, stop codons were added to all three constructs since the V5- and His-tags were not needed.
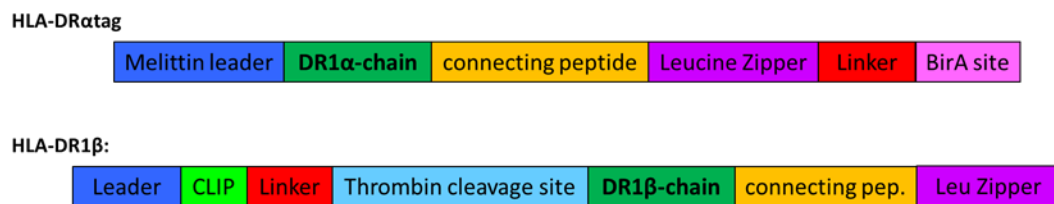
Figure 3.2: Schematic overview of constructs used for expression of HLA-DR1 in insect cells. (Honey bee) melittin leader sequence: targeting of protein for secretory pathway. Leucine zipper: stabilizes facilitates assembly of DR1 α- and β-chain. BirA site: biotinylation site. CLIP: peptide in MHC II groove, placeholder for peptide of interest. Thrombin cleavage site: removal of CLIP peptide prior to peptide exchange.

3.4. Aims

The first aim for this project was to generate baculovirus constructs encoding HLA-DR1αtag and HLA-DR1β, respectively, using the BaculoDirect^{TM} system and use them to generate high titre baculovirus stocks for heterologous protein expression in insect cells.

The second aim was to over express soluble HLA-DR1^{CLIP}tag in *Sf9* insect cells and to purify them from the supernatant using immunoprecipitation.

The third aim was to exchange the bound CLIP peptide from HLA-DR1tag with a peptide of interest and to validate it by staining antigen specific CD4$^+$ T-cell clones with pMHC-II tetramer reagents.

3.5. Results

3.5.1. Generation of BaculoDirect^{TM} constructs

Constructs for HLA-DR1αtag chain, HLA-DR1β and HLA-DR4β chains were cloned into the pENTR11 vector as described in Chapter 2.1.4 and verified by sequencing (full DNA and amino acid sequences shown in **Supplementary figure S**1 and **Supplementary figure S**2). The resulting entry vectors were then used to assemble the baculovirus expression vectors. An LR reaction was conducted using the generated pENTR11 plasmids as entry clones and the BaculoDirect^{TM} linear DNA. As a positive control the pCat entry clone (provided with the BaculoDirect^{TM} expression kit) was used while in the negative control the HLA-DR1β pENTR11 entry clone was used but replacing the LR enzyme mix by TE buffer. A PCR was performed as described in 2.1.4.5 using a forward primer binding to the polyhedron promotor region and a reverse primer binding within the V5 tag. Electrophoresis showed that the LR reaction was successful for all three constructs (Figure 3.3).
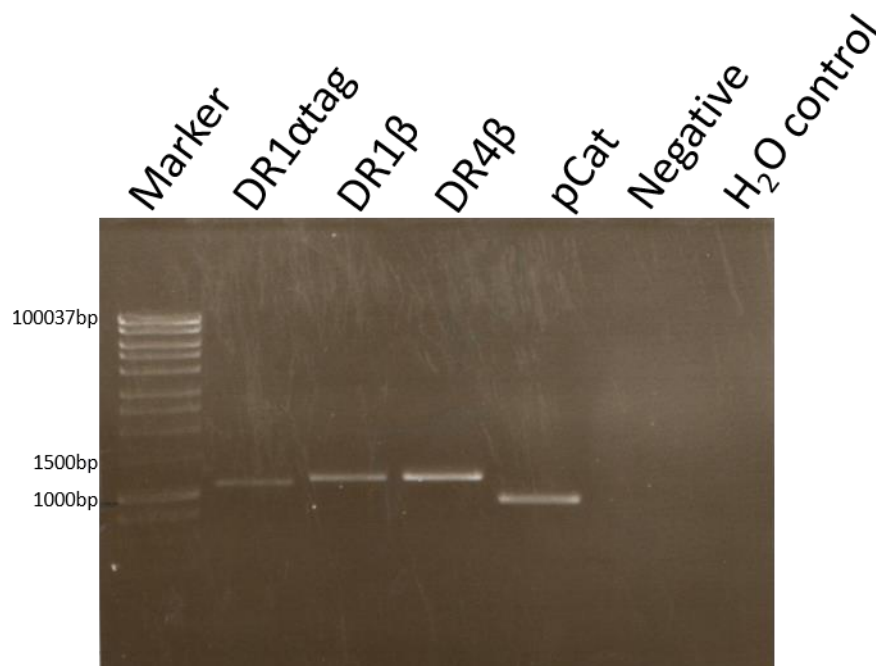
Figure 3.3: Verifying LR reaction using PCR. PCR products were analysed on a 1% agarose gel. Negative: LR reaction where ddH$_2$O was used instead of an entry vector. H$_2$O control: PCR conducted with ddH$_2$O instead of template DNA. Expected sizes of correct bands were: DR1αtag – 1040bp; DR1β – 1090bp, DR4β – 1090bp; pCat – 870bp.

3.5.2. *Generation of high titre viral stocks for protein expression*

The four baculovirus expression vectors generated in 3.5.1 were then used to generate baculoviral stocks. *Sf9* cells (insect cell line derived from the fall army worm, *Spodoptera frugiperda*) were transfected as described in Chapter 2.1.5 and inspected daily under the microscope. However, no clear signs of infection were visible. Supernatants containing the P1 generation of baculoviral stocks were collected 72 h post transduction. P2 viral stocks were generated by seeding *Sf9* cells in suspension culture with P1 viral stock and harvesting the supernatant after 96 h. A small sample of the P2 generation of each construct was taken and viral DNA was isolated as described in Chapter 2.1.7 and amplified by PCR. Sequencing confirmed the presence of each construct in the corresponding viral stocks. In order to generate stocks of larger volume and higher titre,

baculoviral stocks were amplified in several rounds by seeding increasing numbers of *Sf9* cells with viral stocks of the previous round (see Figure 3.4).

Viral titres were measured in round three (P3 viral stocks) or higher. As described in Chapter 2.1.8, an ELISA based method detecting the presence of the viral gp64 protein on the surface of *Sf9* cells was used in order to determine viral titres. Infected cells are visible as blue foci of infection (see Figure 3.4) and the viral titre of each tested stock can be calculated from the number of foci present. Titres are given as infectious units (IFU)/ml.

During the generation of these constructs and high titre viral stocks, I began to focus on HLA-DR1 to characterise new hemagglutinin epitopes as part of chapter 6. Thus, insect cell expression of HLA-DR4 was halted to enable more resources to be directed towards the generation of soluble HLA-DR1 to complement my other research projects.
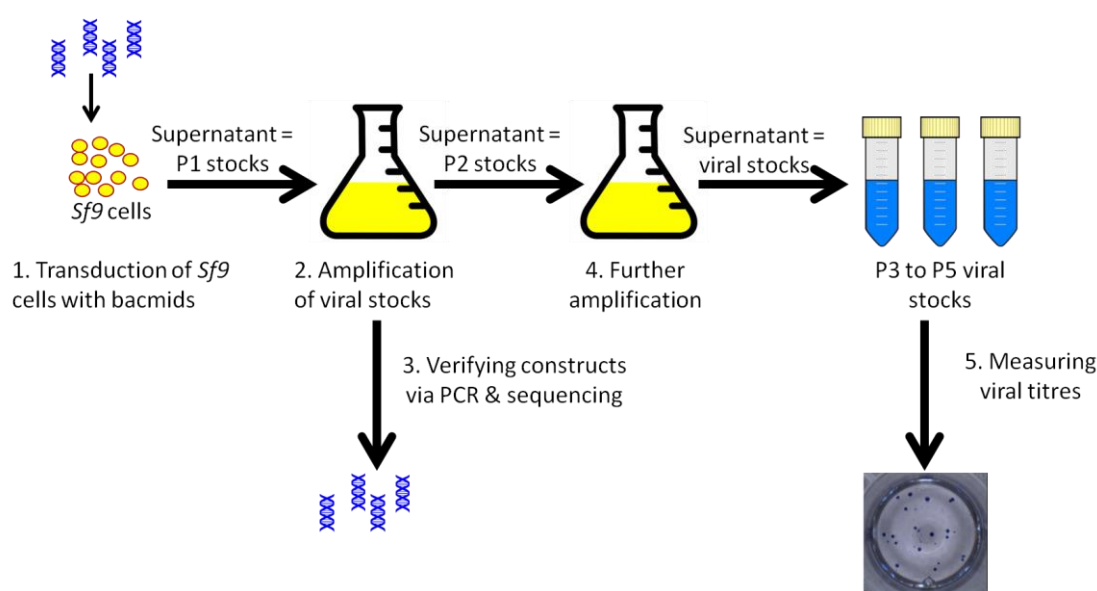
Figure 3.4: Generation of high titre viral stocks and determination of viral titre. High titre viral stocks were generated by successive rounds of infection of *Sf9* cells. Cells infected with baculovirus express gp64 on their cell surface (stained in blue) and form foci of infection.

3.5.3. Expression of HLA-DR1^{CLIP}tag in insect cells and purification from cell supernatants

As a first expression test, *Sf9* cells were infected with equal amounts of HLA-DR1αtag and HLA-DR1β baculovirus. As a starting point a multiplicities of infection (MOI) of 0.5 and 1 were chosen, i.e. one viral particle of each construct per two and one *Sf9* cells, respectively, were added. This was lower than the 3-20 reported in literature (Quarsten et al. 2001; Radu et al. 1998). However, Zhan and colleagues showed that lower MOIs can lead to higher expression of protein (Y. H. Zhang, Enden, and Merchuk 2005). Cells were incubated for 72 h at 27 °C before harvesting supernatants.

Half of each filtered supernatant (~25 ml) was directly applied to a protein A column coated with the pan-HLA-DR antibody L243. HLA-DR1^{CLIP}tag monomers were eluted for each MOI (as described in Chapter 2.1.10) as shown in **Figure** 3**.**5 (red circles). The elution steps yielded ~85 µg (MOI 0.5) and ~60 µg (MOI 1) total.

Loading 25ml of supernatant took 3-4 h. As it was planned to express HLA-DR1^{CLIP}tag in *Sf9* cultures up to 300 ml, loading neat supernatants was not a preferred option in the long term. Previous attempts of concentrating supernatants in spin filtration columns had failed as components of the insect cell media blocked the filter. Therefore, the remaining supernatants were diluted in 200 ml PBS and concentrated first using an ultrafiltration cassette followed by spin filtration columns. Diluting supernatants in PBS circumvented the problem of blocking the columns and allowed concentration down to ~1 ml final volume in the spin filtration columns. Concentrated supernatants were applied to L243 columns and HLA-DR1^{CLIP}tag monomers eluted as before (see **Figure** 3**.**5). The elution steps yielded ~500 µg (MOI 0.5) and ~184 µg (MOI 1) total, which was consistently higher than yields from non-concentrated supernatants.
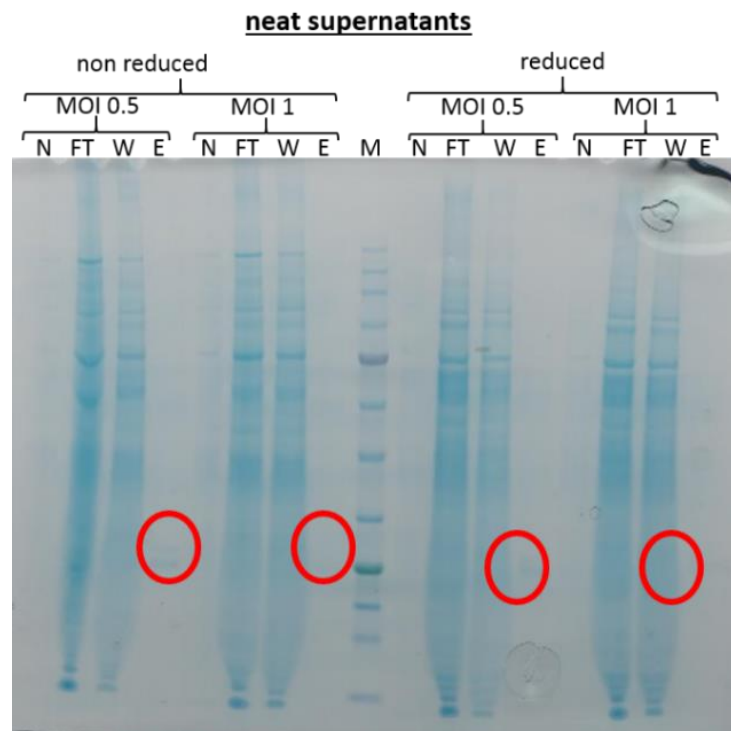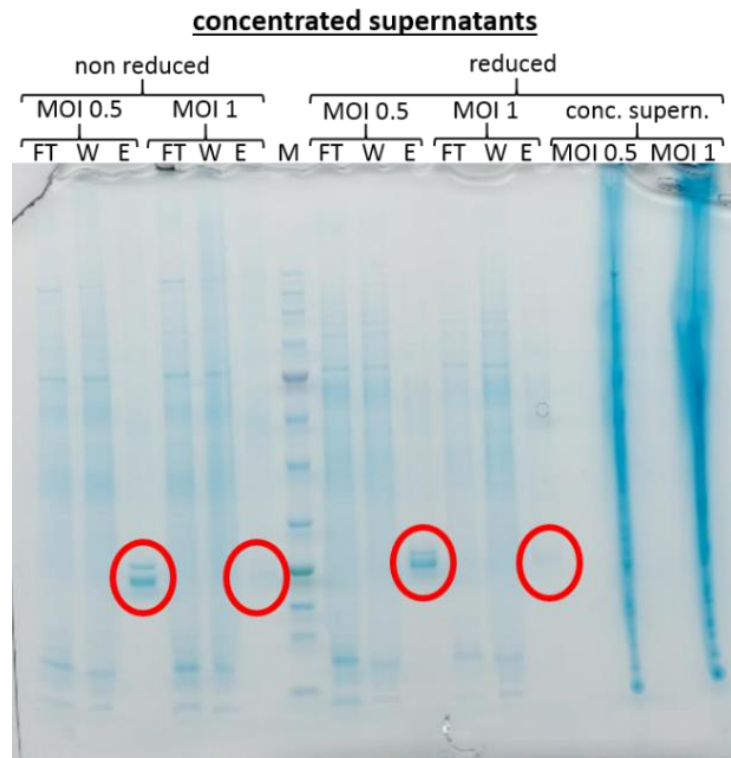
Figure 3.5: Purification of HLA-DR1CLIPtag expressed in Sf9 cells. Samples at different stages of the purification process were analysed on SDS PAGE. Every sample was run both non reduced and reduced by the addition of DTT. Red circles indicate eluted HLA-DR1α- and β-chains. MOI: multiplicity of infection, conc. supern.: concentrated supernatants prior to loading onto L243 column, FT: flow through from L243 column, W: wash steps, E: Eluate.

3.5.4. *Validation of HLA-DR1CLIPtag expressed in insect cells*

In order to assess whether the purified HLA-DR1^{CLIP}tag were correctly refolded and functional, we decided to test for binding of L243 using SPR. L243 binds to the HLA-DR1α-chain and relies on correct conformation of its epitope for binding (Gross et al. 2006). The eluted HLA-DR1^{CLIP}tag from the previous step were pooled and concentrated using a spin filtration column. Monomers were biotinylated and buffer exchanged into BIAcore buffer using size exclusion chromatography as described in Chapter 2.1.11. This had the additional advantage of eliminating any potential aggregates which would interfere with the SPR experiment. HLA-DR1^{CLIP}tag was loaded onto a CM5 dextrose chip using a BIAcore 3000 instrument as described in Chapter 2. Additionally, HLA-DR1^{PKY} monomers refolded in E.coli (as described in Chapter 2.2.4) and HLA-DR4^{CLIP} manufactured in mammalian cells (provided by Dr. Dave Cole) were loaded as positive controls. An irrelevant MHC-I (HLA-A2^{CLG}) was also loaded as negative control. L243 was then injected over the chip surface. Binding could be observed for all three HLA-DR monomers but not for HLA-A2 (see Figure 3.6).
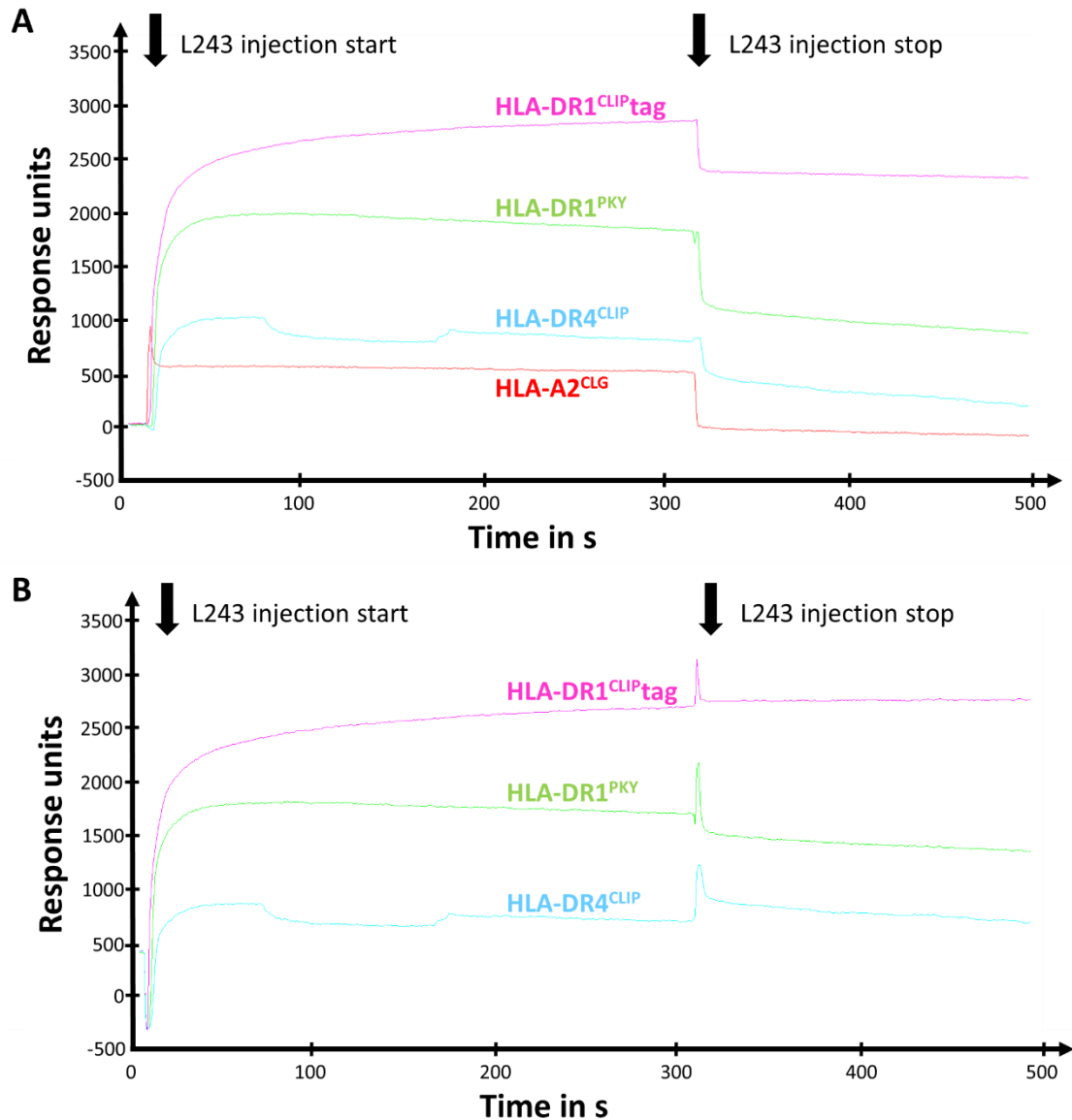
Figure 3.6: Validation of HLA-DR1CLIPtag using SPR. Binding of L243 antibody to HLA-A2CLG (shown in red), HLA-DR1CLIPtag (shown in pink), HLA-DR1PKY (shown in green) and HLA-DR4CLIP (shown in blue) was tested using a biacore machine. A: Traces for all four monomers are shown. B: Traces for HLA-DR1 monomers are shown after subtraction of reference i.e. HLA-A2CLG.

3.5.5. Exchange of bound CLIP peptide by peptides of interest and validation of pMHC-II molecules

The thrombin cleavage site located between the CLIP and the HLA-DR1α-chain allows the exchange of CLIP for any peptide of interest. Following proteolytic cleavage, the peptide of interest is added in excess to the mixture. As CLIP binds MHC-II monomers with low affinity it will be replaced by the peptide of interest. As recorded in the literature, 20U Thrombin is commonly used per mg of MHC-II to be cleaved (Anders et al. 2011) (Pos et al. 2012).

Thrombin cleavage was tested on some of the HLA-DR1^{CLIP}tag expressed in insect cells using a Thrombin cleavage kit as described in Chapter 2.1.11. The cleavage reaction was stopped by adding ABSF (4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride), an irreversible serine protease inhibitor at 1M. The HA$_{306-318}$ peptide (PKYVKQNTLKLAT) derived from influenza hemagglutinin was used as peptide of interest and added over night in 46 times excess. As negative control, PKY was added to non-cleaved HLA-DR1^{CLIP}tag. Both monomers were purified using gel filtration in order to remove any unbound peptide and concentrated using a spin column. Tetramers were assembled and the HLA-DR1$^{HA306-318}$ specific CD4$^+$ T cell clone DC D10 was stained as described in Chapter 2.3.11. The non-cleaved HLA-DR1^{CLIP}tag was used as a negative control. As shown in **Figure** 3.**7**, HLA-DR1$^{HA306-318}$ successfully stained 2C5 as demonstrated in a shift of the MOI while HLA-DR1^{CLIP}tag did not.
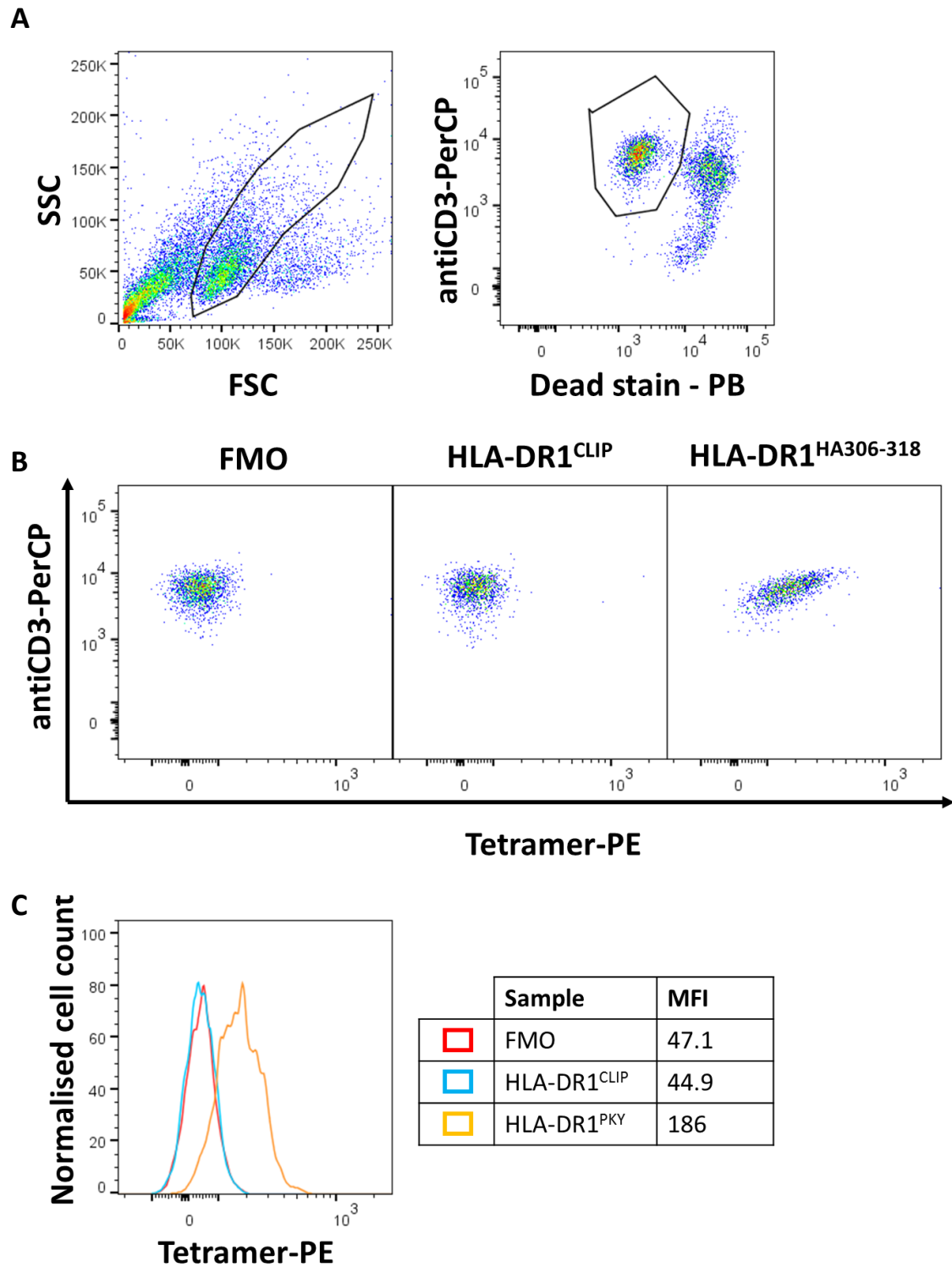
Figure 3.7: Validation of peptide exchange by tetramer staining of a CD4$^+$ T cell clone. A: Gating strategy. B: Dot-plots showing CD4$^+$ T cells stained using peptide exchanged HLA-DR1$^{HA306-318}$ and HLA-DR1^{CLIP} monomers. FMO (Fluorescence Minus One) to tetramer. C: Graph showing shift in MFI (Mean Fluoerescence Itensity).

3.6. Discussion

The ability to generate soluble pMHC-II molecules vastly facilitates investigations involving CD4$^+$ T cells. Foremost, their use in pMHC-II multimer staining for analysis by flow cytometry allows the detection, dissection and monitoring of CD4$^+$ T cell responses in disease settings and vaccine development (Kwok et al. 2002). As the varied roles of CD4$^+$ T cells for successful vaccinations become more evident the need for tools to identify and monitor subpopulations increases. Recent developments in multimer staining techniques, such as dextramers and cross-linking antibody "boost" technology (Dolton et al. 2015; Tungatt et al. 2015), lead to a rapidly expanding field of possible applications of pMHC-II molecules.

Several MHC-II alleles are associated with autoimmune diseases. HLA-DR4 confers an increased risk for rheumatoid arthritis (RA), while HLA-DR1 and HLA-DR10 do so to a lesser extent (Mackie et al. 2012). In the case of type 1 diabetes (T1D), HLA-DR4 and HLA-DR3 lead to increased risk in combination with certain HLA-DQ alleles (C. Nguyen et al. 2013). In both cases, conserved residues within the peptide binding groove allow the presentation of disease associated, MHC-II restricted epitopes (Miyadera and Tokunaga 2015). These are only two examples where the ability to generate soluble pMHC-II molecules allows further study of disease as well as potential treatments.

Due to their many interactions with other players of the immune system, CD4$^+$ T cells are directly and indirectly involved in disease processes such as viral clearance and antibody production in viral infections (Cox et al. 2015). In the case of persistent viral infections such as hepatitis C virus (HCV) infections, HLA-DR3 and HLA-DR5 are associated with increased viral clearance and milder disease symptoms. HLA-DR4, HLA-DR7 and HLA-DR15 on the other hand are associated with chronic infection (Thursz et al. 1999). Being

able to visualize peptide specific CD4[+] T cells by flow cytometry also aids the study of infectious diseases. Again, being able to identify virus or even epitope specific CD4[+] T cells and follow these subpopulations over the course of infection give valuable insight into how the immune system deals with viral pathogens.

While it is possible to express other HLA-DR alleles in *E. coli* and refold *in vitro* (Chen et al. 2013), our own attempts have been unsuccessful (unpublished observations). Flexible expression systems like the one described here allow the generation of different alleles by simply exchanging the HLA-DRB1 gene within the β-chain construct and the thrombin cleavage site allows the pMHC-II molecules to be loaded with any peptide binding to the HLA-DR allele in question. Insect cell based expression systems offer advantages over bacterial expressions systems such as generation of readily refolded proteins while being easier to implement and use than mammalian expression systems. Several modifications to the MHC-II α- and β-chains have been shown to increase stability and secretion of monomers such as Leucine zippers (Scott et al. 1996), a cleavable CLIP linked to the HLA-DR1α-chain (Pos et al. 2012) and addition of a melittin leader sequence (Tessier et al. 1991). In this chapter I have successfully implemented an insect cell based expression system that combines all three approaches.

Many of the more traditional baculovirus expression systems available commercially such as the BAC-to-BAC[TM] system (Invitrogen[TM]) rely on bacmids being assembled in a bacterial system aided by a helper plasmid. This has the advantage of being able to generate baculoviral DNA stocks for transduction of insect cells whenever needed. In the BaculoDirect[TM] system used here, only enough linear baculovirus DNA for two transductions is provided per each LR reaction (five vials à one LR reaction are provided

per kit). On the other hand, the in vitro system used here is straight forward and less time consuming than the traditional approach.

Purification of secreted HLA-DR1^{CLIP}tag from insect cell supernatants presented the next hurdle in this process. Loading higher volumes of neat supernatant onto the L243 column would have resulted in a time consuming process. Concentrating supernatants not only speeded up this process but also resulted in higher yields of protein eluted. This could be due to the monomers being buffer exchanged into PBS leading to increased stability of monomers, the elimination of proteases as well as better binding of the protein to the L243 column.

When expressing HLA-DR1 in *Sf9* cells on a small scale, I noticed that a lower MOI of 0.1 resulted in a higher yield than a MOI of 1. Of course, this observation would have to be confirmed by repeating the experiments while also testing higher MOIs. Nevertheless, this finding was in line with the literature (B. Nguyen et al. 1993; Y. H. Zhang, Enden, and Merchuk 2005; Kioukia et al. 1995). Higher MOIs lead to every cell being infected instantly by more than one viral particle once viral stock is added leading to rapid expression of the desired protein. While this might seem favourable, it also leads to faster exhaustion of the insect cell culture that will quickly be overwhelmed by the propagating viral infection. In addition, higher MOIs go hand in hand with larger volumes of viral stocks added to the cell cultures which results in larger amounts of spent media being added diluting out valuable nutrients and possible introducing toxic by-products. Lower MOIs, on the other hand, lead to only a certain percentage of the cell culture being infected at first which will then generate more viral particles leading to a successive propagation of infection. In the case of the expressing system used here, the successful expression of HLA-DR1^{CLIP}tag relies on co-infection of each insect cell with at least one

baculoviral particle of each construct (i.e. HLA-DR1α-chain and HLA-DR1β-chain, respectively). Taken together, these two mechanisms lead to a delay in protein expression when using low MOIs. However, it also gives the insect cells more time to generate the HLA-DR1^{CLIP}tag heterodimers before the cellular expression systems are fully hijacked by the baculovirus. This approach also reduces the volume of spent media being added to the insect cell culture prolonging the life span of the viral stock itself. As for every other protein, the optimal expression conditions including MOI depend on many factors and need to be adjusted following more in depth experiments than those I was able to conduct in the scope of this chapter.

SPR experiments showed that HLA-DR1^{CLIP}tag manufactured in *Sf9* cells binds to L243 as well as HLA-DR monomers produced using bacterial and mammalian expression systems. Furthermore, tetramer staining of HLA-DR1$^{HA306-318}$ specific CD4$^+$ T cells showed that exchanging CLIP for a peptide of interest (HA$_{306-318}$) generated functional pMHC-II monomers. We have shown in our lab that multimer staining PBMCs can successfully identify even small percentages of cells specific for certain pMHC-IIs (Holland et al. 2015). As described above this presents an extremely valuable tool for studying CD4$^+$ T cell responses in different diseases but can also be used as a diagnostic tool. In addition to multimer staining, soluble pMHC-II molecules can be used in biophysical investigations using a variety of methods such as SPR. This allows the more in depths analysis of TCR/pMHC-II interactions and can be used to modify either component of this interaction in order to strengthen it.

In summary, I have implemented an insect cell based expression system for soluble HLA-DR1 molecules that allows production of monomers suitable for biophysical and cellular analysis of TCR/pMHC-II interactions.

4. Peptide flanking residues length modulates CD4[+] T cell recognition

4.1. Background

Peptides bound to MHC-II are characterised by different length peptide flanking regions (PFR) due to the ability of MHC-II to accommodate peptides that extend beyond the open-ended peptide-binding groove. These different length variants of peptides are then presented to the CD4[+] T cell pool. It is well known that mutations within PFR can influence T cell activation and TCR binding. However, so far the impact of different length variations on T cell activation, peptide binding and TCR repertoire selection has not been studied in detail.

4.2. Introduction

4.2.1. *MHC II restricted epitopes vary in length*

Proteins entering the MHC-II antigen processing pathways are degraded in lysosomes into peptides by cathepsins before being loaded onto the MHC-II (see Chapter 1.3.3). The peptide-binding groove of MHC-II molecules is open ended, allowing longer peptides to extend beyond the groove at both termini. In consequence, MHC-II are not limited to binding peptides of 9-13 aa as seen for MHC-I and can present peptides of at least 25 aa. Sequencing peptides eluted from MHC-II molecules revealed that these different length variants create so-called nested sets (Chicz et al. 1993; Lippolis et al. 2002). Aligning these peptides reveals a shared sequence, typically around 12 aa in length. The 9mer core encoding the anchor residues necessary to bind to the MHC-II is located within this shared sequence. Peptides are extended at either terminus, or indeed at both termini. **Figure** 4**.**1 shows an example of a nested set eluted from HLA-DR4.

Source protein	Residues	Sequence	Length
HLA-A2	28-50	VDD<u>TQFVRFDSDAAS</u>QRMEPRAP	23
	28-48	VDD<u>TQFVRFDSDAAS</u>QRMEPR	21
	28-47	VDD<u>TQFVRFDSDAAS</u>QRMEPP	20
	28-46	VDD<u>TQFVRFDSDAAS</u>QRME	19
	30-48	D<u>TQFVRFDSDAAS</u>QRMEPR	19
	31-49	<u>TQFVRFDSDAAS</u>QRMEPRA	19
	28-44	VDD<u>TQFVRFDSDAAS</u>QR	17
	31-47	<u>TQFVRFDSDAAS</u>QRMEP	17
	31-47	<u>TQFVRFDSDAAS</u>QRM	15
	31-42	<u>TQFVRFDSDAAS</u>	12

Figure 4.1: Nested set of a peptide eluted from HLA-DR4. Peptides are aligned according to their 12 aa overlap. Length of each peptide is given. Adapted from Chicz et al., 1993.

4.2.2. Role of PFR in CD4$^+$ T cell activation, peptide binding stability and TCR gene selection

Crystallographic and functional evidence collected so far suggest that the TCR primarily contacts residues located within P$_{-1}$ and P$_{10}$ of the peptide with the majority of contacts being made with the residues within the 9mer core (Jens Hennecke and Wiley 2002; Stepniak et al. 2005). However, changes in peptide flanking residues have been shown to influence CD4$^+$ T cell activation. In one study, 65% of murine CD4$^+$ T cell hybridomas raised against a hen egg lysozyme (HEL) derived peptide were dependent on the C-terminal PFR (Carson et al. 1997). It was also found that PFR dependent and independent hybridomas showed distinct TCRB gene expression patterns suggesting that PFR can influence TCR gene selection. In a similar study, murine CD4$^+$ T cell hybridomas were raised against a panel of nine hen egg lysozyme (HEL) and glutamate decarboxylase (GAD) derived peptides (Arnold et al. 2002). Depending on the peptide, between 12-100% of hybridomas generated were dependant on PFR. While some peptides generated hybridomas dependent on P$_{-1}$, others generated hybridomas dependent on P$_{11}$. PFR dependent and independent hybridomas recognised wildtype peptides with equal sensitivity. It is also noteworthy, that mice immunised against whole HEL were used to generated hybridomas against HEL derived peptides, suggesting that PFR depending T cell are generated *in vivo* and are not just an artefact of *in vitro* methods.

Sequencing of naturally processed peptide eluted from MHC-II molecules showed that C-terminal PFR are enriched in basic residues such as arginine and lysine, presumably reflecting a bias in the antigen processing machinery towards such peptides (Godkin et al. 2001). This study also showed that substitution of C-terminal PFR within the influenza heamagglutinin (HA) derived peptide HA$_{306-318}$ by arginine lead to increased T cell

activation. The same substitution was subsequently found to increase TCR affinity as well as enhancing pMHC-tetramer staining of CD4[+] T cells (Holland et al. 2015). Sequencing of CD4[+] T cells against the wildtype as well as the modified version of this peptide showed an altered and narrowed TCRB repertoire in response to the modified peptide further proving a role of PFR in TCR gene selection (Cole et al. 2012).

In addition to TCR activation and TCR gene selection, PFRs have also been shown to affect peptide binding stability. Removal of N-terminal PFR from a HEL derived peptide led to substantially lower half-life of murine I-A[k] complexes (Lovitch, Pu, and Unanue 2006). In a different study, the Brownian motion of the same peptides were measured (Haruo Kozono et al. 2015). Missing N-terminal PFR led a higher degree of motion within the peptide and the MHC than peptides with intact PFR leading to reduced binding stability.

In vivo, CD4[+] T cells are presented with different length variation of the same peptide that can differ in their PFR length. Individual T cell clonotypes are likely to preferably recognise certain variants over others. So far, only two studies investigated the impact of PFR on TCR repertoire selection. One study, mentioned above, investigated the effect aa substitutions within the C-terminal flanks (Cole et al. 2012) A second study showed that mutations within N-terminal PFR influences selection of TCRα-genes while C-terminal PFR influence selection of TCRβ-genes (Carson et al. 1997).

4.3. Aims

Despite a growing base of knowledge about the impact of PFR on $CD4^+$ T cell activation, peptide binding stability and TCR repertoire selection, there is still much to be discovered. Although several studies looked at the impact of a point mutation within the C-terminal PFR of the $HA_{306-318}$ epitope, different length variants of this peptide have not been studied to date. For this chapter, a set of PFR length variants of $HA_{306-318}$ was designed and used in a range of experimental approaches to dissect their role in $CD4^+$ T cell biology.

The first aim for this project was to investigate the impact of PFR length variants on T cell activation by testing $CD4^+$ T cell clones for recognition of these variants.

The second aim was to measure peptide binding stability of PFR length variants to HLA-DR1 using a competitive binding assay.

The third aim was to investigate the impact of PFR length variants on TCR repertoire selection by priming PBMCs with a selection of variants and clonotyping $HA_{306-318}$ specific $CD4^+$ T cells.

4.4. Results

4.4.1. *Investigating influence of peptide flanking regions on CD4$^+$ T cell activation*

In order to investigate the effect of PFR on CD4$^+$ T cell activation, three CD4$^+$ T cell clones were tested for their recognition of PFR variants of their cognate influenza hemagglutinin derived epitope, HA$_{306\text{-}318}$ (PKYVKQNTLKLAT) (Lamb et al. 1982). Using the 9mer core as a starting point (shown in grey), a nested set of peptides was designed as shown in Table 4.1. First, the N-terminal PFR was extended one residue at a time (core + 1N, core + 2N, core + 3N). The same was done with the C-terminal PFR (core + 1C, core + 2C core + 3C). Finally, N-terminal and C-terminal PFR were extended simultaneously (core + 1N1C, core + 2N2C, core + 3N3C). The core + 2N2C variant corresponds to HA$_{306\text{-}318}$. All ten peptides were tested for recognition by the CD4$^+$ T cell clones in peptide titrations and INF-γ release was measured as described in Chapter 2.3.9. **Figure** 4.2 shows results for each clone.

Peptide sequence   **Description**

-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	Description
N-terminal			*Core epitope*									*C-terminal*			
			Y	V	K	Q	N	T	L	K	L				core
		K	Y	V	K	Q	N	T	L	K	L				core + 1N
	P	K	Y	V	K	Q	N	T	L	K	L				core + 2N
C	P	K	Y	V	K	Q	N	T	L	K	L				core + 3N
			Y	V	K	Q	N	T	L	K	L	A			core + 1C
			Y	V	K	Q	N	T	L	K	L	A	T		core + 2C
			Y	V	K	Q	N	T	L	K	L	A	T	G	core + 3C
		K	Y	V	K	Q	N	T	L	K	L	A			core + 1N1C
	P	K	Y	V	K	Q	N	T	L	K	L	A	T		core + 2N2C
C	P	K	Y	V	K	Q	N	T	L	K	L	A	T	G	core + 3N3C

Table 4.1: Nested set of peptides based on HA$_{305-318}$. Starting with the 9mer core region, PFR were extended one residue at a time either at one terminus only or both termini at the same time.
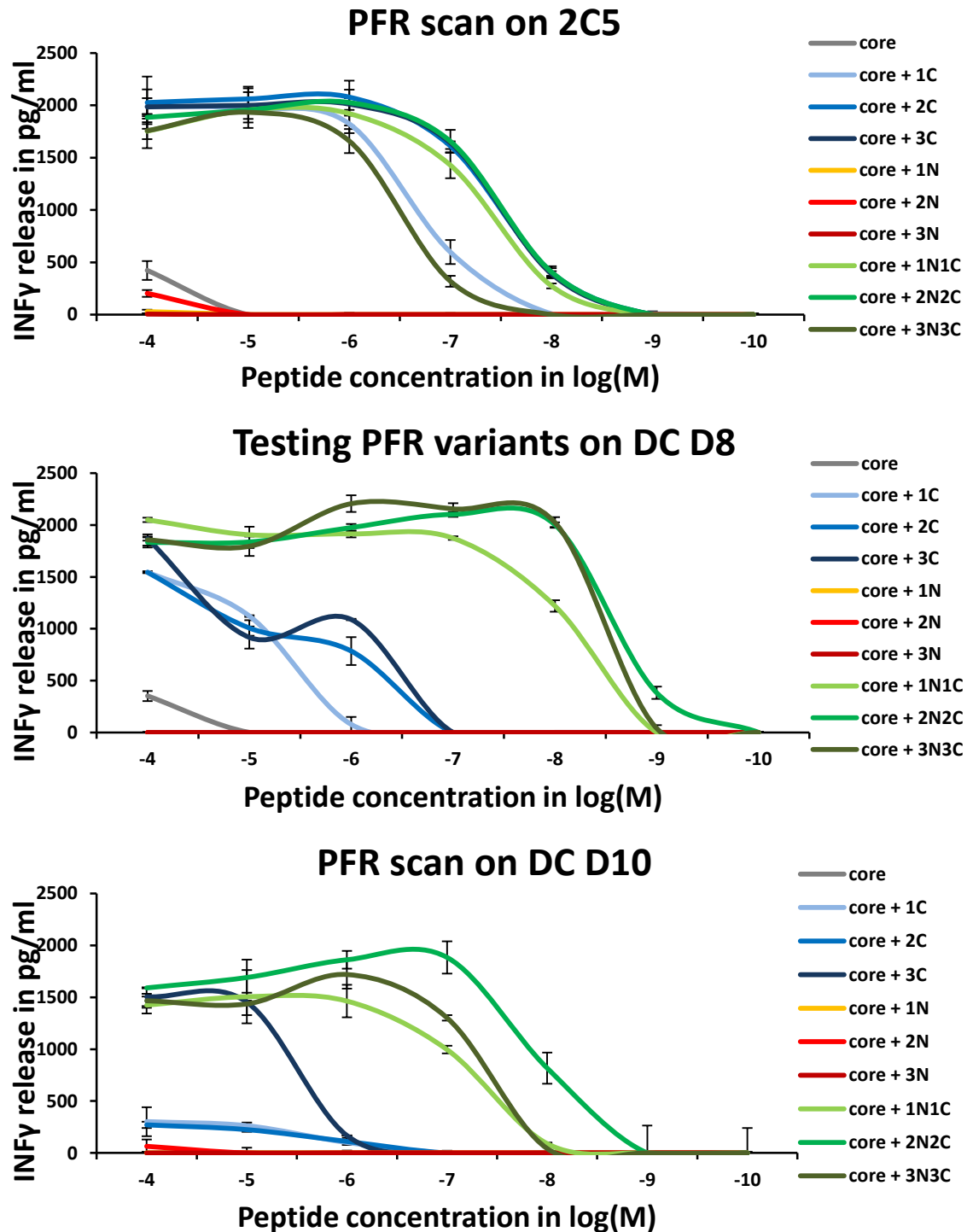
Figure 4.2: Testing PFR length variants on three different CD4⁺ T cell clones. PFR variants were tested in peptide titrations. CD4⁺ T cell clones were incubated overnight with peptides at different concentrations in presence of HOM-2 cells. INFγ release was measured using ELISA. The 9mer core peptide is shown in grey, C-terminal PFR variants in shades of blue, N-terminal PFR variants in shades of red and NC-terminal variants in shades of green. Error bars show standard deviation.

None of the three clones tested recognised the 9mer core on its own or any of the N-terminal PFR variants. 2C5 recognised all three of the C-terminal PFR variants as well as the both-sided ones. None of the N-terminal PFR variants were recognised. Similarly, DC C8 recognised all C-terminal as well as all both-sided PFR variants. However, C-terminal PFR variants were recognised at lower sensitivities. DC D10 also recognised all three both-sided PFR variants. Out of the three C-terminal PFR variants it recognised the core + 3C variant but neither of the remaining two.

Overall, addition of N-terminal PFR and absence of C-terminal PFR did not re-establish recognition. However, addition of C-terminal PFR in absence of N-terminal PFR re-established recognition, although to varying degrees depending on the $CD4^+$ T cell clone with 2C5 > DC C8 > DC D10. Addition of one PFR on both N- and C-termini at the same time (core + 1N1C) outperformed any C-terminal PFR variants. Recognition of the core + 3N3C differed in the three clones, with 2C5 and DC D10 recognising it at lower sensitivities than core + 2N3C.

4.4.2. *Influence of peptide flanking regions on peptide binding to HLA-DR1*

In order to determine whether peptide binding strength to HLA-DR1 influenced peptide recognition, a competitive HLA-DR peptide binding assay based on the assay published by Godkin and colleagues (Godkin et al. 2001) was set up and optimised. The final protocol is described in Chapter 2.2.6. Briefly, HLA-DR1 refolded with a biotinylated CLIP peptide (BT-CLIP) was incubated overnight with competitor peptides at different concentrations before being transferred onto a microtiter plate coated with anti-pan-HLA-DR L243. Since CLIP is known to bind with low affinity, it can be replaced by any higher avidity peptide. In order to facilitate this peptide exchange, it took place in acidic conditions (pH=5). Any free peptide was washed off and remaining HLA-DR1$^{BT-CLIP}$

was detected via ELISA. In wells in which the BT-CLIP peptide has successfully been replaced by the competitor peptide, no BT-CLIP will be left and therefore, none of the substrate will be converted by HRP-Streptavidin beyond background HRP background activity. This leads to low optical density (O.D.) readings for these wells. In wells in which BT-CLIP has not been replaced by the competitor peptide, the substrate will be converted leading to high OD readings. In all assays, the HLA-A2 restricted melanoma peptide ELA was used as negative control (Cole et al. 2010). All ten PFR variants were tested at a range of concentrations. Results are shown in **Figure** 4**.**3.
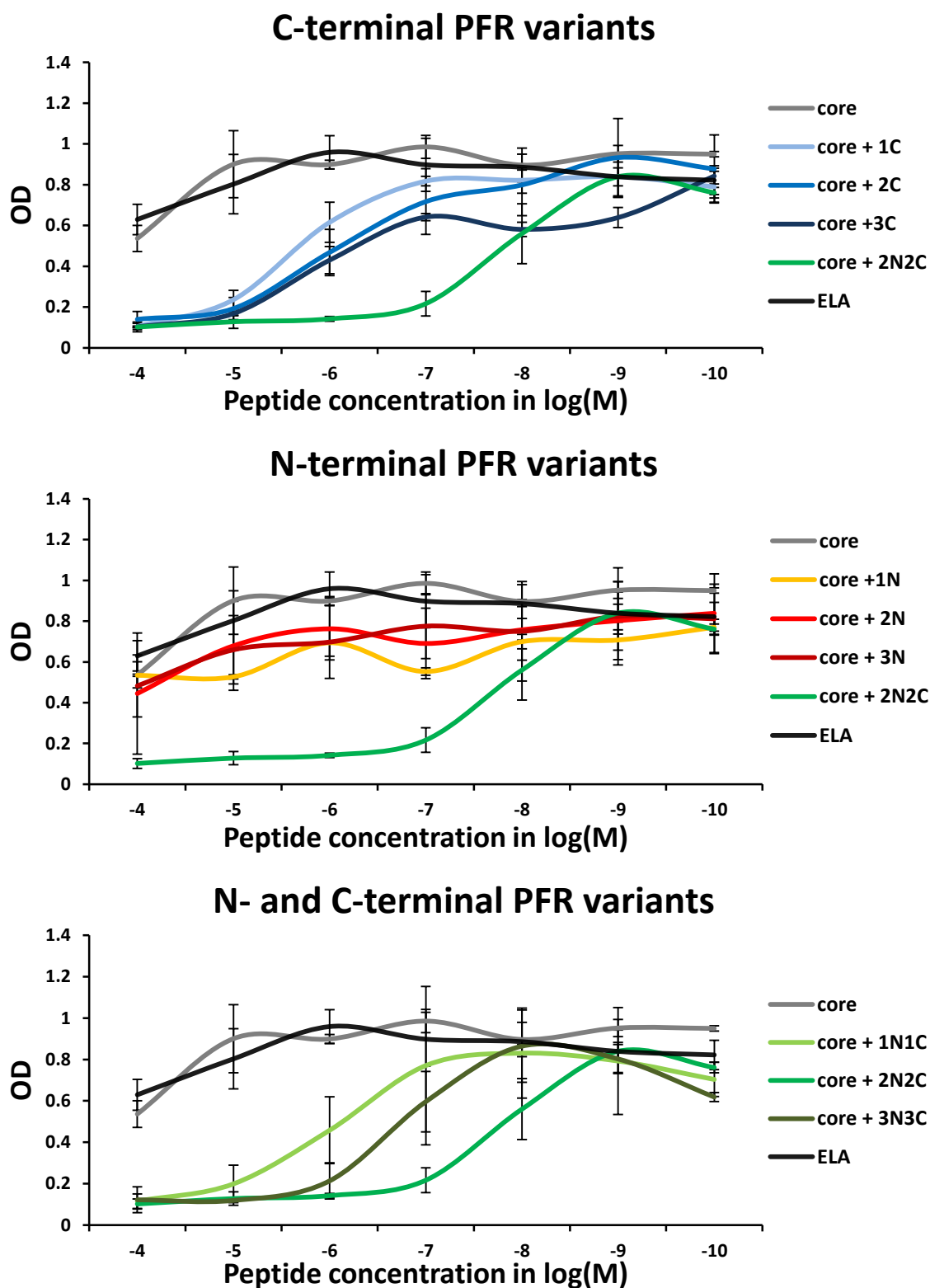
Figure 4.3: Testing PFR length variants in a competitive binding assay. Each peptide was tested for its ability to replace BT-CLIP in the HLA-DR1 binding groove at a range of concentrations. Absolute O.D. readings are shown. ELA (negative control) and core + 2N2C are shown in each graph as reference. Error bars show standard deviation.

As expected, O.D. readings for ELA stayed high at all concentrations indicating that it was unable to replace BT-CLIP in the HLA-DR1 peptide-binding groove. The 9mer core behaved in a similar fashion indicating no binding to HLA-DR1. O.D. readings for C-terminal PFR were lower at higher peptide concentrations compared to the negative control indicating successful binding to HLA-DR1. The N-terminal PFR, on the other hand, were incapable of replacing BT-CLIP as demonstrated by overall high O.D. readings at all concentrations tested. Core + 1N1C was capable of replacing BT-CLIP to a level similar as seen for the C-terminal PFR lengths variants. Core + 2N2C was capable of replacing BT-CLIP at lower concentrations than seen for any of the other PFR variants. Core + 3N3C outperformed core + 1N1C, but not core + 2N2C.

In summary, the 9mer core on its own binds to HLA-DR1 too weakly to replace BT-CLIP in the peptide binding groove. Addition of C-terminal PFR increases peptides binding strength while N-terminal PFR had no influence. Addition of N- and C-terminal PFR simultaneously increased peptide binding strength.

4.4.3. *Influence of peptide binding strength on CD4$^+$ T cell activation*

In order to investigate any relationship between peptide binding strength and CD4$^+$ T cell activation, IC$_{50}$ values were calculated for each of the PFR variants tested in the competitive peptide binding assay as well as EC$_{50}$ values for all peptides tested in T cell activation assays. **Table** 4.2 shows the results for all three CD4$^+$ T cell clones as well as the competitive peptide binding assay as log(IC$_{50}$) and log(EC$_{50}$), respectively.

Peptide	log(IC$_{50}$) Peptide binding assay	log(EC$_{50}$) T cell activation		
		2C5	DC C8	DC D10
core	-0.504	n/a	n/a	n/a
core + 1C	-5.612	-6.745	-5.243	-5.737
core + 2C	-6.162	-7.483	-5.643	-5.814
core +3C	-5.556	-7.523	-4.345	-5.617
core +1N	n/a	n/a	n/a	n/a
core + 2N	3.591	n/a	n/a	n/a
core + 3N	7.430	n/a	n/a	n/a
core + 1N1C	-5.995	-7.397	-8.12	-7.212
core + 2N2C	-7.778	-7.57	-8.906	-7.995
core + 3N3C	-6.669	-6.586	-8.757	-7.31
ELA	0.510			
Spearman's coefficient:		0.769	0.957	0.957

Table 4.2: Influence of peptide binding on CD4$^+$ T cell activation. Log(IC$_{50}$) values of the competitive peptide binding assay correspond to peptide concentrations needed to replace 50% of BT-CLIP. Log(EC$_{50}$) values correspond to peptide concentrations needed to elicit 50% of maximal response by the three CD4$^+$ T cell clones screened. Colour gradient ranges from green (low log(IC$_{50}$)/ log(EC$_{50}$)) to red (high log(IC$_{50}$)/ log(EC$_{50}$)). n/a: log(IC$_{50}$)/ log(EC$_{50}$) could not be determined. Spearman's coefficients were calculated based on the log(EC$_{50}$) values of each CD4$^+$ T cell clone and log(IC$_{50}$) values for each peptide.

Low log(IC$_{50}$) values indicate peptides binding strongly to HLA-DR1 as the BT-CLIP is replaced at low competitor peptide concentrations and are indicated in yellow to green shades. Similarly, low log(EC$_{50}$) indicate peptides activating CD4$^+$ T cells at low concentrations. In some cases, log(IC$_{50}$) or log(EC$_{50}$) values could not be calculated (indicted by n/a). This was the case where INF-γ release dropped to zero at high peptide concentrations (as observed for the core peptide) or could not be observed at all (as observed for the N-terminal PFR variants).

As shown in **Table** 4.2, log(IC$_{50}$) values and log(EC$_{50}$) values follow a same pattern. The 9mer core binds very poorly to HLA-DR1 while log(EC$_{50}$) could not be calculated. As

expected, the HLA-A2 restricted ELA peptide also showed poor binding. Addition of C-terminal led to $\log(IC_{50})$ values in the yellow to light green range with core + 2C being the best binder. This is mirrored in the $\log(EC_{50})$ values with the exception of the $CD4^+$ T cell clone 2C5 where $\log(EC_{50})$ values for core + 2C and core + 3C were in the dark green range, i.e. at the lower end. Addition of N-terminal PFR variants led to $\log(IC_{50})$ values in the red range corresponding to weak binding to HLA-DR1. None of the N-terminal PFR variants elicited any responses in $CD4^+$ T cell clones and no $\log(EC_{50})$ values could be calculated. $\log(IC_{50})$ values for both-sided PFR variants were in the green to dark green range indicating good binding to HLA-DR1. Similarly, $\log(EC_{50})$ values were in the green to dark green range with the exception of core + 3N3C which was in the yellow range for 2C5. This was in line with the results of the T cell activation assay. Spearman's coefficients were calculated in order to test whether the measured $\log(EC_{50})$ correlate with the $\log(IC_{50})$. This showed that the EC_{50}s for all three clones indeed correlated with the $\log(IC_{50})$ of the corresponding peptides. This correlation was stronger for DC C8 and DC D10 than for 2C5.

In summary, the strength of binding of each PFR variant correlated with its ability to elicit INF-γ release in the three $CD4^+$ T cell clones tested indicating that T cell activation depends on peptide binding strength.

4.4.4. HA PFRs influence TCR repertoire selection

The three $CD4^+$ T cell clones studied for this project were initially selected for their recognition of the core + 2N2C peptide i.e. $HA_{306-318}$ (as was the case for the 2C5 clone) or a variant therefore i.e. $HA_{306-318-11R}$ (as was the case for the DC C8 and DC D10 clones). *In vivo* however, $CD4^+$ T cells are presented with nested sets of peptides of varying PFR lengths. It has recently been shown that a point mutation in the C-terminal flank at

position 11 of $HA_{306\text{-}318}$ (PKYVKQNTKLA**R**) alters TCR repertoire selection (Cole et al. 2012). However, the impact of PFR lengths variants of $HA_{306\text{-}318}$on clonotype selection is currently unknown. This would provide new insight into the role of nested sets of peptides on protection against influenza and has wider implications for CD4 T-cell mediated immunity.

In order to investigate the impact of PFR variants on the selection of a TCR repertoire, six PFR variants were chosen for further investigation: core, core + 1C, core + 3C, core + 2N, core + 2N2C, core + 3N3C. The 9mer core was chosen to investigate TCR repertoire selection against the minimal binding epitope. Core + 1 and core + 3C were chosen since addition of C-terminal PFR re-establishes both T cell activation in the three T cell clones tested as well as peptide binding strength. Core + 2N was chosen to include one of the N-terminal PFR variants despite failure to activate the three T cell clones tested and poor binding to HLA-DR1. Core + 2N2C and core + 3N3C were chosen in order to investigate the impact of the addition of PFR at both termini. Short term T cell lines were grown against each of the six peptides using $5x10^6$ PBMC to set up each line. In addition, one further line was grown without the addition of any peptide. On day 14 of culturing, all seven lines were FACS sorted using HLA-DR1$^{core\ +\ 2N2C}$ dextramers. **Figure** 4**.**4A shows the gating strategy used and **Figure** 4**.**4B shows dot plot diagrams from a subset of 100,000 cells of each short term T cell line. **Supplementary Table S**1 shows the total number of cells collected for each individual cell line.
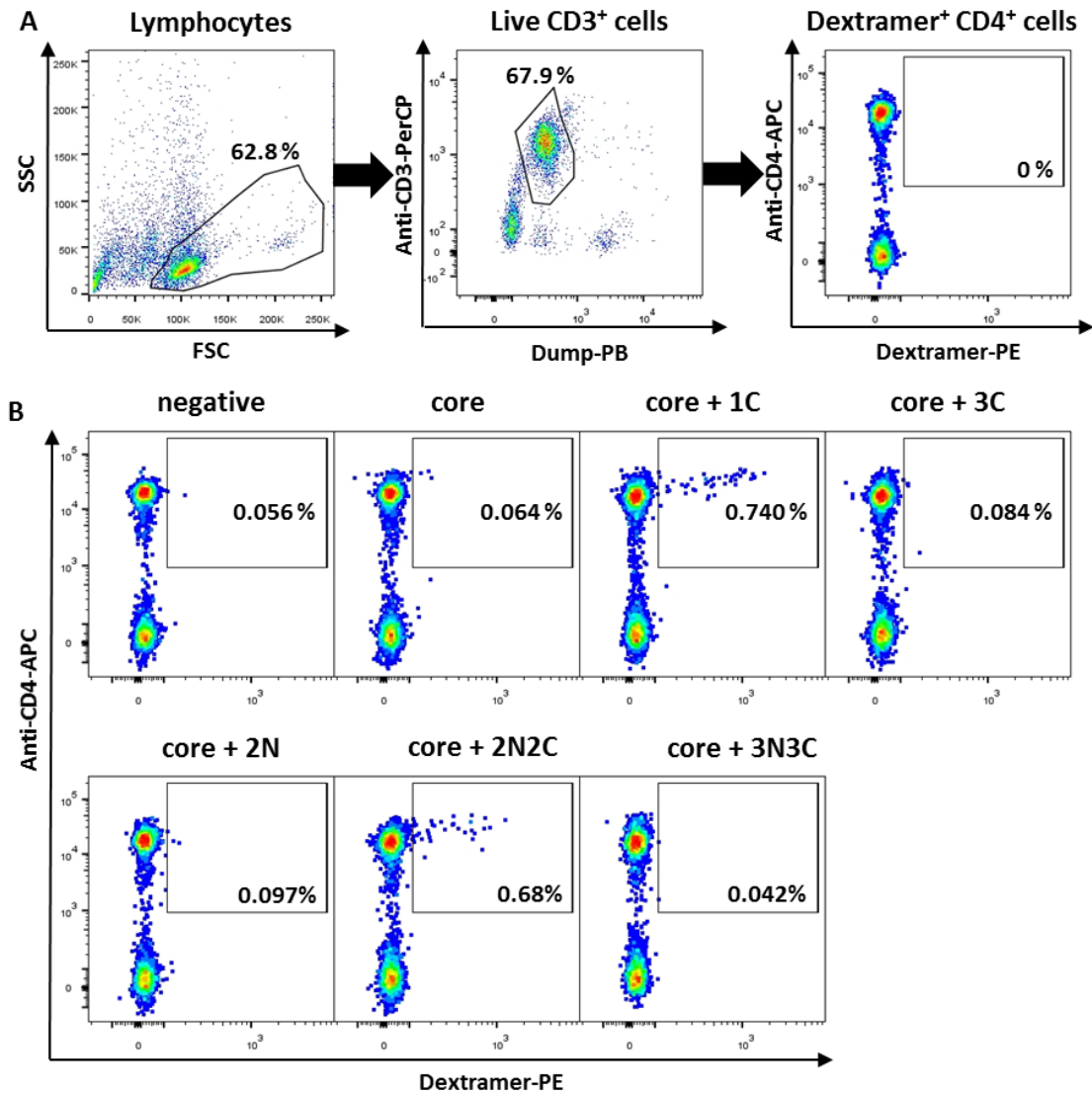
Figure 4.4: Sorting short term T cell lines using dextramers. A: Gating strategy demonstrated using the FMO control. B: Short term T cell lines set up using a selection of PFR variants were sorted using HLA-DR1[HA306-318] dextramers on day 14 of culture. The no peptide control short term line was grown without adding peptide at day 0. Percentages of HLA-DR1[HA306-318] dextramer positive CD4+ T cells are shown. Graphs represent a sample of 100,000 cells of each line. The remainder of each line was then used for FAC sorting.

As shown in **Figure** 4.4B, HLA-DR1[core + 2N2C] specific cells could be identified in the short term T cell line grown in absence of any peptide. Since HLA-DR1[core + 2N2C] specific cells can be detected ex vivo (Holland et al. 2015), it is not surprising that some survived the 14 days culturing period and are still present within the line. As ex vivo responses

were not the focus of this study, these cells were not further analysed. The highest numbers of HLA-DR1$^{core + 2N2C}$ positive cells were detected in lines grown against core + 1C and, not surprisingly, core + 2N2C, followed by core + 3C. The remaining lines yielded lower numbers of HLA-DR1$^{core + 2N2C}$ positive cells (see**Supplementary Table S**1). The mRNA encoding the TCR V$_\beta$ sequences of sorted cells from each cell line was extracted and converted into cDNA using RT-PCR as described in Chapter 2.3.12. cDNA was cloned into a TOPO sequencing vector and transformed into Top10 *E.coli*. Between 70 and 124 colonies were sequenced for each cell line (see Supplementary Table S1). Figure 4.5 shows a summary of all clonotypes that were identified in this study, while Figure 4.6 and Figure 4.7 show the TRB gene and CDR3$_\beta$ usage, respectively. Each PFR variant generated a different set of clonotypes. Comparing the TRB gene usage identified in each line (see **Figure** 4**.**7A) shows that each line displays its own, distinct pattern of clonotypes even though some TRB genes could be identified in more than one line (see **Figure** 4**.**7B). **Figure** 4**.**7 also demonstrates the CDR3$_\beta$ usage pattern was distinctly different for each line. Both core + 2N and core + 2N2C generated the most diverse CDR3$_\beta$ usage patterns.

The TCRB gene usage for core + 2NC has been analysed previously using PBMCs from the same donor (Cole et al. 2012). Four of the ten clonotypes identified in this study, including the three most abundant ones, had been identified by Cole and colleagues, albeit at different frequencies (see Figure 4.5)**.** One additional clonotype identified here encoded the same TRB-CDR3$_\beta$ combination as one of the clonotypes previously identified although in combination with a different TRJ segment. The remaining five clonotypes were unique to this study. However, three of these five unique clonotypes (30% of total) encoded the TRBV28.1 segment which had been identified in five of the thirteen clonotypes (38.46% of total) from the previous study albeit in combination with different

CDR3 loops and TRBJ segments. As a side note, the TCRβ-chain expressed by the 2C5 CD4[+] T cell clone used in this chapter was identified in the core + 2N2C line (indicted by *[1]) as was the TCRβ-chain expressed by another CD4[+] T cell clone, 3A (indicated by *[2]), which has not been used in this chapter. Both clones were originally grown against core + 2N2C, again using blood PBMCs from the same donor and also encode the aforementioned TRBV28.1 segment. This shows that the clonotyping data presented in this thesis matches previous data from the same donor.

Looking at the clonotypes identified in the line grown against the 9mer core, the second most frequent clonotype was also identified for core + 2N2C albeit at lower frequencies (highlighted in bold). The other two clonotypes were unique to the 9mer core. Similarly, the line grown against core + 1C showed its second most frequent clonotype with core + 2N2C. The remaining four clonotypes were unique to core + 1C. The line grown against core + 3C shared its most frequent clonotype with core + 2N2C while the remaining three clonotypes were unique. The line grown against core + 2N did not share any clonotypes with core + 2N2C. It did, however, share one clonotype with core + 3N3C. The remaining six and two clonotypes were unique to core + 2N and core + 3N3C, respectively.

In summary, each PFR variant of HA$_{306-318}$ generated a unique set of clonotypes. Short term CD4[+] T cell lines grown against the 9mer core, core + 1C and core + 3C shared one clonotype each with core + 2N2C. Lines against core + 2N and core + 3N3C did not share any clonotypes with core + 2N2C although, they shared one clonotype between them. These data support the idea that nested sets select a more heterogeneous CD4 T-cell pool than single peptides that could provide broader protection against variable pathogens.
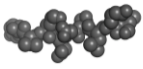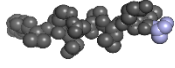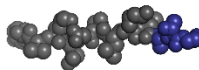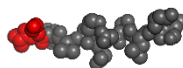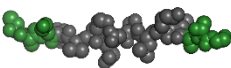
	TRBV	CDR3β	TRBJ	Frequency in %	
				Current study	Cole et al (core + 2N2C)
core	5.5	CASSGLGFNEKLFF	1.4	59.46	
	28.1	**CASGSQGNRFYEQYF**	**2.7**	**24.32**	
	5.5	CASSGLGFNEKLFF	2.4	16.22	
core + 1C	5.1	CASSREGPPDTEAFF	1.1	64.08	
	9.1	CASSMSPGPRESPLHF	1.6	33.01	
	5.8	CASSRGLAGDQETQYF	2.5	1.94	
	5.4	CASSREGPPDTEAFF	1.4	0.97	
core + 3C	24.1	*CATSDESYGYTF*	1.2	66.67	1.96
	18.1	CASSTTTGSHPKTQYF	2.3	16.67	
	21.1	CATSDESYGYTF	2.1	11.11	
	12.3	CASSLSQGDQPQHF	1.5	5.56	
core + 2N	14.1	CASSPWRDTEAFF	1.1	54.05	
	4.3	CASARDRGFEQYF	2.7	10.81	
	5.1	CASSLEGDTEAFF	1.1	8.11	
	20.1	CSARGFEDSSGGELFF	2.2	8.11	
	20.1	CSARDSYEQYF	2.7	8.11	
	<u>7.2</u>	<u>CASSPGAAGEQYF</u>	<u>1.5</u>	<u>5.41</u>	
	7.2	CASSLVGRQPVTYEQYF	2.7	2.70	
	11.2	CASSLVGRQPVTYEQYF	2.7	2.70	
core + 2N2C	24.1	**CATSDESYGYTF**	1.2	60.00	**1.96**
	9.1	CASSMSPGPRESPLHF	1.6	20.00	**13.73**
	4.1	CASSTDRGPYEQYF	2.7	5.45	1.96
	5.1	CASSLIGQTYQETQYF	2.5	3.64	
	6.5	CASSYPSSGGAPQYF	2.3	1.82	
	28.1	CASSSSGRRGNTEAFF	1.1	1.82[2]	*25.49*
	28.1	CASSLAPELAGYTF	1.2	1.82[1]	*(TRBJ2.5)*
	28.1	**CASGSQGNRFYEQYF**	**2.7**	**1.82**	
	29.1	CSVEDPRTDYGYTF	1.2	1.82	
	9.1	CASSREGPPDTEAFF	1.6	1.82	
core + 3N3C	<u>7.2</u>	<u>CASSPGAAGEQYF</u>	<u>2.7</u>	<u>71.43</u>	
	7.9	CASSFPAAGIGDTQYF	2.3	25.00	
	7.2	CASSFPAAGIGDTQYF	2.7	3.57	

Figure 4.5: Clonotypic composition of CD4[+] T cells expanded using different PFR variants. Short term CD4[+] T cell lines were grown against each individual PFR variant and FACS sorted using HLA-[DR1core + 2N2C] (i.e. HLA-[DR1HA306-318]) dextramers. Percentages of each clonotype were calculated for each individual line. Percentages of clonotypes previously identified by Cole and colleagues are shown as well. Clonotypes present in core + 2N2C as well as in other lines are highlighted in bold (Cole et al. 2012). Clonotypes present in two different lines but not in core + 2N2C are underlined. *1 and *2 indicate the TRCβ-chain expressed by 2C5 and 3A, respectively.
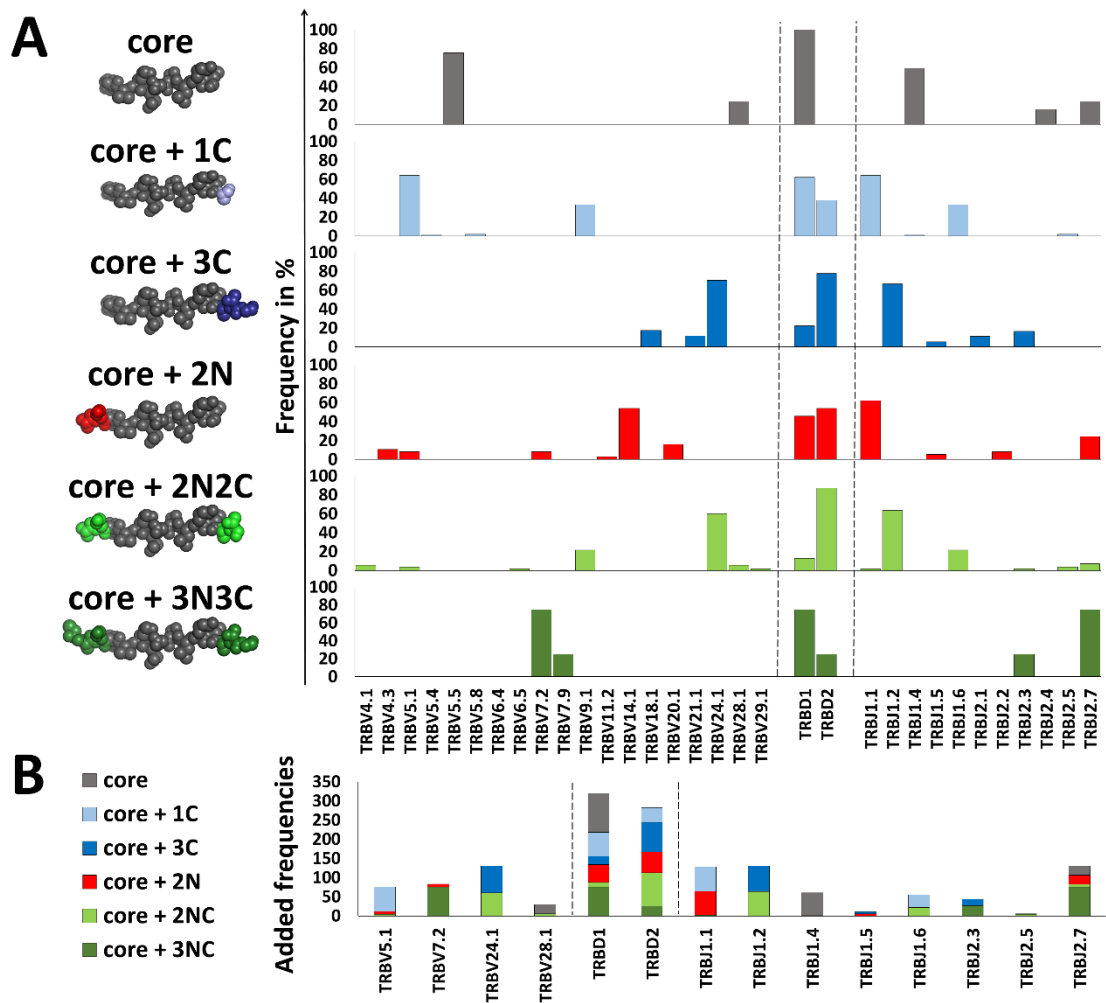
Figure 4.6: TRBV and TRBJ usage in CD4$^+$ T cells expanded using different PFR variants. Short term CD4$^+$ T cell lines were grown against each individual PFR variant and FACS sorted using HLA-DR1$^{core + 2N2C}$ (i.e. HLA-DR1$^{HA306-318}$) dextramers. A: TRB gene usage for each individual short term T cell line is shown. B: Only TRB genes found in more than one short term T cell line are shown.

core

- CASSGLGFNEKLFF
- CASGSQGNRFYEQYF
- CASSGLGFNEKLFF

core + 1C

- CASSREGPPDTEAFF
- CASSMSPGPRESPLHF
- CASSRGLAGDQETQYF

core + 3C

- CATSDESYGYTF
- CASSTTTGSHPKTQYF
- CASSLSQGDQPQHF

core + 2N

- CASSLEGDTEAFF
- CASSPWRDTEAFF
- CASARDRGFEQYF
- CSARGFEDSSGGELFF
- CSARDSYEQYF
- CASSTASGGLQPQHF
- CASSPGAAGEQYF
- CASSLVGRQPVTYEQYF

core + 2N2C

- CATSDESYGYTF
- CASSMSPGPRESPLHF
- CASSTDRGPYEQYF
- CASSLIGQTYQETQYF
- CSVEDPRTDYGYTF
- CASSLAPELAGYTF
- CASSSSGRRGNTEAFF
- CASGSQGNRFYEQYF
- CASSYPSSGGAPQYF

core + 3N3C

- CASSPGAAGEQYF
- CASSFPAAGIGDTQYF

Figure 4.7: CDR3β loop usage of CD4$^+$ T cells expanded using different PFR variants. Each chart shows the CDR3β loops identified in short term T cell lines grown against individual PFR variants and FACS sorted using HLA-DR1$^{core + 2N2C}$ (ie. HLA-DR1$^{HA306-318}$) dextramers.

4.5. Discussion

The aim for this chapter was to investigate the influence of peptide flanking residues on T cell activation, peptide stability and TCR repertoire selection. Testing ten PFR variants on three $CD4^+$ T cell clones showed that addition of C-terminal PFR in absence of N-terminal PFR re-established recognition. Addition of N-terminal PFR in absence of C-terminal PFR, however, did not restore recognition. These results show that in the $HA_{306-318}$ system, C-terminal PFR are essential for T cell recognition. Although N-terminal PFR were non-essential for recognition, addition of PFR at both termini simultaneously improved recognition compared to addition of C-terminal PFR on their own. This suggests some sort of synergistic effect between both termini. Observations from the competitive peptide-binding assay mirrored these results suggesting that peptide binding strength influences $CD4^+$ T cell recognition. Sant'Angelo and colleagues observed a similar effect in an I-Ab restricted peptide where addition of N-terminal PFR increased peptide stability to some degree whereas the addition of C-terminal greatly increased the stability and addition of PFR at both termini increased stability even further (Sant'Angelo et al. 2002). O'Brien and colleagues previously reported that longer peptides bind MHC-II with stronger affinity using an *in silico* approach (O'Brien, Flower, and Feighery 2008). However, they were unable to detect any difference between both termini, suggesting that this phenomenon might only be visible in *in vitro* assays. From the data gathered in this study, no conclusions on the effect of PFR variants on TCR binding could be drawn. Initially, measuring the impact of PFR on TCR affinity using plasmon surface resonance (SPR) was an additional aim for this thesis. Initial experiments showed that variations in PFR lengths indeed impact TCR affinities. However, difficulties in generating enough soluble protein meant that these experiments remain the subject of future investigations.

So far, clonotyping studies have investigated at the impact of PFR by looking at amino acid substitutions (Cole et al. 2012; Carson et al. 1997). Here, the impact of different lengths variants was investigated. Expanding PBMCs with different PFR variants and clonotyping $CD4^+$ T cells specific for the core + 2N2C peptide showed that each variant generated a different set of clonotypes. Apart from core + 2N, all variants generated narrower repertoires than core + 2N2C. The C-terminal variants each shared one clonotype with core + 2N2C at high frequencies. This shows that in the case of $HA_{306-318}$, C-terminal PFR in absence of N-terminal PFR and core + 2N2C generate overlapping TCR repertoires suggesting that TRB gene selection is steered to a certain degree by C-terminal PFR. Without further functional data, it is impossible to say whether these clonotypes depend on the presence of those C-terminal PFR in the same way the $CD4^+$ T cell clones tested here did. Since C-terminal PFR have been shown to stabilize peptide binding in this system, it is entirely possible that their presence is enough for these C-terminal PFR dependant clonotypes to expand. Cole and colleagues showed that a point mutation at the 1st and 2nd C-terminal PFR, respectively leads to a narrower and slightly altered although overlapping set of clonotypes proving that C-terminal PFR do play an important role in the selection of TCRB genes (Cole et al. 2012).

N-terminal PFR in absence of their C-terminal counterparts however, generated a completely different TCR repertoire compared to core + 2N2C. Again, a certain degree of dependence on the C-terminal PFR by the clonotypes identified in the core + 2N2C is a likely explanation for this requiring further functional analysis. The low binding stability of N-terminal PFR presents another potential explanation for this shift in the TCR repertoire. Interestingly, core + 3N3C generated a different set of clonotypes to core + 2N2C showing that extending peptides at both termini can alter the TCR repertoire significantly even if both peptides bind HLA-DR1 with similar stability. One clonotype

was shared between the 9mer core peptide and core + 2N2C suggesting that this clonotype is independent from any PFR. Carson and colleagues used a hen egg lysozyme derived peptide to show that C-terminal PFR influenced TCRB gene selection (Carson et al. 1997). Crystallographic evidence shows that the vast majority of TCR bind pMHC at a roughly diagonal angle with the TCRα-chain located above the N-terminal end of the peptide and the TCRβ-chain above the C-terminus (Markus G Rudolph, Stanfield, and Wilson 2006; Rossjohn et al. 2015). Therefore, the β-chain is more likely to make contacts with the C-terminal end of the peptide and its PFR explaining the effect on TCR gene selection seen by Carson and colleagues. Within the scope of this study, we exclusively looked at TRBβ gene usage and also detected a bias towards two TCR recombination identified both for the C-terminal variants as well as core + 2C2N.

Overall, the TCR repertoire against core + 2N2C identified here showed both similarities with and differences to the repertoire identified by Cole and colleagues. Some clonotypes were identified in both studies at different frequencies while others were unique to each study. Differences in the TCR repertoire are commonly seen following re-infection with heterogeneous pathogens (Sharma and Thomas 2014; Selin and Brehm 2007; A. T. Chen et al. 2012). The individual donating blood for this study has previously been vaccinated using the with the 2011/2012 trivalent influenza vaccine composed of the influenza A strains A/Perth/16/2009 (H3N2) and A/California/7/2009 (H1N1) as well as the influenza B strain B/Brisbane/60/2008. It is very likely that they had contact with heterogeneous influenza strains in the meantime particularly as they reported flu like symptoms some weeks before giving blood for this study. While shifts in the TCR repertoire are commonly seen, they are not always beneficial and can lead to pathology when effective clonotypes are eliminated from or diminish in frequency within the pool of $CD4^+$ T cells (Sharma and Thomas 2014; A. T. Chen et al. 2012).

On the other hand, clonotyping results shown here revealed that different PFR variants lead to partially overlapping but different TCR repertoires thereby broadening the overall repertoire of CD4$^+$ T cell capable of recognising certain epitopes. A broad TCR repertoire has been associated with increased protection against infections (Nikolich-Žugich, Slifka, and Messaoudi 2004; Messaoudi et al. 2002). Naturally processed peptides are characterised by varying lengths PFR which in vivo could contribute to the generation of a broad TCR repertoire. To prove this theory more clonotyping studies will be necessary including expansion of PBMCs against a cocktail of PFR variants.

In summary, the results presented here show that in the HA$_{306-318}$ system, CD4$^+$ T cell activation is likely to be dependent on C-terminal PFR and governed by peptide binding stability. They also show that the CD4$^+$ T cell repertoire of PBMCs raised against different PFR variants resulted in different but overlapping TCR repertoires. A broad TCR repertoire will be able to recognise and respond not only to naturally generated length variants of certain peptides but also heterologous sequences from different virus strains. The inherent cross-reactive nature of TCRs is used here to the advantage of the immune system. This advantage is particularly important in fast evolving pathogens such as influenza A and warrants the additional resources spent on generating many different clonotypes rather than a smaller set of only the most efficient ones.

5. Peptide flanking regions of an HIV derived HLA-DR1 epitope form an unusual hairpin conformation

5.1. Chapter Background

CD4$^+$ T cell responses play a crucial role in the control of HIV infections and have been shown to delay the onset and progress of AIDS. The p24 Gag protein (gag24) constitutes the capsid protein of the virus and contains a highly conserved region which has been shown to induce strong CD4$^+$ T cell responses. Here, the crystal structure of HLA-DR1^{gag24}, combined with functional assays, demonstrates the potential importance of secondary structure in the N-terminal PFR during CD4$^+$ T cell antigen recognition.

5.2. Introduction

5.2.1. *Role of gag24 in HIV infections*

HIV infects $CD4^+$ T cells via their $CD4^+$ co-receptor and use them as host cells for their own reproduction. This loss of $CD4^+$ T cells hampers both cytotoxic $CD8^+$ T cell responses as well antibody production and eventually leads to AIDS, thereby highlighting the important role $CD4^+$ T cells play in the immune system (Rosenberg et al. 1997). HIV specific $CD4^+$ T cells appear to be preferentially targeted by the virus and have been shown to lose their ability to proliferate in response to viral infection (J. D. Wilson et al. 2000; Douek et al. 2002). Nevertheless, $CD4^+$ T cell responses against HIV are crucial for successful control of the virus despite their numbers diminishing early on in the disease progress. High numbers of $CD4^+$ T cells specific for the viral gag24 protein have been shown to inversely correlate with lower viremia (Kaufmann et al. 2004). Patients defined as clinically non-progressors have been shown to exhibit robust anti-HIV $CD4^+$ T cell responses indicating that $CD4^+$ T cells can play an important role in slowing disease progress (Palmer, Boritz, and Wilson 2004). Early treatment with antiviral drugs aids maintenance of $CD4^+$ T cell responses thereby also slowing down disease progress (B. D. Walker et al. 2000).

5.2.2. *The HIV gag24 protein as highly conserved $CD4^+$ T cell antigen*

HIV hijacks the cellular translation machinery and uses it to generate viral proteins including the gag55 protein, a precursor protein important in formation of nascent virions (Freed 1998). Shortly after budding from the cell surface, gag55 is cleaved into four subunits by a virus-internal protease: the gag17 matrix protein (MA), the gag24 capsid

protein (CA), the gag7 nucleocapsid protein (NC) and gag6 (whose function has not been fully elucidated yet). **Figure 5.**1 shows a schematic overview of HIV highlighting gag24.

HIV specific CD4$^+$ T cell responses have been shown to be primarily directed against gag24 and Nef proteins as well as proteins encoded by the Pol gene (C. C. Wilson et al. 2001; Kaufmann et al. 2004). The gag24 protein encodes a segment called the major homology region (MHR) involved in replication and assembly of nascent viral particles. It has been shown to be highly conserved across all retrovirus (Provitera et al. 2001; Mammano et al. 1994). Interestingly, it also encodes a potent CD4$^+$ T cell epitope, DRFYKTLRAEQASA, which has been found to be presented by several MHC-II alleles including HLA-DR1 (Kaufmann et al. 2004; Scriba et al. 2005). The high conservancy of this gag24 peptide makes it a prime target for studying CD4$^+$ T cell responses against HIV.
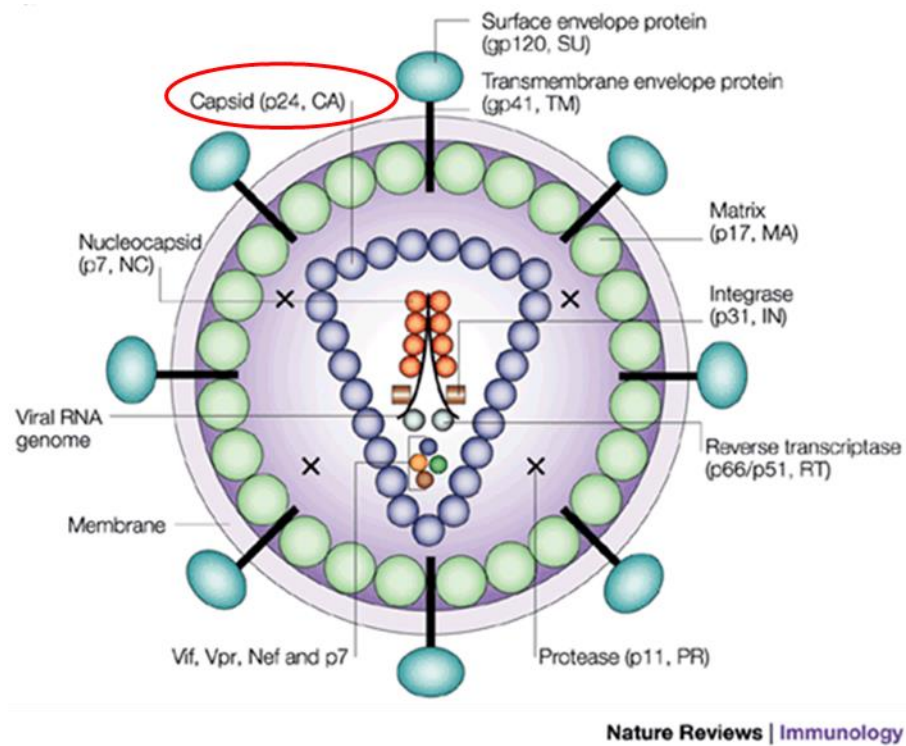
Figure 5.1: Schematic overview of the HI virus. Adapted from Robinson 2002. The gag24 capsid protein is encircled in red.

5.3. Aims

Early HIV research concentrated on anti-viral CD8[+] T cell and humoral responses. More recently, the importance of CD4[+] T cells has been increasingly recognised. In addition, advances in expression systems suitable for soluble pMHC-II molecules enable researchers to use soluble pMHC-II for in depths investigation of CD4[+] T cell responses (Scriba et al. 2005). Our lab has recently solved the structure of HLA-DR1[gag24]. Here, features observed in this structure have been investigated in more detail using modified peptides and testing them for recognition by CD4[+] T cells.

The first aim was to investigate potential TCR contact residues within the gag24 peptide by substituting selected residues with Alanine and testing them for recognition by a gag24 specific CD4[+] T cell clone.

The second aim was to investigate the impact of PFR on CD4[+] T cell activation using PFR length variants and testing them for recognition by HLA-DR1[gag24] specific CD4[+] T cell clones.

5.4. Results

5.4.1. Overview of the HLA-DR1^{gag24} crystal structure

We recently solved the crystal structure of HLA-DR1^{gag24} to 1.89 Å. (see **Figure** 5**.**2) (Cross 2016). Statistics of X-ray diffraction data collection and refinement statistics are shown in Supplementary Table S2. The resolution of 1.89Å allowed to analyse the structure in good detail. Following several rounds of refinements, the reliability factor (R factor) was 0.20, indicating that the refined structure fits well with the collected crystallographic data. The difference between R factor and R_{free} factor (0.26) was 0.06, indicating that no major artificial adjustments have been introduced into the structure that do not fit with the collected crystallographic data. It shows a typical pMHC-II conformation with the peptide adopting a flat conformation within the peptide binding groove (see **Figure** 5**.**2). Electron density around the peptide was unambiguous with the exception of the C-terminal glutamic acid at position P_{11}. Therefore, the position of the sidechain could not be determined with certainty. Figures including P_{11} show its modelled position based on the electron density available for the backbone and COOH terminus.

As shown in **Figure** 5**.**2B, the gag24 peptide binds to HLA-DR1 using the usual anchor residues at position P_1 (tyrosine), P_4 (leucine), and P_9 (alanine). As expected, these three residues point down and bind in pockets located within the β-sheet floor of the binding groove. Residues at positions P_{-3} (asparagine), P_{-1} (phenylalanine), P_2 (lysine), P_5 (arginine), P_7 (glutamate), P_8 (glutamic acid) as well as the C-terminus, however, point upwards while residues in position P_{-2} (arginine), P_3 (threonine), P_6 (alanine), and P_{10} (serine) point toward the walls of the peptide binding groove.
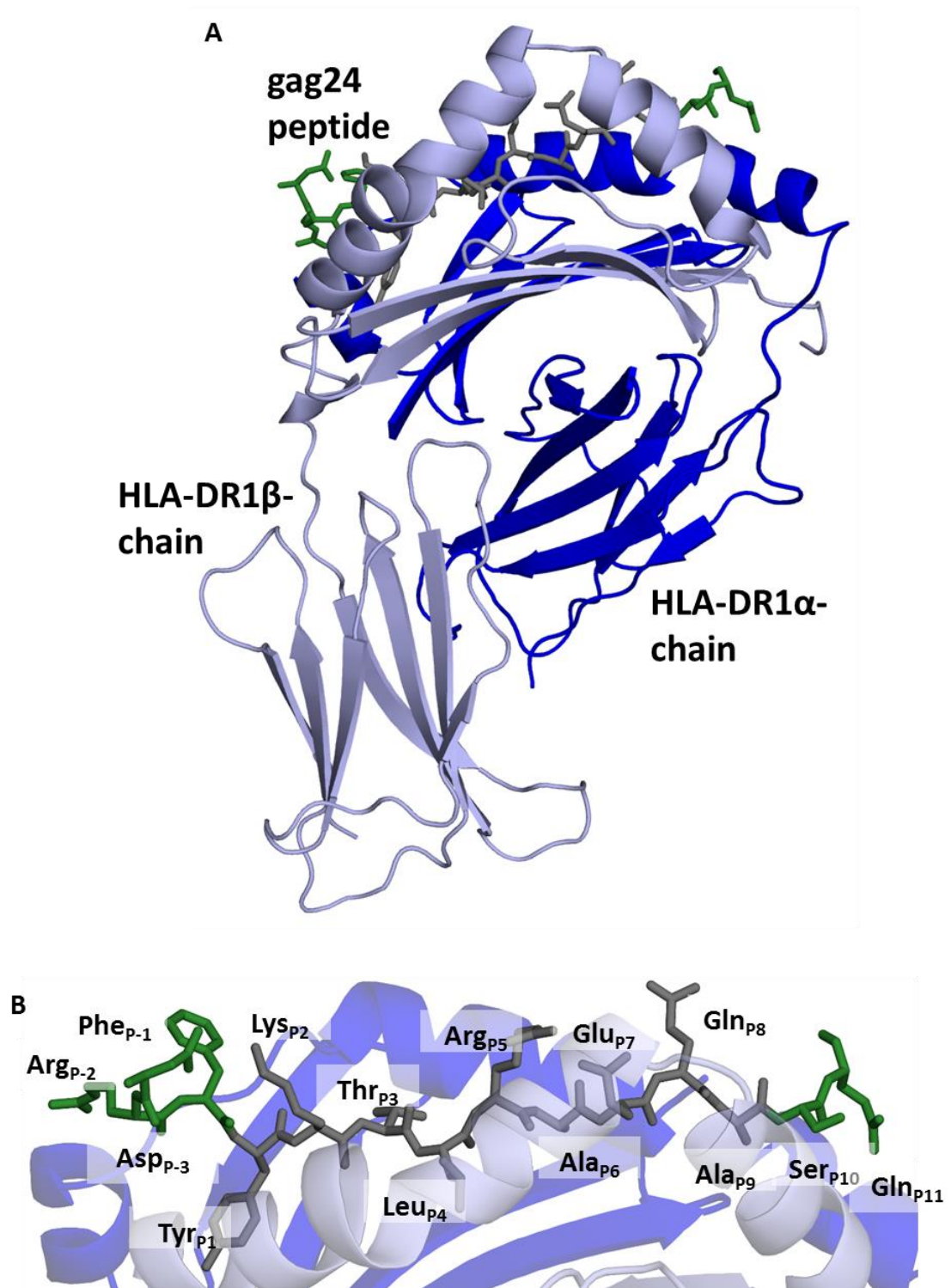
Figure 5.2: Overview of the HLA-DR1gag22 structure. A: Side view of HLA-DR1^{gag24} as a whole. B: Detailed view of the gag24 peptide in the HLA-DR1 peptide binding groove. HLA-DR1α-chain is shown in dark blue, HLA-DR1β-chain is shown in light blue. Peptide core is shown in grey while PFR are shown in dark green.
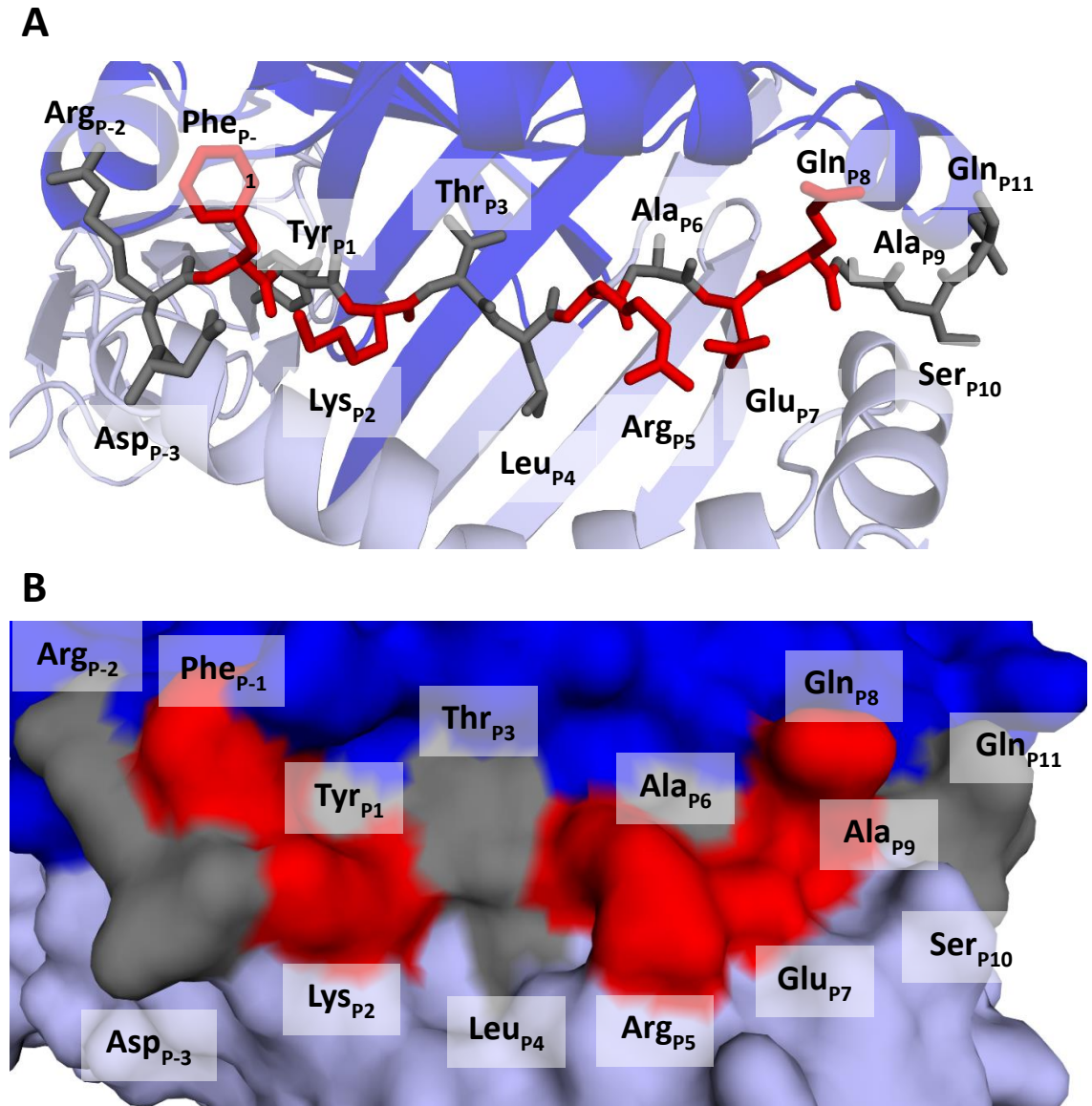
Figure 5.3: HLA-DR1^{gag24} topology as from a birds eye perspective. A: peptide shown in stick representation, MHC-II shown in cartoon representation. B: pMHC-II shown in surface representation. HLA-DR1α-chain is shown in dark blue, HLA-DR1β-chain is shown in light blue. Potential TCR contact residues are shown in red, remaining residues shown in grey.

5.4.2. *Investigating crucial TCR contact positions within the gag24 peptide*

Figure 5.3 shows HLA-DR1^{gag24} from a birds-eye a perspective. The surface representation shown in **Figure** 5.3B is a visualisation of the topology that a TCR binding to HLA-DR1^{gag24} would encounter. This surface is created by the electron densities of all surface exposed residues. While the gag24 peptide binds in the flat fashion typical for MHC-II bound peptides, some residues are solvent exposed and protrude from the HLA-DR1 peptide binding groove (highlighted in red in **Figure** 5.3). Crystal structures of TCR/pMHC complexes have shown that solvent exposed residues are preferentially contacted by TCRs (Rossjohn et al. 2015; Garcia, Teyton, and Wilson 1999).

In order to investigate the role of these solvent exposed residues for CD4$^+$ T cell activation, a set of alanine substitution mutations were designed (see **Table** 5.1). Although P$_3$ was not highly solvent exposed it was included in this study as TCRs have been shown to make contacts with any residues between P$_{-1}$ and P$_{10}$ (O'Brien, Flower, and Feighery 2008; J Hennecke and Wiley 2001). For the same reason, P$_{-3}$ was excluded despite it being solvent exposed. P$_6$ already encoded an Alanine and therefore was not included. These peptides were tested for recognition by the CD4$^+$ T cell clone Ox97 clone 10 recognising the gag24 peptide in a peptide titration assay and INFγ release was measured as described in Chapter 2.3.7 and 2.3.9 (Scriba et al. 2005). Results are shown in **Figure** 5.4.

			Peptide sequence													Description
N-terminal						*Core epitope*							*C-terminal*			
-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11		
D	R	F	Y	K	T	L	R	A	E	Q	A	S	Q		Core + 3N2C
D	R	A	Y	K	T	L	R	A	E	Q	A	S	Q		P$_{-1}$Ala
D	R	F	Y	A	T	L	R	A	E	Q	A	S	Q		P$_2$Ala
D	R	F	Y	K	A	L	R	A	E	Q	A	S	Q		P$_3$Ala
D	R	F	Y	K	T	L	A	A	E	Q	A	S	Q		P$_5$Ala
D	R	F	Y	K	T	L	R	A	A	Q	A	S	Q		P$_7$Ala
D	R	F	Y	K	T	L	R	A	E	A	A	S	Q		P$_8$Ala

Table 5.1: Replacing potential TCR contact residues with alanine. Solvent exposed residues were replaced with alanine (highlighted in bold). The 9mer core is shown in grey, potential TCR contact positions are highlighted in red.
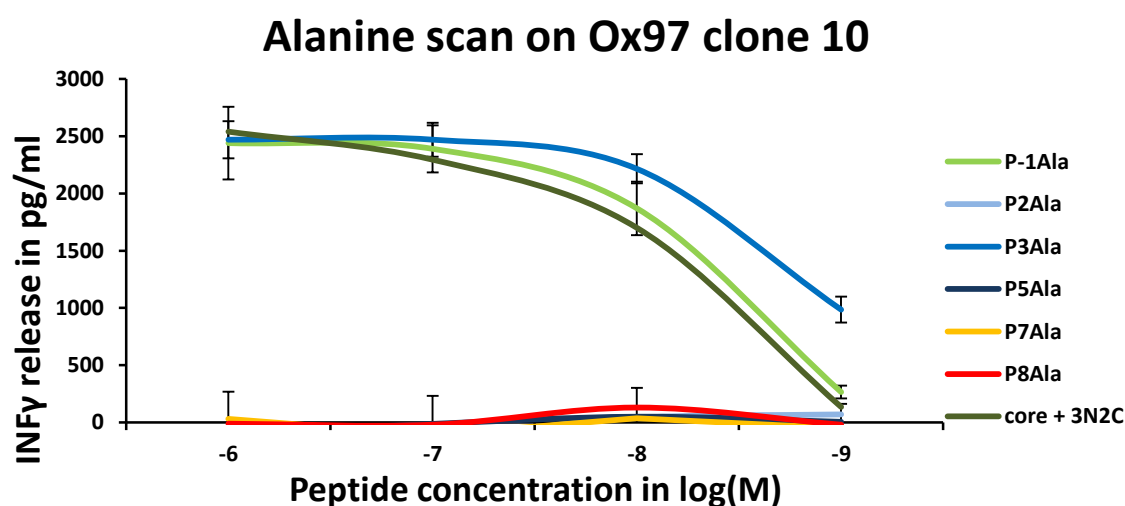
Figure 5.4 Testing alanine substitutions along the gag24 backbone. Alanine mutations were tested in a peptide titration and INFγ release was measured. The wildtype peptide is shown in green. Error bars show standard deviation.

Substituting P$_{-1}$ or P$_3$ by Alanine did not reduce recognition of the wildtype peptide indicating that these residues are no TCR contact residues. Any substitutions at positions P$_2$, P$_5$, P$_7$ or P$_8$ however, led to a complete failure of recognition. This all-or-nothing

response suggests that these peptide positions are essential for recognition by the Ox97 clone 10 and possibly crucial TCR contact points.

5.4.3. *Structural analysis of the N-terminus of the gag24 peptide*

While the core of peptides bound to MHC-II adopts a flat, extended conformation due to its placement in the peptide binding groove, the extruding PFRs are a lot more flexible and can adopt a variety of conformations. In the HLA-DR1^{gag24} structure, the N-terminus forms a hook shape bending back on itself with P_{-2} constituting the apex of this bend (see **Figure** 5.5). **Table** 5.2 shows a summary of all peptide-MHC contacts made by PFR of the gag24 peptide structure (the full list of contacts is shown in Supplementary Table S3). P_{-2} stands out through the high number of contacts made by the arginine side chain with residues of the HLA-DR1α-chain including seven hydrogen bonds. The HLA-DR1α-chain residues involved in these contacts are Arginine$_{α50}$, Phenylalanine$_{α51}$, Alanine$_{α52}$ and Serine$_{α53}$. **Figure** 5.6A shows a selection of these peptide-MHC contacts. The side chain of Aspartate$_{P-3}$, on the other hand, is involved in a network of interactions with both Phenylalanine$_{P-1}$ as well as Leucine$_{P3}$. A selection of these is shown in **Figure** 5.6B. Together, these interactions hold the hooked shaped N-terminus in place. Due to the missing electron density for Glutamate$_{P11}$, no conclusions about peptide-MHC and intra-peptide contacts could be made for the C-terminus of the gag24 peptide.
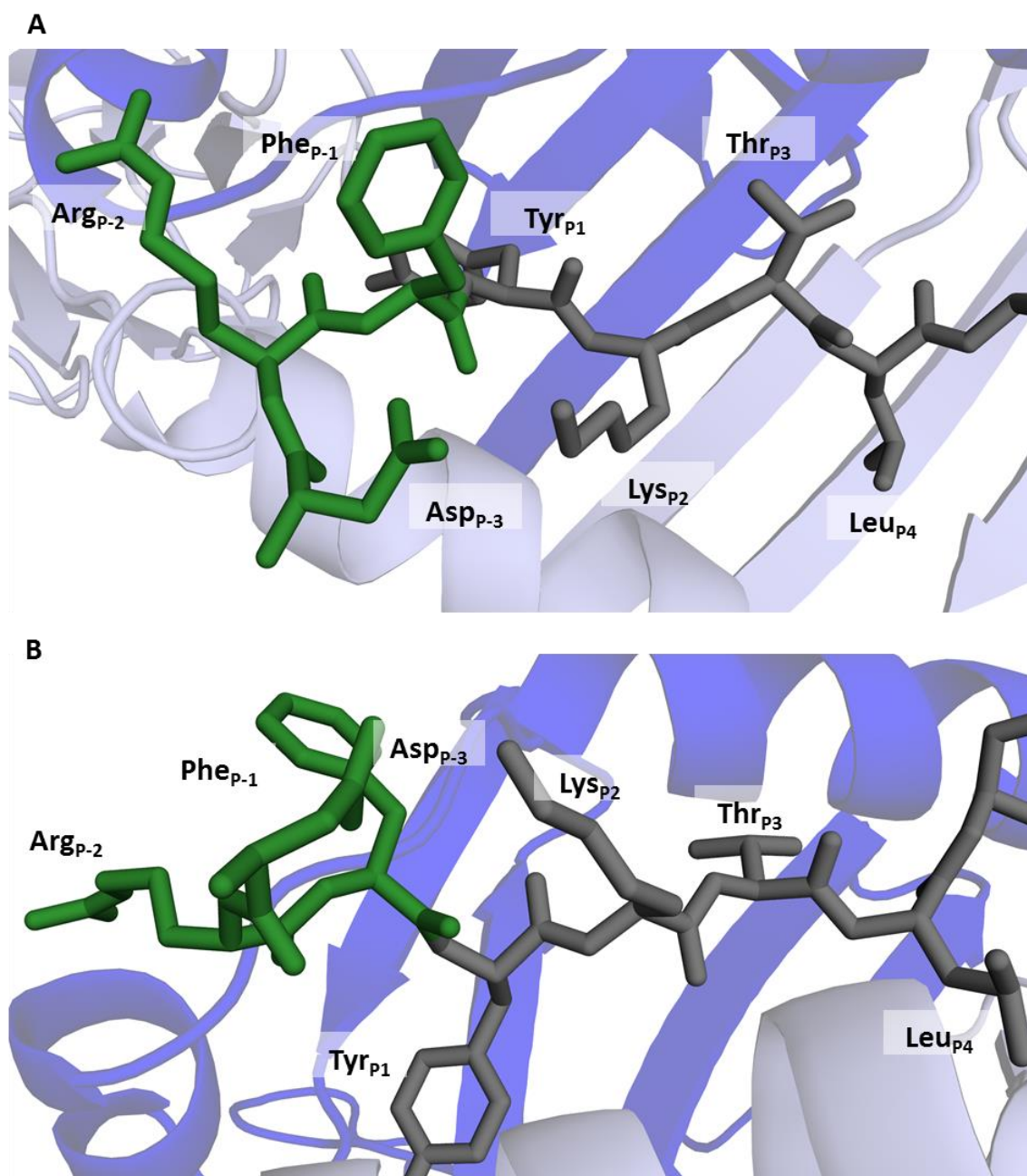
Figure 5.5: The N-terminus of gag24 adopts a hook conformation. A: The gag24 N-terminus from a bird's-eye view. B: The gag24 N-terminus from a side view. Peptide core residues are shown in grey, PFR are shown in dark green. HLA-DR1α-chain shown in dark blue, HLA-DR1β-chain shown in light blue.
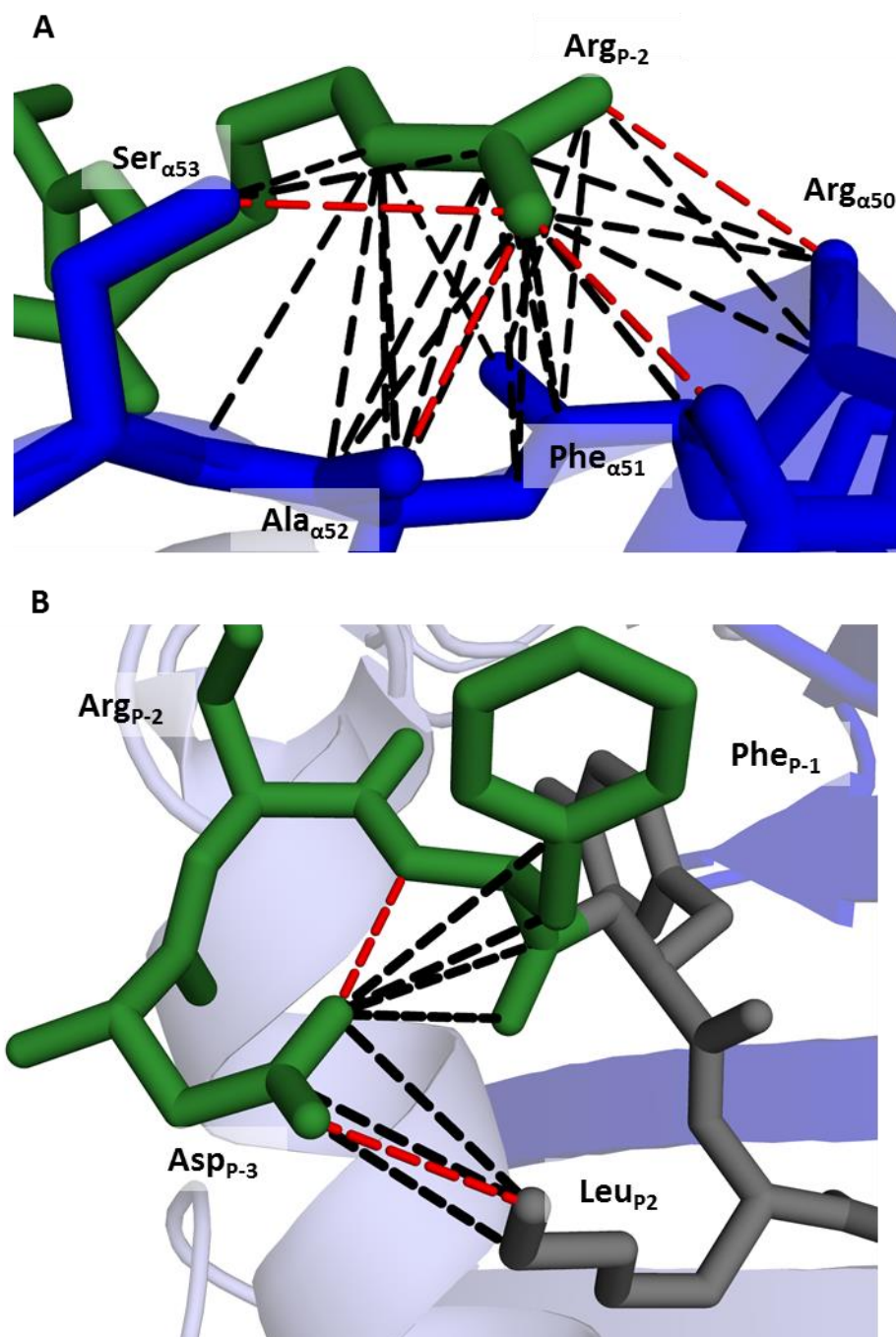
Figure 5.6: The N-terminal hook is held in place by a network of molecular bonds.
A: Contacts made by arginine in P$_{-2}$ with residues of the HLA-DR1α-chain holds this side chain in place. B: Aspartate in P$_{-3}$ is held in place by a network if intra-peptide interactions with phenylalanine in position P$_{-2}$ and leucine in position P$_2$. Core residues are shown in grey, PFR are shown in dark green, HLA-DR1α-chain is shown in dark blue, HLA-DR1β-chain is shown in light blue, van der Vaals interactions are shown in black and hydrogen bonds are shown in red.

Peptide residues		MHC residues			Contacts		
Position	Aa code	Chain	Number	Aa code	VdW	H-bonds	both
-3	ASP	B	81	HIS	2	1	3
			85	VAL	1		1
				total:	**3**	**1**	**4**
-2	ARG	A	49	GLY	2		2
			50	ARG	4	1	5
			51	PHE	6	2	8
			52	ALA	11	2	13
			53	SER	6	2	8
		B	85	VAL	2		2
				total:	**31**	**7**	**38**
-1	PHE	A	53	SER	7		7
			54	PHE	2		2
			55	GLU	5		5
		B	81	HIS	3		3
			85	VAL	1		1
				total:	**18**		**18**
10	SER	A	72	ILE	1		1
			76	ARG	2		2
		B	56	PRO	1		1
			57	ASP	3		3
			60	TYR	3		3
				total:	**10**		**10**

Table 5.2: Summary of PFR-MHC contacts within HLA-DR1[gag24]. Van der Waals (VdW) interactions and hydrogen bonds (H-bonds) between PFR of the gag 24 peptide and MHC residues are listed. No contacts for P_{11} are shown due to missing electron density. Chain A: HLA-DR1α-chain; chain B: HLA-DR1β-chain.

Peptide residues		Peptide residues		Contacts		
Position	Aa code	Position	Aa code	VdW	H-bond	both
-3	ASP	-2	ARG	11	3	14
		-1	PHE	9	1	10
		2	LYS	3	1	4
			total:	**23**	**5**	**28**
-2	ARG	-3	ASP	11	3	14
		-1	PHE	12	1	13
		1	TYR	2		2
			total:	**26**	**3**	**29**
-1	PHE	-3	ASP	9	1	10
		-2	ARG	12	1	13
		1	TYR	12	1	13
		2	LYS	5		5
			total:	**38**	**3**	**41**
10	SER	8	GLN	1		1
		9	ALA	12	1	13
			total:	**13**	**3**	**16**

Table 5.3: Summary of intra-peptide contacts within the gag24 peptide. Van der Waals (VdW) interactions and hydrogen bonds (H-bonds) made by PFR with any other peptide residues are listed. No contacts for P_{11} are shown due to missing electron density.

5.4.4. Investigating the impact of PFR on CD4⁺ T cell activation

In order to investigate the hook shaped N-terminus and its impact on CD4$^+$ T cell activation in more detail, a series of PFR variants were designed (see **Table** 5**.**4). Despite ill-defined structural data for the C-terminus of the gag24 peptide, C-terminal PFR were designed in order to gain more insight. Both sided PFR variants were designed in order to investigate the interplay of N- and C-terminal PFR. Peptides were tested for recognition by the sister clones Ox97 clone 10 and Ox97 clone 11 in a peptide titration and INFγ release was measured as described in Chapter 2.3.7 and 2.3.9. Both clones express the same TCR despite having been cloned independently. Results for both clones are shown in **Figure** 5**.**7.

Peptide sequence															**Description**
N-terminal				*Core epitope*								*C-terminal*			
-3	*-2*	*-1*		*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	
				Y	K	T	L	R	A	E	Q	A			core
				Y	K	T	L	R	A	E	Q	A	S		core + 1C
				Y	K	T	L	R	A	E	Q	A	S	Q	core + 2C
		F		Y	K	T	L	R	A	E	Q	A			core + 1N
	R	F		Y	K	T	L	R	A	E	Q	A			core + 2N
D	R	F		Y	K	T	L	R	A	E	Q	A			core + 3N
		F		Y	K	T	L	R	A	E	Q	A	S		core + 1N1C
	R	F		Y	K	T	L	R	A	E	Q	A	S	Q	core + 2N2C
D	R	F		Y	K	T	L	R	A	E	Q	A	S	Q	core + 3N2C

Table 5.4: Nested set of peptide based on gag24. Starting with the 9mer core region, PFR were extended one residue at a time either at one terminus only or both termini at the same time.
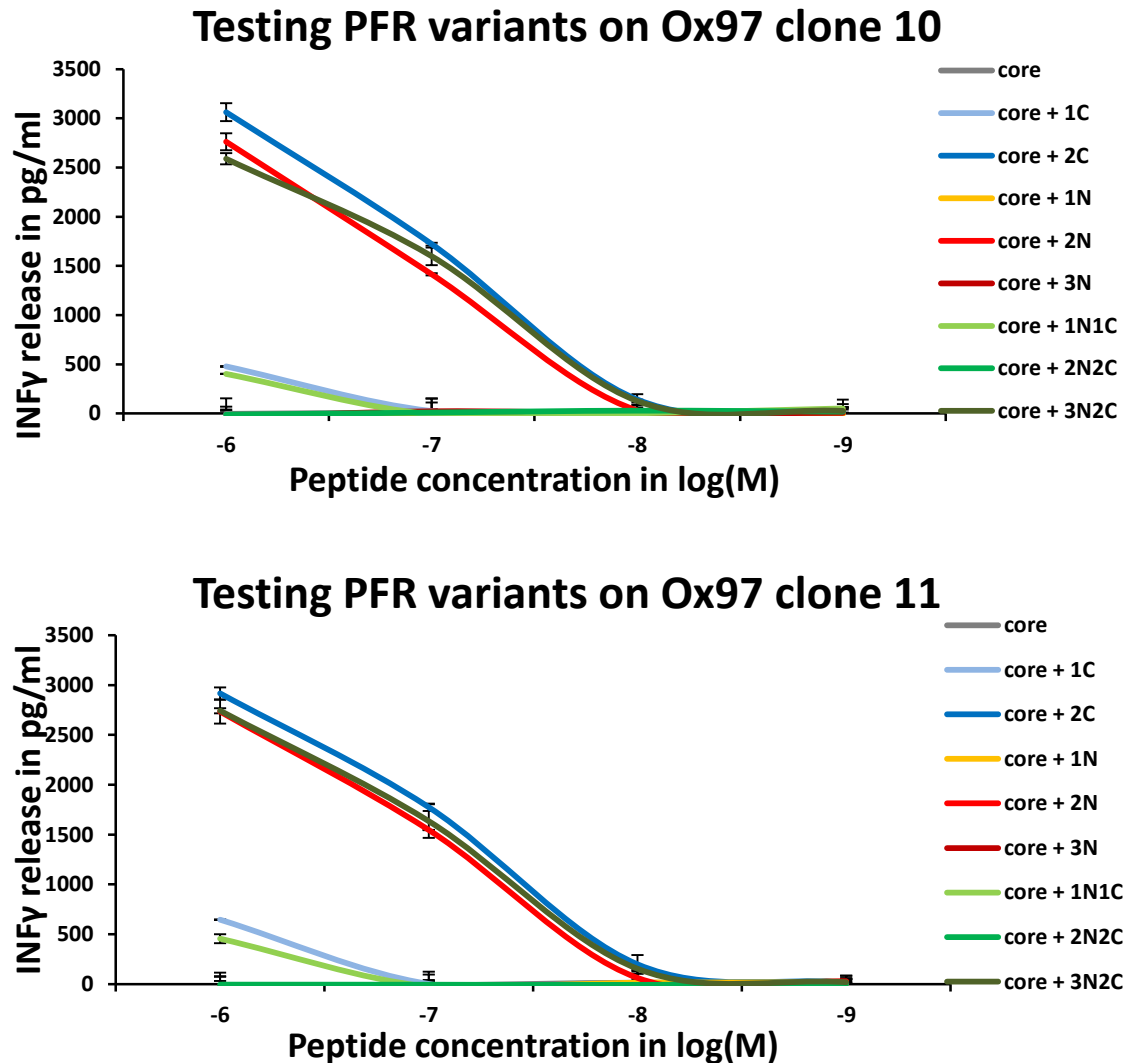
Figure 5.7: Testing PFR length variants on Ox97 clone 10 and clone 11. PFR variants were tested in peptide titrations. CD4[+] T cell clones were incubated overnight with peptides at different concentrations in presence of HOM-2 cells. INFγ release was measured using ELISA. The 9mer core peptide is shown in grey, C-terminal PFR variants in shades of blue, N-terminal PFR variants in shades of red and NC-terminal variants in shades of green. Error bars show standard deviation.

Both clones expanded very poorly in culture, therefore, the overall number of cells and range of concentration was reduced to 9,000 to 11,000 cells/well and 10^{-6} M to 10^{-9} M, respectively. Despite lower cell numbers, the quantity of IFNγ released in response to peptide was comparable what has been observed for other CD4[+] T cell clones such as the

HA$_{306-318}$ specific CD4$^+$ T cell clones used in Chapter 4.There was no measurable difference in the responses between Ox97 clone and Ox97 clone 11. The 9mer core did not elicit any responses even at high peptide concentration demonstrating an important role for the PFR during recognition. In support of this observation, the core + 1C as well as core + 1N1C were also poorly recognised at the concentrations tested. Addition of both C-terminal PFR (core + 2C), however, re-established recognition to the same level as the full length gag24 peptide (core + 3N2C). While addition of Phenylaline$_{P-1}$ on its own (core + 1N) was not enough to re-establish recognition, addition of both Arginine$_{P-2}$ and Phenylalanine$_{P-1}$ (core + 2N) re-established recognition to the same level as seen for core+ 3N2C. Interestingly, the further addition of Aspartate$_{P-3}$ (core + 3N) completely abolished recognition. Despite core + 2C and core + 2N both being recognised to the same level as the full-length gag24 peptide, core + 2N2C was not recognised at all.

Due to the small range of concentrations used for this assay, no EC$_{50}$ values could be determined. **Table** 5.5 shows IFNγ release measured at peptide concentrations of 10^{-6} M.

peptide	clone 10	clone 11
core	0	0
core + 1C	478	645
core + 2C	3062	2916
core + 1N	0	0
core + 2N	2762	2734
core + 3N	0	0
core + 1N1C	401	455
core + 2N2C	0	0
core + 3N2C (wt)	2589	2743

Table 5.5: Summary CD4$^+$ T cell responses to PFR variants. Maximum INFγ release at 10M^{-6} is shown. Colour gradient ranges from red (low INFγ release) to green (high IFNγ release).

In summary, addition of two C-terminal PFR to the 9mer core in absence of any N-terminal PFR re-established recognition. In absence of C-terminal PFR, addition of one or three N-terminal PFR did not re-establish recognition, while the addition of two did not. Addition of PFR at both termini simultaneously required all three N-terminal and both C-terminal PFR in order to re-establish recognition. These findings demonstrate the critical role that PFRs play during CD4 T cell mediated antigen recognition, but also indicate that the PFRs do not act independently of each other. This observation is consistent with other studies from our laboratory showing that modifications to the peptide can have knock on effects at distal sites (Cole et al. 2010; Madura et al. 2015; Bianchi et al. 2016). This complex interplay suggests that the PFR are playing a dual role affecting both TCR binding and pMHC-II stability. A more thorough examination would be needed to deconvolute these roles and gain full insight into the mechanism(s) by which the PFRs can tune antigen recognition in this system.

In order to investigate whether these observations were due to altered peptide stability, PFR variants were tested in a competitive peptide binding assay described in Chapter 2.2.6 and Chapter 4. However, none of the tested peptides were able to replace the BT-CLIP peptide. This could be partially explained with certain peptides coming out of solution in the acidic binding buffer (pH = 5) which had not been observed for any of the $HA_{306-318}$ variants tested in Chapter 5.

5.5. Discussion

The Gag24 peptide has not been studied in comparable detail as other CD4$^+$ T cell epitopes such as HA$_{306-318}$. Overall, the structure I generated was of high resolution (1.89Å) and good electron density was observed for the peptide except at position 11. The first aim of this chapter was to investigate any potential TCR contact residues. Due to the flat binding conformation adopted by peptides bound to MHC-II, TCR contacts are usually solvent exposed residues protruding above the walls of the peptide binding groove. Replacing suitable candidates by alanine and testing them for CD4$^+$ T cell recognition identified P_2, P_5, P_7 and P_8 as potential crucial TCR contacts. This fits well with the crystal structure of HLA-DR1^{gag24} as these residues are all highly solvent exposed. T cell activation assays on their own are not sufficient to prove that these residues are directly contacted by the TCR. Further biophysical and structural investigations using HLA-DR1^{gag24} specific TCRs such as Ox97 will give more insight into TCR-peptide contacts.

Unlike the core of the peptide, PFR can adopt different conformations without the spatial limitations of the peptide binding groove. They can extend away from the MHC or fold back to make additional contacts with the MHC and other peptide residues, respectively. The N-terminus of the gag24 peptide adopts a hook-like shape, presumably being held in place by intra-peptide interactions made by Aspartate$_{P-3}$ with Phenylalanine$_{P-1}$ and Leucine$_{P3}$, respectively. In addition, Arginine$_{P-2}$ is held in place at the apex of the hook by an extensive network of interactions with the HLA-DR1-chain. Testing PFR variants for recognition gave some insight into the role these residues play in CD4$^+$ T cell recognition. The presence of Phenylaline$_{P-1}$ and Arginine$_{P-2}$ in the absence of any C-terminal PFR was sufficient to yield CD4$^+$ T cell recognition to the same level as the full-

length gag24 peptide. Assuming that $Arginine_{P-2}$ engages in similar contact as seen in the HLA-DR1^{gag24} structure, this can be explained by $Arginine_{P-2}$ successfully tethering the core + 2N peptide to HLA-DR1. This also shows that the missing P_{-3}, P_{10} and P_{11} are unlikely to be critical TCR contact residues. Interestingly, additional extension of the N-terminal by $Aspartate_{P-3}$ led to a complete loss of CD4$^+$ T cell activation. Despite being located at the opposite end of the peptide binding groove, missing C-terminal PFR could have knock-on effects along the peptide preventing the hook from forming and forcing the N-terminus into a position unfavourable for TCR binding and CD4$^+$ T cell recognition. This could be due to conformational changes within the peptide preventing $Phenylalanine_{P-1}$ and $Leucine_{P3}$ from making contacts with the $Aspartate_{P-3}$ side chain or even affecting TCR contacts within the peptide core. In the case of $Leucine_{P3}$, a conformational change could have a double effect altering both a potential TCR contact residue preventing the N-terminus from adopting the hook like shape. Peptides missing PFR have been shown to exhibit increased motion (Haruo Kozono et al. 2015). The missing C-terminal PFR could increase the motion within the gag24 to a degree that prevents the formation of the hook and causes $Aspartate_{P-3}$ to interfere with TCR binding. Further biophysical and structural investigations into the different N-terminal PFR and their effect into will shed more light onto this phenomenon.

In contrast, very little could be deducted about the C-terminal PFR of gag24 due to incomplete structural data. Presence of the full length terminus in absence of any N-terminal PFR (core + 2C) was sufficient for CD4$^+$ T cell recognition. Puzzlingly, the presence of two PFR at both termini (core + 2N2C) completely abolished CD4$^+$ T cell recognition despite each extension by itself (core + 2N and core + 2C, respectively) being recognised just as well as the full length gag24 peptide. It is possible that the presence of two C-terminal PFR impacts the conformation of the N-terminus in absence of the

Aspartate$_{P-3}$ and therefore in absence of any stabilising effect of the hook-like conformation. This observation was in contrast with the HA$_{306-318}$ system where addition of PFR at both termini had a synergistic effect on CD4$^+$ T cell activation, presumably due to increased peptide stability (see Chapter 4).

A similar hook shaped conformation has previously been observed in the C-terminal PFR of another HLA-DR1 restricted, gag24 derived epitope (Zavala-Ruiz et al. 2004). A glycine located at the apex of the hook generated a hinge along which the PFR could turn and form the loop. Replacing this glycine by a proline and thereby forcing the C-terminal PFR to follow the extended conformation of the core led to a complete loss of CD4$^+$ T cell activation, confirming the important role secondary structures within MHC-II bound peptides can play (Norris et al. 2006). The N-terminal of the pMART-1 phosphopeptide bound to HLA-DR1 has also been shown to adopt a hook-like shape. In contrast to the gag24 derived peptides however, this bend was less acute causing the N-terminal residues to point upwards out of the groove rather than turning back in a hair-pin like structure (Li et al. 2010). Unlike the HLA-DR1gag24 structure presented here, removal of any N-terminal PFR abolished CD4$^+$ T cell recognition of the pMART-1 peptide. Removal of any C-terminal PFR in presence of N-terminal PFR, however, caused little impact on CD4$^+$ T cell recognition again suggesting that the secondary structure of the N-terminus was crucial for CD4$^+$ T cell recognition.

In contrast to the majority of the HIV genome, the MHR of the gag24 protein has been shown to be highly conserved across all retrovirus (Mammano et al. 1994; Provitera et al. 2001). In consequence, CD4$^+$ T cell responses raised against the gag24 epitopes studied here will remain effective once other parts of the virus escape T cell immunity through mutation. This makes gag24 a prime target for both tracking anti-HIV responses

throughout different stages of infection as well as for vaccine research. The high conservancy of gag24 makes it a universal epitope across all HIV strains. Therefore, features identified as important for TCR recognition here, will be present irrespective of which HIV strain the epitope is derived from.

In this chapter, the structure of HLA-DR1^{gag24} served as a basis for investigations into CD4$^+$ T cell recognition of pMHC-II. Several solvent exposed residues were identified as potential TCR contact residues. Dissecting the role of individual residues within the hook-shaped N-terminal PFR gave further insight into how residues outside the 9mer binding core can influence CD4$^+$ T cell recognition. This is the third study showing that secondary structures within PFR can have major impact on CD4$^+$ T cell recognition further confirming the important role PFR can play. This knowledge has implications for the design of peptide vaccines and pMHC-II multimer staining since even seemingly small changes to PFR can have major impact on CD4$^+$ T cell recognition or TCR binding. Results presented here also generated a starting point for a range of further investigations based on the gag24 peptide. Repeating the functional assays using additional CD4$^+$ T cell clones will broaden the insight into potential TCR contact residues and the role of PFR in CD4$^+$ T cell recognition of this highly conserved peptide. SPR analysis of the alanine substitution peptides will give more insight into the importance of the here identified potential TCR contact residues for TCR binding. Additional SPR as well as structural analysis of selected PFR variants will add further knowledge about role of secondary structures within PFR. In summary, this study has added to our knowledge of PFR in CD4$^+$ T cell immunity while also opening future avenues of investigation.

6. CD4$^+$ T cell responses to influenza hemagglutinin

6.1. Chapter Background

CD4$^+$ T cells play a crucial role in the anti-viral immune response by stimulating B cells to produce antibodies and cytotoxic T cells to eliminate infected host cells. In order to fulfil this role, CD4$^+$ T cells must be able to specifically recognise the intracellular pathogens through their T cell receptor binding peptide fragments of virus proteins presented on MHC-II molecules on the surface of professional antigen presenting cells as well as infected host cells. Identifying which peptides are recognised enables us to study CD4$^+$ T cell responses in more detail. For example, pMHC-II-multimer staining enables the visualisation of specific CD4$^+$ T cell populations using flow cytometry in order to track them throughout an on-going infection or following vaccination.

6.2. Introduction

6.2.1. *Influenza hemagglutinin is an important target for CD4$^+$ T cell responses*

Hemagglutinin (HA) plays a crucial role in the life cycle of the Influenza A virus by mediating its attachment to α(2,6) linked sialic acids on the surface of host cells. The substrate-binding site is located within the globular head domain (see Figure 6.1). Following uptake into the host cell the pH in virus containing endosomes is lowered in order to create an optimal milieu for host proteases. This change in pH however, also leads to a change in conformation within the HA protein resulting in the C-terminal fusion peptide to insert itself into the endosome membrane and ultimately, its fusion with the viral membrane envelope. Thus, the viral RNA is released into the cytoplasm (Samji 2009; Das et al. 2010). HA is an important target for the immune system as well as for anti-viral drugs. Blocking of the sialic acid binding site by neutralising antibodies and small molecule inhibitors, respectively, prevent the virus from infecting host cells (Das et al. 2010).

Viral proteins are degraded by host cell proteases and enter the MHC-II antigen presentation pathway. Despite the important role of CD4$^+$ T cells in fighting influenza A infections, only few epitopes within the Influenza hemagglutinin have been characterised in detail. The best characterised MHC-II restricted epitope in the HA$_{308-316}$ epitope (Lamb and Green 1983) (see Chapter 4 for more details). To date, 415 MHC-II restricted HA epitopes are registered in the Immune Epitope Database (IEDB) (Vita et al. 2015). On closer inspection, the majority of these are identical or of overlapping sequences, stem from different strains showing point mutations within the sequence or have been shown to bind to more than one HLA allele.
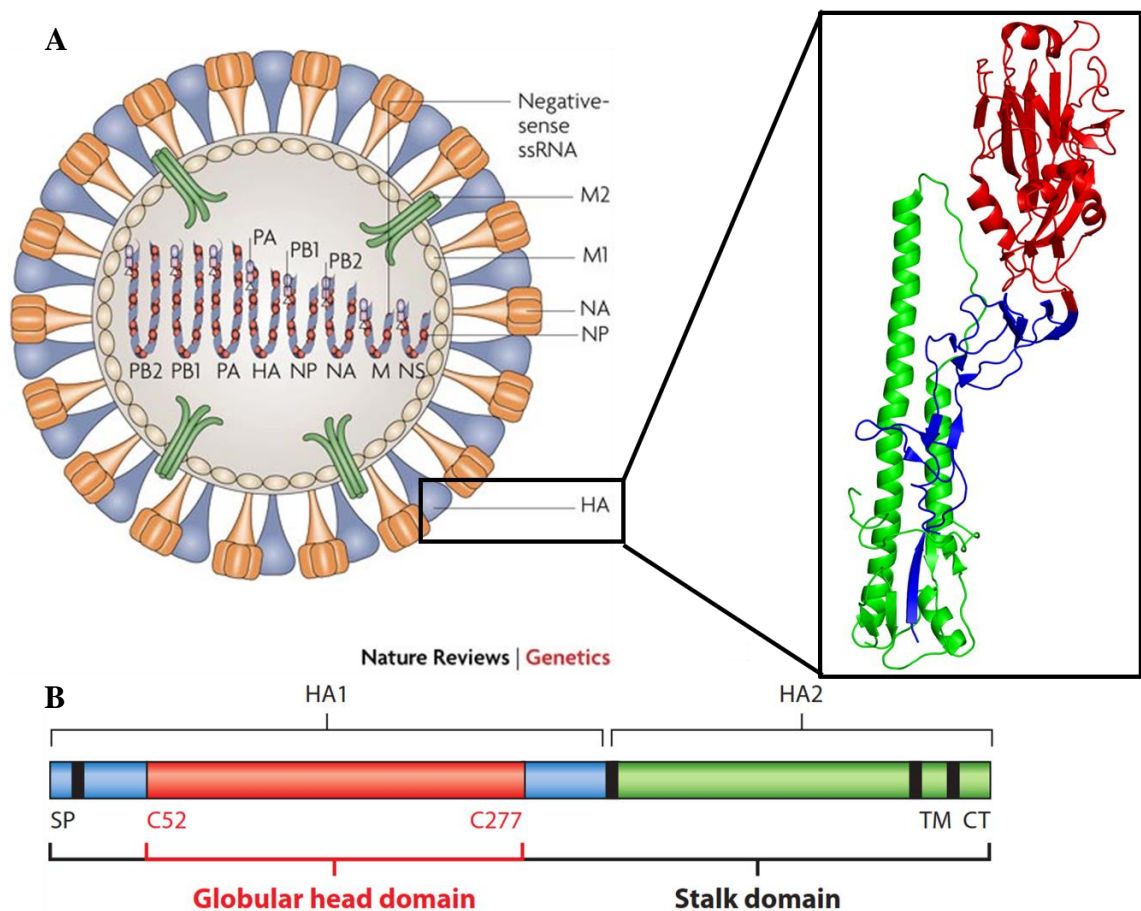
Figure 6.1: Influenza hemagglutinin consists of two subunits, HA1 and HA2. (A) HA1 forms both the globular head domain (coloured red in the cartoon representation), and the fusion peptide (coloured in blue). HA2 forms the stalk or stem domain (coloured in green). PDB code: 4O5N (B) Schematic overview of both HA subunits. Taken from Pica & Palese 2013; Nelson & Holmes 2007.

6.2.2. *Epitope mapping using overlapping peptide libraries*

Well characterised HA epitopes play an important role in monitoring and studying Influenza infections in general and for vaccine studies in particular. Investigating CD4[+] T cell responses against specific epitopes using functional assays such as ELISPot or intracellular cytokine staining (ICS) as well as monitoring CD4[+] T cell populations using multimer staining and flow cytometry can give valuable insight into anti-viral responses

(Long et al. 2013). Different approaches allow the identification of T cell epitopes. Bioinformatic methods rely on the prediction of potential epitopes based on either peptide binding patterns i.e. known anchor residues for certain HLA alleles or proteolytic cleavage sites (Mettu, Charles, and Landry 2016). Eluting peptides from MHC-II and identifying their sequence through mass spectrometry is another option and allows for the identification of different length variants at the same time (Lippolis et al. 2002; Godkin et al. 1998; Strug et al. 2008). The most established method, however, is based on dividing the sequence of the antigen of interest into overlapping peptides of defined length. These can then be tested for T cell recognition by measuring cytokine production in response to peptide stimulations as well as by multimer staining. In order to keep the size of these assays reasonable, peptides are usually combined in a more manageable number of pools. Once responses against certain pools have been detected, peptides are tested for individual responses (Fiore-Gartland et al. 2016).

6.3. Aims

Gelder and colleagues identified five regions within the HA protein from a H3N2 strain which elicited strong responses in HLA-DR1$^+$ individuals (C. M. Gelder et al. 1995). These regions span over 40aa in lengths. For this chapter of my thesis I set out to further characterise peptides within these regions using overlapping peptide libraries and to test their HLA-DR1 restriction.

The first aim was to identify peptides within these five regions using overlapping peptide libraries and testing peripheral blood mononucleated cells (PBMCs) from three healthy donors for responses.

The second aim was to confirm the HLA-DR1 restriction of the peptides identified in the first step.

The third aim was to investigate the conservancy of these potential new epitopes in comparison with the well characterised HA$_{306-318}$ epitope.

6.4. Results

6.4.1. *Identifying potential epitopes using overlappoing peptides*

Gelder and colleagues studied responses against HA peptide pools as well as individual peptides and compared them against the HLA-type of their donors. They identified five regions which consistently elicited responses in HLA-DR1[+] individuals using peptides from the H3N2 strain A/Beijing/32/92 (C. M. Gelder et al. 1995). The three healthy donors I used for this study were vaccinated with the 2011/2012 trivalent influenza vaccine which contained the influenza A strains A/Perth/16/2009 (H3N2) and A/California/7/2009 (H1N1) as well as the influenza B strain B/Brisbane/60/2008. The first step was to find the equivalent regions within the A/Perth/16/2009 strain (shown in **Table** 6**.**1) using the sequence provided in the NCBI protein database (accession number AHX37629.1). Peptides were arranged in a matrix and sorted into 10 pools as shown **Figure** 6**.**2B.

	Peptide sequence	**Position**
Region 1	**CYPYDVPDYASLRSLVASSGTLEFNNESFNWT**	*113-144*
Pep 1	CYPYDVPDYASLRSLV	113-128
Pep 2	DVPDYASLRSLVASSG	117-132
Pep 3	YASLRSLVASSGTLEF	121-136
Pep 4	RSLVASSGTLEFNNES	125-140
Pep 5	ASSGTLEFNNESFNWT	129-144
Region 2	**LIGKTNEKFHQIEKEFSEVEGRIQDLEKYVED**	*400-431*
Pep 6	LIGKTNEKFHQIEKEF	400-415
Pep 7	TNEKFHQIEKEFSEVE	404-419
Pep 8	FHQIEKEFSEVEGRIQ	408-423
Pep 9	EKEFSEVEGRIQDLEK	412-427
Pep 10	SEVEGRIQDLEKYVED	416-431
Region 3	**HHPGTDKDQIFLYAQASGRITVSTKRSQQTVS**	*199-230*
Pep 11	HHPGTDKDQIFLYAQA	199-214
Pep 12	TDKDQIFLYAQASGRI	203-218
Pep 13	QIFLYAQASGRITVST	207-222
Pep 14	YAQASGRITVSTKRSQ	211-226
Pep 15	SGRITVSTKRSQQTVS	215-230
Region 4	**QDLEKYVEDTKIDLWSYNAELLVALENQHTID**	*423-454*
Pep 16	QDLEKYVEDTKIDLWS	423-438
Pep 17	KYVEDTKIDLWSYNAE	427-442
Pep 18	DTKIDLWSYNAELLVA	431-446
Pep 19	DLWSYNAELLVALENQ	435-450
Pep 20	YNAELLVALENQHTID	439-454
Region 5	**QNVNRITYGACPRYVKQNTLATGMRNVPEKQT**	*311-342*
Pep 21	QNVNRITYGACPRYVK	311-326
Pep 22	RITYGACPRYVKQNTL	315-330
Pep 23	GACPRYVKQNTLATGM	319-334
Pep 24	RYVKQNTLATGMRNVP	323-338
Pep 25	QNTLATGMRNVPEKQT	327-342

Table 6.1: Sequences of overlapping peptide libraries used in this study. The full sequence of each region is shown as well as each individual peptide. Peptides are 16aa long with a 4aa overlap. Positions are taken from NCBI database (accession number: AHX37629.1).

Short-term T cell lines were generated against each peptide pool and tested in an IFN-γ ELISpot on day 12-16 as described in Chapter 2.3.8. As a positive control, short term cell lines were set up against the $HA_{306-318}$ epitope and tested in an ELISpot alongside. All lines were set up in triplicate which were then combined prior to the ELISpot assay and split across two wells. One well was stimulated over night with the same peptide pool as the line was set up with initially, while the other well was left unstimulated in order to measure background. Peptide pools which generated at least 20 responses i.e. 20 spot forming cells (SFC) per 100,000 PBMCs (after subtraction of background) and at least a 50% increase in responses compared to background were considered as a positive response. **Figure** 6**.**2A shows an example of one of these assays. Here, peptide pools 1, 3, 4, 6 and 9 generated positive responses as did the positive control peptide $HA_{306-318}$. Since each peptide is present in two pools, potential hits can be identified as both pools containing the peptide will generate a positive response as demonstrated in **Figure** 6**.**2B. In this case, peptides 12, 13, 14, 16, 18 and 19 were identified as potential hits.

A
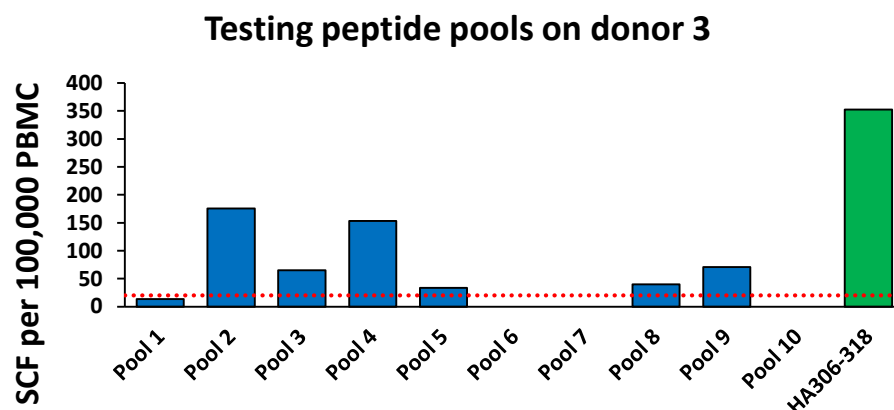
Testing peptide pools on donor 3

B

Peptide pool matrix

	Pool 1	Pool 2	Pool 3	Pool 4	Pool 5
Pool 6	1	2	3	4	5
Pool 7	6	7	8	9	10
Pool 8	11	12	13	14	15
Pool 9	16	17	18	19	20
Pool 10	21	22	23	24	25

Figure 6.2: Measuring CD4+ T cell responses against overlapping peptide pools. (A) ELISpot results of one assay (Donor 3 - Round 2) are shown. Responses against individual peptide pools shown in blue. Responses against the control peptide (PKY) are shown in green. The red dotted line indicates the minimum threshold of 20 SFC per 100,000 PBMCs. (B) Peptides 1-25 were sorted into ten peptide pools with each peptide present in two pools. Positive hits from the assay shown in (A) are highlighted in light blue. Potential new epitopes would cause both pools to elicit a response and are therefore located at the intersection of two pools (highlighted in blue).

This was repeated twice for each donor and results are shown in **Table** 6.2A. Since the aim was to identify epitopes restricted to HLA-DR1, peptides pools eliciting responses across all donors i.e. at least once in each donor where of most interest. **Table** 6.2B shows the peptide pool matrix with potential new epitopes highlighted in blue. Peptides 12, 13 and 14 all are located within region 3 while peptides 17, 18 and 19 are located within region 4.

A

Round	Donor 1 R1	Donor 1 R2	Donor 2 R1	Donor 2 R2	Donor 3 R1	Donor 3 R2	% positive responses
Pool 1	●				●		50.00
Pool 2	●			●			50.00
Pool 3		●		●		●	50.00
Pool 4	●	●	●	●	●	●	100.00
Pool 5	●	●		●			50.00
Pool 6	●	●				●	50.00
Pool 7							0.00
Pool 8		●	●	●			50.00
Pool 9	●	●	●	●		●	83.33
Pool 10	●		●				33.33
HA$_{306-318}$	●	●	●		●	●	83.33

B

Peptide pool matrix

	Pool 1	Pool 2	Pool 3	Pool 4	Pool 5
Pool 6	1	2	3	4	5
Pool 7	6	7	8	9	10
Pool 8	11	12	13	14	15
Pool 9	16	17	18	19	20
Pool 10	21	22	23	24	25

Table 6.2: Summary of two rounds of testing overlapping peptide pools on all three donors. (A) Positive responses are highlighted in light blue. Percentages of positive responses for each pool are shown and highlighted in different shades of green. Absolute responses are shown in Supplementary Table S **5**. (B) Pools which elicited positive responses in at least 50% of cases and at least once in each donor are highlighted in light blue. Potential epitopes are highlighted in blue.

As evident from **Table** 6.**2**, there was great variability between individual assays. The two pools that elicited the most consistent responses across all three donors were pools 4 (100% positive responses) and 9 (83.33% positive responses), pinpointing peptide 19 as potential hit. HA$_{306-318}$ also generated positive responses in 83.33% of cases. However, peptide pool 3 also generated positive responses at least once in each donor. Therefore, peptide 18 was also considered as a potential hit. Each peptide was then tested

individually to remove any bias from HLA-II binding competition between different
peptide.
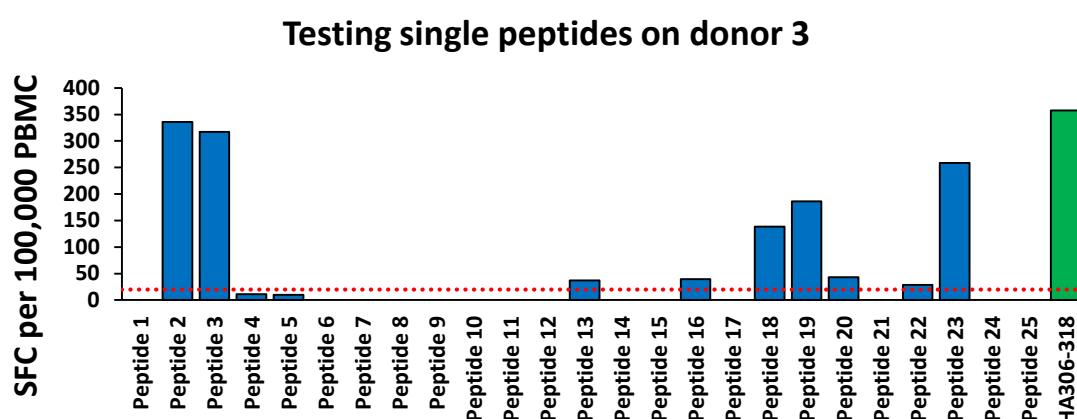
Figure 6.3: Measuring CD4$^+$ T cell responses against single peptides. ELISpot results
of one assay (Donor 3 - Round 3) are shown. Responses against individual peptides shown
in blue. Responses against the control peptide (PKY) are shown in green. The red dotted
line indicates the minimum threshold of 20 SFC per 100,000 PBMCs.

Figure 6.3 shows the results of an IFN-γ ELISpot using PBMCs from donor 3 testing each
individual peptide. As predicted from data gained testing peptide pools using PBMCs
from the same donor (see **Figure** 6**.**2) peptides 3, 16, 18 and 19 elicited positive responses.
However, 20 and 23 also elicited responses as predicted by combining data from all three
donors (see **Table** 6**.**2). Peptides 13 and 20 also elicited responses which could not have
been predicted from testing peptide pools on this donor. When looking at data from testing
peptides pools on donor 1 and 2 however, peptide 13 did appear as potential hit in round
2 for each donor (in both cases both pool 3 and 8 elicited positive responses). Other hits
such as peptide 4 (predicted from testing pools on donor 3, see **Figure** 6**.**2) and peptides

21 and 24 (predicted from combining data from all three donors, see **Table** 6**.**2) did not

elicit responses when testing single peptides in donor 3 on this occasion.

		Donor 1			Donor 2			Donor 3			% positive responses
	Round	1	2	3	1	2	3	1	2	3	
Region 1	Peptide 1	■	■		■						33.33
	Peptide 2			■	■	■				■	44.44
	Peptide 3	■	■	■	■	■	■		■	■	88.89
	Peptide 4		■			■					22.22
	Peptide 5										0.00
Region 2	Peptide 6								■		11.11
	Peptide 7										0.00
	Peptide 8							■	■		22.22
	Peptide 9										0.00
	Peptide 10										0.00
Region 3	Peptide 11										0.00
	Peptide 12	■	■			■	■				44.44
	Peptide 13	■	■	■				■	■	■	66.67
	Peptide 14					■					11.11
	Peptide 15					■					11.11
Region 4	Peptide 16		■				■	■	■		44.44
	Peptide 17						■	■			22.22
	Peptide 18	■	■		■	■	■	■	■	■	88.89
	Peptide 19	■	■		■	■		■	■		66.67
	Peptide 20		■	■			■		■		44.44
Region 5	Peptide 21		■						■		22.22
	Peptide 22		■	■		■			■		55.56
	Peptide 23		■	■	■	■			■	■	66.67
	Peptide 24		■								11.11
	Peptide 25										0.00
	HA$_{306\text{-}318}$		■	■	■	■	■	■	■	■	88.89

Table 6.3: Summary of three rounds of testing single peptides on all three donors. Positive responses are highlighted in light blue. Percentages of positive responses for each pool are shown and highlighted in different shades of green. Absolute responses are shown in Supplementary Table S **6**.

Table 6.3 shows results of IFN-γ ELISpots screening all 25 individual peptides on all three donors. In order to be considered a potential HLA-DR1 restricted hit, peptides were identified which generated positive responses in at least 5 out of the 9 assays in total (corresponding of at least 50% positive responses overall) and in 2 out of 3 rounds for each donor.

Peptide 3 consistently generated positive responses across all three donors (88.89% positive responses). Overall, region 2 only elicited sporadic responses in donor 3 with peptide 8 being a potential hit. However, no responses were observed for the other two donors. Peptide 13 was a potential hit for donors 1 and 3 but only elicited one positive response in donor 2 and was therefore dismissed. Peptide 18 generated responses in 88.89% of cases across all three donors. Interestingly, peptide 19, which overlaps peptide 18 by 12 aa, elicited positive responses in donor 2 and 3 (100% postive responses) but not donor 1.

While peptide 22 generated consistent responses in donor 1 and 2 but not donor 3, its overlapping neighbour, peptide 23, generated consistent responses across all three donors (66.67% positive responses). At first, this was not surprising since peptide 23 contains the full sequence of the $HA_{306-318}$ homologue. However, at this point it was realised that one amino acid within region 5 had been left out when ordering the peptide library. Peptides 20-25 should span the following sequence: QNVNRITYGAC<u>PRYVKQNT**K**LAT</u>GMRNVPEKQT. Yet, Lysine$_{331}$ (highlighted in red) was missing in region 5 of the overlapping peptide library and therefore in peptides 22 to 25.

Based on these results peptides 3 and 18 were taken forward further for more detailed analysis. Although peptide 19 did not elicit any responses in donor 1, it was taken further

as well due to its overlap with peptide 18. Region 5 was eliminated from further experiments due to the missing Lysine in position 331. This decision resulted in peptide 23 being left out from further analysis.

6.4.2. HLA-DR restriction of novel epitopes as registered on the IEDB

The Immune Epitope Database (IEDB) was searched in order to verify whether previous studies had identified any regions that included peptides 3, 18 and 19 as CD4 T-cell targets.

Table 6**.**4 summarizes these findings. In order not to simplify the search, the "search for substrings" option were used which only shows results for peptides that span the query sequences in part or whole. Any sequences homologues and overlapping sequences that do not include the whole of the query sequence were ignored.

Searching for the peptide sequence of peptide 3 (YASLRSLVASSGTLEF) revealed that this peptide had been used in 5 other studies in addition to the paper published by Gelder and colleagues which formed the basis for this study. One of these was a follow-up study looking at HLA-DR7 restriction of the some of the peptides used in their original study (C. Gelder et al. 1998). Here, they showed that peptide 3 binds to HLA-DR7 although it did not elicit responses in an INF-γ ELISpot. Two successive papers describe a CD4[+] T cell clone recognising a longer version of this this epitope, however, HLA-restriction was not investigated (Lamb et al. 1982; Lamb and Green 1983). The most recent study used a shortened version of peptide 3 in a vaccine study in H2[d] BALB/c mice (Valkenburg et al. 2014).

When searching for the peptide sequence spanned by the overlapping peptides 18 and 19 (DTKIDLWSYNAELLVALENQ) on the IEDB, 8 additional studies were identified.

Again, this peptide was used in the follow-up study by Gelder and colleagues (C. Gelder et al. 1998). Here, they confirmed recognition of this peptide by HLA-DR15[+] and HLA-DR7[+] donors, respectively. It was also shown to bind to HLA-DR13 in addition to HLA-DR7. Three additional studies looked at shorter versions of this epitopes recognised by murine T cells (Staneková et al. 2013; Ishizuka et al. 2009; Jackson, Drummer, and Brown 1994). A forth study looking at T cell responses in mice found no responses against peptide 3 (Crowe et al. 2006). In an unpublished study, Harndahl and colleagues identified a 9mer peptide within peptide 18 (IEDB Reference ID: 1019353) as potentially binding to the MHC I allele HLA-B39 using a previously published peptide binding assay (Sylvester-Hvid et al. 2002). One study by Babon and colleagues identified an 17mer sequence located within peptide 18/19 as novel epitope using an approach similar to the one used here which was based on overlapping peptides and detection of CD4[+] T cell responses based on INF-γ ELISpots (Babon et al. 2009). In a further unpublished study, Yang and colleagues identified a 14mer epitope (IEDB Reference ID: 113533) as being HLA-DR15 restricted using a tetramer guided epitope mapping strategy previously published by the same group (Novak et al. 2001). Interestingly they also identified a H1 derived sequence homologue as being both HLA-DR1 and HLA-DR15 restricted (Yang et al. 2013).

In summary, peptide 3 has been shown to bind to HLA-DR7. Peptide 18/19 was identified as being HLA-DR7 and HLA-DR15 restricted while a H1 sequence homologue was identified as being HLA-DR1 and HLA-DR15 restricted. All three donors used for this present study are HLA-DR1[+]. While Donor 2 is also HLA-DR15[+] which could explain the robust recognition of peptides 18 and 19, neither of the other two donors were positive for HLA-DR7 or HLA-DR15. However, peptides can be presented by more than one HLA alleles depending on their anchor residues. The best studied example is HA$_{306-318}$ which

is known to be recognised in the context of HLA-DR1, HLA-DR4, HLA-DR5 and HLA-DR15 (Woody et al. 1982; Z. L. Zheng et al. 1991; P. A. Roche et al. 1990) It is therefore entirely possible that peptide 18 is also HLA-DR1 restricted. In order to confirm this I tested all three peptides for their HLA-DR1 restriction using antibody blocking assays and HLA-DR1 homozygous antigen presenting cells.

Peptide 3: YASLRSLVASSGTLEF			
IEDB ID	**Sequence**	**MHC restriction**	**Reference**
10850	DYASLRSLVASSGTLEFINE GFNWTGVTQNGGSSAC	MHC-II	Lamb et al. 1982
		MHC-II	Lamb et al. 1983
73443	YASLRSLVASSGTLEF	MHC-II	Gelder et al. 1995
		HLA-DR7	Gelder et al. 1998
7431	CYPYDVPDYASLRSLVASS GTLEFINEDFNWT	MHC-II	Gelder et al. 1995
73442	YASLRSLVASSGTLE	negative	Crowe et al. 2006
226494	LRSLVASSG	H2-d (murine class II)	Valkenburg et al. 2014

Peptide 18/19: DTKIDLWSYNAELLVALENQ			
IEDB ID	**Sequence**	**MHC restriction**	**Reference**
1055	AELLVALEN	H2-IA[d] (murine class II)	Staneková et al. 2013; Jackson et al. 1994
31200	KIDLWSYNAELLVALE	Negative	Gelder et al. 1995
		HLA-DR7 HLA-DR15	Gelder et al. 1998
50489	QDLEKYVEDTKIDLWSYNA ELLVALENQHTIDLTDS	MHC-II	Gelder et al. 1995
62653	SYNAELLVAL	H2[d] (murine class I)	Saikh et al. 1995
		H2-K[d] (murine class II)	Staneková et al. 2013
1035198	DTKIDLWSYNAELLV	Negative	Crowe et al. 2006
1035199	LWSYNAELLVALENQ	Negative	Crowe et al. 2006
113533	IDLWSYNAELLVAL	HLA-DR15	Young et al. (unpublished)
124892	YNAELLVAL	HLA-B39	Harndahl et al. (unpublished)
129078	KIDLWSYNAELLVALEN	n/a	Babon et al. 2009
179692	WSYNAELLVA	HLA-A2 HLA-A68	Ishizuka et al. 2009
	Sequence homologue		
188707	LDIWTTYNAELLVLLLENERTL	HLA-DR15	Yang et al. 2013

Table 6.4: HLA restrictions of peptides 3 and peptide 18/19 identified by previous studies. Sequences of peptides 3 and 18, respectively were used as search queries on the IEDB. Only sequences spanning at least part of the peptides were included with exception of regions 2 and 4 identified in the original study by Gelder and colleagues (C. M. Gelder et al. 1995). Differences in homologous sequences highlighted in red.

6.4.3. *Confirming HLA-DR1 restriction of new epitopes*

In order to confirm that peptides 3, 18 and 19 were recognised in the context of HLA-DR1, short term T cell lines were tested against two different lines of antigen presenting cells. HOM-2 is a B-LCL from a donor homozygous for HLA-DR1 and does not express any other HLA-DR alleles. Only peptides binding to HLA-DR1 will be presented by HOM-2 cells and therefore elicit higher responses than unpulsed HOM-2. Short term T cell lines were added, incubated overnight and the INF-γ ELISpot developed the following day. In the same experiment the pan-HLA-DR antibody L243 was used to block presentation through any HLA-DR allele. L243 was added to T cell lines prior to adding peptide.

6.4.3.1. *Optimising ELISpot protocol in order to reduce alloresponses*

Previous experiments demonstrated that incubating ELISPot plates overnight resulted in high responses against unpulsed HOM/2 cells (data not shown). Thus short incubation times were used prior to developing the IFN-γ ELISpot. **Figure** 6.4 shows the results of testing different time points on two different donors. In both cases a short term T cell line against $HA_{306-318}$ was used as this epitope is known to be presented by HLA-DR1.
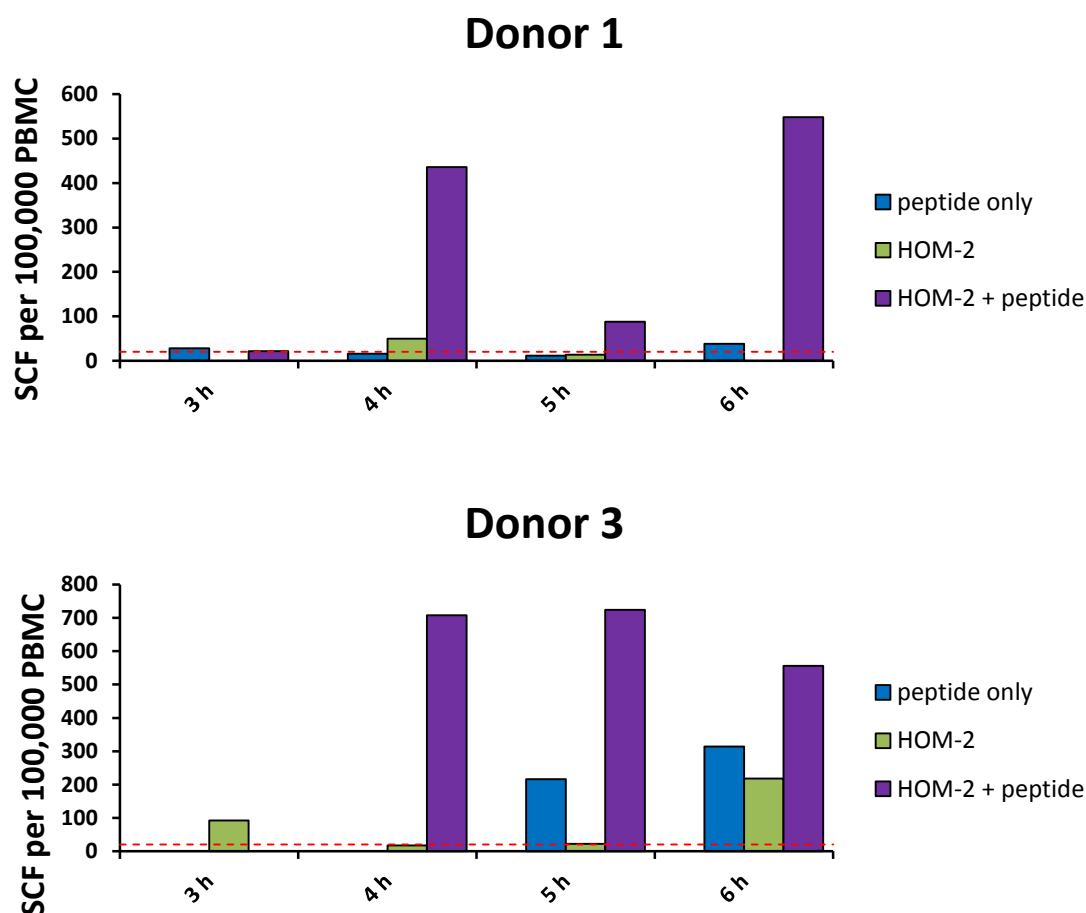
Figure 6.4: Assessing responses against HOM-2 cells at different time points. T cell lines from two different donors recognising the $HA_{306-318}$ epitope were used. Responses against peptide only were measured (shown in dark blue) as well as responses against HOM-2 cells following incubation with peptide (shown in purple and brown, respectively). Control responses against unpulsed HOM-2 were measures as well (shown in green and yellow, respectively).

For Donor 1, responses against unpulsed HOM-2 cells stayed low across all four time points. From 4 h onwards, responses against peptide pulsed HOM-2 cells were substantially higher than against unpulsed HOM-2 cells. Positive responses against peptide only were only detected after 6 h. For donor 3, some responses against unpulsed HOM-2 were observed at 3 h followed by with a decrease at 4 h and 5 h. At 6 h, responses against unpulsed HOM-2 cells were at their highest. At the 4 h and 5 h time points,

responses against peptide pulsed HOM-2 were significantly higher than against unpulsed HOM-2 with a less pronounced difference at 6 h. Similarly to donor 1,positive responses against peptide only were observed from 5 h onwards and stayed lower than responses against peptide pulsed HOM-2.

In summary, an incubation time of 6 h seemed ideal for donor 1 as alloresponses against HOM-2 were still low while responses against peptide only were detectable. For donor 3, 5 h were a better incubation time as alloresponses against HOM-2 seemed to surge at 6 h. In both donors, responses against peptide pulsed HOM-2 cells were significantly higher than against peptide only and were detected earlier on. This confirms that HOM-2 are well suited as antigen presenting cells and elicit stronger responses than antigen presenting cells still present in the short term T cell lines after 12 days or more in culture.

6.4.3.2. *Confirming HLA-DR1 restriction using optimised ELISpot protocol*

Based on these results, short term T cell lines grown against peptides 3, 18 and 19 were tested for their HLA-DR1 restriction using an incubation time of 5 h for all three donors. Results are shown in Figure 6.5.

Donor 1

Donor 2

Donor 3
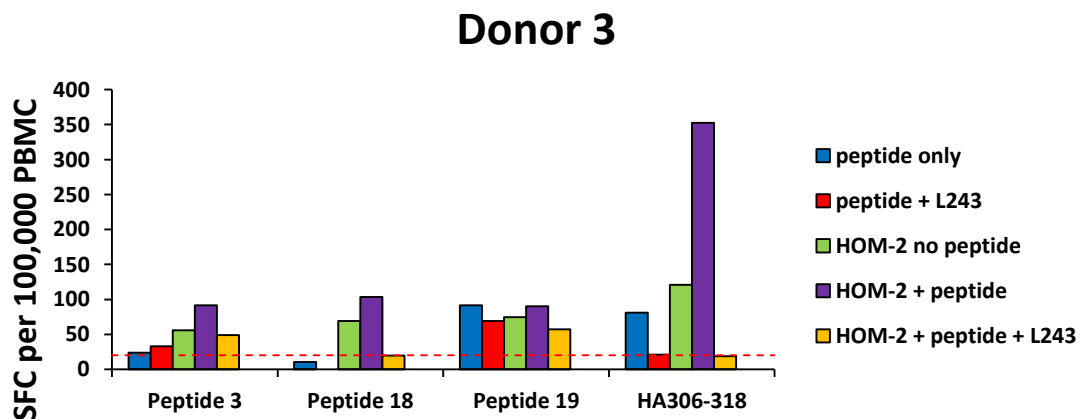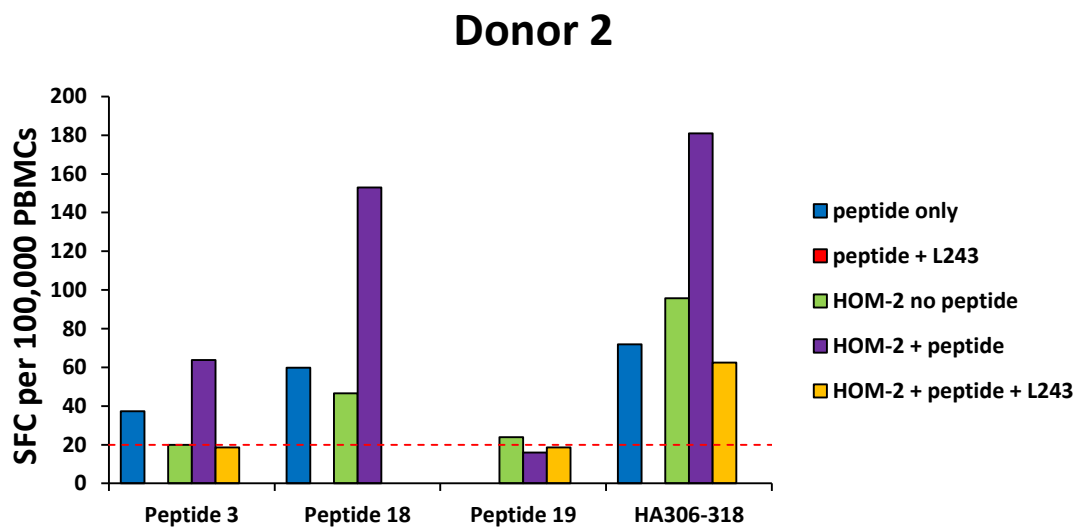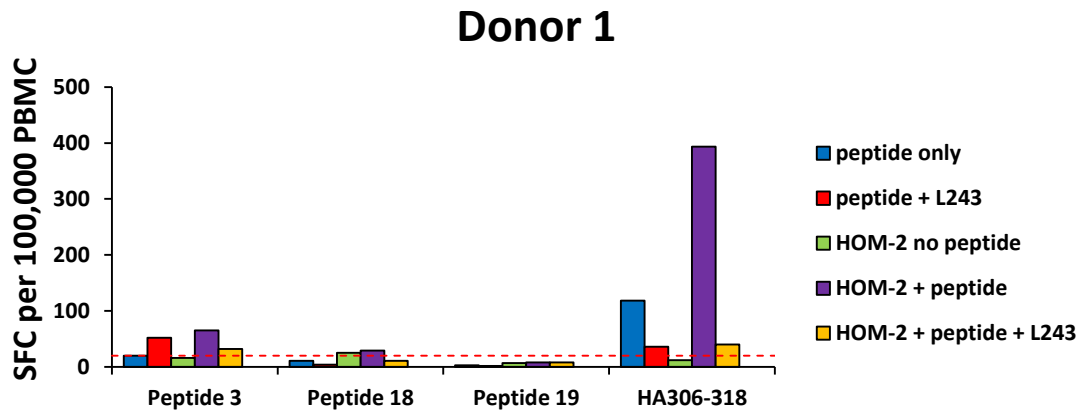
Figure 6.5: Testing HLA-DR1 restriction of novel epitopes using a shorter incubation time of 5h. Positive responses to individual peptides were measured as before (shown in dark blue). The anti-pan-MHC-II antibody L243 was added to short term T cell lines or HOM-2 cells prior to adding peptide (shown in red and yellow, respectively). CD4$^+$ T cells lines were added to peptide pulsed HOM-2 cells (shown in purple). HOM-

2 cells that had not been incubated with peptide were added as negative control (shown in green).

Looking at responses against across all donors, $HA_{306-318}$ showed an HLA-DR1 restricted pattern in all three donors. Responses against peptide only as well as peptide pulsed HOM-2 cells decreased upon addition of L243 and responses against peptide pulsed HOM-2 were markedly higher than against unpulsed HOM-2 cells. When peptide only was used, peptide 3 induced responses above threshold in donor 2 but not donor 1 and 3. However, in all three donors, responses against peptide pulsed HOM-2 were higher than against unpulsed and dropped upon addition of L243, indicating HLA-DR1 restriction. Peptide 18 did not induce any responses above threshold in donor 1. In donor 2 and 3, however, it showed a similar profile as peptide 3, indicating HLA-DR1 restriction. Peptide 19 did not show any HLA-DR1 restricted pattern in any of the three donors.

In summary, HLA-DR1 restriction of peptides 3 and 18 was confirmed. Peptide 19 on the other hand did not appear to be HLA-DR1 restricted. In an attempt to confirm these results, all three peptides were tested alongside $HA_{306-318}$ in a competitive HLA-DR1 binding assay as described in Chapter 2. However, none of the three novel epitopes were able to replace the biotinylated CLIP peptide (data not shown). However, this was also observed for the HLA-DR1 restricted HIV gag24 peptide and its variants in a previous assay (as described in Chapter 5) and therefore did not rule out HLA-DR1 restriction.

6.4.4. Identifying the 9mer binding core.

We routinely generate soluble MHC for bioinformatic analysis, multimer staining and crystallographic studies. Solving the crystal structure p-MHC-II can confirm HLA restriction and identify the 9mer core epitope. Soluble HLA-DR1^{peptide3} and HLA-

DR1[peptide 18] were refolded and purified as described in Chapter 2.2.4. Both complexes refolded in a manner seen for other HLA-DR1 restricted peptides. **Figure** 6.6A shows a SDS PAGE following immunoprecipitation using an L243 coated protein A column which specifically binds correctly refolded HLA-DR monomers. **Figure** 6.6B shows A SDS PAGE following further purification through size exclusion using a gel filtration column. Crystal trays were set up for each complex using two different crystallisation screens routinely used for crystallisation of pMHC molecules (Bulek et al. 2012; Newman et al. 2005). However, neither screen resulted in crystals. Using crystals from a different HLA-DR1 complex as crystallisation seeds did not produce crystals either. This does not contradict any of the results described above as not every pMHC complex forms crystals within the first few trials.
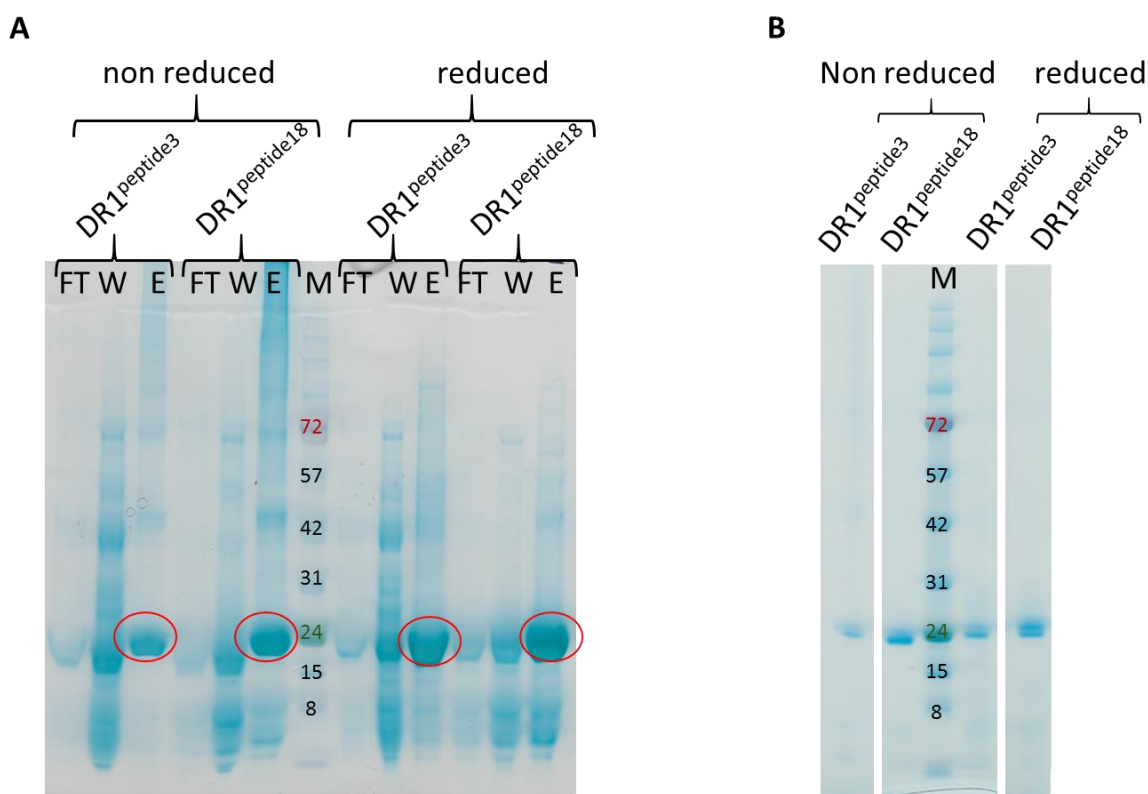
Figure 6.6: Refolding peptide 3 and peptide 18 bound to HLA-DR1. Molecular weights are given in kDa. A: SDS PAGE following purification using L243 antibody column. Red circles indicate HLA-DR1 bands. FT: flow through; W: wash; E: Eluate; M: Marker. B: SDS PAGE following purification of gel filtration column.

6.4.5. *Assessing conservancy of peptides 3 and 18*

As one of the external proteins, HA is a prime target of both neutralising antibodies and T cell mediated immunity and therefore under pressure to mutate in an attempt to evade the immune response. As shown in Figure 6.7, peptides 3 is located within the globular head domain while peptides 18 and 19 and $HA_{306-318}$ are located in the stem of the HA protein. In comparison to the globular head domain, the stem shows a higher degree of conservancy, probably due to its crucial role in fusing the viral and endosomal membranes (Lee et al. 2013). Mutations within functionally important regions of the protein are likely to compromise the virus' ability to infect host cells and are therefore eliminated through natural selection. Furthermore, T cell epitopes located within conserved regions play an important role in inducing protection against infection with heterologous Influenza A strains (Alam and Sant 2011). Thus, investigations were performed to assess the degree of conservancy of peptides 3, 18 and as well as $HA_{306-318}$.
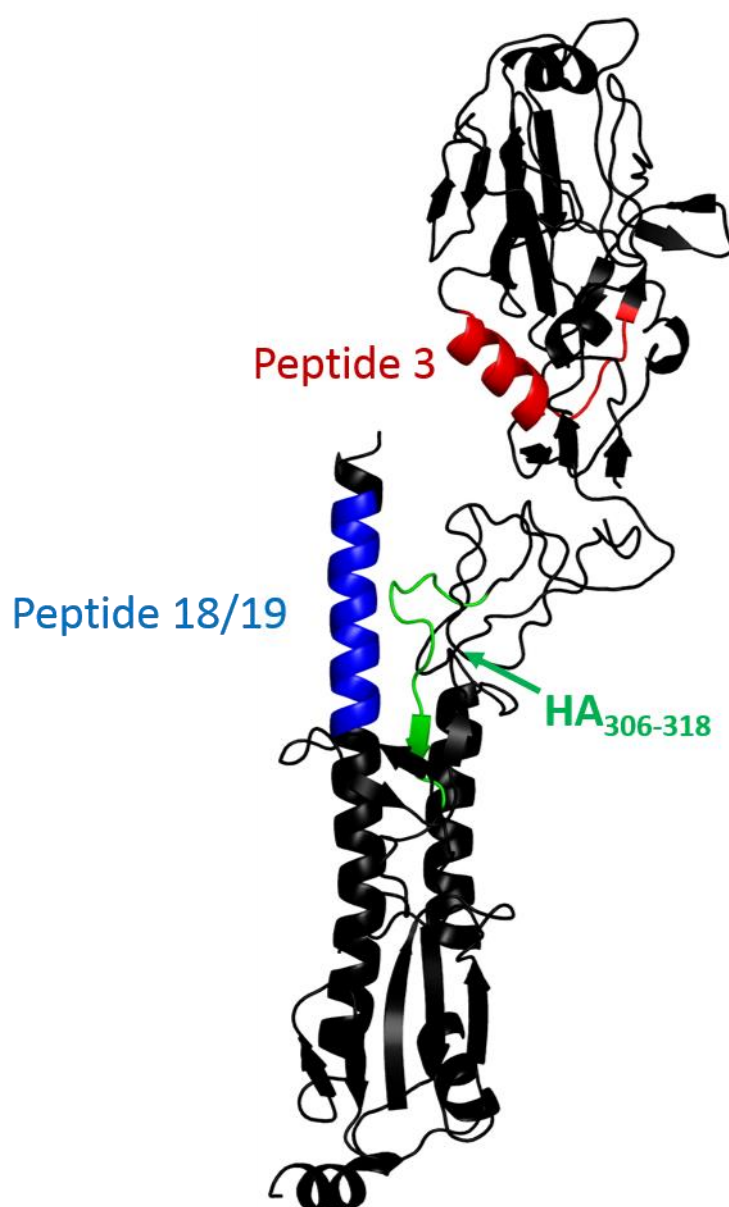
Figure 6.7: Highlighting positions of peptides within influenza HA. Peptide 3 (shown in read) is located in the globular head domain. Peptide 18 (shown in blue) is located in the stem domain while $HA_{306-318}$ (shown in green) forms part of the fusion peptide.

Working datasets were generated of all complete or near-complete HA sequences from H3N2 influenza strains available on the Influenza Research Database (fludb.org). At the time of writing, there were 4999 sequences available. Using the inbuilt multiple sequence alignment tool, all sequences were aligned generating the resulting output file for any following analysis. Using BioEdit a consensus sequence was generated. **Table** 6**.**5 shows

the sequences of peptides 3, 18 and HA$_{306-318}$ aligned with the consensus sequence. While peptides 3 and 18 match the consensus sequence to 100%, HA$_{306-318}$ deviates by one amino acid: the Lysine in position 307 of the consensus sequence is replaced by an Arginine.

Using the epitope conservancy tool available on the IEDB website, the frequency of sequences encoding exactly the same sequences as peptides 3, 18 and HA$_{306-318}$. Out of the 4999 sequences, 92.38% and 96.20% matched peptides 3 and 18 to 100%, respectively. HA$_{306-318}$ however, only matched 2.5% of all sequences rising to 57.83% for the consensus sequence. Conservancies of each individual aa ranged from 98.06% to 100% for peptide 3 and from 99.06% to 99.98% for peptide 18. In the case of HA$_{306-318}$, conservancies ranged from 2.56% to 99.96% with the lowest conservancy being observed for the Lysine at position 307. When looking at the consensus sequences, conservancies ranged from 61.55% to 99.96% with the value for Threonine$_{313}$ rising to 99.84%. All values are shown in **Supplementary Table S** 7, **Supplementary Table S** 8 and **Supplementary Table S** 9.

Peptide 3:

Peptide sequence	Y	A	S	L	R	S	L	V	A	S	S	G	T	L	E	F
Consensus:	Y	A	S	L	R	S	L	V	A	S	S	G	T	L	E	F

peptide 18:

Peptide sequence	D	T	K	I	D	L	W	S	Y	N	A	E	L	L	V	A
consensus	D	T	K	I	D	L	W	S	Y	N	A	E	L	L	V	A

HA$_{306-318}$:

Peptide sequence	P	K	Y	V	K	Q	N	T	L	K	L	A	T
consensus	P	R	Y	V	K	Q	N	T	L	K	L	A	T

Table 6.5: Comparing peptide sequences with the consensus sequence. Differences are highlighted in red.

Shannon entropies were also calculated for all three peptides using BioEdit. Shannon entropies reflect the degree of conservancy of each position of the peptide. High entropy values indicate low degrees of conservancy while low entropies indicate high conservancy at a given position. Figure 6.8 shows Shannon entropies for all three peptides. Absolute values are shown in **Supplementary Table S** 7, **Supplementary Table S** 8 and **Supplementary Table S** 9. Again, peptides 3 and 18 showed higher conservancies overall (i.e. lower Shannon entropies) than HA306-318. Although two positions within peptide 3 are 100% conserved and therefore have a Shannon entropy value of 0, peptide 18 is more conserved overall i.e. maximal Shannon entropies are lower than for peptide 3.
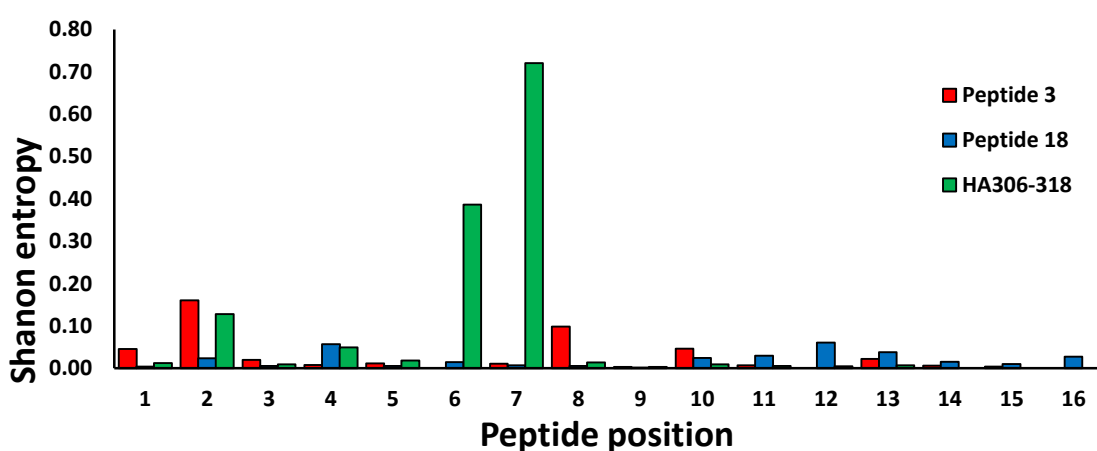
Figure 6.8: Shannon entropies of peptides 3, 18 and HA306-318. Shannon entropies are given for each position within the individual peptide. Peptide 3 is shown in red, peptide 18 is shown in blue and HA$_{306-318}$ is shown in green.

In order to generate a more visual representation of the conservancy of each positions, sequences logos were generated for all three peptides using the WebLogo online tool from

the University of Berkeley, shown in **Figure** 6.9 (Crooks et al. 2004). WebLogos were generated using the aligned sequence data from multiple sequences, in this case the 4999 H3 sequences identified previously. The height of each letter is proportionate to its frequency across all sequences. WebLogos for peptide 3, 18 and $HA_{306-318}$ are shown in **Figure** 6.9. Again, peptides 3 and 18 were highly conserved across all positions. $HA_{306-318}$ on the other hand showed some degree of variation positions 311 and 312. Despite Glutamic Acid and Asparagine still being the most prevalent at their respective positions, their main alternatives, Histidine and Serine are clearly visible. The empty space above the most prevalent amino acids (Glutamic Acid and Asparagine, respectively) indicates that there is even more variation although none of these additional alternative aa are prevalent enough to generate their own letter in the WebLogo.

In summary, peptides 3 and 18 are highly conserved while $HA_{306-318}$ shows some degree of variation in general and around the centre of the peptide, particularly.

Figure 6.9: Sequence logos of peptides 3, 18 and HA306-318. Conservancies are calculated using all HA3 sequences from multiple alignment. The height of the letter corresponds to its conservancy across all sequences. Colours indicate physical properties of amino acids (acidic, polar in green; basic, polar in pink; positively charged in blue; negatively charged in red; hydrophobic in black).

6.5. Discussion

The aim for this chapter was to identify HLA-DR1 restricted epitopes within the five regions of HA previously identified by Gelder and colleagues. Overlapping peptides libraries were used in order to identify any positive hits. Each peptide was present in 2 different pools, allowing its identification when both pools elicited positive responses. Using this approach peptides 18and 19 were identified as potential hits (see **Table** 6**.**2). Single peptides were then used to set up short-term T cell lines. Interestingly, this approach identified peptides 3, 18 and 23 as positive hits in all three donors, as well as peptide 19 in donor 2 and 3 (see **Table** 6**.**3). Peptides 3 and 23 were not identified using peptide pools. Peptides 18 and 19, were identified using both approaches. This could be due to certain peptides within a pool binding to MHC-II more strongly than others, therefore "blocking" access for the lower binders. Another explanation are antagonistic peptides which downregulate T cell responses thereby inhibiting responses against other peptides within the same pool. The inter-assay variabilities observed in both approaches are a common feature when screening overlapping peptides and has been observed in other experiments. Repeating the screens in several donors allows for the detection of overall trends and therefore the successful identification of candidate peptides. This identified peptides 3, 18 and 19 as potentially HLA-DR1 restricted epitopes.

Protein regions in HA that include peptides 3 and 18, defined here, have been previously identified in a number of different studies, confirming their relevance during CD4$^+$ T cell mediated anti-influenza responses. However, none of these previous studies have identified either of the peptides as being HLA-DR1 restricted (see **Table** 6**.**4). During my first attempt to confirm their HLA-DR1 restriction, the main obstacle were high responses against the HLA-DR1 homozygous HOM-2 BLCLs

masking any increase in responses against peptide pulsed versus unpulsed HOM-2 (see **Figure** 6.5). Using a shorter incubation time resolved this problem. Using this approach, the HLA-DR1 restriction of peptides 3 and 18 was confirmed. Peptide 18 has also been found to bind HLA-DR7 and HLA-DR15 in other studies, making it a potentially universal epitope (see **Table** 6.4). It shares this feature with $HA_{306-318}$ which also has been describes to bind multiple HLA-DR alleles. Due to differences in the peptide binding pockets, HLA preferentially bind and present different peptides. However, many peptides have been shown to encode anchor residues suitable for more than one HLA allele (Z. L. Zheng et al. 1991).

T cell responses against conserved epitopes play an important role in cross-protection against other strains of the same pathogen, an important feature for successful vaccines protective against multiple strains (Alam and Sant 2011). Considering the overall variability of H3N2 strains, their high conservancy makes peptides 3 and 18 attractive candidates for vaccine development. Conserved peptides are also highly important for diagnostic tools such as pMHC-II multimer staining. Being able to identify and track T cell populations raised against different strains allows to dissect the bigger picture of anti-Influenza T cell responses. Peptides 3 and 18 are highly conserved across all positions while $HA_{306-318}$ shows some variation in sequence, particularly at positions 311 and 312. Looking at the structure of the HA1.7 TCR bound to HLA-DR1$^{HA306-318}$, both positions of the peptide are contacted by the TCR (J. Hennecke et al. 2000). Any mutations at these positions could potentially impact TCR binding and as a result T cell recognition leading to the virus escaping the immune response as long as the function of the HA protein is not compromised. The lower conservation at positions 311 and 312 indicate a certain degree of pressure on the virus to mutate these positions in order to escape. Judging from

the high conservation across peptides 3 and 18, this evolutionary pressure was not observed. This could be due to both sequences being crucial for the functioning of the protein. Any mutations could compromise the virus and are therefore eliminated through natural selection. Mutation studies investigating the impact of such substitutions are possible but require resources we do not have access to in our group. Another explanation would be that peptides 3 and 18 are not naturally processed and presented to CD4$^+$ T cells. Testing short term T cell lines grown against each peptide and tested for recognition of whole protein processed by antigen presenting cells would provide an answer to this. Unfortunately, these experiments were no possible within the scope of my PhD.

In summary, two potential candidate epitopes within the influenza hemagglutinin protein were identified, confirming findings by other groups. In addition, their HLA-DR1 restriction was confirmed. Being highly conserved across all H3N2, they potentially play an important role in cross-strain protection and therefore have important applications in vaccine development and diagnostics. Once confirmed to be processed and presented naturally at the cell surface, these two epitopes will be useful tools in dissecting the role of CD4$^+$ T cells in influenza A infections.

7. Discussion

CD4$^+$ T cells are at the heart of the immune system. They recognise peptides bound to MHC-II molecules on the cell surface of antigen presenting cells. When activated CD4$^+$ T cells orchestrate other players of the immune system. Despite growing knowledge on TCR/pMHC-II interactions, some caveats remain, particularly concerning peptide-flanking residues. The work presented in this thesis contributes to closing this gap. In chapter 3, an insect cell based expression system for HLA-DR1 was established and optimised. Baculovirus expression vectors for both HLA-DR1α-chain and HLA-DR1β-chain were successfully assembled and co-expressed in *Sf9* cells. Correctly refolded HLA-DR1^{CLIP} was purified from the supernatant of these cells using immunoprecipitation and size exclusion chromatography. Following successful cleavage of the CLIP peptide, and exchange for HA$_{306\text{-}318}$ peptide the refolded protein was confirmed as being functional bu pHLA-II tetramer staining of a cognate CD4$^+$ T cell clone. Having no prior experience with insect cells, this project presented a steep but gratifying learning curve. Insect cell expression systems are already used for generating various MHC-II alleles (Pos et al. 2012; Quarsten et al. 2001). The approach developed here combines several useful features of other expression systems such as a leucine zipper to enhance chain pairing and the cleavable CLIP and the mellitin leader sequence (Scott et al. 1996; Pos et al. 2012; Tessier et al. 1991) allowing easy exchange with a peptide of interest. Easy to use and straight forward expression systems like the one implemented here will greatly help the advancement of research into CD4$^+$ T cells.

Chapter 4 dissected the role of peptide flanking regions using the well-studied influenza HA$_{306\text{-}318}$ peptide as an experimental system. Using a nested set of PFR variants, their impact of these on CD4$^+$ T cell activation, peptide binding stability and TRB gene

selection was investigated. C-terminal PFR proved to be crucial for activation of CD4$^+$ T cells clones specific for the HA$_{306-318}$ peptide. Stabilizing effects of the C-terminal PFR are most likely the underlying cause for this phenomenon. TRB clonotyping showed that each individual PFR variant selected a very different set of TCR clonotypes. The results add valuable knowledge to the diverse roles of PFR heterogeneity in CD4$^+$ T cell immunity. Since naturally processed peptides exhibit different length PFR, their impact on TCR clonotype selection might be of importance for vaccine development. While other studies primarily investigated the effects of amino acid substitutions or complete removal of PFR, this study presented here concentrated on effects of altering the length of PFR (Arnold et al. 2002; Godkin et al. 2001; Holland et al. 2015). The two other studies investigating the impact of PFR on TCR gene selection confirm our observation that changes in PFR lead to a change in TRB repertoire selection (Carson et al. 1997; Cole et al. 2012). Again, these studies investigated the effects of single amino acid substitutions. The effects of different length PFR as shown here, had not been previously investigated. Differences in PFR occur naturally as a result of the MHC-II antigen processing machinery. The results presented in this chapter suggest that PFR and their impact on CD4$^+$ T cell activation are an inherent feature of CD4$^+$ T cell immunity and should be investigated into more detail.

In chapter 5, a recently solved structure of a HIV gag24 derived peptide presented by HLA-DR1 was used to predict and confirm potential TCR contact residues within the peptide using alanine substitution variants. These findings have wide reaching impact due to this epitope being highly conserved across all lentivirus and therefore likely to be present in most HLA-DR1$^+$ HIV$^+$ patients (Mammano et al. 1994; Provitera et al. 2001). This feature also makes in an interesting target for vaccine development. Next, the role of the PFR in this experimental system was dissected, particularly the hook-shaped N-

terminal PFR. The interplay between both termini appeared to be more complex than what had been observed in the $HA_{306-318}$ experimental system described chapter 4. However, my data confirmed the importance of secondary structures observed within peptide bound to MHC-II seen in other experimental systems (Norris et al. 2006; Zavala-Ruiz et al. 2004; Li et al. 2010). With the growing database of pMHC-II structures, such observations are likely to become more common, particularly for longer peptides. Based on these observations and the other two examples, it is probable that secondary structure are an inherent and import feature of PFR. Mutations within PFR could lead to a change in the secondary structure, thereby abrogating $CD4^+$ T cell recognition, by altering TCR contacts or peptide-MHC-II binding, in a novel way.

In chapter 6, potential epitopes within regions of influenza A HA which had been previously identified to elicit responses in HLA-DR1$^+$ individuals, were mapped and analysed in further detail. Using PBMCs of three HLA-DR1$^+$ volunteers who had been recently immunized with the trivalent influenza vaccine, three potential epitopes were identified. All three peptides had been identified in other studies, confirming these results Next, the HLA-DR1 restriction of two of these epitopes was confirmed using a HLA-DR1 homozygous cell line as APC in conjunction with a blocking antibody. Using bioinformatical tools, it was shown that these peptides are highly conserved across all H3N2 strains, making them potentially interesting targets for vaccines. Efforts to develop universal vaccines concentrate on generating broad neutralising antibodies against the conserved stalk region of the HA protein and would likely be able to induce $CD4^+$ T cell responses against epitopes located within the stalks such as peptide 18 identified here (Khanna et al. 2014; Hashem 2015). Identifying novel epitopes aid the study of anti-viral $CD4^+$ T cell responses by looking at a broader range of epitopes (Long et al. 2013).

7.1. Future considerations

7.1.1. *The importance of being able generate soluble pMHC-II of various HLA alleles*

Despite having successfully expressed and purified HLA-DR1 using *Sf9* cells, there is still room for further optimization of this system. Further experiments using a wider range of MOIs would help maximise yield by determining the optimal amount of virus to be used. Our current set-up allows for expression in multiple of 300 ml cultures simultaneously. Once the expression conditions for these bigger scale cultures are optimised, purification methods would have to be adjusted accordingly. While we are having great success with our current L243 coated protein A columns, an FPLC based system would save time and effort. This additional purification step using immunoprecipitation greatly improved our in-house protocol for refolding pMHC-II from inclusion bodies. Proteins expressed in *E. coli* are better suited for crystal studies due to their lack of post-translational modifications. Therefore, further optimisations of the insect cell expression system became less urgent. Nevertheless, this system provides a number of advantages for the production of soluble pMHC-II for multimer staining, as detailed below.

Various MHC-II alleles have been associated with increased risk for certain diseases or protection thereof (Mackie et al. 2012; Koeleman et al. 2004; Thursz et al. 1999). Being able to quickly and easily generate soluble pMHC-II for multimer staining would greatly facilitate research into the involvement of such alleles in different disease settings. HLA-DR alleles express the same or at least a highly similar α-chain. The mix-and-match approach used in the insect cell based expression system for HLA-DR1 in chapter 3 is easily adaptable by simply swapping the HLA-DR1β baculoviral construct for the β-chain of choice. In fact, the baculoviral vector encoding the HLA-DR4β-chain has already been

generated and resulting HLA-DR4 monomers await testing by SPR based experiments using HLA-DR4 restricted TCRs. Since L243 binds to the HLA-DRα-chain, it can be used for the purification of any HLA-DR heterodimer. Similarly, any other human and murine MHC-II allele could be expressed by assembling expression vectors for the according α- and β-chains. However, these alleles would require a different purification approach, possibly using cleavable purification tags such as the V5- and His-tags already encoded in the BaculoDirect^{TM} system used here. Being able to cleave the covalently linked CLIP peptide and replacing it with any peptide of interest further increases the potential applications for this system.

7.1.2. *Further investigations into the role of PFR for CD4$^+$ T cell immunity*

The study presented in Chapter 4 dissected the role of PFR by using nested sets of peptides differing in the length. It would be interesting to repeat this using different experimental models. It would also be interesting to measure changes in TCR affinity to different PFR variants using SPR. This route of investigation was started during the final months of this PhD. Preliminary results showed that TCRs bind PFR variants with different affinities. However, due to time constraints, these experiments have not been completed. pMHC multimer staining of CD4$^+$ T cells using PFR variants could also give more insight into their effect on TCR/pMHC-II interactions. In addition, solving the crystal structures of TCRs in complex with these PFR variants bound to HLA-DR1 to determine how TCRs adapt to the new challenge of altered PFR.

The clonotyping study also opened up several new avenues of investigation. It would be interesting to study the clonotypes generated by the different PFR variants in more details to answer some of the following questions: Are clonotypes generated against the C-terminal variants capable of recognising N-terminal variants and *vice versa*? Do

clonotypes generated against C-terminal PFR preferentially contact C-terminal residues of the peptide and *vice versa*? Are clonotypes generated against the 9mer core peptide still dependant on PFR? Further functional, biophysical and structural studies would answer these questions. Repeating the clonotyping study using a cocktail of PFR variants would simulate an *in vivo* situation where nested sets of peptide are presented simultaneously would give.

Despite $HA_{306-318}$ being a popular experimental system, no nested sets of naturally processed peptides has been characterised to date for this protein. In order to do fill this caveat, such an experiment has recently been started by eluting peptides off the surface of the HLA-DR1$^+$ homozygous HOM-2 cell line pulsed with HA protein from the influenza A X-31 strain commonly used in laboratory research. These peptides await sequencing. Once naturally processed PFR variants have been identified, it would be informative to repeat this study in order to gain more insight into CD4$^+$ T cell responses against naturally occurring PFR variants.

7.1.2.1. *Further investigations into the secondary structure within PFR*

Based on the results of the alanine scan on the gag24 peptide, a number of potential TCR contact residues were identified. Further investigations using SPR and structural analysis of the Ox97 TCR in complex with HLA-DR1^{gag24} will confirm these results. It would be useful to generate additional HLA-DR1^{gag24} specific CD4$^+$ T cell clones in order to identify patterns in CD4$^+$ T cell recognition of this highly conserved peptide. Clonotypic analysis of lines grown against PFR variants would give more insight into the impact of PFR on TCR gene selection.

The observation of the hook like secondary structure adopted by the N-terminal PFR opens up a plethora of possible future investigations. The competitive peptide-binding assay did not generate any data for this experimental system. It would be useful to try out other approaches in order generate data on how disruption of these secondary structures influences peptide binding stability. For example, a cell based competitive peptide binding assay such as the one by described by Weenink and colleagues (Weenink, Milburn, and Gautam 1997). Here, APCs were incubated in presence of both BT-CLIP and competitor peptides. Binding of the competitor peptide was assessed by flow cytometry following staining with fluorochrome-labelled streptavidin. Although similar to the competitive assay used in this thesis, it does not rely on the solubility of peptides at low pH. Measuring changes in TCR affinity for different PFR variants will give more information about the impact of missing PFR on CD4$^+$ T cell interactions. As for HA$_{306-318}$, some soluble TCRs and pMHC-II have already been generated for this purpose. Again, these experiments have not been completed due to time constraints. In conjunction with solving the structure of these variants bound to HLA-DR1^{gag24} in presence or absence of the Ox97 TCR, these data would generate a fuller picture of the impact of secondary structures in MHC-II restricted peptides.

7.1.2.2. *Confirming natural processing of conserved HA peptides*

The two HLA-DR1 restricted peptides 3 and 18 are both highly conserved and therefore interesting targets for the development of a universal vaccine. However, they would need to be naturally processed in order to function as such. Further experiments using APCs pulsed with whole HA protein and testing them for recognition by CD4$^+$ T cells raised against each peptide would give answers. Tracking epitope 3 and 18 specific CD4$^+$ T cells throughout Influenza A infections or following vaccination using pMHC-II multimers

would give more insight into anti-influenza A CD4$^+$ T cell responses in particular and anti-viral CD4$^+$ T cell responses in general.

7.2. Concluding remarks

The focus of this thesis can be summarised under the topic of dissecting TCR/pMHC-II interactions. Future studies of this kind will greatly benefit from versatile manufacturing systems for soluble pMHC-II such as the insect cell based system described here. The confirmation of the HLA-DR1 restriction of two potential HA derived epitopes opens new avenues of research into $CD4^+$ T cell mediated, anti-influenza immunity. Identifying and tracking specific $CD4^+$ T cell responses against multiple epitopes will broaden our understanding of the complex dynamics underlying our immune system. The complex and crucial roles of PFR have become evident in both studies looking at their impact on $CD4^+$ T cell activation. Their role in stabilising the $HA_{306-318}$ peptide in the peptide-binding groove might appear simple but has wide reaching impact for TCR gene selection. The impact of secondary structure elements such as formed by the N-terminus of the gag24 peptide illustrates how seemingly small variations in PFR length can abolish $CD4^+$ T cell recognition. Both studies underline the need for more in-depth investigations into PFR as inherent features in pMHC-II recognition.

In summary, this thesis has added novel knowledge to our understanding of $CD4^+$ T cells as key players in the immune system, in particular the important role of PFR in antigen recognition while opening new avenues of research.

References

Alam, S., and A. J. Sant. 2011. "Infection with Seasonal Influenza Virus Elicits CD4 T Cells Specific for Genetically Conserved Epitopes That Can Be Rapidly Mobilized for Protective Immunity to Pandemic H1N1 Influenza Virus." *Journal of Virology* 85 (24). American Society for Microbiology: 13310–21. doi:10.1128/JVI.05728-11.

Alarcon, Balbino, Diana Gil, Pilar Delgado, and Wolfgang W. A. Schamel. 2003. "Initiation of TCR Signaling: Regulation within CD3 Dimers." *Immunological Reviews* 191 (1). Munksgaard International Publishers: 38–46. doi:10.1034/j.1600-065X.2003.00017.x.

Altman, J D, P A Moss, P J Goulder, D H Barouch, M G McHeyzer-Williams, J I Bell, A J McMichael, and M M Davis. 1996. "Phenotypic Analysis of Antigen-Specific T Lymphocytes." *Science (New York, N.Y.)* 274 (5284): 94–96. http://www.ncbi.nlm.nih.gov/pubmed/8810254.

Anders, Anne-Kathrin K, Melissa J Call, Monika-Sarah S E D Schulze, Kevin D Fowler, David A Schubert, Nilufer P Seth, Eric J Sundberg, and Kai W Wucherpfennig. 2011. "HLA-DM Captures Partially Empty HLA-DR Molecules for Catalyzed Removal of Peptide." Journal Article. *Nat Immunol* 12 (1). Nature Publishing Group: 54–61. doi:ni.1967 [pii]10.1038/ni.1967.

Andreatta, Massimo, Edita Karosiene, Michael Rasmussen, Anette Stryhn, Søren Buus, and Morten Nielsen. 2015. "Accurate Pan-Specific Prediction of Peptide-MHC Class II Binding Affinity with Improved Binding Core Identification." *Immunogenetics* 67 (11–12): 641–50. doi:10.1007/s00251-015-0873-y.

Arnold, Paula Y, Nicole L La Gruta, Tim Miller, Kate M Vignali, P Scott Adams, David L Woodland, and Dario A A Vignali. 2002. "The Majority of Immunogenic Epitopes Generate CD4+ T Cells That Are Dependent on MHC Class II-Bound Peptide-Flanking Residues." *Journal of Immunology (Baltimore, Md. : 1950)* 169 (2). American Association of Immunologists: 739–49. doi:10.4049/JIMMUNOL.169.2.739.

Babon, Jenny Aurielle B, John Cruz, Laura Orphin, Pamela Pazoles, Mary Dawn T Co, Francis A Ennis, and Masanori Terajima. 2009. "Genome-Wide Screening of Human T-Cell Epitopes in Influenza A Virus Reveals a Broad Spectrum of CD4(+) T-Cell Responses to Internal Proteins, Hemagglutinins, and Neuraminidases." Journal Article. *Hum Immunol* 70 (9). Elsevier Inc.: 711–21. doi:S0198-8859(09)00144-X [pii]10.1016/j.humimm.2009.06.004.

Barré-Sinoussi, F, J C Chermann, F Rey, M T Nugeyre, S Chamaret, J Gruest, C Dauguet, et al. 1983. "Isolation of a T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS)." *Science (New York, N.Y.)* 220 (4599). American Association for the Advancement of Science: 868–71. doi:10.1126/science.6189183.

Barré-Sinoussi, Françoise, Anna Laura Ross, and Jean-François Delfraissy. 2013. "Past, Present and Future: 30 Years of HIV Research." *Nature Reviews. Microbiology* 11 (12). Nature Publishing Group: 877–83. doi:10.1038/nrmicro3132.

Bassing, Craig H, Wojciech Swat, and Frederick W Alt. 2002. "The Mechanism and Regulation of Chromosomal V(D)J Recombination." *Cell* 109 (2): S45–55. doi:10.1016/S0092-8674(02)00675-X.

Benes, Petr, Vaclav Vetvicka, and Martin Fusek. 2008. "Cathepsin D--Many Functions of One Aspartic Protease." *Critical Reviews in Oncology/hematology* 68 (1). NIH Public Access: 12–28. doi:10.1016/j.critrevonc.2008.02.008.

Beringer, Dennis X, Fleur S Kleijwegt, Florian Wiede, Arno R van der Slik, Khai Lee Loh, Jan Petersen, Nadine L Dudek, et al. 2015. "T Cell Receptor Reversed Polarity Recognition of a Self-Antigen Major Histocompatibility Complex." *Nature Immunology* 16 (11): 1153–61. doi:10.1038/ni.3271.

Betts, Michael R, David R Ambrozak, Daniel C Douek, Jason M Brenchley, Joseph P Casazza, a Richard, Louis J Picker, and Sebastian Bonhoeffer. 2001. "Analysis of Total Human Immunodeficiency Responses : Relationship to Viral Load in Untreated HIV Infection Analysis of Total Human Immunodeficiency Virus (HIV) -Specific CD4+ and CD8+ T-Cell Responses : Relationship to Viral Load in Untreated HIV Infect." *Journal of Virology* 75 (24): 11983–91. doi:10.1128/JVI.75.24.11983.

Bianchi, Valentina, Anna Bulek, Anna Fuller, Angharad Lloyd, Meriem Attaf, Pierre J Rizkallah, Garry Dolton, Andrew K Sewell, and David K Cole. 2016. "A Molecular Switch Abrogates Glycoprotein 100 (gp100) T-Cell Receptor (TCR) Targeting of a Human Melanoma Antigen." *The Journal of Biological Chemistry* 291 (17). American Society for Biochemistry and Molecular Biology: 8951–59. doi:10.1074/jbc.M115.707414.

Boulter, J. M., M. Glick, P. T. Todorov, E. Baston, M. Sami, P. Rizkallah, and B. K. Jakobsen. 2003. "Stable, Soluble T-Cell Receptor Molecules for Crystallization and Therapeutics." Journal Article. *Protein Eng* 16 (9): 707–11. doi:10.1093/protein/gzg087.

Brodsky, F M. 1984. "A Matrix Approach to Human Class II Histocompatibility Antigens: Reactions of Four Monoclonal Antibodies with the Products of Nine Haplotypes." *Immunogenetics* 19 (3): 179–94. http://www.ncbi.nlm.nih.gov/pubmed/6200433.

Bulek, Anna M, Florian Madura, Anna Fuller, Christopher J Holland, Andrea J a Schauenburg, Andrew K Sewell, Pierre J Rizkallah, and David K Cole. 2012. "TCR/pMHC Optimized Protein Crystallization Screen." *Journal of Immunological Methods* 382 (1–2). Elsevier B.V.: 203–10. doi:10.1016/j.jim.2012.06.007.

Carson, R.T. T, K M Vignali, D.L. L Woodland, D.A.A. A Vignali, P.M. Allen, D.J. Strydom, E.R. Unanue, et al. 1997. "T Cell Receptor Recognition of MHC Class II–Bound Peptide Flanking Residues Enhances Immunogenicity and Results in Altered TCR V Region Usage." *Immunity* 7 (3). Elsevier. http://www.ncbi.nlm.nih.gov/pubmed/9324359.

Chen, Alex T., Markus Cornberg, Stephanie Gras, Carole Guillonneau, Jamie Rossjohn, Andrew Trees, Sebastien Emonet, et al. 2012. "Loss of Anti-Viral Immunity by Infection with a Virus Encoding a Cross-Reactive Pathogenic Epitope." Edited by Christopher M. Walker. *PLoS Pathogens* 8 (4). Public Library of Science: e1002633. doi:10.1371/journal.ppat.1002633.

Chen, Shuming, Yili Li, Florence R Depontieu, Tracee L McMiller, A Michelle English, Jeffrey Shabanowitz, Ferdynand Kos, et al. 2013. "Structure-Based Design of Altered MHC Class II-Restricted Peptide Ligands with Heterogeneous Immunogenicity." *Journal of Immunology (Baltimore, Md. : 1950)* 191 (10). NIH Public Access: 5097–5106. doi:10.4049/jimmunol.1300467.

Chicz, R M, R G Urban, J C Gorga, D A Vignali, W S Lane, and J L Strominger. 1993. "Specificity and Promiscuity among Naturally Processed Peptides Bound to HLA-DR Alleles." Journal Article. *J Exp Med* 178 (1): 27–47. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8315383.

Chicz, R M, R G Urban, W S Lane, J C Gorga, L J Stern, D A Vignali, and J L Strominger. 1992. "Predominant Naturally Processed Peptides Bound to HLA-DR1 Are Derived from MHC-Related Molecules and Are Heterogeneous in Size." Journal Article. *Nature* 358 (6389): 764–68. doi:10.1038/358764a0.

Chlabicz, Michał, Marek Gacko, Anna Worowska, and Radosław Łapiński. 2012. "Cathepsin E (EC 3.4.23.34) — a Review." *Folia Histochemica et Cytobiologica* 49 (4): 547–57. doi:10.5603/FHC.2011.0078.

Christie, J M, C J Healey, J Watson, V S Wong, M Duddridge, N Snowden, W M Rosenberg, K A Fleming, H Chapel, and R W Chapman. 1997. "Clinical Outcome of Hypogammaglobulinaemic Patients Following Outbreak of Acute Hepatitis C: 2 Year Follow up." Journal Article. *Clin Exp Immunol* 110 (1): 4–8. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9353141.

Coffin, John M, Stephen H Hughes, and Harold E Varmus. 1997. "Course of Infection with HIV and SIV." Cold Spring Harbor Laboratory Press.

Cole, David K, Emily S J Edwards, Katherine K Wynn, Mathew Clement, John J Miles, Kristin Ladell, Julia Ekeruche, et al. 2010. "Modification of MHC Anchor Residues Generates Heteroclitic Peptides That Alter TCR Binding and T Cell Recognition." *Journal of Immunology (Baltimore, Md. : 1950)* 185 (4). Europe PMC Funders: 2600–2610. doi:10.4049/jimmunol.1000629.

Cole, David K, Kathleen Gallagher, Brigitte Lemercier, Christopher J Holland, Sayed Junaid, James P Hindley, Katherine K Wynn, et al. 2012. "Modification of the Carboxy-Terminal Flanking Region of a Universal Influenza Epitope Alters CD4$^+$ T-Cell Repertoire Selection." Journal Article. *Nature Communications* 3 (January). Nature Publishing Group: 665. doi:10.1038/ncomms1665.

Cox, Freek, Matthijs Baart, Jeroen Huizingh, Jeroen Tolboom, Liesbeth Dekking, Jaap Goudsmit, Eirikur Saeland, et al. 2015. "Protection against H5N1 Influenza Virus Induced by Matrix-M Adjuvanted Seasonal Virosomal Vaccine in Mice Requires Both Antibodies and T Cells." Edited by Florian Krammer. *PloS One* 10 (12). Public Library of Science: e0145243. doi:10.1371/journal.pone.0145243.

Crooks, Gavin E, Gary Hon, John-Marc Chandonia, and Steven E Brenner. 2004. "WebLogo: A Sequence Logo Generator." *Genome Research* 14 (6). Cold Spring Harbor Laboratory Press: 1188–90. doi:10.1101/gr.849004.

Cross, Deborah. 2016. "Unique Hairpin Binding Mode in Antigen Presentation of HIV-1 Epitope."

Crowe, Sherry R, Shannon C Miller, Deborah M Brown, Pamela S Adams, Richard W Dutton, Allen G Harmsen, Frances E Lund, Troy D Randall, Susan L Swain, and David L Woodland. 2006. "Uneven Distribution of MHC Class II Epitopes within the Influenza Virus." *Vaccine* 24: 457–67. doi:10.1016/j.vaccine.2005.07.096.

Das, K, J M Aramini, L C Ma, R M Krug, and E Arnold. 2010. "Structures of Influenza A Proteins and Insights into Antiviral Drug Targets." Journal Article. *Nat Struct Mol*

Biol 17 (5): 530–38. doi:nsmb.1779 [pii]10.1038/nsmb.1779.

Delamarre, Lélia, Margit Pack, Henry Chang, Ira Mellman, and E Sergio Trombetta. 2005. "Differential Lysosomal Proteolysis in Antigen-Presenting Cells Determines Antigen Fate." *Science (New York, N.Y.)* 307 (5715). American Association for the Advancement of Science: 1630–34. doi:10.1126/science.1108003.

Denzin, Lisa K., and Peter Cresswell. 1995. "HLA-DM Induces Clip Dissociation from MHC Class II Aβ Dimers and Facilitates Peptide Loading." *Cell* 82 (1). Cell Press: 155–65. doi:10.1016/0092-8674(95)90061-6.

Doherty, P C, J M Riberdy, and G T Belz. 2000. "Quantitative Analysis of the CD8+ T-Cell Response to Readily Eliminated and Persistent Viruses." Journal Article. *Philos Trans R Soc Lond B Biol Sci* 355 (1400): 1093–1101. doi:10.1098/rstb.2000.0647.

Dolton, Garry, Katie Tungatt, Angharad Lloyd, Valentina Bianchi, Sarah M. Theaker, Andrew Trimby, Christopher J. Holland, et al. 2015. "More Tricks with Tetramers: A Practical Guide to Staining T Cells with Peptide-MHC Multimers." *Immunology* 146 (1): 11–22. doi:10.1111/imm.12499.

Douek, Daniel C., Jason M. Brenchley, Michael R. Betts, David R. Ambrozak, Brenna J. Hill, Yukari Okamoto, Joseph P. Casazza, et al. 2002. "HIV Preferentially Infects HIV-Specific CD4+ T Cells." *Nature* 417 (6884). Nature Publishing Group: 95–98. doi:10.1038/417095a.

Drake, J R, P Webster, J C Cambier, and I Mellman. 1997. "Delivery of B Cell Receptor-Internalized Antigen to Endosomes and Class II Vesicles." *The Journal of Experimental Medicine* 186 (8). The Rockefeller University Press: 1299–1306. http://www.ncbi.nlm.nih.gov/pubmed/9334369.

Driessen, Christoph, Ana-Maria Lennon-Duménil, and Hidde L. Ploegh. 2001. "Individual Cathepsins Degrade Immune Complexes Internalized by Antigen-Presenting Cells via Fcγ Receptors." *European Journal of Immunology* 31 (5). WILEY-VCH Verlag GmbH: 1592–1601. doi:10.1002/1521-4141(200105)31:5<1592::AID-IMMU1592>3.0.CO;2-K.

Fiore-Gartland, Andrew, Bryce A. Manso, David P. Friedrich, Erin E. Gabriel, Greg Finak, Zoe Moodie, Tomer Hertz, et al. 2016. "Pooled-Peptide Epitope Mapping Strategies Are Efficient and Highly Sensitive: An Evaluation of Methods for Identifying Human T Cell Epitope Specificities in Large-Scale HIV Vaccine Efficacy Trials." Edited by Shan Lu. *PLOS ONE* 11 (2). Public Library of Science: e0147812. doi:10.1371/journal.pone.0147812.

Frayser, M, A K Sato, L Xu, and L J Stern. 1999. "Empty and Peptide-Loaded Class II Major Histocompatibility Complex Proteins Produced by Expression in Escherichia Coli and Folding in Vitro." *Protein Expression and Purification* 15 (1): 105–14. doi:10.1006/prep.1998.0987.

Freed, Eric O. 1998. "HIV-1 Gag Proteins: Diverse Functions in the Virus Life Cycle." *Virology* 251 (1). Academic Press: 1–15. doi:10.1006/viro.1998.9398.

Garcia, K C, L Teyton, and I A Wilson. 1999. "Structural Basis of T Cell Recognition." Journal Article. *Annu Rev Immunol* 17: 369–97. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10358763.

Gelder, C M, K I Welsh, a Faith, J R Lamb, and B A Askonas. 1995. "Human CD4+ T-Cell Repertoire of Responses to Influenza A Virus Hemagglutinin after Recent

Natural Infection." Journal Article. *J Virol* 69 (12): 7497–7506. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=189688&tool=pmcentr ez&rendertype=abstract.

Gelder, Colin, Miles Davenport, Martin Barnardo, Tim Bourne, Jonathan Lamb, Brigitte Askonas, Adrian Hill, and Ken Welsh. 1998. "Six Unrelated HLA-DR-Matched Adults Recognize Identical CD4 ✶ T Cell Epitopes from Influenza A Haemagglutinin That Are Not Simply Peptides with High HLA-DR Binding Affinities." *International Immunology* 10 (2): 211–22.

Godeau, F, C Saucier, and P Kourilsky. 1992. "Replication Inhibition by Nucleoside Analogues of a Recombinant Autographa Californica Multicapsid Nuclear Polyhedrosis Virus Harboring the Herpes Thymidine Kinase Gene Driven by the IE-1(0) Promoter: A New Way to Select Recombinant Baculoviruses." *Nucleic Acids Research* 20 (23): 6239–46. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=334511&tool=pmcentr ez&rendertype=abstract.

Godfrey, Dale I, Adam P Uldrich, James McCluskey, Jamie Rossjohn, and D Branch Moody. 2015. "The Burgeoning Family of Unconventional T Cells." *Nature Immunology* 16 (11). Nature Research: 1114–23. doi:10.1038/ni.3298.

Godkin, A. J., K. J. Smith, A. Willis, M. V. Tejada-Simon, J. Zhang, T. Elliott, and A. V. S. Hill. 2001. "Naturally Processed HLA Class II Peptides Reveal Highly Conserved Immunogenic Flanking Region Sequence Preferences That Reflect Antigen Processing Rather Than Peptide-MHC Interactions." *The Journal of Immunology* 166 (11). American Association of Immunologists: 6720–27. doi:10.4049/jimmunol.166.11.6720.

Godkin, A J, M P Davenport, A Willis, D P Jewell, and A V Hill. 1998. "Use of Complete Eluted Peptide Sequence Data from HLA-DR and -DQ Molecules to Predict T Cell Epitopes, and the Influence of the Nonbinding Terminal Regions of Ligands in Epitope Selection." *Journal of Immunology (Baltimore, Md. : 1950)* 161 (2). American Association of Immunologists: 850–58. http://www.ncbi.nlm.nih.gov/pubmed/9670963.

Gross, Ulrike, Anja K. Schroder, Romney S. Haylett, Sabine Arlt, and Lothar Rink. 2006. "The Superantigen Staphylococcal Enterotoxin A (SEA) and Monoclonal Antibody L243 Share a Common Epitope but Differ in Their Ability to Induce Apoptosis via MHC-II." *Immunobiology* 211 (10): 807–14. doi:10.1016/j.imbio.2006.05.006.

Guo, H., F. Santiago, K. Lambert, T. Takimoto, and D. J. Topham. 2011. "T Cell-Mediated Protection against Lethal 2009 Pandemic H1N1 Influenza Virus Infection in a Mouse Model." *Journal of Virology* 85 (1). American Society for Microbiology: 448–55. doi:10.1128/JVI.01812-10.

Hashem, Anwar M. 2015. "Prospects of HA-Based Universal Influenza Vaccine." *BioMed Research International* 2015. Hindawi Publishing Corporation: 414637. doi:10.1155/2015/414637.

Hayward, Andrew C, Ellen B Fragaszy, Alison Bermingham, Lili Wang, Andrew Copas, W John Edmunds, Neil Ferguson, et al. 2014. "Comparative Community Burden and Severity of Seasonal and Pandemic Influenza: Results of the Flu Watch Cohort Study." *The Lancet Respiratory Medicine* 2 (6). Elsevier: 445–54. doi:10.1016/S2213-2600(14)70034-7.

Hennecke, J., Andrea Carfi, Don C. Wiley, H. Acha-Orbea, DJ. Mitchell, L. Timmermann, DC. Wraith, et al. 2000. "Structure of a Covalently Stabilized Complex of a Human Alphabeta T-Cell Receptor, Influenza HA Peptide and MHC Class II Molecule, HLA-DR1." *The EMBO Journal* 19 (21). EMBO Press: 5611–24. doi:10.1093/emboj/19.21.5611.

Hennecke, J, and D C Wiley. 2001. "T Cell Receptor-MHC Interactions up Close." Journal Article. *Cell* 104 (1): 1–4. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11163234.

Hennecke, Jens, and Don C. Wiley. 2002. "Structure of a Complex of the Human A/β T Cell Receptor (TCR) HA1.7, Influenza Hemagglutinin Peptide, and Major Histocompatibility Complex Class II Molecule, HLA-DR4 (DRA0101 and DRB10401)." *The Journal of Experimental Medicine* 195 (5). Rockefeller University Press: 571–81. doi:10.1084/jem.20011194.

Hewitt, E W, A Treumann, N Morrice, P J Tatnell, J Kay, and C Watts. 1997. "Natural Processing Sites for Human Cathepsin E and Cathepsin D in Tetanus Toxin: Implications for T Cell Epitope Generation." *Journal of Immunology (Baltimore, Md. : 1950)* 159 (10): 4693–99. http://www.ncbi.nlm.nih.gov/pubmed/9366392.

Holland, Christopher J., David K. Cole, and Andrew Godkin. 2013. "Re-Directing CD4+ T Cell Responses with the Flanking Residues of MHC Class II-Bound Peptides: The Core Is Not Enough." *Frontiers in Immunology* 4 (July): 1–9. doi:10.3389/fimmu.2013.00172.

Holland, Christopher J, Garry Dolton, Martin Scurr, Kristin Ladell, Andrea J Schauenburg, Kelly Miners, Florian Madura, et al. 2015. "Enhanced Detection of Antigen-Specific CD4+ T Cells Using Altered Peptide Flanking Residue Peptide-MHC Class II Multimers." *Journal of Immunology (Baltimore, Md. : 1950)* 195 (12): 5827–36. doi:10.4049/jimmunol.1402787.

Homa, F L, C A Baker, D R Thomsen, and A P Elhammer. 1995. "Conversion of a Bovine UDP-GalNAc:polypeptide, N-Acetylgalactosaminyltransferase to a Soluble, Secreted Enzyme, and Expression in Sf9 Cells." Journal Article. *Protein Expr Purif* 6 (2): 141–48. doi:S1046-5928(85)71017-0 [pii] 10.1006/prep.1995.1017.

Ishizuka, Jeffrey, Kristie Grebe, Eugene Shenderov, Bjoern Peters, Qiongyu Chen, Yanchun Peng, Lili Wang, et al. 2009. "Quantitating T Cell Cross-Reactivity for Unrelated Peptide Antigens." *Journal of Immunology (Baltimore, Md. : 1950)* 183 (7). NIH Public Access: 4337–45. doi:10.4049/jimmunol.0901607.

Jackson, David C., Heidi E. Drummer, and Lorena E. Brown. 1994. "Conserved Determinants for CD4+ T Cells within the Light Chain of the H3 Hemagglutinin Molecule of Influenza Virus." *Virology* 198 (2). Academic Press: 613–23. doi:10.1006/viro.1994.1073.

Kaufmann, Daniel E, Paul M Bailey, John Sidney, Bradford Wagner, Philip J Norris, Mary N Johnston, Lisa A Cosimi, et al. 2004. "Comprehensive Analysis of Human Immunodeficiency Virus Type 1-Specific CD4 Responses Reveals Marked Immunodominance of Gag and Nef and the Presence of Broadly Recognized Peptides." *Journal of Virology* 78 (9). American Society for Microbiology: 4463–77. doi:10.1128/JVI.78.9.4463-4477.2004.

Keele, Brandon F, Fran Van Heuverswyn, Yingying Li, Elizabeth Bailes, Jun Takehisa,

Mario L Santiago, Frederic Bibollet-Ruche, et al. 2006. "Chimpanzee Reservoirs of Pandemic and Nonpandemic HIV-1." *Science (New York, N.Y.)* 313 (5786). NIH Public Access: 523–26. doi:10.1126/science.1126531.

Khanna, Madhu, Sachin Sharma, Binod Kumar, and Roopali Rajput. 2014. "Protective Immunity Based on the Conserved Hemagglutinin Stalk Domain and Its Prospects for Universal Influenza Vaccine Development." *BioMed Research International* 2014. Hindawi Publishing Corporation: 546274. doi:10.1155/2014/546274.

Kioukia, N., A.W. Nienow, A.N. Emery, and M. Al-rubeai. 1995. "Physiological and Environmental Factors Affecting the Growth of Insect Cells and Infection with Baculovirus." *Journal of Biotechnology* 38 (3): 243–51. doi:10.1016/0168-1656(94)00128-Y.

Kirschke, Heidrun, and Bernd Wiederanders. 1994. "[34] Cathepsin S and Related Lysosomal Endopeptidases." *Methods in Enzymology* 244: 500–511. doi:10.1016/0076-6879(94)44036-0.

Klein, Ludger, Maria Hinterberger, Gerald Wirnsberger, and Bruno Kyewski. 2009. "Antigen Presentation in the Thymus for Positive Selection and Central Tolerance Induction." *Nature Reviews Immunology* 9 (12). Nature Publishing Group: 833–44. doi:10.1038/nri2669.

Koeleman, B P C, B A Lie, D E Undlien, F Dudbridge, E Thorsby, R R P de Vries, F Cucca, B O Roep, M J Giphart, and J A Todd. 2004. "Genotype Effects and Epistasis in Type 1 Diabetes and HLA-DQ Trans Dimer Associations with Disease." *Genes and Immunity* 5 (5). Nature Publishing Group: 381–88. doi:10.1038/sj.gene.6364106.

Korn, Thomas, Estelle Bettelli, Mohamed Oukka, and Vijay K. Kuchroo. 2009. "IL-17 and Th17 Cells." *Annual Review of Immunology* 27 (1). Annual Reviews: 485–517. doi:10.1146/annurev.immunol.021908.132710.

Koup, R A, J T Safrit, Y Cao, C A Andrews, G McLeod, W Borkowsky, C Farthing, and D D Ho. 1994. "Temporal Association of Cellular Immune Responses with the Initial Control of Viremia in Primary Human Immunodeficiency Virus Type 1 Syndrome." *Journal of Virology* 68 (7). American Society for Microbiology (ASM): 4650–55. http://www.ncbi.nlm.nih.gov/pubmed/8207839.

Kozono, Haruo, Yufuku Matsushita, Naoki Ogawa, Yuko Kozono, Toshihiro Miyabe, Hiroshi Sekiguchi, Kouhei Ichiyanagi, et al. 2015. "Single-Molecule Motions of MHC Class II Rely on Bound Peptides." *Biophysical Journal* 108 (2). Elsevier: 350–59. doi:10.1016/j.bpj.2014.12.004.

Kozono, H, J White, J Clements, P Marrack, and J Kappler. 1994. "Production of Soluble MHC Class II Proteins with Covalently Bound Single Peptides." *Nature* 369 (6476): 151–54. doi:10.1038/369151a0.

Kropshofer, H, A B Vogt, G Moldenhauer, J Hammer, J S Blum, and G J Hämmerling. 1996. "Editing of the HLA-DR-Peptide Repertoire by HLA-DM." *The EMBO Journal* 15 (22). European Molecular Biology Organization: 6144–54. http://www.ncbi.nlm.nih.gov/pubmed/8947036.

Kwok, William W, Nancy A Ptacek, Andrew W Liu, and Jane H Buckner. 2002. "Use of Class II Tetramers for Identification of CD4+ T Cells." *Journal of Immunological Methods* 268 (1): 71–81. doi:10.1016/S0022-1759(02)00201-6.

Laidlaw, Brian J, Nianzhi Zhang, Heather D Marshall, Mathew M Staron, Tianxia Guan,

Yinghong Hu, Linda S Cauley, Joe Craft, and Susan M Kaech. 2014. "CD4(+) T Cell Help Guides Formation of CD103(+) Lung-Resident Memory CD8(+) T Cells during Influenza Viral Infection." *Immunity*, October. doi:10.1016/j.immuni.2014.09.007.

Lamb, J R, D D Eckels, P Lake, J N Woody, and N Green. 1982. "Human T-Cell Clones Recognize Chemically Synthesized Peptides of Influenza Haemagglutinin." Journal Article. *Nature* 300 (5887): 66–69. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=6982419.

Lamb, J R, and N Green. 1983. "Analysis of the Antigen Specificity of Influenza Haemagglutinin-Immune Human T Lymphocyte Clones: Identification of an Immunodominant Region for T Cells." *Immunology* 50 (4). Wiley-Blackwell: 659–66. http://www.ncbi.nlm.nih.gov/pubmed/6197356.

Lamb, J R, B J Skidmore, N Green, J M Chiller, and M Feldmann. 1983. "Induction of Tolerance in Influenza Virus-Immune T Lymphocyte Clones with Synthetic Peptides of Influenza Hemagglutinin." *The Journal of Experimental Medicine* 157 (5). The Rockefeller University Press: 1434–47. http://www.ncbi.nlm.nih.gov/pubmed/6189936.

Lanier, Lewis L. 2013. "Shades of Grey — the Blurring View of Innate and Adaptive Immunity." *Nature Reviews Immunology* 13 (2). Nature Publishing Group: 73–74. doi:10.1038/nri3389.

LeBien, Tucker W, Thomas F Tedder, MD. Cooper, MN. Alder, A. Tiselius, EA. Kabat, A. Fagraeus, et al. 2008. "B Lymphocytes: How They Develop and Function." *Blood* 112 (5). American Society of Hematology: 1570–80. doi:10.1182/blood-2008-02-078071.

Lee, Jong-Soo, Mohammed Y.E. Chowdhury, Ho-Jin Moon, Young-Ki Choi, Melbourne R. Talactac, Jae-Hoon Kim, Min-Eun Park, Hwa-Young Son, Kwang-Soon Shin, and Chul-Joong Kim. 2013. "The Highly Conserved HA2 Protein of the Influenza a Virus Induces a Cross Protective Immune Response." *Journal of Virological Methods* 194 (1): 280–88. doi:10.1016/j.jviromet.2013.08.022.

Li, Yili, Florence R Depontieu, John Sidney, Theresa M Salay, Victor H Engelhard, Donald F Hunt, Alessandro Sette, Suzanne L Topalian, and Roy A Mariuzza. 2010. "Structural Basis for the Presentation of Tumor-Associated MHC Class II-Restricted Phosphopeptides to CD4+ T Cells." *Journal of Molecular Biology* 399 (4). NIH Public Access: 596–603. doi:10.1016/j.jmb.2010.04.037.

Liang, Xiquan, Lansha Peng, Chang-Ho Baek, and Federico Katzen. 2015. "BioTechniques - Single Step BP/LR Combined Gateway Reactions." Accessed February 20. http://www.biotechniques.com/BiotechniquesJournal/2013/November/Single-step-BPLR-combined-Gateway-reactions/biotechniques-348086.html.

Lippolis, John D, Forest M White, Jarrod a Marto, Chance J Luckey, Timothy N J Bullock, Jeffrey Shabanowitz, Donald F Hunt, and Victor H Engelhard. 2002. "Analysis of MHC Class II Antigen Processing by Quantitation of Peptides That Constitute Nested Sets." Journal Article. *J Immunol* 169 (9): 5089–97. doi:10.4049/jimmunol.169.9.5089.

Long, Heather M, Odette L Chagoury, Alison M Leese, Gordon B Ryan, Eddie James,

Laura T Morton, Rachel J M Abbott, Shereen Sabbah, William Kwok, and Alan B Rickinson. 2013. "MHC II Tetramers Visualize Human CD4+ T Cell Responses to Epstein-Barr Virus Infection and Demonstrate Atypical Kinetics of the Nuclear Antigen EBNA1 Response." *The Journal of Experimental Medicine* 210 (5). The Rockefeller University Press: 933–49. doi:10.1084/jem.20121437.

Lovitch, Scott B, Zheng Pu, and Emil R Unanue. 2006. "Amino-Terminal Flanking Residues Determine the Conformation of a Peptide-Class II MHC Complex." *Journal of Immunology (Baltimore, Md. : 1950)* 176: 2958–68. doi:10.4049/jimmunol.176.5.2958.

Maartens, Gary, Connie Celum, SR Sharon R Lewin, PM Sharp, BH Hahn, J Hemelaar, E Gouws, et al. 2014. "HIV Infection: Epidemiology, Pathogenesis, Treatment, and Prevention." *The Lancet* 384 (9939). Elsevier: 258–71. doi:10.1016/S0140-6736(14)60164-1.

Mackie, S L, J C Taylor, S G Martin, P Wordsworth, S Steer, A G Wilson, J Worthington, P Emery, J H Barrett, and A W Morgan. 2012. "A Spectrum of Susceptibility to Rheumatoid Arthritis within HLA-DRB1: Stratification by Autoantibody Status in a Large UK Population." *Genes and Immunity* 13 (2). Nature Publishing Group: 120–28. doi:10.1038/gene.2011.60.

Madura, Florian, Pierre J Rizkallah, Christopher J Holland, Anna Fuller, Anna Bulek, Andrew J Godkin, Andrea J Schauenburg, David K Cole, and Andrew K Sewell. 2015. "Structural Basis for Ineffective T-Cell Responses to MHC Anchor Residue-Improved "heteroclitic" Peptides." *European Journal of Immunology* 45 (2). Wiley-Blackwell: 584–91. doi:10.1002/eji.201445114.

Mammano, F, A Ohagen, S Höglund, and H G Göttlinger. 1994. "Role of the Major Homology Region of Human Immunodeficiency Virus Type 1 in Virion Morphogenesis." *Journal of Virology* 68 (8). American Society for Microbiology (ASM): 4927–36. http://www.ncbi.nlm.nih.gov/pubmed/8035491.

Messaoudi, Ilhem, Jose A. Guevara Patiño, Ruben Dyall, Joël LeMaoult, and Janko Nikolich-Žugich. 2002. "Direct Link Between Mhc Polymorphism, T Cell Avidity, and Diversity in Immune Defense." *Science* 298 (5599).

Mettu, Ramgopal R., Tysheena Charles, and Samuel J. Landry. 2016. "CD4+ T-Cell Epitope Prediction Using Antigen Processing Constraints." *Journal of Immunological Methods* 432: 72–81. doi:10.1016/j.jim.2016.02.013.

Michie, Alison M, and Juan Carlos Zúñiga-Pflücker. 2002. "Regulation of Thymocyte Differentiation: Pre-TCR Signals and β-Selection." *Seminars in Immunology* 14 (5): 311–23. doi:10.1016/S1044-5323(02)00064-7.

Miles, J J, R M Brennan, and J M Burrows. 2009. "T Cell Receptor Bias in Humans." Journal Article. *Current Immunology Reviews* 5 (1): 10–21.

Mitrović, Ana, Bojana Mirković, Izidor Sosič, Stanislav Gobec, and Janko Kos. 2016. "Inhibition of Endopeptidase and Exopeptidase Activity of Cathepsin B Impairs Extracellular Matrix Degradation and Tumour Invasion." *Biological Chemistry* 397 (2): 165–74. doi:10.1515/hsz-2015-0236.

Miyadera, Hiroko, and Katsushi Tokunaga. 2015. "Associations of Human Leukocyte Antigens with Autoimmune Diseases: Challenges in Identifying the Mechanism." *Journal of Human Genetics* 60 (11). Nature Publishing Group: 697–702. doi:10.1038/jhg.2015.100.

Moir, Susan, and Anthony S Fauci. 2009. "B Cells in HIV Infection and Disease." *Nature Reviews. Immunology* 9 (4). NIH Public Access: 235–45. doi:10.1038/nri2524.

Nelson, Martha I., and Edward C. Holmes. 2007. "The Evolution of Epidemic Influenza." *Nature Reviews Genetics* 8 (3). Nature Publishing Group: 196–205. doi:10.1038/nrg2053.

Newman, Janet, David Egan, Thomas S. Walter, Ran Meged, Ian Berry, Marouane Ben Jelloul, Joel L. Sussman, et al. 2005. "Towards Rationalization of Crystallization Screening for Small- to Medium-Sized Academic Laboratories: The PACT/JCSG+ Strategy." *Acta Crystallographica Section D Biological Crystallography* 61 (10). International Union of Crystallography: 1426–31. doi:10.1107/S0907444905024984.

Nguyen, B, K Jarnagin, S Williams, H Chan, and J Barnett. 1993. "Fed-Batch Culture of Insect Cells: A Method to Increase the Yield of Recombinant Human Nerve Growth Factor (rhNGF) in the Baculovirus Expression System." *Journal of Biotechnology* 31 (2): 205–17. http://www.ncbi.nlm.nih.gov/pubmed/7764302.

Nguyen, Cao, Michael D Varney, Leonard C Harrison, and Grant Morahan. 2013. "Definition of High-Risk Type 1 Diabetes HLA-DR and HLA-DQ Types Using Only Three Single Nucleotide Polymorphisms." *Diabetes* 62 (6): 2135–40. doi:10.2337/db12-1398.

Nikolich-Žugich, Janko, Mark K. Slifka, and Ilhem Messaoudi. 2004. "The Many Important Facets of T-Cell Repertoire Diversity." *Nature Reviews Immunology* 4 (2). Nature Publishing Group: 123–32. doi:10.1038/nri1292.

Nobusawa, E., and K. Sato. 2006. "Comparison of the Mutation Rates of Human Influenza A and B Viruses." *Journal of Virology* 80 (7). American Society for Microbiology: 3675–78. doi:10.1128/JVI.80.7.3675-3678.2006.

Norris, Philip J, Jennifer D Stone, Nadezhda Anikeeva, John W Heitman, Ingrid C Wilson, Dale F Hirschkorn, Margaret J Clark, et al. 2006. "Antagonism of HIV-Specific CD4+ T Cells by C-Terminal Truncation of a Minimum Epitope." *Molecular Immunology* 43 (9). NIH Public Access: 1349–57. doi:10.1016/j.molimm.2005.09.004.

Novak, E. J., A. W. Liu, J. A. Gebe, B. A. Falk, G. T. Nepom, D. M. Koelle, and W. W. Kwok. 2001. "Tetramer-Guided Epitope Mapping: Rapid Identification and Characterization of Immunodominant CD4+ T Cell Epitopes from Complex Antigens." *The Journal of Immunology* 166 (11). American Association of Immunologists: 6665–70. doi:10.4049/jimmunol.166.11.6665.

O'Brien, Cathal, Darren R Flower, and Conleth Feighery. 2008. "Peptide Length Significantly Influences in Vitro Affinity for MHC Class II Molecules." *Immunome Research* 4. BioMed Central: 6. doi:10.1186/1745-7580-4-6.

O'callaghan, C A, M F Byford, J R Wyer, B E Willcox, B K Jakobsen, A J McMichael, and J I Bell. 1999. "BirA Enzyme: Production and Application in the Study of Membrane Receptor-Ligand Interactions by Site-Specific Biotinylation." *Analytical Biochemistry* 266 (1): 9–15. doi:10.1006/abio.1998.2930.

Palmer, B. E., E. Boritz, and C. C. Wilson. 2004. "Effects of Sustained HIV-1 Plasma Viremia on HIV-1 Gag-Specific CD4+ T Cell Maturation and Function." *The Journal of Immunology* 172 (5). American Association of Immunologists: 3337–47. doi:10.4049/jimmunol.172.5.3337.

Pamer, Eric, and Peter Cresswell. 1998. "MECHANISMS OF MHC CLASS I–RESTRICTED ANTIGEN PROCESSING." *Annu. Rev. Immunol* 16: 323–58.

Pica, N, and P Palese. 2013. "Toward a Universal Influenza Virus Vaccine: Prospects and Challenges." Journal Article. *Annu Rev Med* 64: 189–202. doi:10.1146/annurev-med-120611-145115.

Pos, Wouter, Dhruv K Sethi, Melissa J Call, Monika-Sarah E D Schulze, Anne-Kathrin Anders, Jason Pyrdol, and Kai W Wucherpfennig. 2012. "Crystal Structure of the HLA-DM-HLA-DR1 Complex Defines Mechanisms for Rapid Peptide Selection." *Cell* 151 (7). Elsevier Inc.: 1557–68. doi:10.1016/j.cell.2012.11.025.

Price, D A, P J Goulder, P Klenerman, A K Sewell, P J Easterbrook, M Troop, C R Bangham, and R E Phillips. 1997. "Positive Selection of HIV-1 Cytotoxic T Lymphocyte Escape Variants during Primary Infection." *Proceedings of the National Academy of Sciences of the United States of America* 94 (5). National Academy of Sciences: 1890–95. http://www.ncbi.nlm.nih.gov/pubmed/9050875.

Provitera, P, A Goff, A Harenberg, F Bouamr, C Carter, and S Scarlata. 2001. "Role of the Major Homology Region in Assembly of HIV-1 Gag." *Biochemistry* 40 (18): 5565–72. http://www.ncbi.nlm.nih.gov/pubmed/11331022.

Pulendran, Bali, and Rafi Ahmed. 2011. "Immunological Mechanisms of Vaccination." *Nature Immunology* 12 (6). NIH Public Access: 509–17. http://www.ncbi.nlm.nih.gov/pubmed/21739679.

Quarsten, H, S N McAdam, T Jensen, H Arentz-Hansen, Molberg Ø, K E Lundin, and L M Sollid. 2001. "Staining of Celiac Disease-Relevant T Cells by Peptide-DQ2 Multimers." *Journal of Immunology (Baltimore, Md. : 1950)* 167 (9): 4861–68. http://www.ncbi.nlm.nih.gov/pubmed/11673490.

Radu, C G, B T Ober, L Colantonio, a Qadri, and E S Ward. 1998. "Expression and Characterization of Recombinant Soluble Peptide: I-A Complexes Associated with Murine Experimental Autoimmune Diseases." *Journal of Immunology (Baltimore, Md. : 1950)* 160 (12): 5915–21. http://www.ncbi.nlm.nih.gov/pubmed/9637504.

Riberdy, Janice M., John R. Newcomb, Michael J. Surman, James A. Barbosat, and Peter Cresswell. 1992. "HLA-DR Molecules from an Antigen-Processing Mutant Cell Line Are Associated with Invariant Chain Peptides." *Nature* 360 (6403). Nature Publishing Group: 474–77. doi:10.1038/360474a0.

Riese, Richard J, Paula R Wolf, Dieter Brömme, Lisa R Natkin, José A Villadangos, Hidde L Ploegh, Harold A Chapman, et al. 1996. "Essential Role for Cathepsin S in MHC Class II–Associated Invariant Chain Processing and Peptide Loading." *Immunity* 4 (4). Elsevier: 357–66. doi:10.1016/S1074-7613(00)80249-6.

Robinson, Harriet L. 2002. "VACCINES: NEW HOPE FOR AN AIDS VACCINE." *Nature Reviews Immunology* 2 (4). Nature Publishing Group: 239–50. doi:10.1038/nri776.

Roche, P A, P Cresswell, Paul A Roche, and Peter Cresswell2. 1990. "HLA-DR. Hemagglutinin-Derived Peptide to Purified High-Affinity Binding of an Influenza HIGH-AFFINITY BINDING OF AN INFLUENZA HEMAGGLUTININ-DERIVED PEPTIDE TO PURIFIED HLA-DR1." *J Immunol The Journal of Immunology at Cardiff Univ on September* 144 (5): 1849–56. http://www.jimmunol.org/content/144/5/1849.

Roche, Paul a., and Kazuyuki Furuta. 2015. "The Ins and Outs of MHC Class II-Mediated

Antigen Processing and Presentation." *Nature Reviews Immunology* 15 (4). Nature Research: 203–16. doi:10.1038/nri3818.

Rosenberg, E S, J M Billingsley, A M Caliendo, S L Boswell, P E Sax, S A Kalams, and B D Walker. 1997. "Vigorous HIV-1-Specific CD4+ T Cell Responses Associated with Control of Viremia." *Science (New York, N.Y.)* 278 (5342): 1447–50. http://www.ncbi.nlm.nih.gov/pubmed/9367954.

Rossjohn, Jamie, Stephanie Gras, John J. Miles, Stephen J. Turner, Dale I. Godfrey, and James McCluskey. 2015. "T Cell Antigen Receptor Recognition of Antigen-Presenting Molecules." *Annual Review of Immunology* 33 (1). Annual Reviews : 169–200. doi:10.1146/annurev-immunol-032414-112334.

Rothenberg, Ellen V., Jonathan E. Moore, and Mary A. Yui. 2008. "Launching the T-Cell-Lineage Developmental Programme." *Nature Reviews Immunology* 8 (1). Nature Publishing Group: 9–21. doi:10.1038/nri2232.

Rudolph, M G, R L Stanfield, and I A Wilson. 2006. "How TCRs Bind MHCs, Peptides, and Coreceptors." Journal Article. *Annu Rev Immunol.* http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt =Citation&list_uids=16411852.

Rudolph, Markus G, Robyn L Stanfield, and Ian a Wilson. 2006. "How TCRs Bind MHCs, Peptides, and Coreceptors." *Annual Review of Immunology* 24 (January): 419–66. doi:10.1146/annurev.immunol.23.021704.115658.

Saikh, Kamal U, John D Martin, A H Nishikawa, and Susan B Dillon. 1995. "Influenza A Virus-Specific H-2 D Restricted Cross-Reactive Cytotoxic T Lymphocyte Epitope(s) Detected in the Hemagglutinin HA2 Subunit of A/Udorn/72." *VIROLOGY* 214: 445–52.

Samji, Tasleem. 2009. "Influenza A: Understanding the Viral Life Cycle." Journal Article. *Yale J Biol Med* 82 (4): 153–59. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2794490&tool=pmcent rez&rendertype=abstract.

Sant'Angelo, Derek B., Eve Robinson, Charles A. Janeway, Jr., and Lisa K. Denzin. 2002. "Recognition of Core and Flanking Amino Acids of MHC Class II-Bound Peptides by the T Cell Receptor." *European Journal of Immunology* 32 (9). WILEY-VCH Verlag: 2510–20. doi:10.1002/1521-4141(200209)32:9<2510::AID-IMMU2510>3.0.CO;2-Q.

Schatz, David G, and Yanhong Ji. 2011. "Recombination Centres and the Orchestration of V(D)J Recombination." *Nature Reviews. Immunology* 11 (4): 251–63. doi:10.1038/nri2941.

Scott, C A, K C Garcia, F R Carbone, I A Wilson, and L Teyton. 1996. "Role of Chain Pairing for the Production of Functional Soluble IA Major Histocompatibility Complex Class II Molecules." *The Journal of Experimental Medicine* 183 (5): 2087–95. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2192579&tool=pmcent rez&rendertype=abstract.

Scott-Browne, James P, Frances Crawford, Mary H Young, John W Kappler, Philippa Marrack, and Laurent Gapin. 2011. "Evolutionarily Conserved Features Contribute to Aβ T Cell Receptor Specificity." *Immunity* 35 (4): 526–35. doi:10.1016/j.immuni.2011.09.005.

Scriba, Thomas J, Marco Purbhoo, Cheryl L Day, Nicola Robinson, Sarah Fidler, Julie Fox, Jonathan N Weber, Paul Klenerman, Andrew K Sewell, and Rodney E Phillips. 2005. "Ultrasensitive Detection and Phenotyping of CD4+ T Cells with Optimized HLA Class II Tetramer Staining." Journal Article. *J Immunol* 175 (10): 6334–43. doi:175/10/6334 [pii].

Selin, Liisa K, and Michael A Brehm. 2007. "Frontiers in Nephrology: Heterologous Immunity, T Cell Cross-Reactivity, and Alloreactivity." *Journal of the American Society of Nephrology : JASN* 18 (8). American Society of Nephrology: 2268–77. doi:10.1681/ASN.2007030295.

Sharma, Shalini, and Paul G Thomas. 2014. "The Two Faces of Heterologous Immunity: Protection or Immunopathology." *Journal of Leukocyte Biology* 95 (3). The Society for Leukocyte Biology: 405–16. doi:10.1189/jlb.0713386.

Shiina, Takashi, Kazuyoshi Hosomichi, Hidetoshi Inoko, and Jerzy K Kulski. 2009. "The HLA Genomic Loci Map: Expression, Interaction, Diversity and Disease." *Journal of Human Genetics* 54 (1). Nature Publishing Group: 15–39. doi:10.1038/jhg.2008.5.

Singer, Alfred, Stanley Adoro, and Jung-Hyun Park. 2008. "Lineage Fate and Intense Debate: Myths, Models and Mechanisms of CD4- versus CD8-Lineage Choice." *Nature Reviews Immunology* 8 (10). Nature Publishing Group: 788–801. doi:10.1038/nri2416.

Spits, Hergen. 2002. "Development of Alphabeta T Cells in the Human Thymus." *Nature Reviews. Immunology* 2 (10): 760–72. doi:10.1038/nri913.

Sridhar, Saranya, Shaima Begom, Katja Hoschler, Alison Bermingham, Walt Adamson, William Carman, Steven Riley, and Ajit Lalvani. 2015. "Longevity and Determinants of Protective Humoral Immunity after Pandemic Influenza Infection." *American Journal of Respiratory and Critical Care Medicine* 191 (3). American Thoracic Society: 325–32. doi:10.1164/rccm.201410-1798OC.

Staneková, Zuzana, Irena Adkins, Martina Kosová, Jana Janulíková, Peter Šebo, and Eva Varečková. 2013. "Heterosubtypic Protection against Influenza A Induced by Adenylate Cyclase Toxoids Delivering Conserved HA2 Subunit of Hemagglutinin." *Antiviral Research* 97 (1): 24–35. doi:10.1016/j.antiviral.2012.09.008.

Stepniak, Dariusz, L. Willemijn Vader, Yvonne Kooy, Peter A. van Veelen, Antonis Moustakas, Nikolaos A. Papandreou, Elias Eliopoulos, Jan Wouter Drijfhout, George K. Papadopoulos, and Frits Koning. 2005. "T-Cell Recognition of HLA-DQ2-Bound Gluten Peptides Can Be Influenced by an N-Terminal Proline at P-1." *Immunogenetics* 57 (1–2). Springer-Verlag: 8–15. doi:10.1007/s00251-005-0780-8.

Stern, Lawrence J., and Don C. Wiley. 1992. "The Human Class II MHC Protein HLA-DR1 Assembles as Empty Aβ Heterodimers in the Absence of Antigenic Peptide." *Cell* 68 (3): 465–77. doi:10.1016/0092-8674(92)90184-E.

Strug, Iwona, J Mauricio Calvo-Calle, Karin M Green, John Cruz, Francis A Ennis, James E Evans, and Lawrence J Stern. 2008. "Vaccinia Peptides Eluted from HLA-DR1 Isolated from Virus-Infected Cells Are Recognized by CD4+ T Cells from a Vaccinated Donor." *Journal of Proteome Research* 7 (7). NIH Public Access: 2703–11. doi:10.1021/pr700780x.

Suzuki, Yoshiyuki, and Masatoshi Nei. 2002. "Origin and Evolution of Influenza Virus Hemagglutinin Genes." *Molecular Biology and Evolution* 19 (4). Oxford University

Press: 501–9. http://www.ncbi.nlm.nih.gov/pubmed/11919291.

Sylvester-Hvid, C., N. Kristensen, T. Blicher, H. Ferre, S.L. Lauemoller, X.A. Wolf, K. Lamberth, M.H. Nissen, L.O. Pedersen, and S. Buus. 2002. "Establishment of a Quantitative ELISA Capable of Determining Peptide - MHC Class I Interaction." *Tissue Antigens* 59 (4). Blackwell Science, Ltd: 251–58. doi:10.1034/j.1399-0039.2002.590402.x.

Szabo, Susanne J., Brandon M. Sullivan, Stanford L. Peng, and Laurie H. Glimcher. 2003. "M OLECULAR M ECHANISMS R EGULATING T H 1 I MMUNE R ESPONSES." *Annual Review of Immunology* 21 (1). Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA : 713–58. doi:10.1146/annurev.immunol.21.120601.140942.

Tessier, Daniel C., David Y. Thomas, Henry E. Khouri, France Laliberié, and Thierry Vernet. 1991. "Enhanced Secretion from Insect Cells of a Foreign Protein Fused to the Honeybee Melittin Signal Peptide." *Gene* 98 (2): 177–83. doi:10.1016/0378-1119(91)90171-7.

Thursz, Mark, Rhiannon Yallop, Robert Goldin, Christian Trepo, and Howard C Thomas. 1999. "Influence of MHC Class II Genotype on Outcome of Infection with Hepatitis C Virus." *The Lancet* 354 (9196): 2119–24. doi:10.1016/S0140-6736(99)91443-5.

Tungatt, Katie, Valentina Bianchi, Michael D Crowther, Wendy E Powell, Andrea J Schauenburg, Andrew Trimby, Marco Donia, et al. 2015. "Antibody Stabilization of Peptide-MHC Multimers Reveals Functional T Cells Bearing Extremely Low-Affinity TCRs." *Journal of Immunology (Baltimore, Md. : 1950)* 194 (1): 463–74. doi:10.4049/jimmunol.1401785.

Turk, Vito, Veronika Stoka, Olga Vasiljeva, Miha Renko, Tao Sun, Boris Turk, and Dušan Turk. 2012. "Cysteine Cathepsins: From Structure, Function and Regulation to New Frontiers." *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1824 (1): 68–88. doi:10.1016/j.bbapap.2011.10.002.

Valkenburg, Sophie A, Olive T W Li, Polly W Y Mak, Chris K P Mok, John M Nicholls, Yi Guan, Thomas A Waldmann, J S Malik Peiris, Liyanage P Perera, and Leo L M Poon. 2014. "IL-15 Adjuvanted Multivalent Vaccinia-Based Universal Influenza Vaccine Requires CD4+ T Cells for Heterosubtypic Protection." *Proceedings of the National Academy of Sciences of the United States of America* 111 (15). National Academy of Sciences: 5676–81. doi:10.1073/pnas.1403684111.

Verma, R, E Boleti, and A J George. 1998. "Antibody Engineering: Comparison of Bacterial, Yeast, Insect and Mammalian Expression Systems." Journal Article. *J Immunol Methods* 216 (1–2): 165–81. doi:S0022-1759(98)00077-5 [pii].

Vita, Randi, James A Overton, Jason A Greenbaum, Julia Ponomarenko, Jason D Clark, Jason R Cantrell, Daniel K Wheeler, et al. 2015. "The Immune Epitope Database (IEDB) 3.0." *Nucleic Acids Research* 43 (Database issue). Oxford University Press: D405-12. doi:10.1093/nar/gku938.

Walker, Bruce D., Eric S. Rosenberg, Marcus Altfeld, Samuel H. Poon, Mary N. Phillips, Barbara M. Wilkes, Robert L. Eldridge, Gregory K. Robbins, Richard T. D'Aquila, and Philip J. R. Goulder. 2000. "Immune Control of HIV-1 after Early Treatment of Acute Infection." *Nature* 407 (6803). Nature Publishing Group: 523–26. doi:10.1038/35035103.

Wang, Xin Xiang, Yili Li, Yiyuan Yin, Min Mo, Qian Wang, Wei Gao, Lili Wang, and

Roy a Mariuzza. 2011. "Affinity Maturation of Human CD4 by Yeast Surface Display and Crystal Structure of a CD4-HLA-DR1 Complex." Journal Article. *Proc Natl Acad Sci U S A* 108 (38): 15960–65. doi:1109438108 [pii]10.1073/pnas.1109438108.

Waugh, David S. 2011. "An Overview of Enzymatic Reagents for the Removal of Affinity Tags." *Protein Expression and Purification* 80 (2): 283–93. doi:10.1016/j.pep.2011.08.005.

Weenink, Sarah M, Peter J Milburn, and Anand M Gautam. 1997. "A Continuous Central Motif of Invariant Chain Peptides, CLIP, Is Essential for Binding to Various I-A MHC Class II Molecules." *International Immunology* 9 (2): 317–25.

"WHO | HIV/AIDS." 2016. *WHO*. World Health Organization.

"WHO | Influenza." 2004. *WHO*. World Health Organization.

Wiendl, Heinz, Alfred Lautwein, Meike Mitsdörffer, Sabine Krause, Stella Erfurth, Wolfgang Wienhold, Matthias Morgalla, et al. 2003. "Antigen Processing and Presentation in Human Muscle: Cathepsin S Is Critical for MHC Class II Expression and Upregulated in Inflammatory Myopathies." *Journal of Neuroimmunology* 138 (1): 132–43. doi:10.1016/S0165-5728(03)00093-6.

Willcox, B E, G F Gao, J R Wyer, C A O'Callaghan, J M Boulter, E Y Jones, P A van der Merwe, J I Bell, and B K Jakobsen. 1999. "Production of Soluble Alphabeta T-Cell Receptor Heterodimers Suitable for Biophysical Analysis of Ligand Binding." Journal Article. *Protein Sci* 8 (11): 2418–23. doi:10.1110/ps.8.11.2418.

Wilson, C C, B Palmer, S Southwood, J Sidney, Y Higashimoto, E Appella, R Chesnut, A Sette, and B D Livingston. 2001. "Identification and Antigenicity of Broadly Cross-Reactive and Conserved Human Immunodeficiency Virus Type 1-Derived Helper T-Lymphocyte Epitopes." *Journal of Virology* 75 (9). American Society for Microbiology (ASM): 4195–4207. doi:10.1128/JVI.75.9.4195-4207.2001.

Wilson, J D, N Imami, A Watkins, J Gill, P Hay, B Gazzard, M Westby, and F M Gotch. 2000. "Loss of CD4+ T Cell Proliferative Ability but Not Loss of Human Immunodeficiency Virus Type 1 Specificity Equates with Progression to Disease." *The Journal of Infectious Diseases* 182 (3). Oxford University Press: 792–98. doi:10.1086/315764.

Woody, J N, J R Lamb, D D Eckels, P Lake, A H Johnson, R J Hartzman, Jonathan R Lamb, et al. 1982. "ANTIGEN-SPECIFIC HUMAN T LYMPHOCYTE CLONES: INDUCTION, ANTIGEN SPECIFICITY, AND MHC RESTRICTION OF INFLUENZA VIRUS-IMMUNE CLONES'." *J Immunol The Journal of Immunology at Cardiff Univ on September THE JOURNAL OF IMMUNOLOGY CoDyright* 128 (1): 233–38. http://www.jimmunol.org/content/128/1/233.

Yang, J, E James, T J Gates, J H DeLong, R E LaFond, U Malhotra, and W W Kwok. 2013. "CD4+ T Cells Recognize Unique and Conserved 2009 H1N1 Influenza Hemagglutinin Epitopes after Natural Infection and Vaccination." Journal Article. *Int Immunol* 25 (8): 447–57. doi:dxt005 [pii]10.1093/intimm/dxt005.

Yin, L, E Huseby, J Scott-Browne, K Rubtsova, C Pinilla, F Crawford, P Marrack, S Dai, and J W Kappler. 2011. "A Single T Cell Receptor Bound to Major Histocompatibility Complex Class I and Class II Glycoproteins Reveals Switchable TCR Conformers." Journal Article. *Immunity* 35 (1): 23–33. doi:S1074-7613(11)00180-4 [pii] 10.1016/j.immuni.2011.04.017.

Zavala-Ruiz, Zarixia, Iwona Strug, Bruce D Walker, Philip J Phillip J Norris, and Lawrence J Stern. 2004. "A Hairpin Turn in a Class II MHC-Bound Peptide Orients Residues Outside the Binding Groove for T Cell Recognition." *Proceedings of the National Academy of Sciences of the United States of America* 101 (36): 13279–84. doi:10.1073/pnas.0403371101.

Zhang, T, Y Maekawa, J Hanba, T Dainichi, B F Nashed, H Hisaeda, T Sakai, et al. 2000. "Lysosomal Cathepsin B Plays an Important Role in Antigen Processing, While Cathepsin D Is Involved in Degradation of the Invariant Chain Inovalbumin-Immunized Mice." *Immunology* 100 (1). Wiley-Blackwell: 13–20. http://www.ncbi.nlm.nih.gov/pubmed/10809954.

Zhang, Xin, Sharon Ing, Austin Fraser, Minzi Chen, Omar Khan, Jerald Zakem, William Davis, and Robert Quinet. 2013. "Follicular Helper T Cells: New Insights into Mechanisms of Autoimmune Diseases." *The Ochsner Journal* 13 (1). Ochsner Clinic, L.L.C. and Alton Ochsner Medical Foundation: 131–39. http://www.ncbi.nlm.nih.gov/pubmed/23531878.

Zhang, You H., Giora Enden, and José C. Merchuk. 2005. "Insect cells–Baculovirus System: Factors Affecting Growth and Low MOI Infection." *Biochemical Engineering Journal* 27 (1): 8–16. doi:10.1016/j.bej.2005.05.013.

Zheng, Jian, Yinping Liu, Yu-Lung Lau, and Wenwei Tu. 2013. "Γδ-T Cells: An Unpolished Sword in Human Anti-Infection Immunity." *Cellular & Molecular Immunology* 10 (1). Nature Publishing Group: 50–57. doi:10.1038/cmi.2012.43.

Zheng, Z L, S M Colón, F C Gaeta, J J Sidney I Krieger, R W Karr, H M Grey, W Y Yu, et al. 1991. "SINGLE AMINO ACID CHANGES IN DR AND ANTIGEN DEFINE RESIDUES CRITICAL FOR PEPTIDE-MHC BINDING AND T CELL RECOGNITION1." *J Immunol The Journal of Immunology at Cardiff Univ on September THE JOURNAL OF IMMUNOLOGY* 146 (7): 2331–40. http://www.jimmunol.org/content/146/7/2331.

Zhu, Jinfang, and William E Paul. 2010. "Peripheral CD4+ T-Cell Differentiation Regulated by Networks of Cytokines and Transcription Factors." *Immunological Reviews* 238 (1). NIH Public Access: 247–62. doi:10.1111/j.1600-065X.2010.00951.x.

Zhu, Jinfang, Hidehiro Yamane, and William E. Paul. 2010. "Differentiation of Effector CD4 T Cell Populations (*)." Journal Article. *Annu Rev Immunol* 28 (1). Annual Reviews : 445–89. doi:10.1146/annurev-immunol-030409-101212.

Supplementary Figures and Tables

```
  1 M Y I Y A D P S P A M I K E E H V I I Q
  1 ATGTACATCTACGCTGATCCCAGCCCCGCCATGATCAAGGAGGAGCACGTCATCATCCAG

 21 A E F Y L N P D Q S G E F M F D F D G D
 61 GCCGAGTTCTATCTCAACCCCGACCAATCCGGCGAGTTTATGTTCGACTTCGACGGCGAT

 41 E I F H V D M A K K E T V W R L E E F G
121 GAGATCTTCCACGTGGACATGGCTAAGAAGGAGACCGTCTGGCGCCTCGAAGAGTTCGGC

 61 R F A S F E A Q G A L A N I A V D K A N
181 CGCTTCGCCAGCTTTGAGGCTCAGGGCGCTCTGGCTAACATCGCCGTCGATAAGGCCAAC

 81 L E I M T K R S N Y T P I T N V P P E V
241 CTGGAGATCATGACCAAGCGCTCCAACTACACCCCCATCACCAACGTGCCCCCCGAGGTC

101 T V L T N S P V E L R E P N V L I C F I
301 ACCGTGCTGACCAACTCCCCTGTGGAGCTGCGCGAACCCAACGTCCTGATCTGCTTCATC

121 D K F T P P V V N V T W L R N G K P V T
361 GACAAGTTCACTCCCCCCGTGGTCAACGTGACTTGGCTGCGCAACGGCAAACCCGTGACC

141 T G V S E T V F L P R E D H L F R K F H
421 ACCGGTGTCTCCGAGACCGTGTTCCTGCCCCGCGAGGATCACCTCTTCCGTAAGTTCCAC

161 Y L P F L P S T E D V Y D C R V E H W G
481 TACCTCCCCTTCCTGCCCTCCACCGAGGACGTCTACGACTGTCGCGTCGAACACTGGGGC

181 L D E P L L K H W E F D A E S A Q S K V
541 CTCGACGAGCCTCTGCTGAAGCACTGGGAGTTCGACGCTGAGAGCGCCCAATCCAAGGTC

201 D G G G G L T D T L Q A E T D Q L E D
601 GACGGTGGCGGTGGCGGTCTGACTGACACCCTCCAGGCCGAGACCGATCAGCTGGAAGAC

221 K K S A L Q T E I A N L L K E K E K L E
661 AAAAAGTCCGCCCTCCAGACTGAGATCGCCAACCTCCTCAAGGAGAAGGAGAAACTGGAG

241 F I L A A Y G G S G G S G L N D I F E A
721 TTCATCCTGGCTGCTTATGGCGGTTCCGGTGGCAGCGGTCTGAACGACATCTTCGAGGCC

261 Q K I E W H E * *
781 CAGAAGATCGAATGGCACGAGTAATAA
```

Supplementary figure S1: Sequence of HLA-DR1αtag construct. Mellitin tag sequence is highlighted in blue, HLA-DR1α chain in grey, connecting peptide in yellow, leucine zipper in purple, linker peptide in red and biotinylation sequence in pink.

```
  1  M   Y   I   Y   A   D   P   S   P   A   M   P   V   S   K   M   R   M   A   T
  1  ATGTACATCTACGCTGATCCCTCCCCCGCTATGCCCGTGTCCAAGATGCGTATGGCTACC

 21  P   L   L   G   G   S   G   G   S   L   V   P   R   G   S   G   G   S   G   G
 61  CCCCTGCTGGGCGGTTCCGGTGGTTCTCTGGTGCCTCGTGGTTCTGGTGGTTCCGGCGGT

 41  S   G   D   T   R   P   R   F   L   W   Q   L   K   F   E   C   H   F   F   N
121  TCAGGCGACACCCGTCCCCGTTTCTTGTGGCAGCTGAAGTTCGAGTGCCACTTCTTCAAC

 61  G   T   E   R   V   R   L   L   E   R   C   I   Y   N   Q   E   E   S   V   R
181  GGCACCGAGCGTGTGCGTCTGCTCGAGCGTTGCATCTACAACCAGGAAGAGTCCGTCCGT

 81  F   D   S   D   V   G   E   Y   R   A   V   T   E   L   G   R   P   D   A   E
241  TTCGACTCCGACGTGGGCGAGTACCGTGCTGTGACCGAGCTGGGTCGTCCCGACGCTGAG

101  Y   W   N   S   Q   K   D   L   L   E   Q   R   R   A   A   V   D   T   Y   C
301  TACTGGAACTCCCAGAAGGACTTGCTCGAACAGCGTCGTGCTGCTGTGGACACCTACTGC

121  R   H   N   Y   G   V   G   E   S   F   T   V   Q   R   R   V   E   P   K   V
361  CGTCACAACTACGGTGTCGGCGAGTCCTTCACCGTGCAGCGTCGCGTCGAGCCCAAGGTC

141  T   V   Y   P   S   K   T   Q   P   L   Q   H   H   N   L   L   V   C   S   V
421  ACCGTGTACCCCTCCAAGACCCAGCCCCTGCAGCACCACAACCTGCTCGTGTGCTCCGTG

161  S   G   F   Y   P   G   S   I   E   V   R   W   F   R   N   G   Q   E   E   K
481  TCCGGTTTCTACCCCGGTTCCATCGAAGTGCGTTGGTTCCGTAACGGTCAAGAAGAGAAG

181  A   G   V   V   S   T   G   L   I   Q   N   G   D   W   T   F   Q   T   L   V
541  GCTGGCGTCGTGTCCACCGGCCTGATCCAGAACGGCGACTGGACCTTCCAGACCCTGGTC

201  M   L   E   T   V   P   R   S   G   E   V   Y   T   C   Q   V   E   H   P   S
601  ATGCTGGAAACCGTGCCCCGTTCCGGCGAGGTGTACACTTGCCAGGTCGAGCACCCCTCC

221  V   T   S   P   L   T   V   E   W   R   A   E   S   A   Q   S   K   V   D   G
661  GTGACCTCCCCCCTGACCGTGGAATGGCGTGCTGAGTCCGCTCAGTCCAAGGTGGACGGT

241  G   G   G   G   R   I   A   R   L   E   E   K   V   K   T   L   K   A   Q   N
721  GGTGGTGGCGGTCGTATCGCTCGTCTGGAAGAGAAGGTCAAGACCCTGAAGGCTCAGAAC

261  S   E   L   A   S   T   A   N   M   L   R   E   Q   V   A   Q   L   K   Q   K
781  TCCGAGCTGGCTTCCACCGCTAACATGCTGCGCGAGCAGGTCGCCCAGCTGAAGCAAAAG

281  V   M   N   Y   *   *
841  GTCATGAACTACTAATAA
```

Supplementary figure S2: Sequence of HLA-DR1β construct. Mellitin tag sequence is highlighted in blue, CLIP in green, Thrombin cleavage site in light blue, linker peptides in red, HLA-DR1β chain in grey, connecting peptide in yellow and leucine zipper in purple.

T cell line	Total number of cells sorted	Number of colonies sequenced
no peptide	153	n/a
core	260	76
core + 1C	857	118
core + 3C	358	525
core + 2N	107	74
core + 2N2C	1272	124
core +3N3C	137	63

Supplementary Table S1: Total number of cell sorted for each short term T cell line.
Numbers correspond to cells collected from all PBMCs in each cell line.

Refinement statistics:

Resolution (Å)	1.89
No reflections used	35543
No reflections in R_{free} set	1775
R_{crys} (no cut-off) (%)	0.20
R_{free}	0.26

Supplementary Table S2: X-ray data diffraction data acquisition and refinement statistics of HLA-DR1[gag24].

Peptide residues			MHC residues				Contacts	
Position	Aa code	Atom	Chain	Position	Aa code	Atom	Distance	Type
-3	ASP	CB	B	81	HIS	CE1	3.86	VdW
		O				CE1	3.64	VdW
		O				NE2	3.92	HB
		O	B	85	VAL	CG2	3.64	VdW
-2	ARG	NH2	A	49	GLY	C	3.88	VdW
		NH2				O	2.98	VdW
		CZ	A	50	ARG	O	3.89	VdW
		NH1				C	3.86	VdW
		NH1				O	3.21	VdW
		NH2				C	3.8	VdW
		NH2				O	3.81	HB
		CB	A	51	PHE	O	3.71	VdW
		CD				O	3.91	VdW
		NE				O	3.68	HB
		CZ				C	3.71	VdW
		CZ				O	3.71	VdW
		NH1				C	3.94	VdW
		NH1				O	3.88	HB
		NH2				C	3.87	VdW
		NE	A	52	ALA	O	3.87	HB
		NE				C	3.68	VdW
		NE				CA	3.92	VdW
		CZ				O	3.71	VdW
		CZ				C	3.89	VdW
		CZ				N	3.87	VdW
		NH2				O	2.83	VdW
		NH2				C	3.41	VdW
		NH2				N	3.64	HB
		NH2				CA	3.93	VdW
		O				CB	3.7	VdW
		O				C	3.77	VdW
		O				CA	3.49	VdW
		CD	A	53	SER	OG	3.75	VdW
		NE				CB	3.84	VdW
		NE				OG	2.66	VdW
		NE				N	3.89	HB
		CZ				OG	3.24	VdW
		NH2				OG	3.04	VdW
		O				N	3.04	VdW
		O				O	3.64	HB
		C	B	85	VAL	CG1	3.97	VdW
		O				CG1	3.43	VdW
-1	PHE	CA	A	53	SER	O	3.35	VdW
		CE1				CB	3.86	VdW

		CZ				CB	3.66	VdW
		CE2				CB	3.67	VdW
		CD2				CB	3.88	VdW
		CD2				C	3.94	VdW
		C				O	3.69	VdW
		CD2	A	54	PHE	CA	3.94	VdW
		CD2				N	3.8	VdW
		CE2	A	55	GLU	CD	3.78	VdW
		CE2				OE1	3.46	VdW
		CE2				CG	3.93	VdW
		CD2				OE1	3.93	VdW
		CD2				CG	3.8	VdW
		O	B	81	HIS	CE1	3.87	VdW
		O				NE2	2.83	VdW
		O				CD2	3.66	VdW
		O	B	85	VAL	CG2	3.68	VdW
1	TYR	OH	A	31	ILE	CD1	3.78	VdW
		OH				CG2	3.97	VdW
		CG	A	32	PHE	CE1	3.97	VdW
		CD2				CE1	3.89	VdW
		CE2	A	52	ALA	CB	3.71	VdW
		N	A	53	SER	O	3.01	VdW
		CD2				O	3.93	VdW
		CB	A	54	PHE	CD1	3.77	VdW
		O				CE1	3.85	VdW
		O				CD1	3.55	VdW
		CA	B	82	ASN	OD1	3.35	VdW
		CD1				CG	3.68	VdW
		CD1				OD1	3.68	VdW
		CE1				O	3.8	VdW
		C				OD1	3.55	VdW
		CZ	B	85	VAL	CG1	3.86	VdW
		CE2				CG1	3.66	VdW
		CD2				CG1	3.83	VdW
		OH	B	86	GLY	N	3.73	HB
		OH				CA	3.44	VdW
		OH	B	89	PHE	CE1	3.99	VdW
		OH				CZ	3.81	VdW
2	LYS	O	A	24	PHE	CZ	3.56	VdW
		O	B	78	TYR	CD1	3.42	VdW
		O				CE1	3.09	VdW
		CG	B	81	HIS	NE2	3.68	VdW
		CG				CD2	3.5	VdW
		CE				CE1	3.73	VdW
		CE				NE2	3.58	VdW
		N	B	82	ASN	CG	3.66	VdW

						OD1	2.82	VdW
		N				ND2	3.72	HB
		CA				OD1	3.86	VdW
		CB				OD1	3.99	VdW
		O				CG	3.9	VdW
		O				OD1	3.93	HB
		O				ND2	3.03	VdW
3	THR	CA	A	9	GLN	OE1	3.69	VdW
		CG2				NE2	3.72	VdW
		C				OE1	3.76	VdW
		OG1	A	22	PHE	CZ	3.95	VdW
		OG1	A	54	PHE	CE1	3.18	VdW
		OG1				CZ	3.15	VdW
		CG2	A	62	ASN	ND2	3.69	VdW
4	LEU	N	A	9	GLN	NE2	3.87	HB
		N				OE1	2.96	VdW
		N				CD	3.82	VdW
		CA				OE1	3.89	VdW
		CB				OE1	3.81	VdW
		O				NE2	3.23	VdW
		O	A	62	ASN	CG	3.97	VdW
		O				ND2	2.99	VdW
		CB	B	13	PHE	CD1	3.97	VdW
		CD2				CD1	3.49	VdW
		CD2				CE1	3.96	VdW
		O				CZ	3.82	VdW
		O				CE2	3.66	VdW
		CD1	B	70	GLN	NE2	3.56	VdW
		CG	B	71	ARG	NH1	3.89	VdW
		N	B	78	TYR	CD1	3.83	VdW
		N				CE1	3.41	VdW
		CB				CE1	3.87	VdW
		CB				CZ	3.96	VdW
		CD1				CG	3.94	VdW
		CD2				CE2	3.62	VdW
		CD2				CD2	3.61	VdW
5	ARG	CA	A	62	ASN	OD1	3.59	VdW
		C				OD1	3.81	VdW
		C	B	71	ARG	NH2	3.79	VdW
		O				CZ	3.5	VdW
		O				NH1	3.58	HB
		O				NH2	2.57	VdW
6	ALA	N	A	62	ASN	CG	3.55	VdW
		N				OD1	3.05	VdW
		N				ND2	3.53	HB
		CA				OD1	3.99	VdW

		CB				CG	3.73	VdW
		CB				OD1	3.69	VdW
		CB				ND2	3.9	VdW
		CB	A	65	VAL	CB	3.97	VdW
		CB				CG1	3.56	VdW
		C				CG1	3.93	VdW
		O				CG1	3.67	VdW
		CB	A	66	ASP	OD1	3.71	VdW
		CA	B	11	LEU	CD2	3.92	VdW
		CA				CD2	3.92	VdW
		CB				CD2	3.55	VdW
		CB				CD2	3.55	VdW
7	GLU	C	A	65	VAL	CG1	3.87	VdW
		O				CG1	3.75	VdW
		C	A	69	ASN	ND2	3.86	VdW
		O				CG	3.8	VdW
		O				OD1	3.98	HB
		O				ND2	2.77	VdW
		CB	B	61	TRP	NE1	3.63	VdW
		CB				CE2	3.88	VdW
		O				CZ2	3.73	VdW
		CG	B	67	LEU	CD1	3.4	VdW
		CG				CD2	3.61	VdW
		CD				CD1	3.79	VdW
		OE2				CD1	3.89	VdW
8	GLN	CA	A	69	ASN	OD1	3.79	VdW
		C				OD1	3.8	VdW
		O	B	60	TYR	CE2	3.66	VdW
		O				CD2	3.97	VdW
		C	B	61	TRP	NE1	3.8	VdW
		O				CD1	3.82	VdW
		O				NE1	2.8	VdW
		O				CE2	3.68	VdW
		O				CZ2	3.89	VdW
9	ALA	N	A	69	ASN	CG	3.57	VdW
		N				OD1	2.87	VdW
		N				ND2	3.72	HB
		CA				OD1	3.66	VdW
		CB				CG	3.66	VdW
		CB				OD1	3.28	VdW
		CB	A	73	MET	CE	3.74	VdW
		CA	B	57	ASP	OD1	3.69	VdW
		CB				OD1	3.49	VdW
		C				OD1	3.86	VdW
		CA	B	61	TRP	CZ2	3.96	VdW
10	SER	O	A	72	ILE	CG2	3.54	VdW

		C	A	76	ARG	NH2	3.98	VdW
		O				NH2	2.93	VdW
		CB	B	56	PRO	O	3.55	VdW
		N	B	57	ASP	OD1	3.06	VdW
		CA				OD1	3.92	VdW
		CB				OD1	3.66	VdW
		N	B	60	TYR	CD2	3.79	VdW
		CA				CD2	3.84	VdW
		CB				CB	3.9	VdW
11	GLN	CB	A	72	ILE	CG1	3.86	VdW
		CB				CG2	3.79	VdW
		CB				CD1	3.79	VdW
		NE2				CG2	3.66	VdW
		O				CD1	3.73	VdW
		NE2	A	76	ARG	NE	3.34	VdW
		NE2				CZ	3.92	VdW
		NE2				NH2	3.56	HB

Supplementary Table S3: List of all peptide-MHC contacts made within HLA-DR1gag24. Chain A: HLA-DR1α-chain; chain B: HLA-DR1β-chain; VdW: Van der Waals; HB: Hydrogen bond.

Peptide residues			Peptide residues			Contacts	
Position	Aa code	Atom	Position	Aa code	Atom	Distance	Type
-3	ASP	N	-2	ARG	N	3.51	VdW
		CA			N	2.47	HB
					CA	3.88	VdW
		CB			N	3.55	VdW
					N	3.63	VdW
					C	3.72	VdW
					N	3.03	HB
					CA	3.89	VdW
		C			C	3.42	VdW
					N	1.34	PB
					CA	2.48	VdW
					CB	3.66	VdW
		O			C	3.59	VdW
					N	2.26	HB
					CA	2.83	VdW
			-1	PHE	N	3.92	VdW
		OD			CG	3.89	VdW
					CD	3.66	VdW
					N	2.7	HB
					CA	3.39	VdW
					CB	3.21	VdW
					C	3.98	VdW
					O	3.64	VdW
					N	3.44	VdW
					N	3.68	VdW
		CG	2	LYS	NZ	3.57	VdW
					NZ	3.6	VdW
		OD			CE	3.94	VdW
					NZ	3.13	HB
-2	ARG	N	-3	ASP	CG	3.63	VdW
					OD	3.03	HB
					N	3.51	VdW
					CA	2.47	VdW
					CB	3.55	VdW
					C	1.34	PB
					O	2.26	HB
		CA			OD	3.89	VdW
					CA	3.88	VdW
					C	2.48	VdW
					O	2.83	VdW
					C	3.66	VdW
		C			OD	3.72	VdW
					C	3.42	VdW
					O	3.59	VdW

			-1	PHE	N	2.86	VdW
					N	2.5	VdW
					CA	3.88	VdW
	CB				N	3.63	VdW
	CG				N	3.83	VdW
					CD	3.99	VdW
					N	1.35	PB
					CA	2.47	VdW
					CB	3.71	VdW
					C	3.24	VdW
					O	3.75	VdW
	O				N	2.26	HB
					CA	2.83	VdW
					C	3.44	VdW
			1	TYR	N	3.97	VdW
					N	3.76	VdW
-1	PHE		3	ASP	CG	3.92	VdW
					OD	2.7	HB
					C	3.44	VdW
					O	3.68	VdW
	CA				OD	3.39	VdW
					OD	3.21	VdW
	CG				OD	3.89	VdW
	CD				OD	3.66	VdW
					OD	3.98	VdW
					OD	3.64	VdW
	N		-2	ARG	CG	3.83	VdW
					C	1.35	PB
					O	2.26	HB
					N	2.86	VdW
					CA	2.5	VdW
					CB	3.63	VdW
					C	2.47	VdW
					O	2.83	VdW
					CA	3.88	VdW
					C	3.71	VdW
					C	3.99	VdW
					C	3.24	VdW
					O	3.44	VdW
					C	3.75	VdW
			1	TYR	N	3.47	VdW
					N	2.4	VdW
					CA	3.77	VdW
	CB				O	3.91	VdW
					N	3.37	VdW
	C				O	3.14	VdW

					N	1.33	PB
					CA	2.43	VdW
					CB	3.74	VdW
					C	2.91	VdW
		O			O	3.46	VdW
					N	2.25	HB
					CA	2.76	VdW
					C	2.98	VdW
			2	LYS	NZ	3.89	VdW
					N	3.82	VdW
					CE	3.8	VdW
					CD	3.78	VdW
					N	3.48	VdW
1	TYR	N	-2	ARG	C	3.97	VdW
					O	3.76	VdW
			-1	PHE	N	3.47	VdW
					CA	2.4	VdW
					CB	3.37	VdW
					C	1.33	PB
					O	2.25	HB
		CA			CA	3.77	VdW
					C	2.43	VdW
					O	2.76	VdW
		CB			C	3.74	VdW
		C			C	2.91	VdW
					O	2.98	VdW
		O			CB	3.91	VdW
					C	3.14	VdW
					O	3.46	VdW
					C	3.82	VdW
					O	3.48	VdW
		N	1	TYR	O	2.24	HB
					N	3.51	VdW
					CA	2.41	VdW
					CB	3.32	VdW
					C	1.32	PB
			2	LYS	N	3.51	VdW
					CA	3.81	VdW
					N	2.41	VdW
					N	3.32	VdW
					CB	3.63	VdW
					CA	2.44	VdW
					C	3.36	VdW
					O	3.77	VdW
					CG	3.93	VdW
					CD	3.86	VdW

					Atom	Dist	Type
					N	1.32	VdW
					CA	2.79	VdW
					C	3.66	VdW
					CD	3.89	VdW
					N	2.24	HB
			3	THR	N	3.59	VdW
2	LYS	CE	-3	ASP	OD	3.94	VdW
		NZ			CG	3.57	VdW
					OD	3.6	VdW
					OD	3.13	HB
			-1	PHE	O	3.78	VdW
					O	3.8	VdW
					CB	3.89	VdW
		CA	1	TYR	O	2.79	VdW
					CA	3.81	VdW
					C	2.44	VdW
		CB			C	3.63	VdW
		CG			C	3.93	VdW
		CD			O	3.89	VdW
					C	3.86	VdW
		C			O	3.66	VdW
					C	3.36	VdW
		O			C	3.77	VdW
			3	THR	N	2.43	VdW
					CA	3.84	VdW
					N	3.25	VdW
					C	3.28	VdW
					O	3.82	VdW
					N	1.34	PB
					CA	2.49	VdW
					CB	3.67	VdW
					OG	3.88	VdW
					C	3.56	VdW
					N	2.27	HB
					CA	2.88	VdW
			4	LEU	N	3.99	VdW
					N	3.89	VdW
3	THR	N	2	LYS	CB	3.25	VdW
					CA	2.43	VdW
					C	1.34	PB
					O	2.27	HB
					N	3.59	VdW
					CA	3.84	VdW
					C	2.49	VdW
					O	2.88	VdW
		CB			C	3.67	VdW

		Atom			Atom	Dist	Type
		OG			C	3.88	VdW
					C	3.28	VdW
					O	3.56	VdW
					C	3.82	VdW
			4	LEU	N	3.42	VdW
		CA			CA	3.78	VdW
					N	2.45	VdW
					N	3.48	VdW
		CG			C	3.89	VdW
					O	3.43	VdW
					N	3.54	VdW
		C			C	3.02	VdW
					O	3.45	VdW
					CA	2.41	VdW
					N	1.33	PB
					CB	3.66	VdW
		O			C	3.1	VdW
					O	3.79	VdW
					CA	2.69	VdW
					N	2.21	HB
			5	ARG	N	3.76	VdW
					N	3.4	HB
4	LEU	N	2	LYS	C	3.99	VdW
					O	3.89	VdW
			3	THR	CG	3.54	VdW
					C	1.33	PB
					O	2.21	HB
					N	3.42	VdW
					CA	2.45	VdW
					CB	3.48	VdW
					C	2.41	VdW
					O	2.69	HB
					CA	3.78	VdW
					C	3.66	VdW
					CG	3.89	VdW
					C	3.02	VdW
					O	3.1	VdW
					CG	3.43	VdW
					C	3.45	VdW
					O	3.79	VdW
		N	5	ARG	N	3.5	VdW
		CA			N	2.46	VdW
					CA	3.84	VdW
		CB			N	3.43	VdW
		CG			N	3.98	VdW
		C			N	1.32	PB

					CA	2.42	VdW
					CB	3.68	VdW
					C	3.16	VdW
					O	3.43	VdW
		O			N	2.23	HB
					CA	2.69	VdW
					C	3.14	VdW
					O	3.6	VdW
			6	ALA	N	3.69	VdW
5	ARG	N	3	THR	C	3.76	VdW
					O	3.4	HB
			4	LEU	C	1.32	PB
					O	2.23	HB
					CA	2.46	VdW
					N	3.5	VdW
					CG	3.98	VdW
					CB	3.43	VdW
		CA			C	2.42	VdW
					O	2.69	VdW
					CA	3.84	VdW
		CB			C	3.68	VdW
					C	3.16	VdW
					O	3.14	VdW
					C	3.43	VdW
					O	3.6	VdW
			6	ALA	N	3.64	VdW
					N	2.45	VdW
					CA	3.86	VdW
					N	3.27	VdW
		CG			N	3.68	VdW
		C			O	3.39	VdW
					CB	3.72	VdW
					N	1.33	PB
					CA	2.47	VdW
					C	3.1	VdW
		O			O	3.76	VdW
					N	2.24	HB
					CA	2.79	VdW
					C	3.21	VdW
			7	GLU	OE	3.7	VdW
		CD			OE	3.8	VdW
		NE			OE	2.86	HB
		CZ			OE	3.2	VdW
		NH			OE	3.12	HB
					N	3.75	VdW
6	ALA	N	4	LEU	O	3.69	VdW

			5	ARG	N	3.64	VdW
					CA	2.45	VdW
					CB	3.27	VdW
					CG	3.68	VdW
					C	1.33	PB
					O	2.24	HB
		CA			CA	3.86	VdW
					C	2.47	VdW
					O	2.79	VdW
		CB			C	3.72	VdW
		C			C	3.1	VdW
					O	3.21	VdW
		O			C	3.39	VdW
					O	3.76	VdW
			7	GLU	N	3.53	VdW
					N	2.46	VdW
					CA	3.81	VdW
					N	3.36	VdW
					N	1.32	VdW
					CA	2.41	VdW
					CB	3.58	VdW
					C	3.33	VdW
					O	3.75	VdW
					N	2.23	HB
					CA	2.72	VdW
					C	3.49	VdW
					OE	3.95	VdW
			8	GLN	N	3.99	VdW
7	GLU	N	5	ARG	O	3.75	VdW
					NE	2.86	HB
					CZ	3.2	VdW
					CG	3.7	VdW
					CD	3.8	VdW
					NH	3.12	HB
			6	ALA	O	2.23	HB
					CB	3.36	VdW
					N	3.53	VdW
					CA	2.46	VdW
					C	1.32	PB
		CA			O	2.72	VdW
					CA	3.81	VdW
					C	2.41	VdW
		CB			C	3.58	VdW
		OE			O	3.95	VdW
		C			O	3.49	VdW
					C	3.33	VdW

					C	3.75	VdW
			8	GLN	N	3.61	VdW
					N	2.43	VdW
					CA	3.82	VdW
					N	3.27	VdW
		OE			N	3.83	VdW
					CB	3.73	VdW
					C	3.04	VdW
					O	3.34	VdW
					N	1.33	VdW
					CA	2.46	VdW
		O			C	3.06	VdW
					O	3.55	VdW
					N	2.24	HB
					CA	2.81	VdW
			9	ALA	N	3.83	VdW
					N	3.43	VdW
8	GLN	N	6	ALA	O	3.99	VdW
			7	GLU	OE	3.83	VdW
					N	3.61	VdW
					CA	2.43	VdW
					CB	3.27	VdW
					C	1.33	PB
					O	2.24	HB
					CA	3.82	VdW
					C	2.46	VdW
					O	2.81	VdW
					C	3.73	VdW
					C	3.04	VdW
					O	3.06	VdW
					C	3.34	VdW
					O	3.55	VdW
			9	ALA	N	3.5	VdW
		CA			N	2.4	VdW
					CA	3.77	VdW
		CB			N	3.37	VdW
		C			N	1.33	PB
					CA	2.42	VdW
					CB	3.69	VdW
					C	3.06	VdW
					O	3.11	VdW
		O			N	2.25	HB
					CA	2.75	VdW
					C	3.12	VdW
					O	3.31	HB
			10	SER	N	3.94	VdW

9	ALA		7	GLU	C	3.83	VdW
					O	3.43	VdW
		N	8	GLN	CB	3.37	VdW
					C	1.33	PB
					O	2.25	HB
					N	3.5	VdW
					CA	2.4	VdW
		CA			C	2.42	VdW
					O	2.75	VdW
					CA	3.77	VdW
		CB			C	3.69	VdW
					C	3.06	VdW
					O	3.12	VdW
					C	3.11	VdW
					O	3.31	VdW
			10	SER	N	3.63	VdW
					N	2.43	VdW
					CA	3.8	VdW
					N	3.1	VdW
					N	1.33	VdW
					CA	2.42	VdW
					CB	3.66	VdW
					C	3	VdW
					O	3.35	VdW
					N	2.23	HB
					CA	2.78	VdW
					C	3.06	VdW
					O	3.64	VdW
		C	11	GLN	N	3.81	VdW
		O			N	3.45	VdW
10	SER	N	8	GLN	O	3.94	VdW
			9	ALA	N	3.63	VdW
					CA	2.43	VdW
					CB	3.1	VdW
					C	1.33	VdW
					O	2.23	HB
					CA	3.8	VdW
					C	2.42	VdW
					O	2.78	VdW
					C	3.66	VdW
					C	3	VdW
					O	3.06	VdW
					C	3.35	VdW
					O	3.64	VdW
		N	11	GLN	N	3.49	VdW
		CA			N	2.43	VdW

					CA	3.88	VdW
		CB			N	3.35	VdW
		OG			N	3.12	HB
					OE	3.65	VdW
		C			N	1.34	PB
					CA	2.55	VdW
					CB	3.07	VdW
					C	3.53	VdW
					CG	3.89	VdW
					CD	3.61	VdW
					OE	3.28	VdW
		O			N	2.26	HB
					CA	2.98	VdW
					CB	2.85	VdW
					CG	3.43	VdW
					CD	2.97	VdW
					OE	2.95	VdW
					NE	3.49	VdW
11	GLN	N	9	ALA	C	3.81	VdW
					O	3.45	VdW
			10	SER	N	3.49	VdW
					CA	2.43	VdW
					CB	3.35	VdW
					OG	3.12	HB
					C	1.34	
					O	2.26	HB
		CA			CA	3.88	VdW
					C	2.55	VdW
					O	2.98	VdW
		CB			C	3.07	VdW
					O	2.85	VdW
		CG			C	3.89	VdW
					O	3.43	VdW
		CD			C	3.61	VdW
					O	2.97	VdW
		OE			OG	3.65	VdW
					C	3.28	VdW
					O	2.95	VdW
		NE			O	3.49	VdW
		C			C	3.53	VdW

Supplementary Table S4: Intrapeptide contacts within the gag24 peptide. VdW: Van der Waals; HB: Hydrogen bond. PB: peptide bond.

| | Donor 1 | | Donor 2 | | Donor 3 | | |
| | Round | | Round | | Round | | |
	1	2	1	2	1	2	% positive responses
Pool 1	81		4	13	70	51	50.00
Pool 2	127	231		175			50.00
Pool 3		182		65		98	50.00
Pool 4	252	130	179	152	103	95	100.00
Pool 5	103	59		33	9		50.00
Pool 6	266	133				50	50.00
Pool 7		3					0.00
Pool 8		222	36	39		6	50.00
Pool 9	45	45	105	70		86	83.33
Pool 10	29		103				33.33
HA$_{306\text{-}318}$	182	202	304	352		211	83.33

Supplementary Table S 5: Absolute values of INF-γ ELISpots on all three donors using peptides pool. Numbers are given in SFC per 100,000 PBMCs.

Region	Peptide	Donor 1 Round			Donor 2 Round			Donor 3 Round			% positive responses
		1	2	3	1	2	3	1	2	3	
Region 1	Peptide 1	40	47		49				16		33.33
	Peptide 2	8	4	112		118	38			336	44.44
	Peptide 3	28	124	280	25	347	23		172	317	88.89
	Peptide 4	29	5	14	7	36				11	22.22
	Peptide 5		3		4					10	0.00
Region 2	Peptide 6		9			3	8	15	21		11.11
	Peptide 7	3			19		8		12		0.00
	Peptide 8	3	5					24	32		22.22
	Peptide 9	8			3	15					0.00
	Peptide 10	7				7					0.00
Region 3	Peptide 11							4			0.00
	Peptide 12	203	479	246		305	5				44.44
	Peptide 13	40	45	25	20	121	12		81	37	66.67
	Peptide 14	4	11		57			16			11.11
	Peptide 15	1			23	4					11.11
Region 4	Peptide 16		116	15	12		232		85	39	44.44
	Peptide 17					7	166	44			22.22
	Peptide 18	28	140		108	149	61	29	24	139	88.89
	Peptide 19	3	20	20	148	601	247	55	306	186	66.67
	Peptide 20		233	25		188			17	43	44.44
Region 5	Peptide 21		51						24		22.22
	Peptide 22		31	23		279	195		11	28	55.56
	Peptide 23	16	21	72	113	3	63		36	258	66.67
	Peptide 24	5	32	5		12					11.11
	Peptide 25	0	1	11							0.00
	HA$_{306-318}$		118	100	44	390	137	21	294	358	88.89

Supplementary Table S 6: Absolute values of IFN-γ ELISpots on all three donors using single peptides. Numbers are given in SFC per 100,000 PBMCs.

Peptide 3

Overall conservancy: 92.38% (4618/4999)

	Shannon entropy	Conservancy	Consensus sequence
Tyr121	0.0456	99.22%	Tyr
Ala122	0.1603	96.66%	Ala
Ser123	0.0201	99.74%	Ser
Leu124	0.0076	99.92%	Leu
Arg125	0.0112	99.86%	Arg
Ser126	0.0000	100.00%	Ser
Leu127	0.0106	99.86%	Leu
Val128	0.0988	98.06%	Val
Ala129	0.0035	99.96%	Ala
Ser130	0.0460	99.28%	Ser
Ser131	0.0073	99.92%	Ser
Gly132	0.0000	100.00%	Gly
Thr133	0.0218	99.72%	Thr
Leu134	0.0065	99.92%	Leu
Glu135	0.0038	99.96%	Glu
Phe136	0.0000	100.00%	Phe

Supplementary Table S 7: Conservancy and Shannon entropies for peptide 3. Positions numbered according to the HA sequence of A/Perth/16/2009 (H3N2). Overall conservancy reflects the proportion of H3N2 strains matching the full sequence of peptide 3 to 100%. Shannon entropies reflect the overall variation at any given position. Conservancy values for individual positions reflect the proportion of H3N2 strains matching this individual residue.

Peptide 18

Overall conservancy:		96.20% (4759/4999)	
	Shannon entropy	**Conservancy**	**Consensus sequence**
Asp431	0.0038	99.96%	Asp
Thr432	0.0234	99.72%	Thr
Lys433	0.0057	99.94%	Lys
Ile434	0.0570	99.12%	Ile
Asp435	0.0054	99.94%	Asp
Leu436	0.0144	99.82%	Leu
Trp437	0.0073	99.92%	Trp
Ser438	0.0057	99.94%	Ser
Tyr439	0.0019	99.98%	Tyr
Arg440	0.0242	99.70%	Arg
Ala441	0.0299	99.60%	Ala
Glu442	0.0606	99.06%	Glu
Leu443	0.0380	99.50%	Leu
Leu444	0.0157	99.78%	Leu
Val445	0.0098	99.88%	Val
Ala446	0.0277	99.64%	Ala

Supplementary Table S 8: Conservancy and Shannon entropies for peptide 18. Positions numbered according to the HA sequence of A/Perth/16/2009 (H3N2). Overall conservancy reflects the proportion of H3N2 strains matching the full sequence of peptide 18 to 100%. Shannon entropies reflect the overall variation at any given position. Conservancy values for individual positions reflect the proportion of H3N2 strains matching this individual residue.

HA$_{306-318}$

Overall conservancy: **2.5% (125/4999)**

Overall conservancy (consensus): **57.83% (2891/4999)**

	Shannon entropy	Conservancy	Consensus sequence	Conservancy (consensus)
Phe306	0.0121	99.86%	Phe	
Lys307	0.1281	2.56%	Arg	97.34%
Tyr308	0.0090	99.90%	Tyr	
Val309	0.0497	99.20%	Val	
Lys301	0.0179	99.78%	Lys	
Gln311	0.3865	87.14%	Gln	
Asp312	0.7207	61.33%	Asp	
Thr313	0.0141	99.84%	Thr	
Leu314	0.0035	99.96%	Leu	
Lys315	0.0090	99.90%	Lys	
Leu316	0.0054	99.94%	Leu	
Ala317	0.0051	99.94%	Ala	
Thr318	0.0073	99.92%	Thr	

Supplementary Table S 9: Conservancy and Shannon entropies for HA306-318. Positions for HA$_{306-318}$ are numbered according to HA sequence of the Influenza A/Texas/1/77 (H3N2) strain. Overall conservancy reflects the proportion of H3N2 strains matching the full of HA$_{306-318}$ to 100%. Overall conservancy of the consensus sequence reflects the proportions of H3N2 strains matching the full consensus sequence to 100%. Shannon entropies reflect the overall variation at any given position. Conservancy values for individual positions reflect the proportion of H3N2 strains matching this individual residue. Differences in the consensus sequence highlighted in red.