

LIFE SCIENCES RESEARCH NETWORK WALES
RHWYDWAITH GWYDDORAU BYWYD CYMRU

RESOLVING THE 'CORE' OF INFLUENZA INFECTION

ALEXANDER LIAM GREENSHIELDS-WATSON

A THESIS

SUBMITTED TO

CARDIFF UNIVERSITY

IN CANDIDATURE FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

MMXVII



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 fulfilment 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. 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 for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed (candidate) Date

Summary

Background - The internal elements of the Influenza-A virus, exhibit high levels of conservation and offer a more consistent target for the immune system amidst the diversity of potential strains. T-cell responses to these proteins have been shown to correlate with protection and decreased symptom severity during infection. Yet the epitopes and T-cell receptor (TCR) repertoires that underpin these important responses have not been analysed in detail at the molecular level.

Results - These responses were analysed by the development of an HLA-DR1 restricted epitope mapping platform in chapter 3, followed by its application in finding DR1-restricted epitopes within the three internal proteins in chapter 4. Two of these epitopes were analysed by X-ray crystallography to understand their presentation and complement HLA-binding algorithm data.

Identification of epitopes that gave robust and reproducible responses allowed analysis of responding T-cell populations by HLA-multimer staining on flow cytometry and subsequent clonotypic analysis of TCR repertoires in chapter 5. The clonotypic repertoire data was interpreted and then information in response to a single epitope was aligned with structural data in chapter 6 to further understand the molecular interactions that shape these responses.

Conclusions - This work generated several novel HLA-DR1 restricted epitopes, crystal structures and TCR repertoire information that both expands existing knowledge of CD4+ T-cell responses, and confirms the potential of the conserved influenza proteins as targets in future vaccination research.

Acknowledgements

I would like to thank Dr David Cole and Prof Andrew Godkin for plucking me from industry and giving me the opportunity to learn, create and think freely over the past three years. I have enjoyed the experience immensely, and hopefully become a better scientist in the process. Thank you for sharing your time, attention, enthusiasm and knowledge. Dave, good luck with the expanding empire, I'm going to miss the banter and the pearls of wisdom, but I'm sure we will be in steady contact from your new base over the coming years. Andy, I'm already looking forward to the new challenges.

Special acknowledgements go to several members of the lab that contributed directly to this project: Dr Meriem Attaf, many thanks for the clonotypic sequencing work, brainstorming sessions and various philosophical discussions on life and immunology. Dr Garry Dolton, many thanks for the countless nuggets of cell culture wisdom, knowledge of the force and staining tips. Dr Pierre Rizkallah, many thanks for the crystallographic help and good company at Heathfields lunches. Many thanks to Prof Gallimore and Prof Sewell for the many stimulating lab meetings, feedback during presentations and the discussions on all things immunological.

In the office of the Cole-Crew: Aaron, Jade, Georgie, Andrea and Master Bruce. You are a wonderful team of humans, and it's been a pleasure to share an office. Thank you for being a sounding board of understanding and sanity in each presentational, experimental and written venture. Comrade Hywel, I hope it was fun, class-II has not been the same since. Thanks to everyone on the second and third floor for the help, banter and nights out, it has not been boring.

To Mum, Dad, Ellie, Gran and Grandad: I would neither be here, nor doing work like this without your love and support. Thank you for every text, phone call and all the times at home. To mi Guapa, Alicia, thank you for the shelter from the storm and for giving me something to look forward to everyday. I think we found some way to make the time pass.

*Pray, remember, that I leave you all my theory complete,
Lacking only certain data, for your adding as is meet;
And remember, men will scorn it, 'tis original and true,
And the obloquy of newness may fall bitterly on you.*

From “The Old Astronomer (To his Pupil)”

Sarah Williams 1868

Table of Contents

1	Introduction	1
1.1	Understanding the Immune System.....	1
1.1.1	Innate Immunity.....	1
1.1.2	Adaptive Immunity.....	2
1.1.3	Generation of Receptor Diversity.....	5
1.1.4	Thymic Selection.....	9
1.1.5	The HLA-Restriction Paradox	13
1.1.6	T-cell Activation.....	15
1.1.7	CD8+ T-cells	20
1.1.8	Chemical and Lipid Mediated T-cell Recognition.....	20
1.2	CD4+ T-cells: Conductors of the Immune Response	21
1.2.1	CD4+ Subsets: Protection and Disease	21
1.2.2	Plasticity	26
1.2.3	Understanding CD4+ T-cells.....	26
1.3	HLA Class-II	27
1.3.1	HLA Structure.....	27
1.3.2	Class-II Expression and Antigen Processing.....	30
1.3.3	Cross-Presentation	30
1.3.4	HLA-DM and HLA-DO.....	31
1.3.5	Class-II Peptide Presentation	36
1.4	The Influenza Virus	37
1.4.1	Seasonal	37
1.4.2	Pandemic	37
1.4.3	Strain Classification.....	38
1.4.4	Viral Structure	39
1.4.5	Life Cycle	42
1.4.6	Antigenic Drift.....	42
1.4.7	Antigenic Shift	43

1.4.8	Original Antigenic Sin.....	44
1.5	Aims and Hypothesis of the Thesis	45
2	Methods.....	47
2.1	Generation of DR1.APCs	47
2.1.1	List of General Reagents	47
2.1.2	Cell Culture	48
2.1.3	DR1 Construct and Lentivirus Production.....	48
2.1.4	Assessment of DR1 Expression by Flow Cytometry	49
2.1.5	Infection of “Naked” Cell Lines with DR1 Lentivirus.....	49
2.1.6	IFN- γ ELISpot with CD4+ T-cell Clones.....	49
2.2	Screening of Donor PBMC against Peptide Libraries	51
2.2.1	Peptide Libraries	51
2.2.2	Peptide Pools	51
2.2.3	Processing of PBMC.....	51
2.2.4	Generation of Lines through Culture of PBMC with Peptide Pools	52
2.2.5	Pulsing of B-LCL 174.DR1 Transduced APCs	52
2.2.6	IFN- γ ELISpot with PBMC Lines	52
2.2.7	Plate Analysis and Normalisation	53
2.2.8	Binding Algorithms Inputs	53
2.3	Protein Production for Crystals and HLA-Multimers.....	55
2.3.1	List of Reagents	55
2.3.2	HLA-DR1 Plasmids	56
2.3.3	Inclusion Body Production	56
2.3.4	Inclusion Body Purification.....	58
2.3.5	Refolding of HLA Class-II.....	58
2.3.6	Purification of Refolded HLA Class-II.....	59
2.3.7	Biotinylation and QC Shift Assay	60
2.3.8	SDS-PAGE.....	60
2.3.9	TCR-pHLA Complex Formation	61

2.3.10	Protein Crystallography.....	61
2.3.11	X-Ray Crystallographic Sample Preparation and Data Collection.....	61
2.3.12	Structure Solution from DLS Datasets.....	62
2.4	Analysing Epitope-Specific T-cells	63
2.4.1	Culture of PBMC for HLA-Multimer Staining.....	63
2.4.2	Preparation of HLA-Multimers.....	63
2.4.3	Flow Cytometry: HLA-Multimer Staining.....	63
2.4.4	Clonotyping by Next Generation Sequencing.....	64
3	HLA-DR1 Transduced APCs Facilitate Detection of DR1-Restricted T-Cell Responses.....	67
3.1	Abstract.....	67
3.2	Introduction	68
3.3	Aims.....	71
3.4	Results	72
3.4.1	Generation of HLA-DR1 Presenting Cells.....	72
3.4.2	DR1 Presentation to CD4+ T-cell Clones on IFN- γ ELISpot	74
3.4.3	Dissecting DR1 Responses from Polyclonal Lines.....	76
3.4.4	Comparing Expansion Conditions: FCS versus Human Serum	78
3.4.5	Distinguishing between Autologous and DR1 Presentation	81
3.5	Discussion	85
3.5.1	Presentation of Peptide by DR1 in the Absence of HLA-DM.....	85
3.5.2	A More Diverse DR1 Peptide Repertoire?	86
3.5.3	Presence of A2.....	87
3.5.4	Future Work	87
4	Identification of HLA-DR1 Restricted Epitopes within the Internal Influenza Proteins	88
4.1	Abstract.....	88
4.2	Introduction	89
4.3	Aims.....	92

4.4	Results	94
4.4.1	Peptide Libraries and Pooling Matrix Design	94
4.4.2	Overview of Screens in two HLA-DR1+ Donors	95
4.4.3	Identification of Immunogenic Peptides	101
4.4.4	Defining Epitopes within Long Sequences.....	106
4.4.5	Confirmation of Epitope Immunogenicity	115
4.4.6	Confirming Algorithm Predictions Using Crystal Structures.....	118
4.5	Discussion	126
4.5.1	The Precursor Frequency and Sensitivity of Cognate T Cells	126
4.5.2	Culturing with a Peptide Pool versus an Individual Peptide.....	128
4.5.3	Using Algorithms to Define Epitopes	129
4.5.4	Epitope Missed within Highly Immunogenic Peptide Pools	130
4.5.5	Analysis of Three Misidentified Epitopes	131
4.5.6	Future Work	134
5	Epitope-Specific CD4+ T-cells from HLA-DR1 Donors Exhibit Shared Cellular and Genetic Characteristics	135
5.1	Abstract.....	135
5.2	Introduction	136
5.3	Aims.....	138
5.4	Results	139
5.4.1	Flow Cytometry Analysis of Epitope-Specific CD4+ Populations	139
5.4.2	TCR α -Chain Analysis of Epitope-Specific CD4+ T-cells	148
5.4.3	Epitope-Specific TRAV Gene Usage	151
5.4.4	TCR α CDR3 Diversity	158
5.4.5	Public CDR3 Amino Acid Sequences	161
5.4.6	Analysis of Amino Acid Motifs in Response to Each Epitope	163
5.5	Discussion	167
5.5.1	Response Hierarchies and Protective Capacity	167
5.5.2	Explaining Public/Shared Clonotypic Features	168

	5.5.3	Future Work	170
6		Recognition of DR1-PKY by Two Distinct TRAV8-4 TCRs is Mediated by Residues that are Conserved across PKY-Specific TCRα Repertoires.	172
	6.1	Abstract	172
	6.2	Introduction	173
	6.2.1	Features of the TCR-pHLA Class II Protein Complex	173
	6.2.2	Rationalising Repertoire Information using Structural Analysis	174
	6.2.3	PKY as a Model System	175
	6.3	Aims.....	176
	6.4	Results	177
	6.4.1	Comparison of TCR-pHLA Complex Structures	177
	6.4.2	TCR-pHLA Interface.....	184
	6.4.3	Understanding Repertoires with Structural Information	187
	6.4.4	Conserved Interactions with the TCR β -chains and DR1-PKY.....	197
	6.4.5	Summary.....	202
	6.5	Discussion	203
	6.5.1	Implications Beyond TRAV8-4 and TRAV8-6	205
	6.5.2	Conservation of Salt Bridges over Weaker Contacts.....	207
	6.5.3	Future Work	208
7		Discussion.....	209
	7.1	The Basis of Heterosubtypic Protection?	209
	7.2	Implications for Vaccine Design.....	211
	7.3	Broad or Narrow Repertoires, What's Best for Protection?	212
	7.4	Application of Findings to Other Challenges	213
	7.5	Future Work	214
8		Appendix	216
	8.1	Details of Peptides	216
	8.2	Flow Cytometry & Sorting Data	222

8.3	Application of the 50 read threshold cut off.....	227
8.4	CDR2 β Contact Tables	229
9	Bibliography.....	231

List of Figures

<i>Figure 1.1 Adaptive mechanisms of protein recognition</i>	4
<i>Figure 1.2. Schematic representation of (A) TCR and (B) BCR protein structures</i>	5
<i>Figure 1.3. Schematic of V(D)J recombination at the TCR locus</i>	8
<i>Figure 1.4. Schematic representation of the major events in thymic selection</i>	11
<i>Figure 1.5. Events in T-cell activation, Signal-1, Signal-2 and Signal-3</i>	16-19
<i>Figure 1.6. The roles of CD4+ subsets in the inflammatory immune response</i>	24
<i>Figure 1.7. Characteristics of HLA Class-II (specific to HLA-DR1)</i>	28
<i>Figure 1.8. The endosomal class-II presentation pathway and HLA-DM peptide selection</i>	33
<i>Figure 1.9. Loading of peptides onto HLA class-II in the absence of HLA-DM</i>	35
<i>Figure 1.10. The Influenza-A virion and mechanisms of viral evolution</i>	40
<i>Figure 2.1. NetPanMHCII input view</i>	53
<i>Figure 2.2 Parameters for NetPanMHCII</i>	54
<i>Figure 3.1. Deletion in the 721.174 parent line conferring the “naked” phenotype</i>	70
<i>Figure 3.2. Expression levels of HLA-DR across wild type and transduced cell lines on flow cytometry using pan anti-HLA-DR antibody (L243)</i>	73
<i>Figure 3.3. Presentation of HA peptide 306-318 (PKYVKQNTLKLAT) to the cognate CD4+ T-cell clone DCD10 by wild type and transduced antigen-presenting cells on IFN-γ ELISpot</i>	75
<i>Figure 3.4. Using DR1 cells to identify HLA-DR1 restricted T-cell responses from polyclonal PBMC lines</i>	76
<i>Figure 3.5. Comparison of Foetal Calf Serum (FCS) and Human AB Serum in expansion and ELISpot media</i>	79
<i>Figure 3.6. Using a PBS based wash Technique gives a clearer distinction between autologous and DR1 presentation</i>	81

<i>Figure 3.7. Assessment of presentation to DR1 restricted T-Cell clones after peptide pulsing by 174.WT and 174.DR1 APCs under stringent PBS wash conditions.....</i>	<i>83</i>
<i>Figure 3.8. Comparison of autologous presentation and DR1 presentation in a DR1 Negative donor</i>	<i>84</i>
<i>Figure 4.1. Schematic showing the structural representations of three internal influenza proteins.....</i>	<i>90</i>
<i>Figure 4.2. Architecture of pooling matrices for M1, NP and PB-1</i>	<i>94</i>
<i>Figure 4.3. Summary of assays on M1 peptide pools.....</i>	<i>96</i>
<i>Figure 4.4. Summary of assays on NP peptide pools.....</i>	<i>98</i>
<i>Figure 4.5. Summary of assays on PB-1 peptide pools.....</i>	<i>100</i>
<i>Figure 4.6. Individual peptide analyses of regions identified from pool assays on HLA-DR1 IFN-γ ELISpot.....</i>	<i>102</i>
<i>Figure 4.7A. Example of the output of NetPanMHCII 3.1</i>	<i>109</i>
<i>Figure 4.7B. Example with M1 peptide-2 and peptide-3 of how the output length can impact the predicted binding affinities and hence ranking of each potential core within a sequence.....</i>	<i>110</i>
<i>Figure 4.8. Summary analysis of IFN-γ DR1 ELISpot testing on lines cultured with short peptides across 4 HLA-DR1+ donors.....</i>	<i>116</i>
<i>Figure 4.9. Images and details of the successful crystallisation conditions for (A) SGP and (B) QAR.....</i>	<i>119</i>
<i>Figure 4.10. B-factor and electron density representation of each peptide at 0.3 sigma</i>	<i>121</i>
<i>Figure 4.11. Identification of peptide anchor residues and orientation within the binding pockets.....</i>	<i>122</i>
<i>Figure 4.12. Polar interactions between the peptide and side chains of the binding groove</i>	<i>123</i>

<i>Figure 4.13. Orientation of potential TCR contact residues and QAR N-terminal flanking region</i>	<i>125</i>
<i>Figure 5.1. Comparison of HLA-Multimer staining and IFN-γ ELISpot responses in two HLA-DR1 donors</i>	<i>140</i>
<i>Figure 5.2. Regression analysis of flow cytometry statistics (%CD4+ and MFI) with IFN-γ ELISpot data (SFC).....</i>	<i>142</i>
<i>Figure 5.3. Comparison of epitope-specific stains in 5 HLA-DR1 donors</i>	<i>144</i>
<i>Figure 5.4. Comparison of epitope-specific HLA-multimer staining in five HLA-DR1 donors</i>	<i>147</i>
<i>Figure 5.5. Summary of TCR-α chain data in five donors.....</i>	<i>149</i>
<i>Figure 5.6. Analysis of epitope-specific TRAV gene usage in multiple donors</i>	<i>153</i>
<i>Figure 5.7. Clonotypic diversity in response to each epitope across all donor samples</i>	<i>160</i>
<i>Figure 5.8. Amino acid motif analysis of α chain CDR3 regions in all sequences obtained in response to each epitope.....</i>	<i>165</i>
<i>Figure 6.1. HA1.7-DR1-PKY complex structure secondary structural representation and data table.....</i>	<i>180</i>
<i>Figure 6.2. F11-DR1-PKY complex structure secondary structural representation and data table.....</i>	<i>181</i>
<i>Figure 6.3. Structural comparison of F11 and HA.17 complexes by secondary structural overlay alignment.....</i>	<i>182</i>
<i>Figure 6.4. Stabilising polar peptide contacts with the HLA groove for A. HA1.7 complex and B. F11</i>	<i>183</i>
<i>Figure 6.5. TCR Contact Footprint representation over pHLA surface plot.....</i>	<i>185</i>
<i>Figure 6.6. CDR1α Contacts with both peptide and HLA</i>	<i>189</i>
<i>Figure 6.7. CDR2α contacts with the peptide and HLA</i>	<i>191</i>

<i>Figure 6.8. CDR3 A loop contacts with the peptide and corresponding acidic enriched motifs from repertoire data.....</i>	<i>194</i>
<i>Figure 6.9. CDR1β contacts with peptide and HLA</i>	<i>198</i>
<i>Figure 6.10. CDR2β contacts with the HLA.....</i>	<i>199</i>
<i>Figure 6.11. CDR3β contacts with the peptide</i>	<i>200</i>
<i>Figure 6.12. Examples of Gene usage and CDR3 motifs from the previous chapter, which may be understood by the generation and analysis of relevant TCR-pHLA complex structures.....</i>	<i>204</i>

List of Tables

<i>Table 4.1. Summary of identified regions found by pools and individual analysis.....</i>	<i>105</i>
<i>Table 4.2. Summary of immunogenic regions and the corresponding binding predictions by NetPanMHCII 3.1.....</i>	<i>111</i>
<i>Table 4.3. Short peptides for further testing.....</i>	<i>115</i>
<i>Table 4.4. Crystallographic Data Table for DR1-SGP and DR1-QAR.....</i>	<i>119</i>
<i>Table 5.1. Shared TCRα CDR3 amino acid and nucleotide sequences in response to each epitope.....</i>	<i>162</i>
<i>Table 6.1. Genetic and amino acid composition of HA1.7 and F11 TCRs.....</i>	<i>178</i>
<i>Table 6.2. CDR1α contacts with both peptide and HLA.....</i>	<i>190</i>
<i>Table 6.3. CDR2α contacts with the peptide and HLA.....</i>	<i>191</i>
<i>Table 6.4A. Corresponding contact details of all atoms within 4.00 Å between HA1.7 CDR3α and peptide.....</i>	<i>195</i>
<i>Table 6.4B. Corresponding contact details of all atoms within 4.00 Å between F11 CDR3α and peptide.....</i>	<i>196</i>
<i>Table 6.5. CDR1β contacts with peptide and HLA.....</i>	<i>198</i>
<i>Table 6.6. CDR3β contacts with peptide and HLA within 4.00 Å.....</i>	<i>201</i>
<i>Table 6.7. Variation of germline sequences from TRAV8-4 (crystal structure) around the CDR loops of TRAV genes used to respond to PKY in five donors.....</i>	<i>206</i>

1 Introduction

1.1 Understanding the Immune System

In order to protect against disease, the immune system recognises and eliminates threats that have the potential to damage the body. This involves distinguishing between self-antigens and those antigens which indicate the presence of non-self or malignant entities like bacteria, viruses or cancer. Following recognition of a threat, the immune system acts to neutralise and eliminate its source without causing extensive damage to the host.

The primary theme of this project centres on how specific cells of the immune system can mediate recognition of a rapidly evolving pathogen, and the extent to which these recognition mechanisms are conserved across the human population. By understanding the mechanistic basis of protection at the molecular level, it is possible to gain fundamental knowledge that might be applied in the prevention and treatment of disease.

1.1.1 Innate Immunity

The cells and mechanisms of the immune system are divided into two arms: innate and adaptive immunity. Innate immunity provides broad protection against many pathogens utilising fixed mechanisms that do not change in response to evolving threats. Barriers (skin and antimicrobial peptides), molecular pathways (complement), receptors and cells (phagocytic, lytic and cytotoxic) fall under this definition¹.

Bacteria, viruses and parasites are detected by their pathogen associated molecular patterns (PAMPs) such as lipopolysaccharides (LPS) and double stranded RNA. PAMPs activate pathogen recognition receptors (PRRs) such as toll-like and NOD-like receptors that initiate the process of inflammation. Immune cells and coagulation factors are drawn to the site of injury and infection by inflammatory cytokines and chemoattractant molecules in order to neutralise and limit the spread of invading threats.

Neutrophils, macrophages, dendritic cells and basophils can uptake and destroy pathogens by phagocytosis, while natural killer (NK) cells and complement exert cytotoxic effects that can clear infected or damaged cells. Several innate functions such as antigen uptake and presentation on macrophages and dendritic cells, complement activation and NK cell cytotoxicity work in tandem with adaptive functions to target pathogens.

1.1.2 Adaptive Immunity

Innate mechanisms are broad, and cannot focus responses towards unique features expressed by a specific pathogen or malignancy. The direction of vast immune resources at these specific features or “epitopes” is the most effective method to neutralise and eliminate the causative agents of disease. When recognition of epitopes is encoded into a rapidly accessible pool of cells, i.e. memory, then reinfection should no longer pose a threat. This is the role of adaptive immunity.

Adaptive immunity is mediated primarily by T-cells and B-cells, which recognise antigens through the T-cell receptor (TCR), B-cell receptor (BCR) and its soluble form (Antibody). These receptors exhibit high levels of sequence variation, with a single sequence having the potential to bind a protein, peptide or chemical antigen in an extremely specific manner.

The epitope is defined as the exact recognition site, and may be only a few amino acids in length or a single molecule. The TCR and BCR recognise protein epitopes by distinct mechanisms. These mechanisms result in different forms of recognition target and resulting effector function, which define the roles of T- and B-cells in response to disease.

1.1.2.1 BCR and Antibody Recognition

Recognition of conformational or intact protein antigens is achieved by the BCR and antibody. Soluble proteins, free in solution or present at the cell surface can be bound at any sterically accessible site (Fig. 1.1A). The solvent face of an intact protein

presents a unique surface that allows for highly specific receptor binding and adaptive immune targeting.

Yet this surface is dependent on protein stability, and once inside a cell it is protected from BCR recognition and therefore no longer active as a marker of infection or malignancy. Although the epitope can consist of less than ten amino acids, these residues may be distant in terms of linear sequence, and only come together in correctly folded conformations^{2,3}.

1.1.2.2 TCR Recognition and HLA Proteins

In order to maintain adaptive recognition of pathogens or malignancy in the absence of extracellular material and independent of protein conformation, T-cells are essential. They can respond to linear sequences of amino acids derived predominantly from the cytosol (class-I) or the extracellular space and membrane (class-II) presented by the human leukocyte antigen (HLA) proteins⁴ (Fig. 1.1B).

The epitope recognised by the TCR is defined by both the peptide and HLA protein that presents it. This is due to the unique solvent exposed surface formed by the rigidly orientated peptide and surrounding HLA amino acids, both components contact the TCR during binding⁵.

Antigen presentation and TCR recognition will be discussed, in further detail, in later sections. Here, the important point is that T- and B-cells together can exert extensive coverage of antigens at the conformational and linear level. Such vast antigenic coverage will require a large receptor repertoire to take advantage of the many possible epitope-signatures of disease.

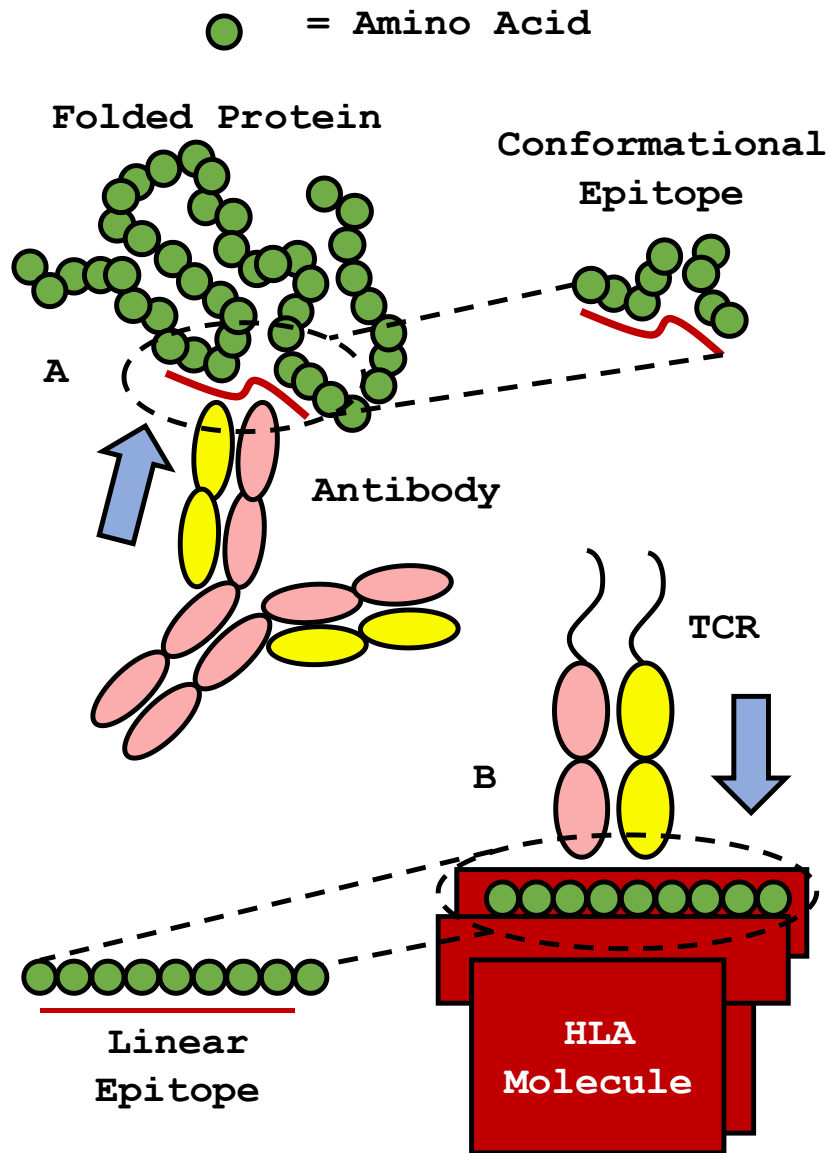


Figure 1.1 Adaptive mechanisms of protein recognition.

(A) Antibodies can recognise non-linear or “conformational” epitopes at the surface of folded protein. In the denatured or linear state such epitopes may not be present, as the amino acids that constitute them are in distant sequence positions. **(B)** HLA molecules hold linear stretches of amino acids in an extended conformation derived from digested extracellular or cytosolic material. Recognition of linear epitopes is independent of the conformation of the parent protein, and solely related to properties of the amino acid sequence.

1.1.3 Generation of Receptor Diversity

The protein structures of T-cell and B-cell receptors are composed of immunoglobulin (Ig) domains, which consist of two stable antiparallel β -sheets in a folded conformation covalently linked by a disulphide bond (the Immunoglobulin “fold”).

The TCR is an $\alpha\beta$ heterodimer of two chains each composed of two Ig domains. Diversity is concentrated in the $\alpha_1 \beta_1$ “variable” region that facilitates antigen recognition, while the “constant” $\alpha_2 \beta_2$ region is membrane proximal, preceding a transmembrane sequence that anchors that TCR to the cell surface (Fig. 1.2A).

The BCR and soluble form are homodimers (Fig. 1.2B), with each monomer consisting of a two Ig-domain light chain (L) disulphide bonded to a four-domain heavy chain (H). Like the TCR, diversity is focused in the variable region (V_H and V_L), while the remaining light chain domain and three heavy chain domains constitute the constant region (C_L , C_H1 , C_H2 , C_H3), which is membrane anchored in the BCR but soluble as an antibody.

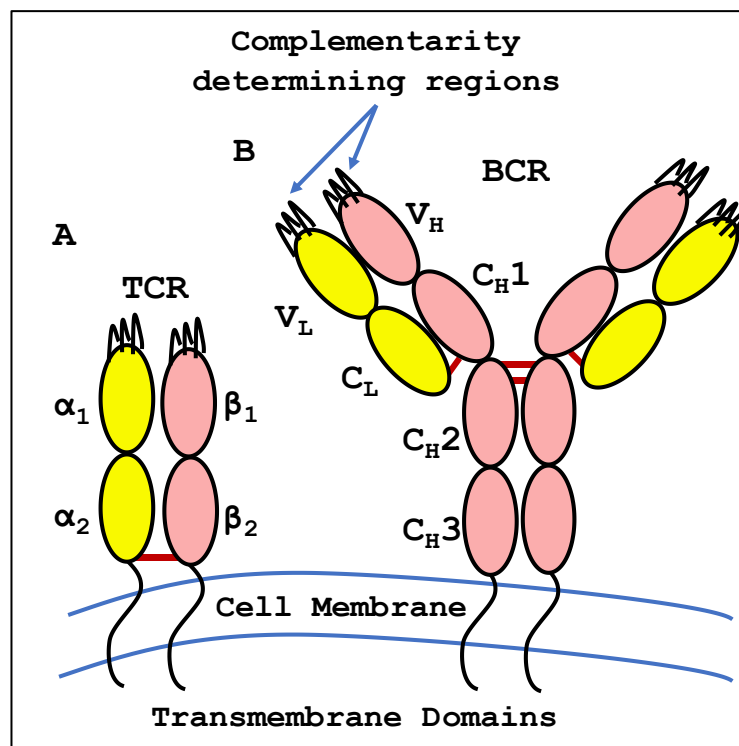


Figure 1.2. Schematic representation of (A) TCR and (B) BCR protein structures (representative of an IgG isotype). Each immunoglobulin domain is represented by an oval, with CDRs shown at the ends of each variable domain (TCR - α_1 , β_1 , BCR - V_L , V_H). Inter-chain disulphide bonds are represented by red lines.

Complementarity determining regions (CDRs) are amino acid loops of variable length which form the contact interface of TCR and BCR interactions with cognate antigen. Each chain has three regions, CDR1, CDR2 and CDR3, the amino acids of which form key contacts with antigen that determine the receptor binding specificity.

The TCR α and BCR light chain sequences are generated from variable and junctional genes, while TCR β and BCR heavy chain sequences are derived from variable, junctional and an additional diversity gene. Many different V, D and J genes are encoded in the human genome, but only one of each is incorporated in a specific chain sequence⁶.

During the development of T- and B-cells, V(D)J segments are recombined to give a single exon that may encode a functional variable sequence⁷ (Fig. 1.3). The term “recombination” is used specifically, as gene segments are not simply joined, but undergo nucleotide sequence alterations at V-J and V-D-J junctions catalysed by the recombinase complex. These involve palindromic (P) and non-templated-encoded (N) nucleotide sequence additions to the ends of gene segments, which occurs before pairing and ligation of V(D)J strands⁸. Additionally, exonuclease activity can delete terminal nucleotides in extensive numbers, eliminating germline or P/N incorporated sequences. These mechanisms are random, and result in somatically encoded amino acids and highly diverse V(D)J joining regions that comprise the hypervariable CDR3 loop of each chain.

Therefore, chain sequence diversity arises from both the number of potential V, D and J gene combinations, and junctional variation created by random nucleotide addition and deletion during recombination (Fig. 1.3). Receptor diversity is further increased through pairing of α and β , or heavy and light chains, as well as somatic hyper mutation in B-cells which will not be discussed here.

The total number of unique receptors that could arise from these processes has been estimated as 5×10^{13} for the BCR (not including somatic hyper mutation) and 10^{18}

for the TCR. This theoretical immune diversity has helped us to evolve alongside the vast diversity of potentially encountered pathogens.

As T-cells are the focus of this project, they shall be described in more detail in the following sections.

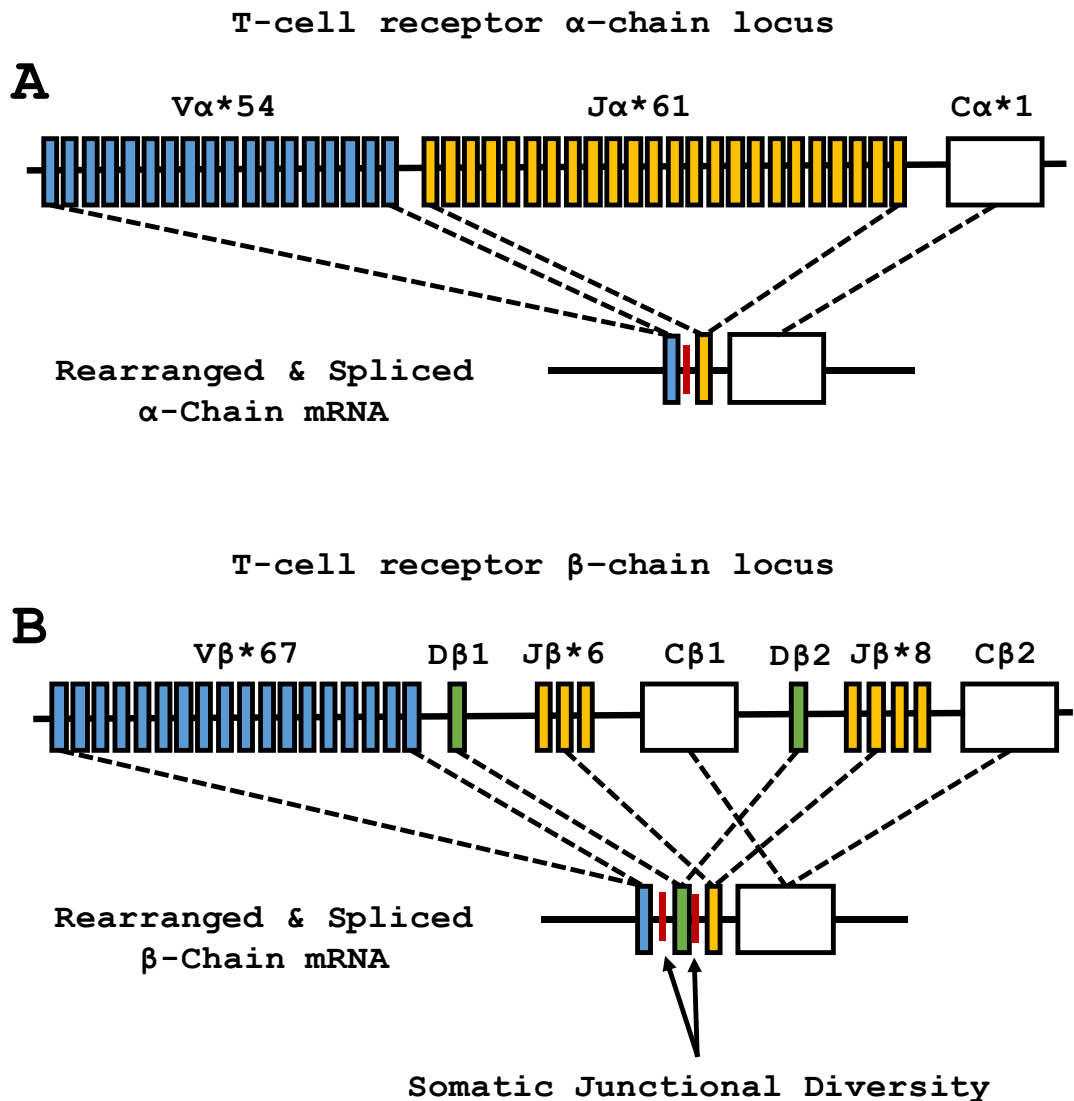


Figure 1.3. Schematic of V(D)J recombination at the TCR locus.

(A) TCR α -chain **(B)** TCR β -chain. V genes are in blue, J genes in yellow, D genes in green and constant genes in white. Red lines represent junctional diversity introduced by the recombinase enzymes during joining of genes. Following rearrangement and splicing one of each gene is incorporated into the final mRNA coding for each chain. Diversity arises from both the different combinations of V(D)J genes as well as the variation in nucleotide additions and deletions at the joining regions.

1.1.4 Thymic Selection

A fundamental property of the immune system is its ability to eliminate disease whilst limiting damage to the host tissues and organs. The process of thymic selection generates a T-cell TCR repertoire that recognises peptide-HLA (pHLA) but does not respond aggressively to self, referred to as “tolerance”.

T-cell precursors migrate from the bone marrow to the thymus where they undergo several stages of development (Fig. 1.4). Commitment to the T-cell lineage is followed by active rearrangement of the TCR β -chain, expression of both CD4 and CD8 co-receptors and rearrangement of the TCR α -chain resulting in a double positive (DP) thymocyte. Upon the generation of a functional $\alpha\beta$ TCR, DP thymocytes that exhibit avidity for self-pHLA molecules on the surface of cortical thymic epithelial cells (cTECs) receive activation and survival signals. Those cells which fail to recognise self-pHLA die by neglect. Activated DP cells increase TCR expression and lose expression of one co-receptor, becoming single positive CD8+ or CD4+ T-cells specific to either HLA class-I, or class-II, respectively.

Recognition of self-peptide on HLA has been shown to facilitate “tonic” signalling of T-cells in the periphery, maintaining a low level of TCR signalling that can be quantified by CD5 expression^{9,10}. Yet if this self-recognition is too efficient it may stimulate effector functions outside of the desired setting, destroying healthy host tissues and leading to pathogenic autoimmunity. In order to mitigate this threat, CD8+ T-cells with too high an avidity for self-antigen receive apoptotic signals and die by negative selection^{11–13}. Expression of many tissue-specific antigens on thymic epithelial cells (that would not otherwise be present in one body location) is driven by the AIRE (autoimmune regulator) transcription factor^{14,15}.

For high avidity CD4+ T-cells, negative selection may occur, but additionally these cells may be driven down the suppressive or regulatory pathway by expression of FOXP3^{16–18}. Such cells, termed “thymic” T-regs (tT_{REG}) do not respond to cognate pHLA

by inflammatory processes, but instead release suppressive cytokines or employ other regulatory mechanisms that negatively regulate surrounding immune cells.

The exact mechanisms of tT_{REG} generation, specifically in relation to TCR avidity thresholds that distinguish negative selection and onset of a regulatory lineage are the subject of many hypotheses^{19,20}. Recent work analysing the epigenetic changes at the FOXP3 locus and super-enhancers^{21,22} has shed light on this relationship, but biophysical and structural evidence for tT_{REG} mediated tolerance mechanisms has yet to be uncovered.

The resulting thymic output is balanced in its ability to recognise pHLA and exert tolerance in response to self-antigen. Disruptions in this balance result in either impaired immune responses or breaking of tolerance and the onset of disease.

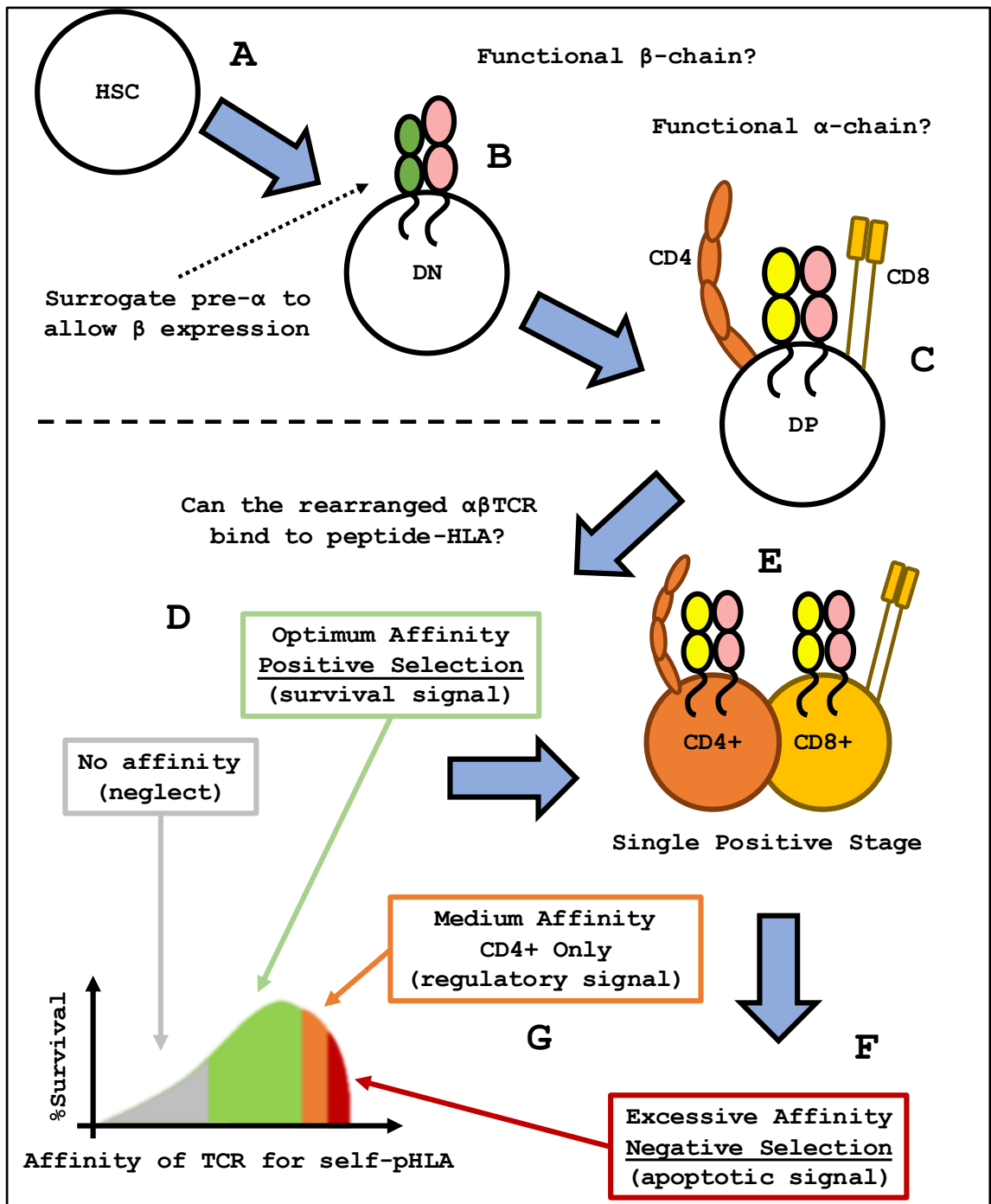


Figure 1.4. Schematic representation of the major events in thymic selection.

(A) Haematopoietic stem cells (HSC) enter the thymus from the bone marrow. (B) Rearrangement and transcription of the TCR β -chain which pairs with a surrogate pre- α chain to allow surface expression in the "double negative" thymocyte (DN). (C) Rearrangement and transcription of the TCR α -chain is followed by expression of both CD4 and CD8 co-receptors in the "double positive" thymocyte (DP). (D) DP thymocytes

are positively selected if they can bind cognate HLA molecules expressed on thymic cortical epithelial cells. Binding stimulates the T-cell through the TCR and triggers proliferation and survival signals. **(E)** Positively selected thymocytes lose the redundant co-receptor to become single positive (SP). **(F)** SP thymocytes which are exposed to diverse repertoires of self-antigen presented by cortical epithelial cells can be negatively selected if they show too high an affinity for self-antigen. **(G)** Those CD4⁺ T-cells which are hypothesised to express a slightly higher affinity for cell antigen are thought to become thymically derived T-regulatory cells (tT_{REG}). This results in a thymic output of T-cells which have an optimum affinity for only pHLA, but do not express too high an affinity for self-antigen.

1.1.5 The HLA-Restriction Paradox

The process of thymic selection creates an effective T-cell repertoire, which is somewhat paradoxical in nature. Put simply, how does this highly diverse population of TCRs, exhibit such limited specificity for *only pHLA molecules* and not the many other potential ligands that could possibly be recognised? Is this exceptional restriction imposed by the thymus, or does it arise before, in the genes that underlie these receptors?

Two models have been proposed to explain this paradox, yet neither is unanimously accepted and new publications present evidence in support of (and opposition to) either theory.

1.1.5.1 Germline-Encoded Model

This model, championed by Arrack, Kappler and Garcia²³, states that the germline contacts between the TCR and the HLA have been “evolutionarily refined” in order to enhance the probability of a successful interaction. In this model, the notion that CDR1 and CDR2 contacts are “opportunistic bystanders” to the CDR3 binding event is rejected. Instead they serve to orientate the TCR complex and allow for a highly specific mode of recognition. The concept is tackled from a structural perspective, arguing that TCR recognition is so focused that it must have evolved, rather than being externally imposed in the thymus, somewhat discounting the other interactions that take place at the immune synapse.

Work in support of this hypothesis involves identification of a germline interaction “codon” that mediates a conserved binding motif by the TRBV8.2 gene^{24–26} in many different systems.

1.1.5.2 Co-receptor Selection Model

The co-receptor model was first proposed by Alfred Singer in 2007²⁷. It rationalises the breadth of TCR diversity and its potential to see non-HLA ligands by postulating that HLA restriction is externally imposed by the CD4 and CD8 co-receptors.

As co-receptor engagement is necessary for signalling, this disables T-cell recognition of non-HLA ligands and gives the thymically generated repertoire its paradoxical specificity²⁸.

The paradox is dealt with by focusing on the cellular interaction of two membranes to facilitate recognition and TCR based signalling. This could be viewed as detracting from the unique structural importance of the TCR in recognition, negating the countless other contact-mediated cellular interactions that do not exhibit the diversity/specificity paradox. The primary piece of work in support of this theory came in the 2007 publication by Van Laetham et al. where knockout of the co-receptors (and HLA molecules) from transgenic mice resulted in widespread recognition of non-HLA ligands by the thymic emigrants²⁷.

Recent evidence involves germline CDR randomised mouse models²⁹ and further analysis of the Singer knockout repertoires³⁰. Other structural evidence is more in opposition to germline selection, rather than in favour of the co-receptor model, by documenting non-canonical binding modes in T-cell recognition^{31,32}. Such systems were very low affinity, in comparison to well-characterised TCR-pHLA interactions, and could possibly be regarded as the exceptions rather than the rule.

1.1.6 T-cell Activation

After leaving the thymus, T-cells migrate to the periphery where exposure to non-self, or malignant-self antigens may occur. This is mediated by professional antigen presenting cells (pAPCs) such as B-cells, dendritic cells and macrophages which endocytose pathogens and drain to the lymphatics for antigen presentation to the naïve and memory T-cell repertoires. In addition to classical adhesion molecules, pAPCs express both pHLA and costimulatory molecules at the cell surface. These are defined as “signal-1” (Fig 1.5A) and “signal-2” (Fig 1.5B) respectively, and are necessary for activation and proliferation to occur. A third event “signal-3” (Fig 1.5C), can determine the differentiation pathway taken by a naïve cell in order to generate functionally distinct CD4+ T-cell subsets.

Once “primed,” a T-cell can respond to antigen through TCR and co-receptor recognition alone (Fig. 1.5A). The receptor complex includes two transmembrane CD3 molecules, and localisation of intracellular TCR- ζ chains, both of which possess immunoreceptor tyrosine based activation motifs (ITAMs) that are readily phosphorylated³³. The intracellular domains of CD4 and CD8 co-receptors bind the src-kinase Lck^{34–36}, which can catalyse ITAM phosphorylation when in close proximity³⁷. The activation event centres on $\alpha\beta$ TCR and co-receptor binding to cognate pHLA with sufficient avidity to allow for phosphorylation of ITAMs to occur. This initiates signalling cascades that result in gene transcription and effector functions^{38,39}.

Either a sufficient TCR to pHLA “dwell time”⁴⁰, the rapid induction of multiple binding events⁴¹ or the segregation of large inhibitory molecules such as CD45⁴² at the immune synapse^{43–45}, are hypothesised to facilitate co-receptor mediated Lck phosphorylation of the intracellular TCR:CD3 complex. These models of T-cell activation are not necessarily exclusive. What is important is that they explain the balance between TCR avidity and cognate pHLA abundance that result in a vast range of dynamic recognition states.

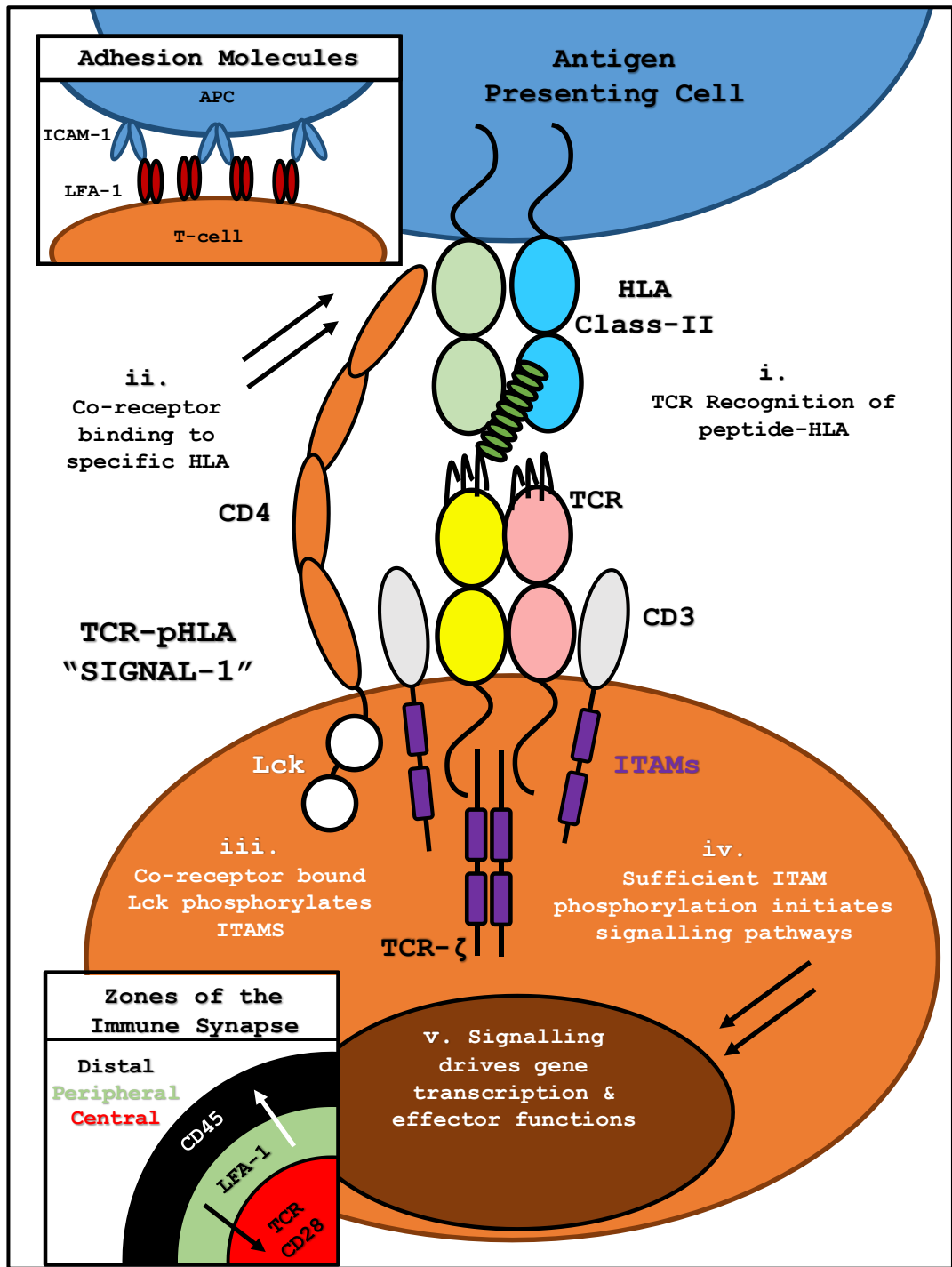


Figure 1.5.A. Events in T-cell activation, "Signal-1," schematic diagram specific for CD4+ T-cells. Full legend on page 19.

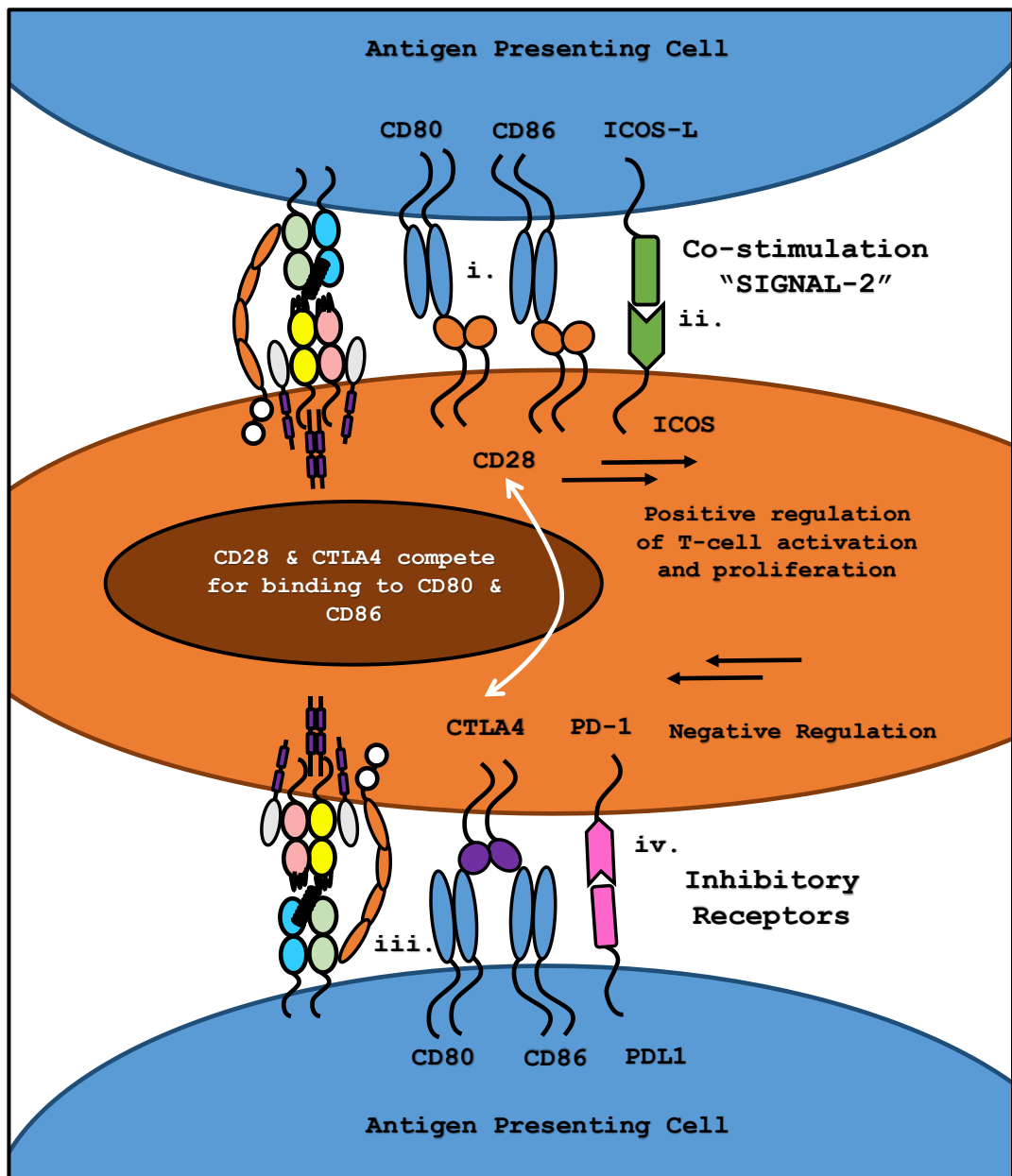


Figure 1.5.B. Events in T-cell activation, "Signal-2," costimulatory and inhibitory molecules at the immune synapse. Full legend on page 19.

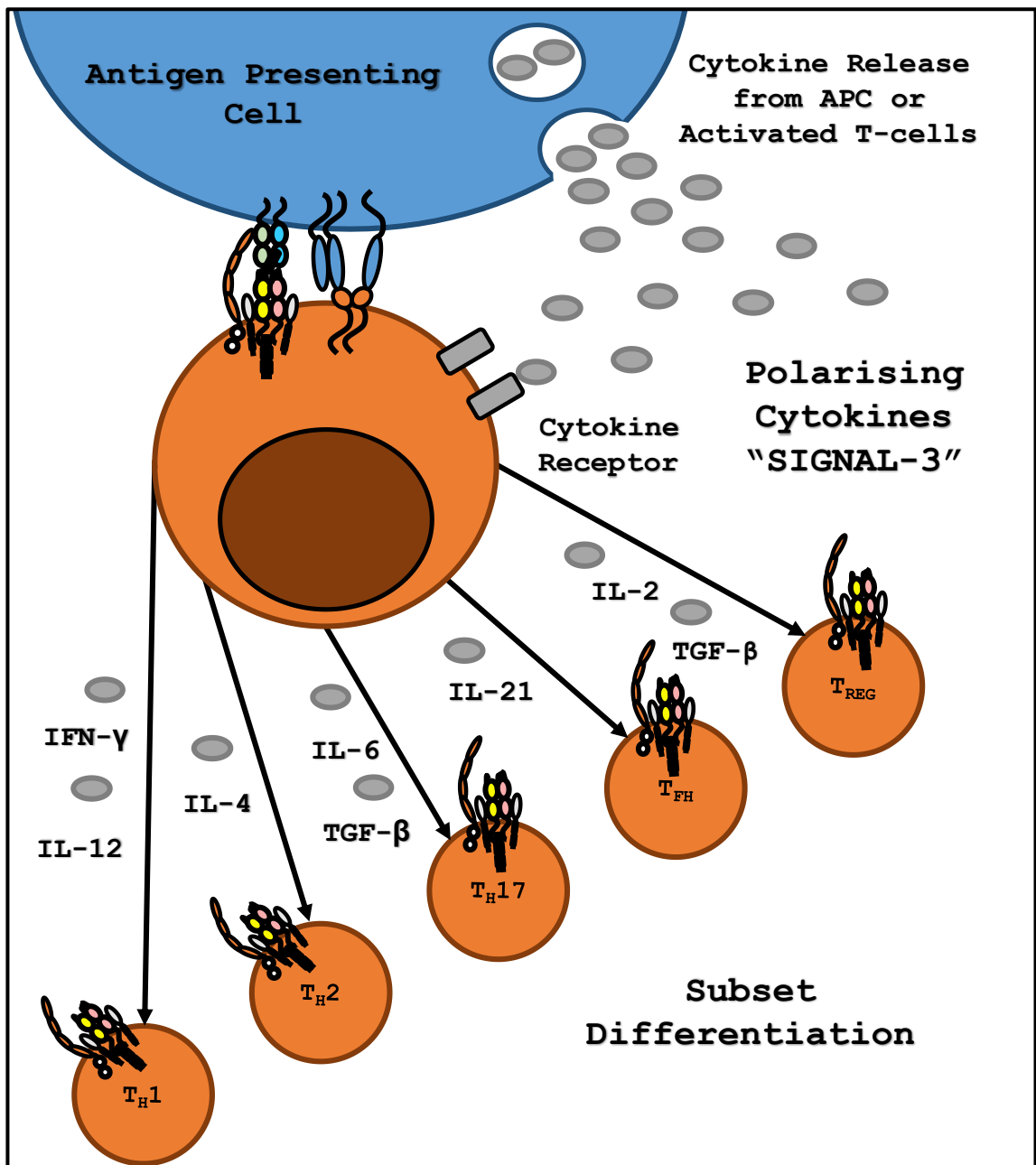


Figure 1.5.C. Events in T-cell activation, "Signal-3," the roles of polarising cytokines in $CD4^+$ T-cell subset differentiation.

During T-cell activation, the release of cytokine from other activated T-cells and APCs can influence the terminal differentiation of effectors into various $CD4^+$ T-cell subsets.

Figure 1.5.A. Events in T-cell activation, “Signal-1,” schematic diagram specific for CD4+ T-cells. (i) TCRs bind to HLA class-II on the surface of an antigen presenting cell. (ii) Binding of the TCR facilitates localisation of the CD4 co-receptor which binds to a site on HLA class-II. (iii) Localisation of CD4 brings the cytosolic tail carrying Lck kinase into close proximity with the CD3 cytoplasmic tail and the TCR-zeta chain, both of which are enriched for ITAMs (immunoreceptor tyrosine –based activation motif). These are phosphorylated by the action of Lck. (iv) ITAM phosphorylation facilitates binding of further kinases (ZAP-70) and the initiation of signalling cascades. (v) Signalling ultimately results in activation of transcription factors and elicitation of effector functions. **(Adhesion molecules insert)** The molecules ICAM-1 and LFA-1 facilitate cell-to-cell contact, loosely forming a zone of the immune synapse between the inner pHLA-TCR interaction zone and outer exclusion zone of the inhibitory CD45 molecule **(Zones of the immune synapse insert)**.

Figure 1.5.B. Events in T-cell activation, “Signal-2,” costimulatory and inhibitory molecules at the immune synapse. (i) CD28 on the surface of T-cells is bound by either CD80 (B7.1) or CD86 (B7.2) on the surface of antigen presenting cells. This interaction occurs alongside the TCR-pHLA interaction and is referred to as “costimulation” or “signal-2.” It is essential for activation of naïve precursors to become activated effectors. Additional examples of costimulatory molecules are ICOS and its ligand ICOS-L (ii), an interaction important in activation and development of T_H2 CD4+ T-cells. Inhibition of T-cell activation, is facilitated by upregulation of inhibitory molecules on the T-cell surface. (iii) CTLA-4 directly competes with CD28 for binding of CD80 and CD86, thus blocking costimulation and inhibiting activation. (iv) PD-1 is upregulated on exhausted T-cells and binds to its APC ligand PDL1, this is thought to deliver inhibitory signals to the T-cell and regulates further activation following TCR-pHLA engagement.

1.1.7 CD8+ T-cells

Following priming, T-cell activation results in proliferation and effector function. Generally, these effector functions involve release of cytokines that may act on surrounding tissues or other immune cells, as well as the direct induction of cell death. Effector functions, in addition to surface molecules and transcription factors are used to group T-cells into distinct subsets. The primary division exists between CD4+ and CD8+ T-cells.

CD8+ T-cells are generally cytotoxic and recognise pathogen infected or malignant cells through cytosolic epitopes presented by HLA class-I. The release of perforin and granzyme, or expression of Fas ligand (FasL) can induce apoptosis, killing the compromised target cell^{46,47}. Cytotoxic CD8+ T-cells are essential in clearance of viral infections and cancers, exhibiting many correlations with protection and recovery from disease⁴⁸⁻⁵⁰ as well as being successfully used in novel cancer immunotherapies⁵¹⁻⁵³. When compared to CD4+ T-cells, they are relatively simple in terms of effector function and subset distribution. This simplicity, and the availability of HLA class-I multimer reagents and crystal structures, means they are better understood at the molecular level.

1.1.8 Chemical and Lipid Mediated T-cell Recognition

As this project deals with protein antigens, two additional classes of T-cells will not feature in this introduction: $\gamma\delta$ T-cells and unconventional or invariant T-cells such as mucosal associated invariant T-cells and invariant natural killer T-cells⁵⁴. These cells have the ability to see chemical and lipid molecules presented by unique HLA-like molecules that are essential for immunity to many types of pathogens⁵⁵. Some of these cells can also recognise peptides, but it is their unconventional recognition that sets them apart from standard $\alpha\beta$ T-cells, making them an attractive target for novel structural and biophysical research.

1.2 CD4+ T-cells: Conductors of the Immune Response

CD4+ T-cells respond to HLA class-II peptides derived from exogenous material by antigen presenting cells. They do not act on the pathogen or malignancy directly (except as a rare cytotoxic subset) but instead elicit effector functions *through* macrophages, dendritic cells, B-cells and CD8+ T-cells. In addition to provoking inflammatory responses, they also act on these same cells to induce suppression, thus limiting inflammation and preventing immunopathology.

1.2.1 CD4+ Subsets: Protection and Disease

CD4+ T-cell subsets are defined by the cytokines they release and the groups of cells they act on. These subsets are hypothesised to arise from polarising cytokines⁵⁶, distinct antigen presenting cell subsets^{57,58} and the quantity of antigen encounter during priming^{59,60}. Each subset is linked to a specific aspect of protection as well as pathology, thus exhibiting the complex balance that CD4+ T-cells must maintain in a healthy immune system (Fig. 1.6).

1.2.1.1 T_H1

T_H1 cells (CD4+, TBX21+) are central to responses against viruses and malignancy. Interactions with dendritic cells can increase antigen presentation and licence APCs⁶¹ for priming of cytotoxic CD8+ T-cells. T_H1 stimulation of macrophages can cause destruction of pathogens contained within intracellular vesicles, and further stimulate presentation⁶².

In the viral setting, secretion of the inflammatory cytokines interferon- γ (IFN- γ) and tumour necrosis factor- α (TNF- α) can drive the localisation of CD8+ tissue resident effectors to sites of infection⁶³, as well as stimulate tissues to maintain an antiviral like state and the involvement of innate cells⁶⁴. Action of IFN- γ on T_{FH} cells stimulates the generation of an antibody response to further control free pathogen in the extracellular space. The T_H1 subset is associated with effective clearance of viral infections and was

one of the first defined subsets⁶⁵. If uncontrolled, these cells are primarily responsible for an IFN- γ , TNF- α “cytokine storm” and immunopathology associated with systemic infections^{66–68}.

1.2.1.2 T_{FH}

The T_{FH} (CD4+, BCL6+) subset, polarised by IL-6 and IL-21, is essential for the generation of an effective antibody response^{69,70}. Following recognition of B-cell presented antigen (cross-presentation), T_{FH} cells release IL-4 and IL-21 to stimulate class-switching and somatic hyper mutation of the BCR, which is subsequently released by plasma cells as antibody. Antibody recognition facilitates antibody-dependant cell-mediated cytotoxicity (ADCC), α -toxin neutralisation, macrophage mediated phagocytosis, complement activation, opsonisation of free pathogen and prevention of viral entry into host cells. Generation of such T_{FH} responses is essential for viral immunity^{71,72} and the primary goal of many vaccines^{73–75}. Unregulated, T_{FH} responses to self-antigens in systemic lupus erythematosus⁷⁶ and other antigens can drive autoimmunity.

1.2.1.3 T_{H2}

T_{H2} (CD4+, GATA3+) cells are polarised by IL-4 in the absence of IFN- γ , and are likely to have evolved in response to helminth infection^{77,78}. In the absence of helminth exposure these cells have pathogenic roles in allergy and delayed type hypersensitivity, releasing IL-4, IL-5, IL-13 and IL-6 in response to innocuous antigens^{79–81}.

1.2.1.4 T_{H17}

First recognised in 2006⁸², T_{H17} cells (CD4+, RORC+)⁸³ are generally characterised by the release of IL17 and IL-22. They are able to strongly activate and recruit neutrophils to sites of inflammation, and are associated with protective antimicrobial and antifungal specificity^{84,85}. They also have a close developmental relationship to suppressive CD4+ subsets^{86–88} which may be linked to their function.

Pathological roles for this subtype are found in gut inflammation⁸⁹, colon cancer^{90–92} and rheumatoid arthritis^{93–95}. Whether they are a symptom or driving factor of pathology is yet to be fully understood.

1.2.1.5 Cytotoxic CD4 T-cells

When attempting to rationalise the highly effective antiviral properties of some CD4+ T-cell subsets, populations able to directly kill infected cells were identified both *in vitro*⁹⁶ and *in vivo*⁹⁷ (2001 and 2008). Cytotoxic (T-Bet+ EOMES+) CD4+ cells^{98,99} are treated with some scepticism due to the inconsistency in how these cells direct their killing activity in the absence of widespread HLA class-II expression.

Work suggests that these are important in cancer and in viral infections, where class-II is upregulated on tumours¹⁰⁰ and respiratory epithelia respectively¹⁰¹. In the absence of class-II expression, such cells may induce cell death indirectly through contact mediated macrophage activation in the presence of IFN- γ ¹⁰⁰.

1.2.1.6 Regulatory T-cells

T_{REG} cells (CD4+, CD25^{HI}, FOXP3+) are essential in the maintenance of peripheral tolerance and prevention of autoimmunity. Although identified as “suppressors” in 1998^{102,103}, the transcriptional basis of suppression, FOXP3, was only identified in 2003¹⁸.

They are polarised through IL-2 and TGF- β ^{104,105} and can be further divided into thymic and peripheral groups depending on where the regulatory phenotype was induced. The mechanistic basis of suppression is highly complex and still to be fully understood. Examples include sequestering of IL-2 by CD25¹⁰⁶, inhibition of DC antigen presentation by CTLA-4¹⁰⁷, direct killing of effector cells¹⁰⁸, release of IL-10¹⁰⁹ and other interactions at the immune synapse.

Although protective against autoimmunity, these cells are enriched at tumour sites and may correlate with tumour growth and inhibition of an effective anti-cancer immune response^{110–114}.

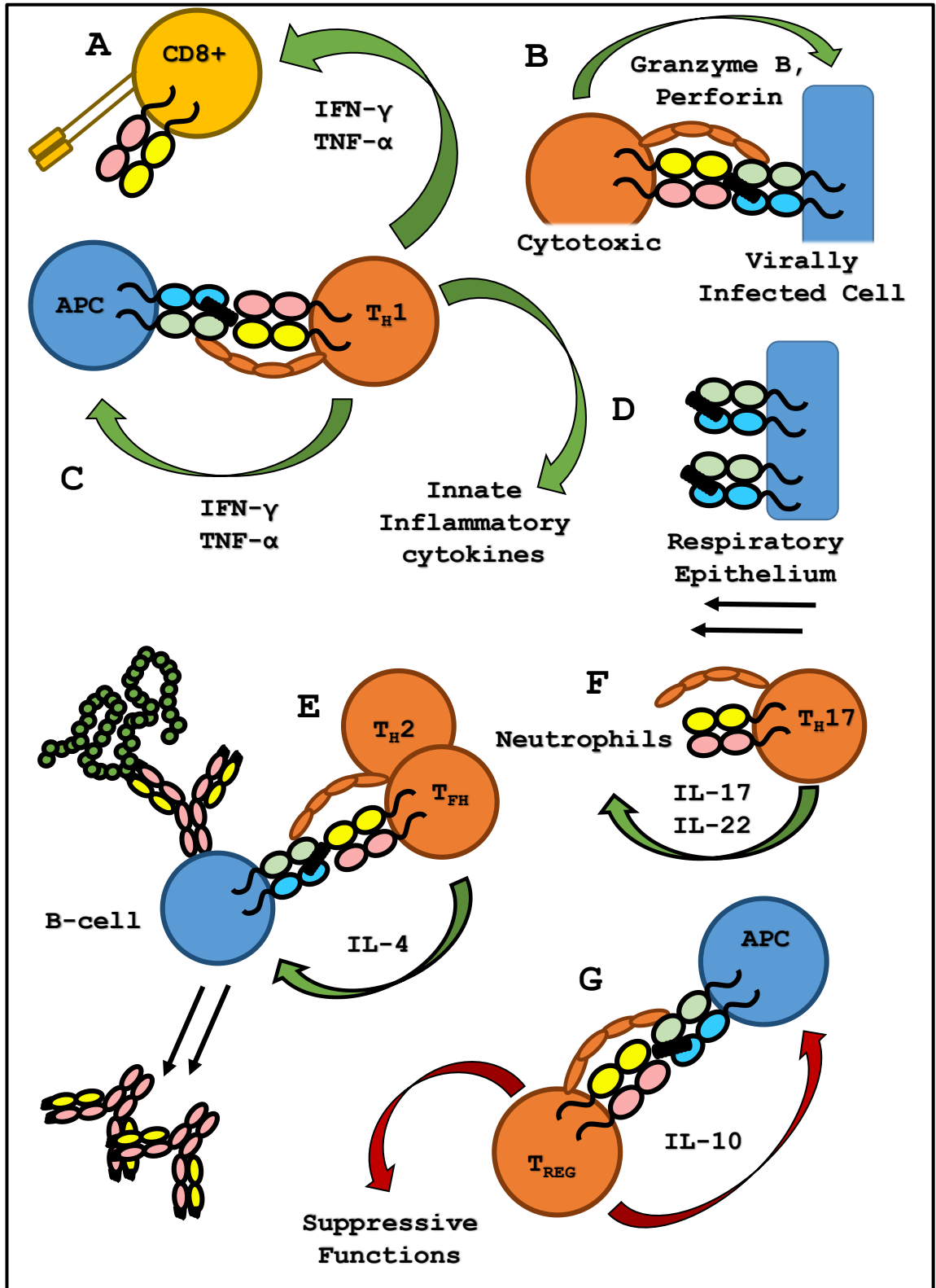


Figure 1.6. The roles of CD4+ subsets in the inflammatory immune response.

Figure 1.6. The roles of CD4+ subsets in the immune response.

(A) T_H1 cells stimulate activation of $CD8+$ T-cells through the release of $IFN-\gamma$ and $TNF-\alpha$. (B) Cytotoxic $CD4+$ T-cells can directly kill virally infected cells (and tumour cells) through recognition of class-II expressed at the cell surface by release of granzyme B and perforin. (C) Interactions with antigen presenting cells stimulate antigen uptake and presentation. (D) In addition, they release innate inflammatory cytokines and chemokines that can maintain an antiviral state and aid innate functions, as well as cause upregulation of class-II expression on epithelial cells. (E) An antibody response is stimulated by the action of T_H2 and T_{FH} cells which respond to B-cell presented class-II epitopes and release of IL-4. These cells are central to germinal centre formation and driving the processes of antibody class switching and somatic hypermutation. (F) T_H17 cells release IL-17 and IL-22 which increase neutrophil migration to the site of inflammation as well as promoting other inflammatory pathways. (G) In response to release of self-antigen in an inflammatory environment, regulatory T-cells can suppress immune responses by acting on antigen presenting cells and surrounding B- and T-cells through the release of IL-10 and other mechanisms.

The actions of these T-cell subsets have both positive and negative outcomes in different disease settings. They are both essential for protection from disease, but simultaneously may lead to pathology, unwanted immune suppression, hypersensitivity and autoimmunity.

1.2.2 Plasticity

The definitions of T helper subsets are not absolute. There are expanding numbers of recently discovered subsets such as T_R1⁸⁶, T_H9^{115,116}, T_H22¹¹⁷ and T_H3¹¹⁸ (not detailed here), as well as examples of subset interchange^{86,119,120} in both mice and humans. Recent findings suggest that lineage differentiation may be reversible under certain conditions, with certain cell surface markers identified that may indicate the former phenotype¹²¹.

This phenotypic “plasticity” may be a powerful mechanism by which the immune system maintains balance and homeostasis in response to dynamic threats.

1.2.3 Understanding CD4+ T-cells

The complexity of CD4+ subsets is paralleled by their protective, detrimental or unassigned roles in many common diseases. Their widespread absence in HIV due to viral targeting, untreated, leads to fatal compromise of the immune system by infections that are controllable in healthy persons. Yet their presence is linked to nearly every serious autoimmune disorder, as well as the failure of immune systems to rid the body of rapidly growing tumours.

In order to understand these cells, investigation at the molecular and genetic level is necessary to uncover the fundamental basis of each immune response. The first step is to look at the peptides they see, and the way in which they are presented. Following this, epitope-specific understanding and the molecular basis of immunity can be probed.

1.3 HLA Class-II

The unifying feature of highly diverse CD4+ T-cell subsets is that they recognise and respond to pHLA class-II through their TCR. Therefore, an understanding of the unique features of class-II mediated antigen processing and presentation is relevant to the elicitation of a variety of immune effector functions.

1.3.1 HLA Structure

In HLA class-I and class-II molecules, peptides are bound between two α helices above a β pleated sheet. This holds the peptide in an elongated conformation that allows linear sequence recognition by the TCR.

Class-I is asymmetric in composition, with a three-domain α chain, of which $\alpha 1$ and $\alpha 2$ form the binding groove, while the $\alpha 3$ domain is membrane proximal and includes a transmembrane region. A monomorphic $\beta 2$ -microglobulin associates with the $\alpha 3$ subunit to complete the structure^{122,123}.

Class-II is comparatively symmetric in composition, with individual α and β chains, each divided into two domains (Fig. 1.7). The $\alpha 1$ and $\beta 1$ domains form the binding groove, while $\alpha 2$ and $\beta 2$ associate with the membrane via their respective transmembrane regions¹²⁴.

Class-I binds peptides between eight and fourteen amino acids in length in a closed groove, accommodating two peptide “anchor” residues within binding pockets. Longer peptides form a peaked or bulged conformation¹²⁵ with the majority of HLA interactions at the termini. Class-I molecules are shown to be highly unstable in the absence of peptide¹²⁶.

HLA class-II molecules present peptides greater than nine amino acids in length in a flat conformation, with up to four anchor residues and significant backbone interactions that stabilise binding^{5,127}. The binding groove is open so there is no known structural limit on peptide length¹²⁸.

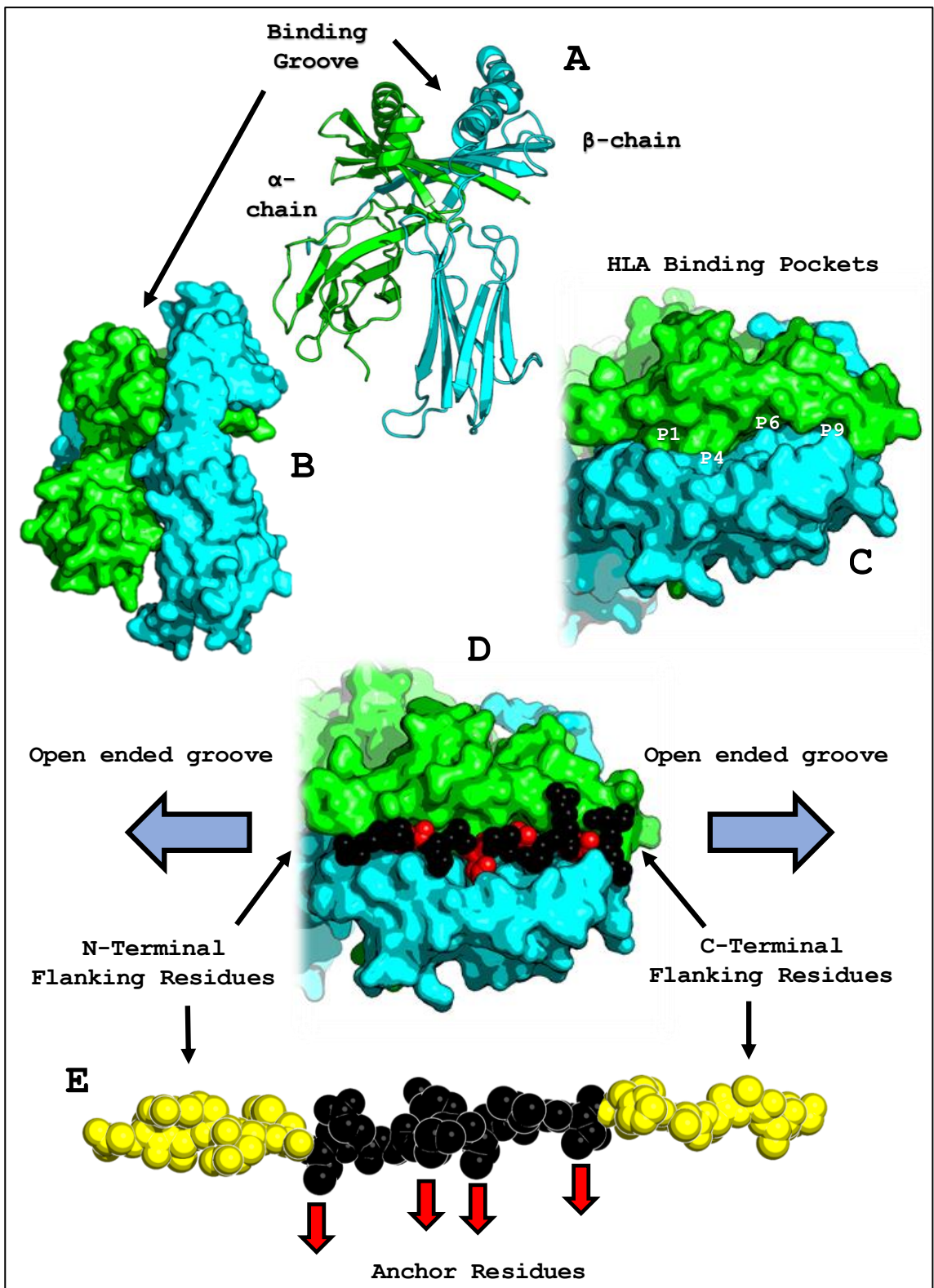


Figure 1.7. Characteristics of HLA Class-II (specific to HLA-DR1)

(A) Secondary structural representation of HLA class-II with two α -helices over a β -pleated sheet form the peptide binding groove. (B) Surface representation in the absence

of peptide with the empty binding groove clearly visible. **(C)** HLA-DR1 peptide binding groove viewed from the top down. Labels correspond to the peptide amino acids or “anchor residues” whose side chains sit within each pocket. **(D)** Peptide within HLA-DR1, anchor residues are coloured red. Arrows indicate that the groove is open and there is no limit on length of peptide which can be bound. **(E)** Definition of peptide flanking residues. Amino acids at the N- and C- termini beyond the first and last anchor residues (marked by red arrows pointing down) are coloured yellow. Amino acids within these regions are referred to as the flanks.

1.3.2 Class-II Expression and Antigen Processing

HLA Class-II expression is mainly limited to professional antigen presenting cells such as B-cells, macrophages and dendritic cells, but it has been shown to be upregulated in the epithelia¹²⁹ and certain cancers^{130,131}.

Class-II peptides are generally, but not exclusively, derived from membrane bound and extracellular sources¹³². Antigenic material captured by endocytosis and denatured through endosomal acidification, pH induced proteolytic cleavage, and disulphide reductase activity. These endosomes fuse with the antigen processing or "MIIC" compartment where they encounter HLA class-II^{133,134}.

Class-II molecules are synthesised from the endoplasmic reticulum with an invariant chain "CLIP" (Ii) bound to the peptide-binding groove that signals for trafficking to the MIIC compartment via the plasma membrane¹³⁵. HLA-DM catalyses the release of CLIP to allow antigenic peptides to bind and be presented at the cell surface¹³⁶.

Components of this pathway, such as the mode of endocytosis¹³⁷, the types of proteolytic enzymes^{138,139} and the expression of HLA-DM¹⁴⁰ vary between APC subsets. This means they can give rise to distinct peptide repertoires depending on cell type, tissue location and inflammatory state¹⁴¹; this may contribute to the diversity of CD4+ T-cell subsets.

1.3.3 Cross-Presentation

A linear view of class-II antigen presentation raises conceptual problems of how extensively CD4+ T-cells mediate the response to antigens that are not prevalent in the extracellular space. If presentation is exclusive to the exogenous pathway, are virally infected cells, or intact tumour cells *incapable* of stimulating an immune response through HLA class-II?

A body of evidence¹⁴²⁻¹⁴⁴ and mechanistic theory pioneered by Laurence Eisenlohr^{145,146} has addressed this question. In 2005 it was shown that viral class-II

restricted epitopes can be completely dependent on the proteasome and TAP (key elements of the cytosolic pathway) for presentation¹⁴⁷. This was followed by a 2015 publication where infection of APCs with the influenza virus was essential to account for the full cohort of CD4+ T-cell antigen specificity observed in response to live attenuated virus¹⁴⁸

While cross presentation is a relatively old observation (1976)¹⁴⁹ and widely accepted concept in respect to CD8+ T-cells¹⁵⁰, it requires further scrutiny in CD4+ T-cells where understanding is limited to a few examples.

1.3.4 HLA-DM and HLA-DO

HLA-DM was first identified as a positive regulator of class-II peptide presentation in a series of publications between 1994-1995^{136,151–154}. It was not until three years later that HLA-DO was identified as a negative regulator of class-II presentation^{155,156}.

Crystal structures of HLA-DM alone^{157,158}, with HLA-DR¹⁵⁹ and with HLA-DO¹⁶⁰ have helped uncover the molecular basis of their function as well as demonstrate their close structural morphology when compared side by side¹²⁷.

HLA-DM facilitates release of the invariant chain from virgin class-II molecules by associating with the class-II α -chain close to the peptide N-terminal face and inducing a conformational change that destabilises the P1 binding pocket to open the binding groove¹⁵⁹. DM stabilises this open or “peptide receptive” conformation through several hydrogen bonds and salt bridge interactions. Upon the release of CLIP, antigenic peptides can freely associate and dissociate with the class-II binding groove in an exchange equilibrium (Fig. 1.8).

Adhering to thermodynamic principles, this equilibrium will tend towards the lowest energy state, i.e. binding of peptides that exhibit the strongest molecular interactions with HLA class-II. As a result, the most stable peptide and class-II combinations will predominate¹⁴¹ and ultimately be trafficked to the surface of cells from the MIIC compartment for presentation.

HLA-DM is upregulated by APCs in response to the T_H1 cytokine IFN- γ , therefore driving presentation of high affinity pathogenic (and self) peptides in the inflammatory state¹⁶¹. Peptide loading in the absence of HLA-DM may also occur (Fig. 1.9). This takes place in the “recycling pathway,” where a peptide receptive conformation is induced by acidification in endosomes close to the cell surface^{162,163}.

In the non-inflammatory state HLA-DO inhibits the activity of DM by binding at the same site as used to bind class-II, thus competing for occupancy¹⁶⁰. This competitive inhibition can be overcome by an increase in the expression of DM, triggered by IFN- γ ¹⁶¹, altering the peptide repertoire in the inflammatory state¹⁶⁴.

As DM and DO are differentially expressed in macrophages and specific subsets of dendritic cells¹⁴⁰, the DM:DO ratio has been used to explain the differences in antigen presentation by distinct populations of APCs¹⁴⁰. The outcome of this interplay is to tightly regulate the potential peptide agonists of highly diverse CD4+ T-cell subsets, thus influencing the resulting immune response.

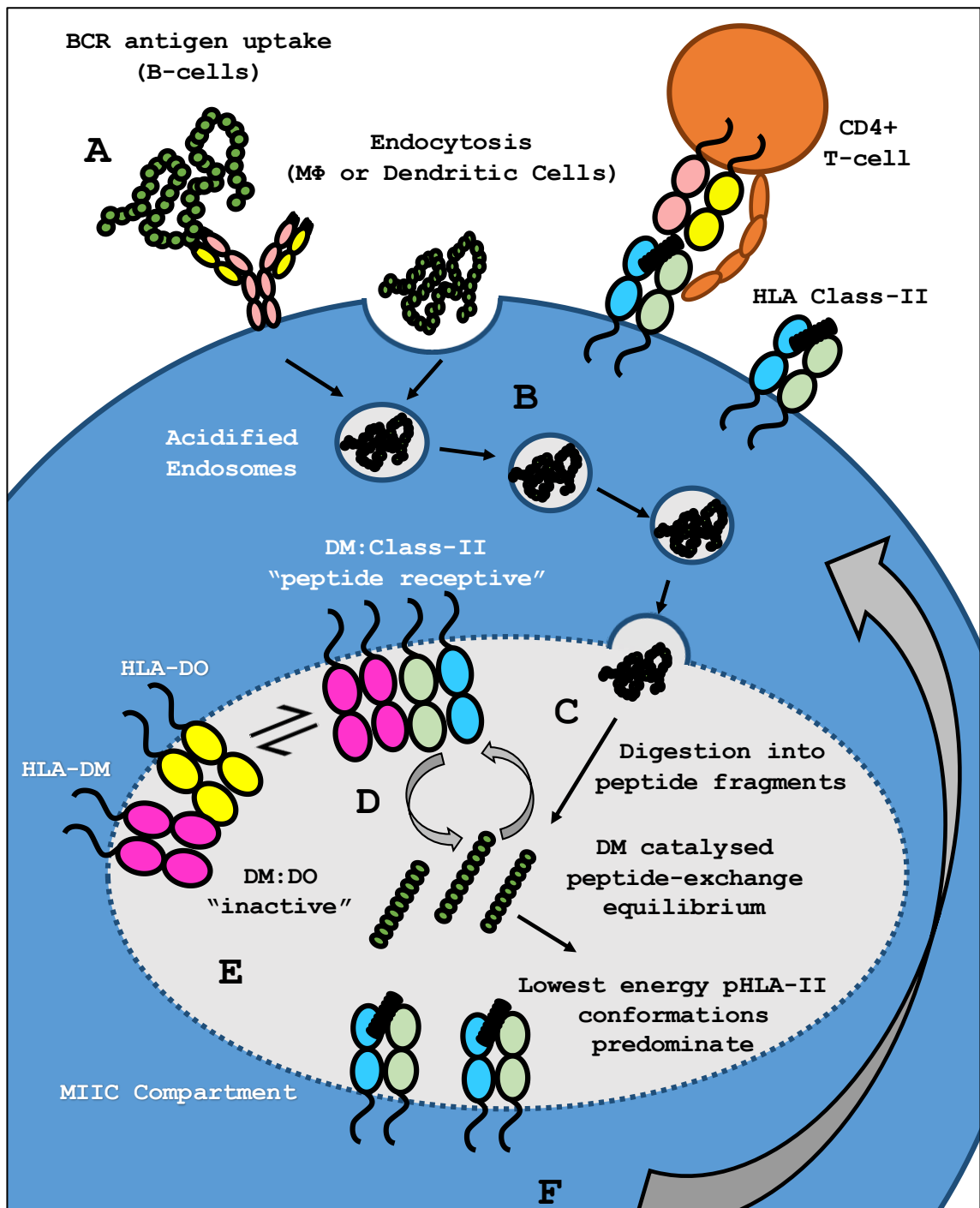


Figure 1.8. The endosomal class-II presentation pathway and HLA-DM peptide selection.

(A) Antigen gets endocytosed from the extracellular space into an acidified endosome. (B) Endosomes are trafficked to the MIIC compartment which is enriched for HLA class-II, molecular chaperones and proteolytic enzymes such as cathepsins. (C) Protein

antigen is digested into peptide fragments which can bind to HLA class-II when in the peptide receptive conformation, reliant on the binding of HLA-DM. (D) The action of DM facilitates rapid peptide exchange within the class-II binding groove which reaches equilibrium with complementary peptides. Strongest binding peptides will show increased occupancy of the groove. (E) HLA-DO competitively inhibits the action of HLA-DM by binding at the class-II site and holding the DM in an inactive form. (F) Upon release of HLA-DM, the stable form of HLA class-II is trafficked to the cell surface for presentation to T-cells.

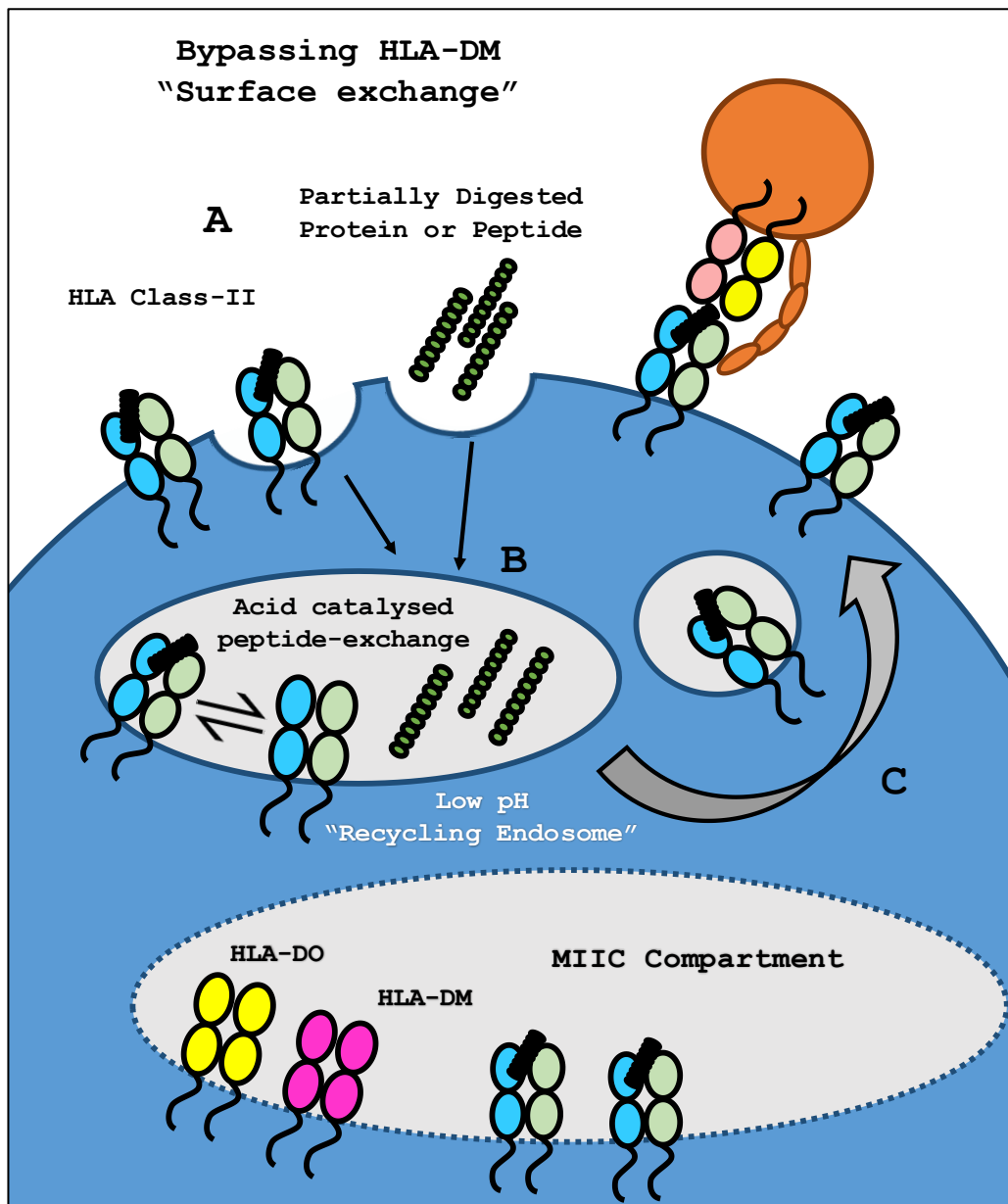


Figure 1.9. Loading of peptides onto HLA class-II in the absence of HLA-DM.

(A) Protein that is not intact, or fully folded can be endocytosed with HLA class-II at the cell surface. **(B)** Endosomes fuse to form the "recycling endosome" which contains both protein material and HLA class-II. As this compartment is acidified it facilitates opening of the class-II binding groove and release and binding of peptide. **(C)** This class-II can traffic back to the cell surface without having been exposed to DM mediated peptide selection. This is relevant where cells are exposed to partially digested extracellular

material or excess of synthetic peptide, and also where knockout cell lines do not express HLA-DM.

1.3.5 Class-II Peptide Presentation

Following processing; peptides are accommodated within the open-ended class-II groove with no known restrictions on maximum presentable length¹⁶⁵ (the minimum being nine residues). This makes the definition of a class-II epitope more complex, as there is no clear point at which amino acids beyond the first (P1) and last (P9) anchors, no longer contribute to either TCR binding or pHLA complex stability. These amino acids are referred to as peptide flanking residues (PFRs).

Crystallographic^{166,167} and biophysical¹⁶⁸ studies have highlighted interactions between the flanks and cognate TCRs to occur at position P-1 (minus one) and P11 amino acids, but investigation into highly extended flanks has not been carried out. The natural peptide repertoire presented by HLA class-II does not provide insight into either the necessity or superfluity of flanking residues.

Acidic elution of peptides from class-II molecules on the surfaces of antigen presenting cells has identified diverse “nested sets” of peptides that differ in length but contain a common nine amino acid core encompassing the four anchor residues (P1, P4, P6 & P9)^{169,170}.

This suggests that antigen processing and presentation does not produce a strictly controlled length of peptide, unlike HLA class-I, and the resulting contribution of these length variations to pHLA stability, T-cell activation and repertoire selection is largely unknown.

1.4 The Influenza Virus

Influenza is an acute viral infection that can vary in severity from mild illness to hospitalisation and death, representing a continual burden to healthcare and society. The hallmark of Influenza is its highly infectious nature, coupled with strain diversity, the capacity for rapid viral evolution and existence of zoonotic reservoirs that harbour sleeping threats.

1.4.1 Seasonal

Circulating strains of the virus are defined as “seasonal flu.” They are variable and in a state of continual evolution driven by immune selection pressure. The majority of the human population is thought to encounter these on a frequent basis, and have some pre-existing immunity, the efficacy of which varies markedly from person to person¹⁷¹. Peak incidence of infection occurs during winter, and necessitates the provision of an annual flu vaccine to mitigate the impact on healthcare and vulnerable patients.

In the UK between week 40 of 2015 and week 17 of 2016, “flu season,” there were 2,462 influenza confirmed hospital admissions and 209 deaths¹⁷². Each year the virus is responsible for between 250,000 to 500,000 global deaths¹⁷³ and is particularly dangerous to those who are immunocompromised.

1.4.2 Pandemic

Pandemic strains of the virus are those to which a lack of pre-existing immunity results in geographically widespread infection and disease. These strains do not circulate in society and are likely to originate in zoonotic reservoirs, of which wild fowl, domestic poultry and swine are documented to have crossed the species barrier into humans¹⁷⁴.

Crossing the species barrier requires close contact with the animal or carcass, the probability of which is increased by intensive farming practices. After infection has occurred, airborne transmission between humans is essential for a pandemic to be

possible. This catastrophic step is held in check until a zoonotic virus is able to mutate its entry protein and facilitate easy transmission via human respiratory epithelium¹⁷⁵. In zoonotic reservoirs, the virus is an infection of the gut, and thus enters cells by a different sialic acid moiety¹⁷⁴.

Between April 2009 and August 2010, the swine flu pandemic was estimated to have caused 284,500 deaths, only 20 % of which were in individuals over 65 years old¹⁷⁶. In comparison, 35 % of hospitalisations in 2015/2016 associated with seasonal flu in the UK concerned those over 65 years old¹⁷².

1.4.3 Strain Classification

The term, Influenza, specifically refers to three distinct genera of the Orthomyxoviridae family: Influenza-A, -B and -C. Influenza-B is isolated to humans and can cause seasonal epidemics (13.8 % of UK hospitalisations 2015/2016), while Influenza-C results in mild infection and is not deemed necessary to vaccinate against. Neither -B nor -C is associated with pandemics, or has the rapid evolutionary capacity and seasonal impact of -A (86.2 % of UK hospitalisations 2015/2016).

Influenza-A is highly variable and further divided into subtypes based on the surface antigens haemagglutinin and neuraminidase. The subtype (p)H1N1 was responsible for the 1918 Spanish flu and 2009 swine flu pandemics, while H3N2 was responsible for the 1968 Hong Kong outbreak. Currently H1N1, pH1N1 and H3N2 are widely circulating in society and account for seasonal infections^{172,173}.

“Avian flu” strains including H5N1¹⁷⁷ and H7N9¹⁷⁸ have been tipped as sources of the next global pandemic, having been responsible for sporadic outbreaks of extremely high mortality^{179,180}. So far, human-to-human transmission of these strains is yet to be observed in more than a few isolated cases¹⁸¹

1.4.4 Viral Structure

The influenza virion is composed of eleven proteins encoded on eight strands of negative sense viral RNA. Three external proteins: Haemagglutinin (HA), neuraminidase (NA) and the matrix-2 ion channel (M2) mediate viral entry, exit and pH mediated breakdown of the virion, respectively¹⁸².

The eight remaining influenza proteins are internal. These consist of nucleoprotein (NP), matrix-1 (M1), the polymerase complex (PA, PB-1, PB-2, PB1-F2) and non-structural proteins (NS1 and NS2).

NP binds viral RNA in a deep groove lined with basic amino acids, forming an extended oligomeric ribonucleoprotein structure^{183,184}. The purpose is to pack the RNA into a higher order structure, facilitate exit from the host cell nucleus (following transcription) and form interactions with the polymerase complex¹⁸⁵.

The M1 protein is essential to virion structure, forming a polymeric helical inner coat or “net” that associates with HA and NA cytoplasmic tails¹⁸⁶ and the ribonucleoprotein complex in the cytoplasm¹⁸⁷. This helps to form a stable network of interactions and is hypothesised to facilitate budding and virion formation¹⁸⁸.

The first crystal structure of a fully assembled influenza-A polymerase (PA, PB-1, PB-2 and RNA promoter) was obtained in 2014^{189,190}. The polymerase complex can synthesise both positive sense mRNA for translation, and negative sense RNA for genome replication^{191,192}.

NS1 is a virulence factor, involved in disrupting pathways of innate immunity, but it is not present in the virion¹⁹³. NS2 is present and interacts with the ribonucleoprotein complex to facilitate import and export from the nucleus of the host cell¹⁹⁴

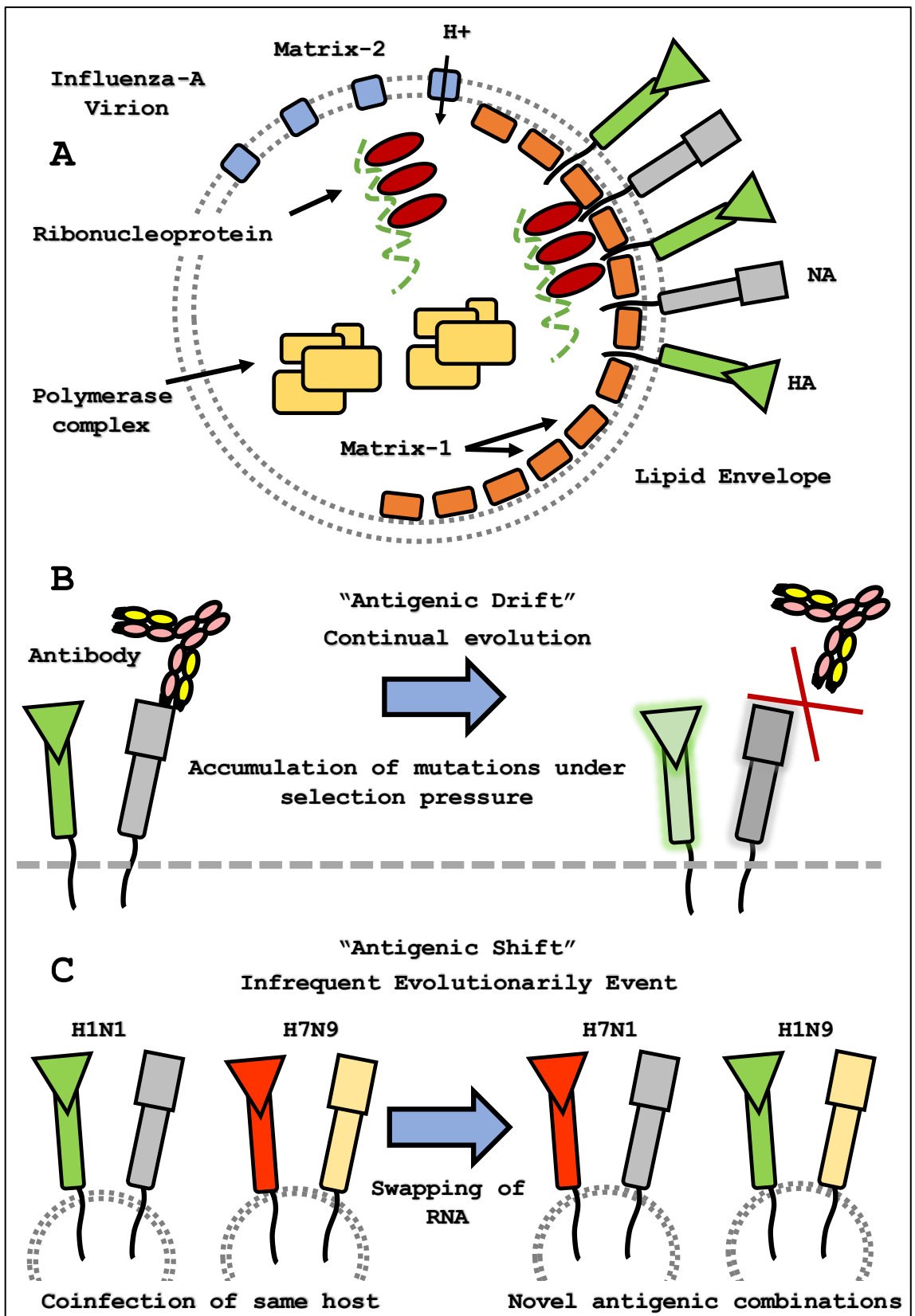


Figure 1.10. The Influenza-A virion and mechanisms of viral evolution.

Figure 1.10. The Influenza-A virion and mechanisms of viral evolution.

(A) Overview of the important components of the influenza virus in this study. Non-structural protein is not detailed. **(B)** Schematic representation of antigenic drift, the process which results in the incidence of seasonal epidemics of flu. Surface proteins HA and NA gradually mutate to maintain or increase viral fitness in response to immune and replicative selection pressures. **(C)** Antigenic shift requires coinfection of multiple strains in a single host in order to allow swapping of RNA strands and the expression of new combinations of previously unseen viral subtypes. This event can lead to pandemic strains of influenza under certain circumstances.

1.4.5 Life Cycle

The virus enters host cells by HA mediated binding of sialic acid and the initiation of endosomal uptake. Acidification of the endosome causes the release of virion contents and the synthesis of positive sense RNA from the viral negative sense genome by the accompanying polymerase. Hijacking of host machinery results in translation of viral proteins and replication of the viral genome. These components are assembled at lipid rafts in the cell membrane where newly formed virions can leave the cell in a process known as “budding,” which involves cleavage of bound sialic acid by NA^{182,195}.

The external proteins HA, NA and M2 serve as antibody targets for the immune system. An effective antibody repertoire can prevent HA-mediated entry to the cell, and opsonise free virus in the periphery thus neutralising its infectious capacity and facilitating phagocytic destruction¹⁹⁶. Breaching of this serological protection is the first step in a pathological infection.

Serological specificity to HA and NA is responsible for the commonly used viral nomenclature. These proteins were grouped according to the antibodies by which they were recognised, for example H1, H2, H3 and N1, N2, N3. Within the H1 group there may be thousands of sequence variants, but these were historically seen by the same “H1 antibodies.” Today genetic sequencing provides a much more complex, but accurate, picture of influenza phylogeny.

1.4.6 Antigenic Drift

Antigenic drift describes the minor mutations of viral proteins that accumulate over successive replication events and person-to-person transmission. Mutations that increase viral fitness in response to pre-existing immune selection pressures are positively selected and become prevalent in the gene pool (Fig. 10B). This advantageous phenomenon is possible because of polymerase infidelity, which imparts a coding error every ten thousand nucleotides¹⁹⁷.

Antigenic drift is largely responsible for the persistence of seasonal influenza in the population, and the limited effectiveness of predictive based vaccine design by the WHO^{198,199}. Although all proteins should be equally impacted, it has been shown that some proteins are much more prone to mutation than others²⁰⁰. These include the external proteins HA and NA, which are under the greatest antibody-based selection pressure, and are perhaps more tolerant to mutation due to their singular roles.

As the internal proteins are not subject to the antibody response, this may explain their conservation. In addition, many of these proteins exhibit poly-functional and interdependent roles. The combined importance of M1, NP and the polymerase complex to virion structure, organisation, stabilisation and replication of the viral genome as well as virion assembly prior to budding may limit their mutational tolerance, even under T-cell based selection pressures.

1.4.7 Antigenic Shift

Pandemic events are primarily associated with a drastic change in viral protein expression that abrogates the pre-existing antibody response. This is known as antigenic shift and occurs when different Influenza-A subtypes (potentially interspecies) co-infect the same host and assimilate complete strands of each other's genetic material²⁰¹, termed reassortment (Fig. 1.10C).

This can give rise to novel antigenic combinations that are not recognised by the human population. If the new strain exhibits transmission and infectivity rates that are comparable to seasonal strains, then global pandemic may ensue.

This was the mechanism hypothesised to explain the origin of swine flu in 2009, where up to three strains were thought to have recombined in the domesticated pig population^{202,203}. The origin of the virus was subject to well publicised debate^{204,205}, and generated a global outcry for monitoring of influenza in swine populations.

1.4.8 Original Antigenic Sin

One of the most fascinating immunological observations is the concept of original antigenic sin. First described in 1953 by Thomas Francis Jr²⁰⁶, the concept states that when an individual is first exposed to influenza-A, they produce a balanced antibody response to viral antigens. When infected later in life with virus that has undergone antigenic drift or shift, the immune response is fundamentally unbalanced, with antibody skewed towards the epitopes first recognised during primary infection.

Most significant is that if a third variant of virus infects the same individual, their immune response is still unbalanced in favour of the epitopes of primary infection. Epitopes encountered in the secondary infection still do not appear to be evenly adopted into the immune response. Hence, human immune systems appear forever “tainted” by their first encounter, and the analogy of original sin becomes clear. Several recent studies have found evidence that both supports²⁰⁷ as well as refutes this concept²⁰⁸.

Although associated with humoral immunity, observations of original antigenic sin in T-cells are prevalent^{209,210}. Yet distinguishing the molecular basis of antigenic sin from that of T-cell cross-reactivity^{211,212} is an experimental and conceptual challenge. Harvesting the benefits of the latter, while negating the wastefulness of the former could be a goal in forthcoming T-cell vaccination research.

1.5 Aims and Hypothesis of the Thesis

The interplay between the immune system and the influenza virus is a fascinating topic in immunology and disease. The periodic incidence and pandemic potential discussed in previous sections are made possible by the viral evolution mechanisms that evade the immune response. Seasonal vaccination is largely successful at preventing epidemics, but offers little protection against novel pandemic strains or in situations where the vaccine strains have been incorrectly matched to the environmental strain (for example in 2015/2016).

The internal elements of the virus, discussed in section 1.4.4, exhibit high levels of conservation and offer a more consistent target for the immune system amidst the diversity of potential strains. T-cell responses to these proteins have been shown to correlate with protection and decreased symptom severity during infection^{48,213}. Yet the epitopes and T-cell receptor repertoires that underpin these important responses have not been analysed in detail at the molecular level.

CD4+ T-cell recognition and responses to three of these internal proteins, M1, NP and PB-1 are the subject of this thesis. The questions driving this investigation have focused on how many robust and reproducible HLA-restricted epitopes were contained within these proteins, and to what extent they were recognised in multiple individuals of the same HLA-type? Following this, what did the magnitude and clonotypic architecture of responding cell populations look like, and given their importance in protection were such response characteristics shared across the population?

Hypothesis:

CD4+ T-cell responses to the internal proteins of the Influenza virus will be detectable in multiple individuals and generate strong cellular and genetic signatures shared across the population.

Specific aims:

- 1) To isolate HLA-restricted CD4+ T-cell responses to the conserved internal proteins and characterise these to the epitope level.
- 2) To examine the nature of responses to these epitopes in multiple individuals sharing a common HLA allele.
- 3) To analyse the underlying clonotypic characteristics of these responses in order to understand the genes and amino acids that mediate CD4+ T-cell immunity and assess whether these are shared across the population.
- 4) To align clonotypic data with structural data to explain the conserved features observed in epitope-specific repertoires.

2 Methods

2.1 Generation of DR1.APCs

2.1.1 List of General Reagents

A5 media: RPMI 1640 (Gibco), 5 % human AB serum (heat inactivated, acquired from Cardiff University Hospital Wales), 2 mM L-glutamine, 100 IU/mL penicillin, 100 µg/mL streptomycin

R10 media: RPMI 1640 (Gibco), 10 % FCS (heat inactivated, Gibco), 2 mM L-glutamine, 100 IU/mL penicillin, 100 µg/mL streptomycin

R5 media: same as above with 5 % FCS (Gibco)

R0 media: same as above in the absence of FCS

D10 media (HEK293T cells only): Dulbecco's Modified Eagle Medium (DMEM; Gibco), 10% FCS, 2 mM L-glutamine, 100 IU/mL penicillin, 100 µg/mL streptomycin

CD4+ T-cell media: RPMI 1640 media, 40 IU/mL IL-2 (human recombinant, Proleukin), 10 % FCS, 2 mM L-glutamine, 100 IU/mL penicillin & 100 µg/mL streptomycin, 0.02 M HEPES, 1mM non-essential amino acids, 1mM sodium pyruvate

All above reagents acquired from Life Technologies unless stated

FACS buffer: PBS, 2 % FCS

TE Buffer: 10 mM Tris, 1mM EDTA, pH 8.0

HEPES buffer: 2.5 mM HEPES, pH 7.3

Calcium chloride solution: 2.5 M CaCl₂

HEPES buffered saline: 280 mM NaCl, 50 mM HEPES, 1.5 mM Na₂HPO₄, pH 7.0

ELISpot plate: Millipore multiscreen filter plate MSIPS4510

IFN-γ ELISpot kit: Mabtech basic kit, IFN-γ capture mAb (1-D1K), biotinylated detection mAb (7-B6-1), Streptavidin-ALP

Anti-HLA-DR antibody: L243 (0.05 mg/ml, Biolegend)

2.1.2 Cell Culture

Human (Epstein-Barr virus transformed) B Lymphoblastoid cell lines (B-LCL), T2 and 721.174 (full description and origin detailed in Chapter 3) were cultured in R10 media at 37 °C, 5 % CO₂ and replenished with fresh media regularly (3-5 days, split 70 %). T-cell clones were cultured in CD4+ T-cell media (3-5 days, split 50 %). PBMC lines were cultured in A5 media, controlled for human AB serum batch. Lines were only exposed to a single batch of AB serum, mixing of batches, or culture in different batches of serum was strictly avoided.

2.1.3 DR1 Construct and Lentivirus Production

HLA-DR1 gene construct was ordered from Genewiz. The construct contained HLA-DR1 α and β domains detailed in section 2.3.2, separated by a P2A self-cleavage site, and Age1 and Sal1 restriction sites. This construct was cloned into the pRRLSIN.cPPT.PGK-GFP.WPRE transfer vector (2nd generation), before calcium chloride transfection with the packaging vectors pMD2.G and pCMV-dR8.74 (obtained from Dr. Bruno Laugel of Cardiff University) into human embryonic kidney 293T cells.

Virus was packaged using 2 million 293T cells in 20 mL D10 medium (estimated at 60-80% confluency assessed 4 hours before transfection). 550 μ L TE buffer was combined with pMD2.G (13 μ g), pCMV-dR8.74 (24 μ g) and HLA-DR1 transfer vector (18.75 μ g) and 190 μ L CaCl₂. 1.9 mL of HEPES buffered saline was added dropwise with agitation (vortex) and incubated (15 - 25 min, RT) to facilitate precipitate formation.

Virus solution was added to 293T cells dropwise under agitation, and cells were incubated overnight (37 °C, 5 % CO₂). Media was replaced (17 ml split) at 16 h then supernatant collected 24 hours and 48 hours later. Viral containing supernatants were filtered (0.45 μ m), 24 and 48-hour fractions pooled and concentrated by ultracentrifugation at 150,000 g (2 hours, 4 °C) in sterilised ultra-clear ultracentrifuge tubes (Beckman Coulter). The lentivirus pellet was resuspended in 1 mL of R10.

Lentivirus preparations were aliquoted (100 µL), snap frozen and stored in lentiviral designated freezer location.

2.1.4 Assessment of DR1 Expression by Flow Cytometry

All flow cytometry was carried out on a BD Sciences FACS Canto. Cells were stained with primary mouse anti-HLA DR antibody L243 (0.05 mg/ml, Biolegend) for 20 min on ice in FACS buffer. Cells were washed twice (FACS buffer, 400 g, 5 min) before incubation with a secondary goat anti-mouse PE (0.02 mg/ml, BD Pharmingen) for 20 min on ice followed by two washes.

2.1.5 Infection of “Naked” Cell Lines with DR1 Lentivirus

The cell lines 721.174 and T2 (174 x CEM.T2) were kindly provided by Prof Awen Gallimore and Dr Garry Dolton respectively. 5 million of these cells were incubated with 100 µL of DR1 lentivirus in a 24 well plate overnight. Media was removed at 16 hours and replaced with fresh R10 (to minimise exposure to virus). DR1 expression was assessed after one week, as described in section 2.1.4.

Populations were then enriched for DR1 expressing cells by magnetic activated cell sorting (MACS, Miltenyl Biotech) using the L243 antibody, goat anti-mouse FITC (BD Pharmingen) and anti-FITC micro beads (Miltenyl Biotech) following manufacturers protocol. Clones were obtained by limiting dilution of the DR1 enriched populations.

2.1.6 IFN-γ ELISpot with CD4+ T-cell Clones

The T-cell clones DCD10 (specific to a DR1-restricted influenza epitope HA₃₀₆₋₃₁₈) and GD.D104 (specific to a DR1-restricted cancer epitope 5T4_{xxx-xxx} peptide sequence not disclosed)²¹⁴ were “rested” (24 hours, 37 °C, 5 % CO₂) prior to the assay in R-5 media. Peptide pulsing consisted of incubation (2 hours, 37 °C, 5 % CO₂) at chosen peptide concentrations in assay media. Cells were then washed in either R0 or PBS four times in order to remove unbound peptide from the media, thus limiting T-cell to T-cell presentation.

Per ELISpot well 50,000 peptide-pulsed APCs were used. T-cell numbers (200 – 400 per well) and peptide concentrations were varied in order to optimise assay conditions. Overnight incubation was carried out on an ELISpot plate coated with the IFN- γ capture antibody using the Mabtech IFN- γ ELISpot kit. Development of plates was carried out following manufacturer's protocol and spot forming cells are counted using the AID GmbH plate reader and software. For later assays a more advanced plate reader was used (see section 2.2.7).

2.2 Screening of Donor PBMC against Peptide Libraries

2.2.1 Peptide Libraries

Peptide libraries were obtained from GL Biochem (Shanghai) Ltd as 20-mers in the crude form. Peptide sequences overlapped 10 amino acids (for full library details see appendix section 8.1). Original sources of the sequences are as follows:

Matrix Influenza A virus (A/Wilson-Smith/1933(H1N1) 252 amino acids), 24 overlapping peptides.

Nucleoprotein Influenza A virus (A/Ck/HK/96.1/02 (H5N1) 401 amino acids), 39 overlapping peptides.

PB1 Influenza A virus (A/Puerto Rico/8/1934(H1N1) 757 amino acids), 74 overlapping peptides.

2.2.2 Peptide Pools

Lyophilised peptides were re-suspended in DMSO (stock concentration of 20 mg/mL). 1 μ L of the designated peptides were added to PBS to make a final volume of 40 μ L. Individual peptides were always at a concentration of 0.5 mg/mL, whether alone or in a pool. 2 μ L of PBS stock was added to give a working concentration of 10 μ g/mL (estimated as 5 μ M for a 20-mer).

2.2.3 Processing of PBMC

PBMC were obtained from local donors, anonymous donor coding and consent taken and recorded. Assays were carried out in order to ensure each screen (per protein) is carried out on a different bleed. No PBMC from before 2014 was tested.

20-50 ml of fresh blood taken by venepuncture (done by a trained phlebotomist), was transferred into a Falcon tube containing heparin (LEO Laboratories Ltd) at 100 IU/ml of blood. The blood was separated by careful suspension over an equal volume of Ficoll (Lymphoprep, Axis-Shield) and density gradient centrifugation (20 min at 1200 g, brake off). The PBMC layer (at the plasma interface) was gently aspirated and washed

once in R0 (10 min at 700 g, brake on) before being treated with 5-10 ml of RBC lysis buffer (10 min, 37 °C). Cells were washed in R0 to remove lysate (6 min at 300 g) before resuspension in A5 media.

2.2.4 Generation of Lines through Culture of PBMC with Peptide Pools

At day-0 three wells of 200,000 PBMC per condition (from frozen or fresh blood) in 100 µL of A5 media was cultured with peptide or peptide pool at 10 µg/mL (1 µg in 100 µL) in a 96 well plate, with sterile water in the surrounding wells (37 °C, 5 % CO₂). 10 µL of Cell-Kine (Helvetica Healthcare) was added per well at day-3. 100 µL of A5 + IL-2 (40 IU per µL) was added at day-6. 100 µL of media was replaced at day-9. Cells were “rested” until day-14, following which cells were assayed up to day-21. Prior to assay, cells were washed 3 times in PBS before resuspension in A5 media.

2.2.5 Pulsing of B-LCL 174.DR1 Transduced APCs

For assays detailed in chapter 3 and 4, 174.DR1 APCs were pulsed in a 96 well plate at a concentration of 200,000 cells per 100 µL with peptide or peptide pool at 10 µg/mL (1 µg / 100 µL) for a minimum of 2 hours (37 °C, 5 % CO₂) in R0 media. Pulsing concentration and cell numbers replicate those conditions at which cultured cells were expanded (section 2.2.4.) Following pulsing, cells were washed in PBS (150 µL) 3 times to remove unbound peptides before resuspension in A5 media. APCs that were not pulsed with peptide (negative control for ELISpot) were incubated and washed alongside pulsed cells.

2.2.6 IFN-γ ELISpot with PBMC Lines

75,000 PBMC (number based on the initial PBMC line set up) were cultured on anti-IFN-γ coated ELISpot plate (MSIPS4510) coated with anti-IFN-γ capture antibody (1-D1K, Mabtech) with relevant 50,000 peptide pulsed APC in a total volume of 150 µL for 16 hours (37 °C, 5 % CO₂). Plate was washed 5 times in PBS, before development

with relevant antibodies following manufacturer's protocol (Detect: 7-B6-1-Biotin, Streptavidin-ALP).

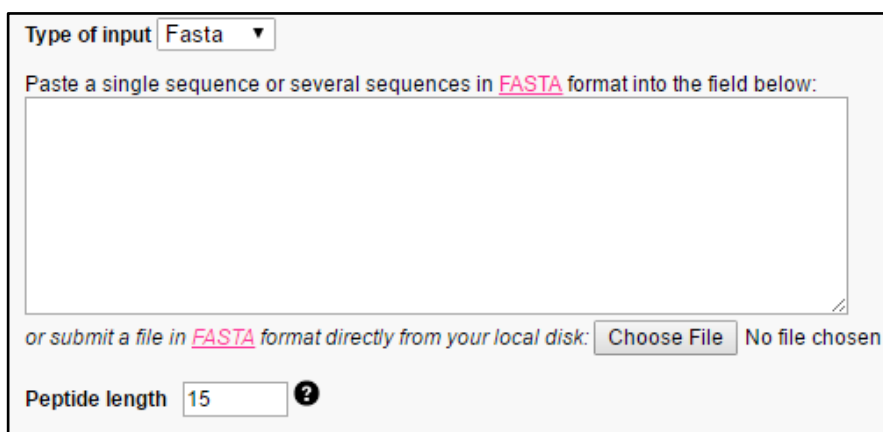
Controls were PHA (Aleré) and PKY ("universal peptide") peptide¹⁶⁸. Tests were run in duplicate, with a single negative control (PBMC and APCs in the absence of peptide or PHA stimulation).

2.2.7 Plate Analysis and Normalisation

Developed plates were imaged and counted using a CTL Immunospot analyser. CTL Single Colour software was used for spot counting and QC. Settings were kept constant for each reading. Assays were normalised for cumulative analysis by division of individual well spots by total number of spots across all wells (minus background). This accounted for inter-assay variation derived from plate sensitivity differences, low or high PBMC numbers or experimental error resulting in an inter-assay difference.

2.2.8 Binding Algorithms Inputs

NetPanMHCII (<http://www.cbs.dtu.dk/services/NetMHCIIpan/>) was used to predict the epitope based on the strongest binding core. Sequences 20-30 amino acids were input in FASTA format. Input length was set at either 11, 13 or 15 amino acids (Fig. 2.1).



Type of input ▼


Paste a single sequence or several sequences in **FASTA** format into the field below:

or submit a file in **FASTA** format directly from your local disk: No file chosen

Peptide length ?


Figure 2.1. NetPanMHCII input view.


HLA-DRB1*0101 was the allele selected, threshold of strong and weak binders were left at default parameters (does not affect output), with “print only strongest binding core” and “sort output by affinity” checked (ticked boxes, Fig. 2.2).

or type a list of molecules names separated by commas (no spaces) 
Max 20 alleles per submission.

DRB1_0101


For the list of available molecule names click here: [List of MHC molecule names.](#)


Alternatively, upload full length Alpha and Beta chain protein sequences: 


Definition of binding peptides: 


Threshold for strong binder (% Rank)


Threshold for weak binder (% Rank)


Turn on filtering options 

Fast mode 

Print only the strongest binding core 

Sort output by affinity 

Exclude offset correction 

Graphical representation of binding registers 

for peptides with %Rank < %


Save predictions to xls file 

Figure 2.2 Parameters for NetPanMHCII.

2.3 Protein Production for Crystals and HLA-Multimers

2.3.1 List of Reagents

TYP media: 16 g/L tryptone, 16 g/L yeast extract, 5 g/L potassium phosphate dibasic & 5 g/L sodium chloride.

LB agar for plates: 15 g/L agar bacteriological (Oxoid), 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl

Carbenicillin (carbenicillin direct) antibiotic: Added at 50 mg/L to TYP or LB agar following a reduction in temperature after autoclave sterilisation.

Bacterial Cell Lysis Buffer: 10 mM Tris, 10 mM MgCl₂ (Acros organics), 150 mM NaCl, 10 % glycerol; pH 8.1

Inclusion Body Wash Buffer: 50 mM Tris, 0.5 % Triton X-100, 100 mM NaCl, 2 mM EDTA; pH 8.1

Inclusion Body Resuspension Buffer: 50 mM Tris, 100 mM NaCl, 2 mM EDTA; pH 8.1

Urea Buffer A: 8 M Urea, 20 mM Tris pH 8.1 and 0.5 mM EDTA; pH 8.1

Urea Buffer B: 8 M Urea, 1 M NaCl, 20 mM Tris pH 8.1 and 0.5 mM EDTA; pH 8.1

HLA Class-II refold buffer: 25 % glycerol, 20 mM Tris, 1 mM EDTA, 20 mM NaCl, 1.48 g/L (13 mM) cysteamine hydrochloride & 0.83 g/L (3.7 mM) cystamine hydrochloride.

Crystal buffer: 10 mM Tris, 10 mM NaCl, pH 7.4

CAPS elution buffer: 50 mM 3-(cyclohexylamino)-1-propanesulfonic acid (CAPS); pH 11.5

Neutralisation buffer: 300 mM sodium phosphate; pH 6

Biomix A – Avidity: 0.5 M bicine buffer; pH 8.3

Biomix B – Avidity: 100 mM ATP, 100 mM MgO(Ac)₂ & 500 μM biotin

d-Biotin Buffer – Avidity: 500 μM d-biotin

Sample loading buffer (4 X): 1 M Tris, 0.008 % bromophenol blue, 10 % SDS, 40 % glycerol; pH 6.8

Reducing sample loading (4 X): 4M dithiothreitol (DTT), 1 M Tris, 0.008 % bromophenol blue, 10 % SDS, 40 % glycerol; pH 6.8

All buffers filtered using a 0.45 µm filter and vacuum pump system prior to use or application on AKTA FPLC. Chemicals obtained from Sigma or Fisher Scientific unless stated.

2.3.2 HLA-DR1 Plasmids

For bacterial expression three plasmids were obtained from Dr Chris Holland of Cardiff University, each was cloned into a pGMT7 E. coli expression vector with an ampicillin resistance gene. The resistance gene allowed growth in carbenicillin supplemented media (50 mg/L). Expression of protein was under control of the lac-operon and could be induced by addition of 0.5 mM Isopropyl β-D-thio-galactoside, (IPTG, Fisher Scientific). DR1α-bt was used to produce pHLA that would be used in HLA-multimers, while non-biotin tagged DR1α was used to produce pHLA for crystallisation.

DR1α: HLA-DRA*01 (Uniprot: P01903, residues [26-207])

DR1α with biotin tag (-bt): as above, with C-terminal Biotinylation signal sequence (GLNDIFEAQKIEWHE) joined via a flexible linker (GSGG)

DR1β: HLA-DRB1*0101 (Uniprot: P04229, residues [30-219])

2.3.3 Inclusion Body Production

Individual plasmids were transformed into Rosetta (DE3) competent BL21 E. coli (Novagen) by heat shock. 1 µL of plasmid DNA was added to 25 µL of BL21 cells and incubated (5 min, ice, gentle agitation). The mixture was heat shocked (42 °C, 2 min)

before transfer back to ice (5 min, gentle agitation). Cells were plated on LB agar carbenicillin plates and incubated overnight (37 °C).

Following overnight incubation, 3 distinct bacterial colonies were selected and transferred to 30 ml of TYP-carbenicillin to form a starter culture. Starter cultures were cultured in an orbital shaker (37 °C, 220 rpm) until the optical density at 600 nm (OD600) reached between 0.4 and 0.6.

Mature starter cultures were transferred to 1 L of TYP-carbenicillin and cultured in an orbital shaker (37 °C, 220 rpm) until the optical density at 600 nm (OD600) reached between 0.4 and 0.6. At an OD600 of 0.4-0.6, protein expression was induced by addition of 0.5 mM IPTG. Following IPTG addition, cultures were maintained for three to four hours in order produce sufficient quantities of protein.

Cells were isolated from cultures by centrifugation (4000 rpm, 20 min), and subsequently resuspended in lysis buffer (40 mL/L of cultured cells). 40 ml lysis resuspensions were sonicated using a Sonopulse HD 2070 with MS73 probe (Bandelin) at 60-90 % power for 20 mins using 2 second intervals. If large volumes of cells were obtained then sonication could be repeated, as well as the addition of extra freeze thaw steps in between sonication. Complete lysis of all cells was essential for purity at later steps.

Lysed cells were treated with DNase (160 µg/mL, 2 hours, 37 °C, agitation) before high speed centrifugation (10,000 rpm, 20 mins). Pellet was resuspended in triton wash buffer, homogenised and centrifuged (10,000 rpm, 20 mins) in order to remove irrelevant material. Triton wash step was repeated until a clear white pellet was obtained after centrifugation and there was an absence of gel-like substances.

Clean pellets were resuspended in resuspension buffer (to remove triton) before centrifugation (10,000 rpm, 20 mins) and solubilisation of the pellet in urea buffer A (40-60 ml per litre of bacterial culture, more was added to large quantities of protein). Addition

of 8 M urea solubilised the aqueous insoluble inclusion bodies which contain the DR1 α or DR1 β protein, enabling further purification steps.

2.3.4 Inclusion Body Purification

Urea solubilised inclusion bodies were filtered using a 0.45 μ m filter and vacuum pump. Inclusion bodies were purified by anion exchange chromatography (5 mL HiTrap Q HP; GE healthcare life sciences) on an AKTApure FPLC system (GE healthcare life sciences). The HiTrap column was equilibrated with urea buffer A (3 column volumes), and inclusion bodies loaded onto the column until saturation (when protein no longer binds to the column).

Protein was eluted using a NaCl salt gradient (urea buffer B, 0-100 % over 10 column volumes of buffer) and collected in 1 ml fractions. Fractions that contained protein, as indicated by UV absorbance (A280) on the AKTApure, were tested for A260/A280 and those with a high DNA content discarded (A260/A280 > 1 indicates DNA contamination).

Remaining protein fractions were analysed by SDS-PAGE (section 2.3.8) to find fractions that contain a band at 20 kDa (individual DR1 α or β chain) and minimal other impurities. Pure fractions were pooled, the pool concentration measured using a NanoDrop ND100 (ThermoScientific) and taken forward for refolding.

2.3.5 Refolding of HLA Class-II

Between 1-4 L of HLA refold buffer (size of refold depends on the amount of protein required) was prepared at 4 °C. The relevant DR1 α -chain and DR1 β -chain were each slowly added at a concentration of 5 mg/L of refold buffer under continuous mixing. Peptide was added at a concentration of 0.5 mg/L of refold buffer. The mixture was stirred magnetically for 1-2 hours before incubation at 4 °C for 3-5 days.

2.3.6 Purification of Refolded HLA Class-II

Following incubation, the refold was filtered at 0.45 μm before concentration using a Vivaflow crossflow concentration cassette (molecular weight cut-off 10 kDa; Sartorius) and peristaltic pump. The refold was concentrated to less than 50 mL (the minimal volume reached by the Vivaflow system) before addition of 500 mL of PBS for buffer exchange. The PBS refold mixture was then concentrated to 50 mL in the Vivaflow system, then further concentrated using 10 kDa MWCO centrifugal filter units (Millipore) to 10 mL.

Correctly refolded pHLA was purified using an immunoaffinity column containing the anti-HLA-DR antibody (L243, in house hybridoma) which binds class-II molecules in their intact conformation. Columns were equilibrated with PBS (15 mL flowed through) before addition of the 10 mL refold concentrate (flowed through the column 3-4 times to ensure binding of pHLA to antibody). Following addition of refold concentrate, the column was washed with PBS (15 mL) to ensure removal of non-specific binders.

Bound protein was eluted using CAPS buffer (pH 11.5, 8 mL total volume, applied to the column in 2 mL fractions) dripped into the equivalent volume of neutralising buffer (pH 6.0) to give a final pH of 7.4. Using a centrifugal filter unit (10 kDa MWCO) the protein was buffer exchanged back into crystal buffer (10 mM Tris, 10 mM NaCl, pH 7.4) and concentrated to 700 μL .

If protein was produced for HLA-multimers then biotinylation was carried out as described in section 2.3.7. If protein was produced for crystallisation then sample was purified by size exclusion column chromatography (gel filtration) on an AKTApure FPLC using a Superdex 200HR SEC column (GE healthcare life sciences), equilibrated and run with crystal buffer. Pure pHLA protein resulted in a single well-defined peak that was checked by SDS-PAGE (section 2.3.8) to give two distinct bands at the 20 kDa marker in reducing conditions.

2.3.7 Biotinylation and QC Shift Assay

Biotin was added to refolded HLA-DR1 for HLA-multimer staining (made using DR1 α -bt inclusion bodies) using a BirA biotinylation kit (Avidity). Refolded pHLA was concentrated to 700 μ L in crystal buffer (PBS cannot be used to high salt concentration) to which 100 μ L Biomix A, 100 μ L Biomix B, 100 μ L d-Biotin 500 μ M and 2 μ L BirA enzyme (Avidity) was added, followed by overnight incubation at room temperature.

Excess biotin was removed by buffer exchange in a centrifugal filter unit (10 kDa MWCO) into PBS and the success of the biotinylation reaction was assessed by a biotin shift assay on SDS-PAGE. 5 μ g of biotinylated pHLA was incubated with 5 μ g of streptavidin, and the formation of complexed pHLA-streptavidin multimeric complexes (a shift resulting in a smear at high molecular weight) was assessed on SDS-PAGE under non-reducing conditions.

2.3.8 SDS-PAGE

The purity and molecular weight of protein samples was analysed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). 5-10 μ g of sample was mixed with an equal volume of sample loading buffer for unreduced conditions (intact disulphide bonds). Analysis under reduced conditions (broken disulphide bonds) required addition of reducing sample loading buffer (with 1 M DTT) and boiling (95 $^{\circ}$ C, 5 min) before running on SDS-PAGE.

Samples were loaded onto Bolt 4-12 % Bis-Tris Plus Gels run in NuPAGE MES SDS running buffer in a Bolt Mini Gel Tank (all ThermoFisher Scientific). The gel tank was connected to a Bio-Rad Powerpac 200 power supply (Bio-rad) and run at 200 V over 25 min (longer if necessary).

2.3.9 TCR-pHLA Complex Formation

The F11 TCR was provided by Aaron Wall (Cardiff University) purified by size exclusion chromatography. Purified DR1-PKY and F11 TCR were mixed at an equimolar ratio to give a total protein concentration of 6 mg/mL in crystal buffer. This mixture was taken forward for crystallography and production of TCR-pHLA complex crystals (described in chapter 6).

2.3.10 Protein Crystallography

Purified pHLA protein (obtained as described in section 2.3.6) was concentrated to the maximum possible concentration, generally between 6-10 mg/mL in crystal buffer. Crystal trays were set up using the TOPS screen²¹⁵ with sitting drop vapour diffusion plates. Each TOPS screen buffer condition was dispensed into corresponding wells of an ARI INTELLI-PLATE 96-2 low volume reservoir plate (Art Robbins Instruments, LLC) using an Art-Robbins Gryphon robot (Art Robbins Instruments, LLC.). From the screen, 60 μ L was dispensed into a mother liquor well, and two dispenses of 200 nL into separate sitting drop wells. 200 nL of protein sample was dispensed into the top well containing 200 nL of a TOPS screen buffer. The plate was then sealed with a ClearVue seal (molecular dimensions) such that each condition was a closed system. Vapour diffusion occurred between the sample well and the mother liquor well to facilitate a change in osmotic conditions that would drive crystal formation. Plates were immediately imaged using a Formulatrix Rock Imager 2 (Formulatrix, Inc.) and incubated at 18 °C, with further images taken at daily intervals to monitor crystal growth.

2.3.11 X-Ray Crystallographic Sample Preparation and Data Collection

Wells containing crystals were opened with a scalpel, and crystals cryo-protected by the addition of 20 % ethylene glycol diluted in the corresponding TOPS screen buffer. Crystals were collected using 20 μ m or 40 μ m mounted loops (Molecular Dimensions),

immediately snap frozen in liquid nitrogen (liq. N₂) and placed in a Diamond light source storage puck stored in a Dewar vessel containing liq. N₂.

Crystals were subject to X-ray diffraction and data collection at Diamond light source (Diot, England) under supervision of Dr. David Cole, or Dr. Pierre Rizkallah (1000 diffraction images taken at 200 ° rotation and 0.2 second exposure time).

2.3.12 Structure Solution from DLS Datasets

Data sets were obtained from the DLS servers in three file formats (3dii, 3d, dials) with details on space group and resolution. Highest resolution data sets were analysed first. Data sets were processed using the program suite CCP4 (www.ccp4.ac.uk). “MATTHEWS” was used to obtain the number of molecules in the asymmetric unit, the details of which were input into “PHASER” with an HLA-DR1 model for molecular replacement to occur. This generated a model from the dataset based on the structure of HLA-DR1 which was subsequently edited using the visual WinCoot software (www.ccp4.ac.uk) and subject to several rounds of refinement using “REFMAC5”.

Refinement was continued until values of R_{WORK} and R_{FREE} reached acceptable levels, below 0.20 and 0.25 respectively, where possible. Model was interrogated visually to ensure the molecular backbone fitted the observed electron density at 1 sigma, and conformed to Ramachandran bonding principles. Final PDB files were visualised using PyMOL graphics software (Schrodinger) and contact tables generated from “NCONT” in CCP4 to define the interaction distances and partners. Types of non-covalent interaction, i.e. Van Der Waals (vdWs), polar, non-polar hydrophobic and salt bridges were inferred based on the details of these tables.

2.4 Analysing Epitope-Specific T-cells

2.4.1 Culture of PBMC for HLA-Multimer Staining

At day-0, five wells of 200,000 PBMC (from frozen or fresh blood) in 100 μ L of A5 media was cultured with peptide at 10 μ g/mL (1 μ g in 100 μ L) in a 96 well plate, with sterile water in the surrounding wells (37 °C, 5 % CO₂). One million total cells were necessary for efficient staining on flow cytometry and for sorting of sufficient numbers for later clonotyping work.

As previously described in section 2.2.4, 10 μ L of Cell-Kine (Helvetica Healthcare) was added per well at day-3. 100 μ L of A5 + IL-2 (40 IU per μ L) was added at day-6. 100 μ L of media was replaced at day-9. Cells were “rested” until day-14, following which cells were stained and assayed by flow cytometry as described in section 1.4.3.

2.4.2 Preparation of HLA-Multimers

All monomers were multimerised on a dextramer backbone^{216,217} (Immudex) following published methodology. Per individual stain, 0.5 μ g of refolded and biotinylated pHLA was incubated with 2 μ L of phycoerythrin (PE)-conjugated dextramer backbone solution (20 min, ice). Solution was spun at high speed (16000 rpm, 30 sec) to pellet insoluble material before use. Multimers could be made on the day of staining, or up to five days before, stored at 4 °C.

2.4.3 Flow Cytometry: HLA-Multimer Staining

PBMC lines cultured as described in section 2.4.1 were combined (estimated as 1 million total cells) then split into three wells (for test, irrelevant HLA class-II multimer and FMO controls) and washed in FACS buffer (400 g, 5 min). Cells were incubated with protein kinase inhibitor 50 nM dasatanib (30 min, 37 °C; Axon Medchem). HLA-multimers (prepared in section 2.4.2) were made up to a total volume of 10 μ L per stain and added directly to PBMC lines following dasatanib incubation without washing (30 min, 4 °C).

Lines were washed (FACS buffer, 400 g, 5 min), and incubated with “boost” antibody²¹⁶, anti-PE added (10 µg/mL, 20 min, 4 °C; clone PE001, BioLegend). Cells were washed twice in PBS buffer, then stained with violet LIVE/DEAD Fixable Dead Cell Stain, Vivid (Life Technologies) (5 min, RT).

The antibody cocktail of remaining stains was added for incubation (20 min, 4 °C): anti-CD8-allophycocyanin-vio770 (clone BW135/80; Miltenyi Biotec), anti-CD4 allophycocyanin (clone M-T466; Miltenyi Biotec), anti-CD3-peridinin chlorophyll protein (PerCP) (clone BW264/56; Miltenyi Biotec); anti-CD19-Pacific blue (clone HIB19; BioLegend); and anti-CD14-Pacific blue (clone M5E2; Bio- Legend). Following cocktail incubation, cells were washed twice in FACS buffer before analysis by flow cytometry.

Cells were sorted on a BD FACS ARIA (BD Biosciences) with the help of central biology services (CBS) Cardiff University. Cells were sorted directly into lysis buffer (Qiagen) supplemented with 0.5 M DTT, and frozen at – 80 °C for later RNA extraction and cDNA isolation described in in section 2.4.4.

2.4.4 Clonotyping by Next Generation Sequencing

2.4.4.1 RNA Extraction and cDNA Synthesis

RNA was extracted from thawed samples using an RNAeasy PLUS micro extraction kit (Qiagen) following manufacturers protocol. Purified RNA was used to produce cDNA using the SMARTER cDNA synthesis kit (Clontech) following manufacturers protocol. 10 µL of purified mRNA was incubated with oligo-dT in a thermal cycler (3 min, 72 °C followed by 2 min, 42 °C) to allow annealing of the oligo-dT primer to the oligo-A mRNA tail. Following annealing, 4 µL 5X First Strand buffer, 0.5 µL 100 mM DTT, 1 µL 20 mM dNTP, 0.5 µL RNase inhibitor and 2 µL SMARTScribe RT and 1 µL oligo-A primer II were added to the sample for a cyclic incubation (90 min, 42 °C followed by 10 min, 70 °C).

2.4.4.2 PCR Amplification of cDNA Product

Two PCR steps were used to amplify the TCR α or TCR β products from cDNA, using a single primer pair.

The first step: 2.5 μ L of cDNA sample with 10 μ L 5x high fidelity (HF) buffer, 0.5 μ L 100 mM DMSO, 1 μ L 20 mM dNTPs, 5 μ L 10X Universal Primer A (forward primer), 1 μ L Primer C α -R1 or C β -R1 (reverse primer) and 0.25 μ L Phusion Taq polymerase were mixed and made up to a final volume of 50 μ L final volume with nuclease free water (Ambion). The mixture was incubated (94 $^{\circ}$ C, 5 min) for the starting denaturation followed by 30 cycles (1 cycle consists of 30 secs at 94 $^{\circ}$ C, 30 secs at 63 $^{\circ}$ C and 3 min at 72 $^{\circ}$ C) before a final extension (72 $^{\circ}$ C, 7 min).

The second step: 2.5 μ L of sample from the step out PCR was mixed with 10 μ L 5x high fidelity (HF) buffer, 0.5 μ L 100 mM DMSO, 1 μ L 20 mM dNTPs, 1 μ L Primer short (forward primer), 1 μ L Primer C α -R2 or C β -R2 (reverse primer) and 0.25 μ L Phusion Taq polymerase and made up to a final volume of 50 μ L final volume with nuclease free water (Ambion).

The mixture was incubated (94 $^{\circ}$ C, 5 min) for the starting denaturation followed by 30 cycles (1 cycle consists of 30 secs at 94 $^{\circ}$ C, 30 secs at 63 $^{\circ}$ C for TCR α -chains and 66 $^{\circ}$ C for TCR β -chains and 3 min at 72 $^{\circ}$ C) before a final extension (72 $^{\circ}$ C, 7 min). Samples were analysed by electrophoresis on a 1% agarose gel (45 min, 90 V). Amplified samples of 400 base pairs were excised and purified using a Gel extraction kit (Nucleospin).

2.4.4.3 Next Generation Sequencing

Purified samples were kindly sequenced by Dr. Meriem Attaf (Cardiff University) using an Illumina MiSeq. NEBNext Ultra Library preparation kit (New England Biolabs, Cambridge, UK) was used to produce samples that were run on an Illumina MiSeq using the MiSeq v2 reagent kit (Illumina).

TCR gene usage was determined based on reference sequences from the ImMunoGenetics (IMGT) database (imgt.org) and all TCR gene segments were designated according to the IMGT nomenclature using MiXCR software. Low quality reads, TCRs with one single read (singletons) and TCRs with CDR3 sequences less than seven amino acids in length were eliminated.

Following processing of raw sequencing data, information was processed and presented using Microsoft Excel. Motifs were visualised using WEBLOGO (<http://weblogo.berkeley.edu/logo.cgi>) without small sample correction set.

3 HLA-DR1 Transduced APCs Facilitate Detection of DR1-Restricted T-Cell Responses

3.1 Abstract

Cell mediated immunity relies on the presentation of antigenic pHLA molecules to T-cells. The nature of an individual's T-cell response to antigen is determined by the T-cell receptor repertoire they are capable of generating and their set of HLA alleles. HLA genes exhibit high levels of polymorphism. Across the population this provides a high level of immunological diversity in responses to the same antigen.

Amidst this diversity, common responses mediated by shared HLA alleles are present. Knowledge of these HLA-restricted epitopes enables further examination of the immune system and aids the development of novel vaccination strategies. To determine the HLA-restriction of an *in vitro* response to peptide, HLA-specific blocking antibodies, HLA-matched presenting cells or HLA-multimer staining can be used. Employing such techniques during the screening of large peptide libraries in multiple donors is a challenge.

To overcome this, a B-cell line shown to be deficient in HLA class-I expression and missing all class-II genes was used. The "naked" line was transduced with HLA-DR1, and shown to successfully present peptides to cognate CD4+ T-cell clones and polyclonal lines. The use of this line during the screening of peptide libraries would allow identification of immune responses attributable to HLA-DR1 epitopes without further restriction assays.

3.2 Introduction

In vivo, antigen is processed into short peptide fragments and loaded onto HLA molecules that are presented on the cell surface. The T-cell receptor (TCR) recognises the unique contact surface formed from both the peptide epitope and HLA amino acid residues^{166,218,219}. Therefore, when characterising the T-cell response to antigenic peptides *in vitro*, it is useful to know the HLA-restriction of the responding cells. Knowledge of HLA-restricted epitopes forms the basis for developing a better understanding of the adaptive immune system. Studies involving pHLA multimer staining²²⁰, clonotypic analysis²²¹, T-cell vaccine design²²² and X-ray crystallography^{166,218,219} often examine responses to well defined HLA-restricted epitopes. Research at the level of an epitope answers questions concerning antigen specific T-cell populations, receptor interactions, antigen processing and molecular information that is limited at the level of whole protein or pathogen.

Finding and characterising HLA-restricted epitopes is a challenge due to the polymorphic nature of HLA genes. Peptides may be presented by up to six different class-I alleles and up to twelve class-II variants in one individual. Identifying the HLA-restriction of each response using HLA-blocking antibodies or matched and mismatched presenting cells becomes technically difficult on large peptide libraries. The work is further complicated when analysing sets of responses in multiple individuals, each with a distinct HLA-type.

Instead of looking at responses to antigen at the level of each individual, it is more relevant to focus on a specific allele that is shared in a significant percentage of the population. In this case, the target epitopes are those most consistently recognised between individuals carrying the allele of interest. The challenge of epitope mapping under this genetic focus, is therefore to detect responses to a single HLA, and bypass those mediated by all others.

In order to do this, the method of *non-autologous* presentation can be employed. In an immunoassay such as an ELISpot, non-autologous APCs of a known HLA-type can present peptide to an expanded T-cell line. If the line and non-autologous APC have one common HLA allele, and elicit a measurable response to peptide, then that response is attributable to presentation by the common allele.

Developing this concept, if the non-autologous presenting cells *expressed only a single HLA*, then the restriction information of any response generated would be implicit. Designing such an APC could be achieved by lentiviral transduction of an HLA construct into a cell line that did not otherwise express surface HLA.

The “naked” cell lines 721.174 and its daughter T2 (174 x CEM.T2) are two examples, both are well characterised in the literature as lines deficient in antigen presentation machinery and surface HLA expression²²³. 721.174 cells were generated by two cycles of γ -radiation exposure followed by selection for the loss of HLA class-II expression²²⁴. Analysis revealed these cells to have a homogeneous deletion in the class-II locus including all HLA-DR, DQ and part of the DP genes²²⁵ (Fig 3.1). The T2 cell line was made through fusion of 721.174 with CEMR.3 as part of an investigation into the trans-acting factors that govern HLA expression in lymphoblast hybrids²²⁶.

Both lines express no class-II and low amounts of HLA-A2 at around 20-50% of wild type levels²²⁶. In the absence of both TAP and HLA-DM, these lines are impaired in their ability to process and present the normal complement of self-antigen via both class-I and class-II pathways.

Concerning the class-II pathway, when the T2 line is transduced with HLA-DR3, no issues in trafficking or surface presentation²²⁷ are observed, instead there is an accumulation of HLA-DR3 with bound CLIP (the invariant chain) in the MIIC compartment and at the cell surface¹³⁵. ***This does not pose a problem in the presentation of short exogenous peptides***^{228,229}, like those used in this study, but it does prevent processing

and presentation of whole, non-denatured protein. Normal function can be restored by transduction with HLA-DM, as shown in multiple studies^{151,230}

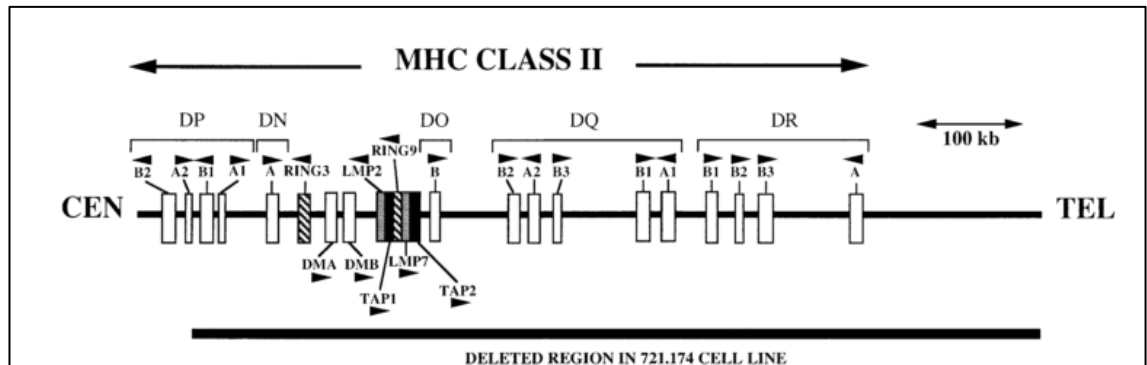


Figure 3.1. Deletion in the 721.174 parent line conferring the “naked” phenotype.
Taken from Fabb²²³ et al. The schematic shows the deletion in chromosome 6 missing from the naked APC 721.174. The deletion covers all class-II genes except the pseudogenes DP-A2 and DP-B1 which are not expressed or necessary for HLA function²³¹.

3.3 Aims

The HLA allele HLA-DRB1*0101 (HLA-DR1) is present in 20 % of the European Caucasian population. In our laboratory, technologies are in place to study DR1-restricted responses at the cellular and molecular level through HLA-multimer staining and X-ray crystallography. These methods require specific knowledge of DR1-restricted epitopes that stimulate cognate CD4+ T-cell populations.

The aim of this chapter was to develop a methodology for isolating DR1-restricted T-cell responses when screening polyclonal PBMC lines against pools of peptide. This would require the generation of an APC line expressing only HLA-DR1, using “naked” APC lines and lentiviral transduction. Such DR1.APCs could be used on an IFN- γ ELISpot in order to reveal DR1-restricted T-cell responses, and eliminate responses mediated by other HLA alleles.

Specific aims:

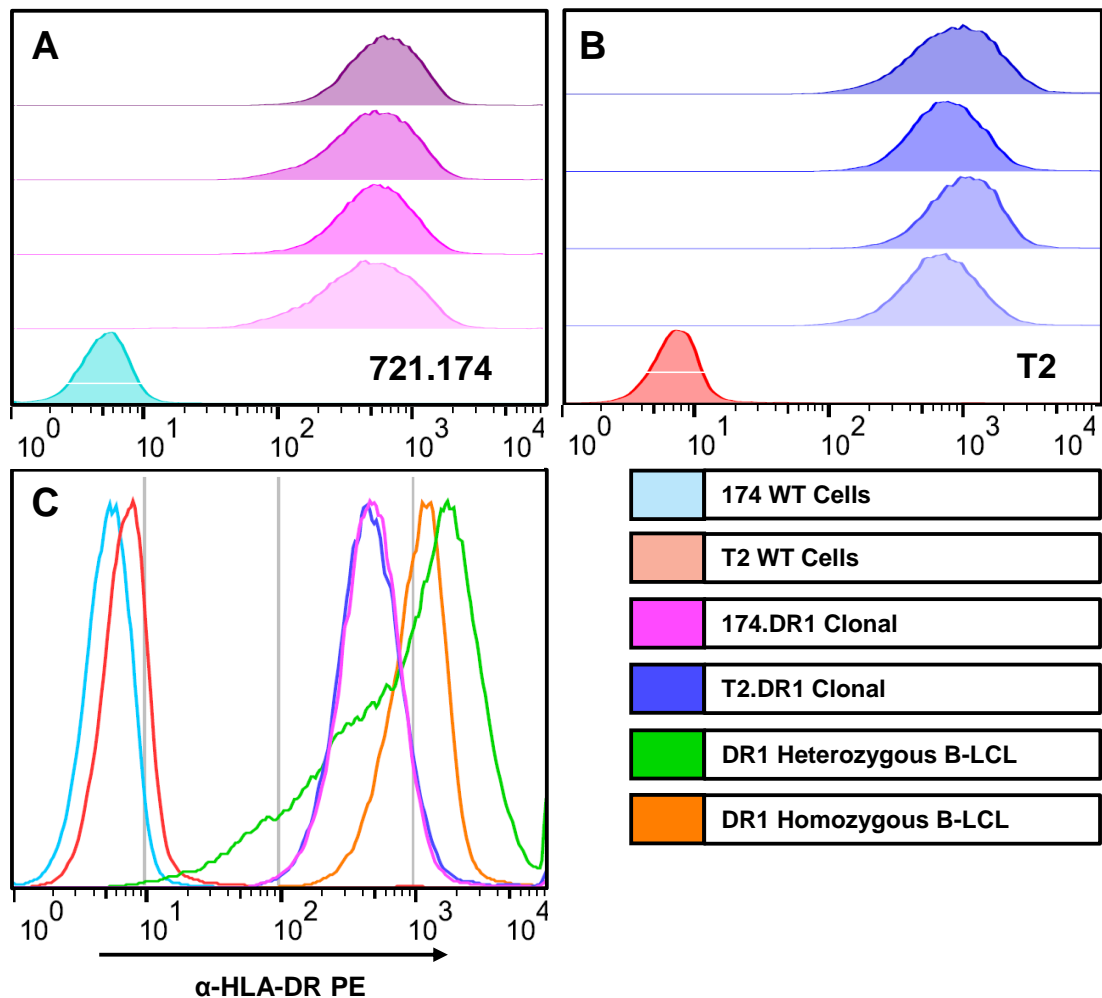
- 1) To transduce “naked” APC lines with an HLA-DR1 lentivirus, isolate clones and verify HLA-DR1 expression.
- 2) To confirm that transduced cells can present HLA-DR1 restricted epitopes to cognate T-cells on IFN- γ ELISpot.
- 3) To optimise HLA-DR1 peptide presentation to polyclonal lines, and compare HLA-DR1 and autologous presentation to identify HLA-DR1 restricted responses.

3.4 Results

3.4.1 Generation of HLA-DR1 Presenting Cells

HLA-DR1 was successfully cloned into a 2nd generation lentiviral construct and packaged virus was harvested from HEK293T cells. Transduction of T2 and 721.174 cells resulted in a mixed population of transduced and non-transduced lines. DR1 expressing cells were enriched by magnetic bead purification prior to limiting dilution and cloning. The expression of four selected DR1 clones was assessed using pan α -**HLA-DR** antibody (L243) via flow cytometry. No HLA-DR1 specific reagents are available, yet as the wild type (non-transduced) lines show no HLA-DR expression, all staining could be attributed to the transduced HLA. Expression of HLA was two logs higher than wild type populations, with a relatively uniform and consistent MFI between the different clones (Fig. 3.2A-B).

In order to assess surface expression levels of HLA-DR relative to other antigen presenting cells, the pan HLA-DR staining of a 174.DR1 and T2.DR1 clone, their corresponding wild type lines, and two immortalised B-LCLs were compared (Fig. 3.2C). The immortalised B-LCLs, one homozygous for HLA-DR1 and one heterozygous were derived from healthy donors and thus expressed wild type levels of HLA and antigen processing genes. Both natural B-LCLs had higher expression than transduced clones. This was observed in previous studies with cells lacking antigen-processing genes, where wild type expression levels are restored upon transduction with HLA-DM^{151,232}. The implications of the absence of HLA-DM in peptide presentation are evaluated in section 3.5.



Cell Type	MFI
174.WT	5.3
T2.WT	7.4
174.DR1 Clonal	467.0
T2.DR1 Clonal	473.0
DR1 Heterozygous B-LCL	780.0
DR1 Homozygous B-LCL	1043.0

Figure 3.2 Expression levels of HLA-DR across wild type and transduced cell lines on flow cytometry using pan anti-HLA-DR antibody (L243). **(A)** HLA-DR expression in the 721.174 wild type cell line and four examples of 721.174 DR1 expressing clones obtained by limiting dilution from the transduced line. **(B)** The same data shown for the T2 lineage. **(C)** Comparison of HLA-DR expression on transduced clones and two B-LCL lines with wild type levels of HLA expression.

3.4.2 DR1 Presentation to CD4+ T-cell Clones on IFN- γ ELISpot

When using an IFN- γ ELISpot assay to detect T-cell responses to an antigen, the lower limit of detection is a single cognate T-cell. This sensitivity means that even the weakest T-cell affinities and smallest responses are visible at minimal cell numbers. Separating relevant responses from background, i.e. the signal to noise ratio, may be challenging, especially when working with polyclonal lines to multiple peptides. Complicating factors such as aberrantly active cells, low intensity responses, and high background associated with strongly responding lines are encountered.

Given these complications, it was logical to test the peptide presentation capacity of the transduced APCs with a cleaner system. This was achieved using a cultured CD4+ T-cell clone (DCD10) specific to an influenza haemagglutinin peptide presented by HLA-DR1¹²⁴. The clone DCD10 recognises the thirteen amino acid peptide HA₃₀₆₋₃₁₈ (PKYVKQNTLKLAT) and is used as a model system in our laboratory for CD4+ T-cell IFN- γ production¹⁶⁸.

In this assay, 50,000 peptide-pulsed APCs were used to form a uniform lawn of presenters in a single well with 300 T-cells. The pulsing of wild type cells elicited a smaller but positive IFN- γ response (> 20 SFC) at very high peptide concentrations of 10^{-3} M, but **no response** at the lower concentration of 10^{-5} M. In comparison, DR1 transduced presenters were consistently able to stimulate cognate T-cells at both concentrations to the same level (Fig 3.3). Presentation by wild type cells at 10^{-3} M was attributed to ineffective washing of non-specifically bound peptide, a problem identified and discussed in later sections.

This assay confirmed the ability of these transduced APCs to present DR1 specific peptide. The next challenge was to isolate restricted responses from polyclonal lines specific to multiple peptides. As no difference in presentation capacity between T2 and 721.174 was apparent, the 174.DR1 clone was used in all future work.

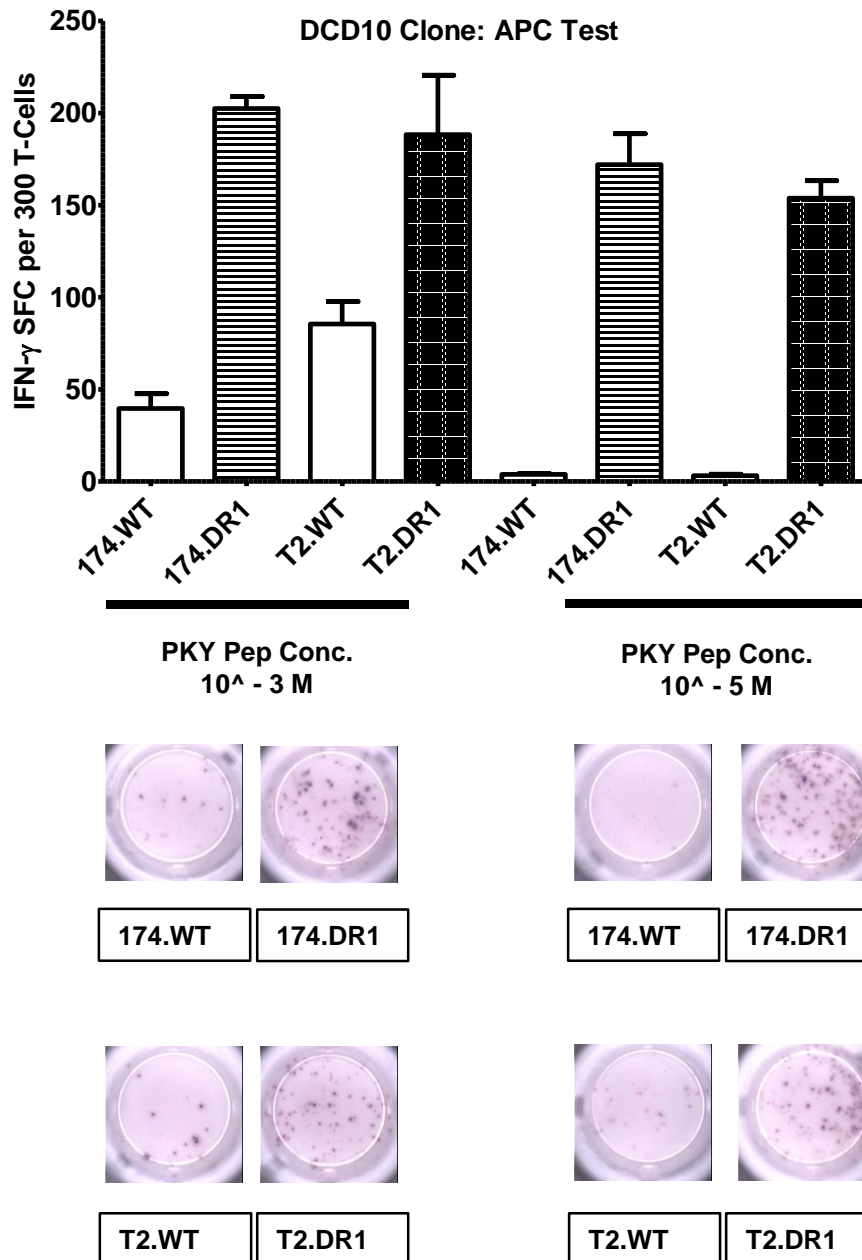
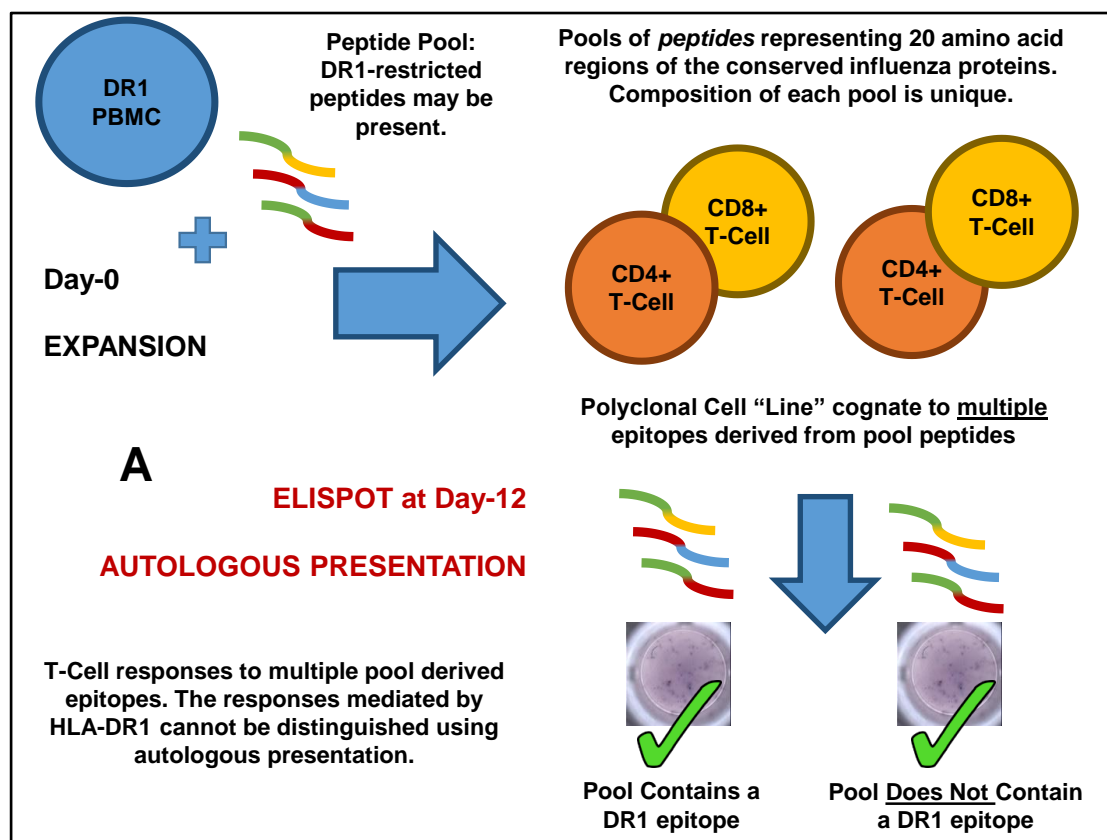


Figure 3.3 Presentation of HA peptide 306-318 (PKYVVKQNTLKLAT) to the cognate CD4+ T-cell clone DCD10 by wild type and transduced antigen-presenting cells on IFN- γ ELISpot. 300 T-cells were plated onto 50,000 peptide pulsed APCs at two different concentrations of HA peptide. Representative images of IFN- γ ELISpot wells shown at each concentration below. Number of spot forming cells for each well were background subtracted and averaged over the number of repeats (mean with SD error bars, $n = 3$).

3.4.3 Dissecting DR1 Responses from Polyclonal Lines

In order for HLA-restriction to be inferred accurately using DR1 presentation in the polyclonal setting, it was necessary to compare with autologous presentation and observe consistent differences (Fig. 3.4A-B). Pools of peptides from the matrix protein were used to expand DR1+ donor PBMC for fourteen days, and lines subsequently tested using both autologous and DR1 presentation methods. The contents of the Matrix (M1) pools is discussed, in detail, in chapter 4. Briefly, each of the ten pools contains a mixture of five different peptides, twenty amino acids in length, representing the sequence of the M1 protein.

At this stage, the M1 peptide pools *served as an optimisation platform* for the development of a successful protocol. With ten distinct pools covering a highly immunogenic protein, the span of results and challenges encountered were representative of future assays.



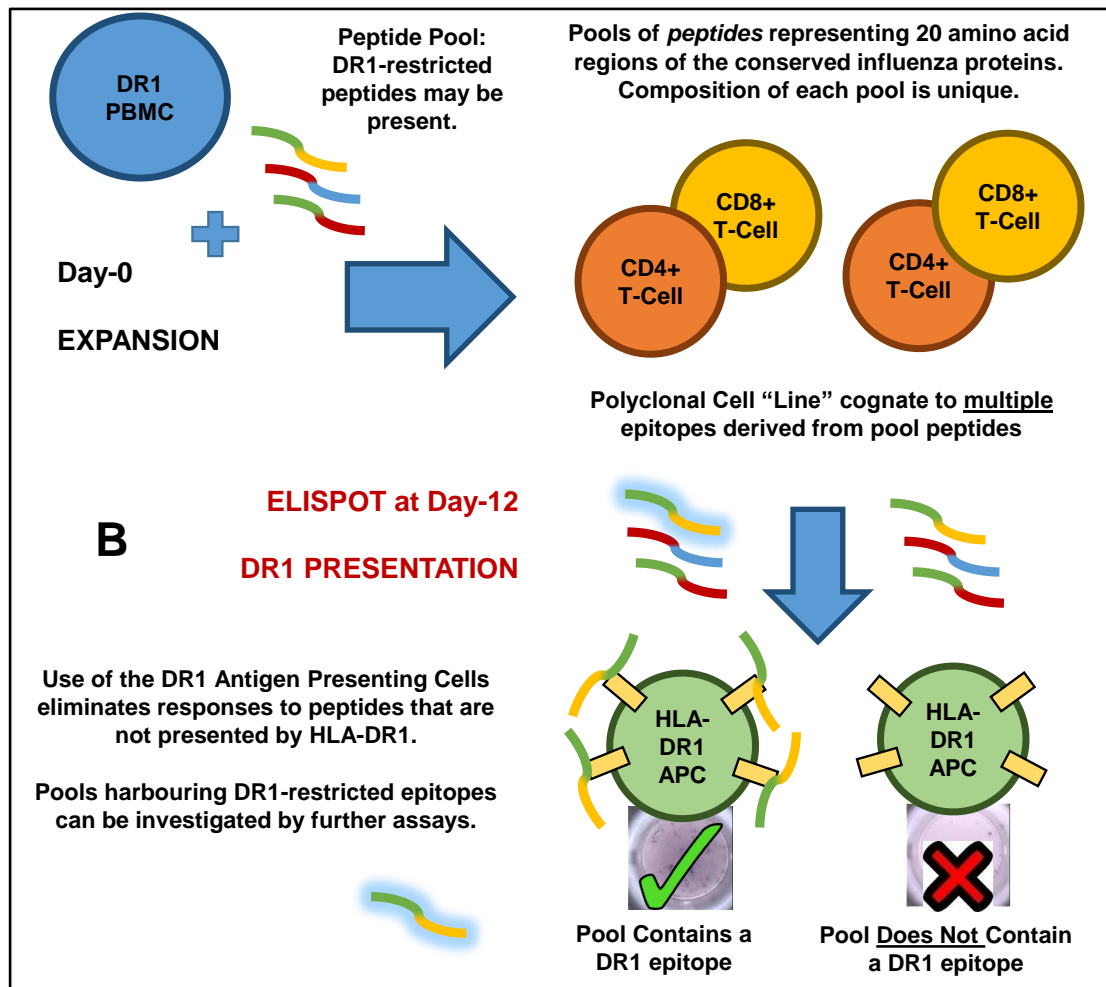


Figure 3.4: Using DR1 cells to identify HLA-DR1 restricted T-cell responses from polyclonal PBMC lines. Culture of PBMC with a pool of distinct peptides results in a polyclonal line. (A) Probing of this line via an autologous presentation ELISpot, will highlight all responses. (B) Probing via a DR1 ELISpot, will reveal those responses to epitopes within the pool restricted to HLA-DR1.

3.4.4 Comparing Expansion Conditions: FCS versus Human Serum

The analysis of a line expanded against a specific peptide or peptide pool requires a background control (same assay conditions, in the absence of peptide restimulation). The background control is an indication of how “active” or stimulated cells are in culture conditions, a significant factor being the serum used during expansion. The majority of studies culture cells in human serum, the source of which is non-commercial and is supplied in batches, which must be tested to ensure efficacy. Some labs use foetal calf serum (FCS) with some success²¹⁴. Due to its increased availability and lower cost, it potentially provides an alternative to human serum.

As part of early optimisation, a side-by-side comparison of both serums was carried out following standard protocol. When human PBMC were cultured with FCS serum, widespread activation was observed on both ELISpot formats (Fig 3.5A, 3.5C). Background controls for each line were displayed alongside tests, with similar levels of activity seen in both. In comparison, background levels in human serum were much lower than their corresponding test wells (Fig 3.5B, 3.5D).

ELISpot assay data are conventionally presented with background subtraction (Fig 3.5E, 3.5F); the resulting signal-above-background value allows inter assay comparison, and forms part of the criteria for a positive result. As shown, with high background SFC values in FCS, the data was meaningless, with most responses failing to meet the 20 SFC per 10^5 significance level (Fig 3.5E). In human serum, the responses were of varying intensity and significance, and some pools demonstrated strong T-cell reactivity (Fig. 3.5F). Such strong response frequencies to the M1 peptide pool were expected based on previous studies^{213,233,234}.

Testing of this protocol change served as an introduction to assay optimisation before more significant experiments were undertaken. All subsequent assays were performed in human AB serum. This serum was not commercially obtained; therefore each new batch was tested and validated before further use.

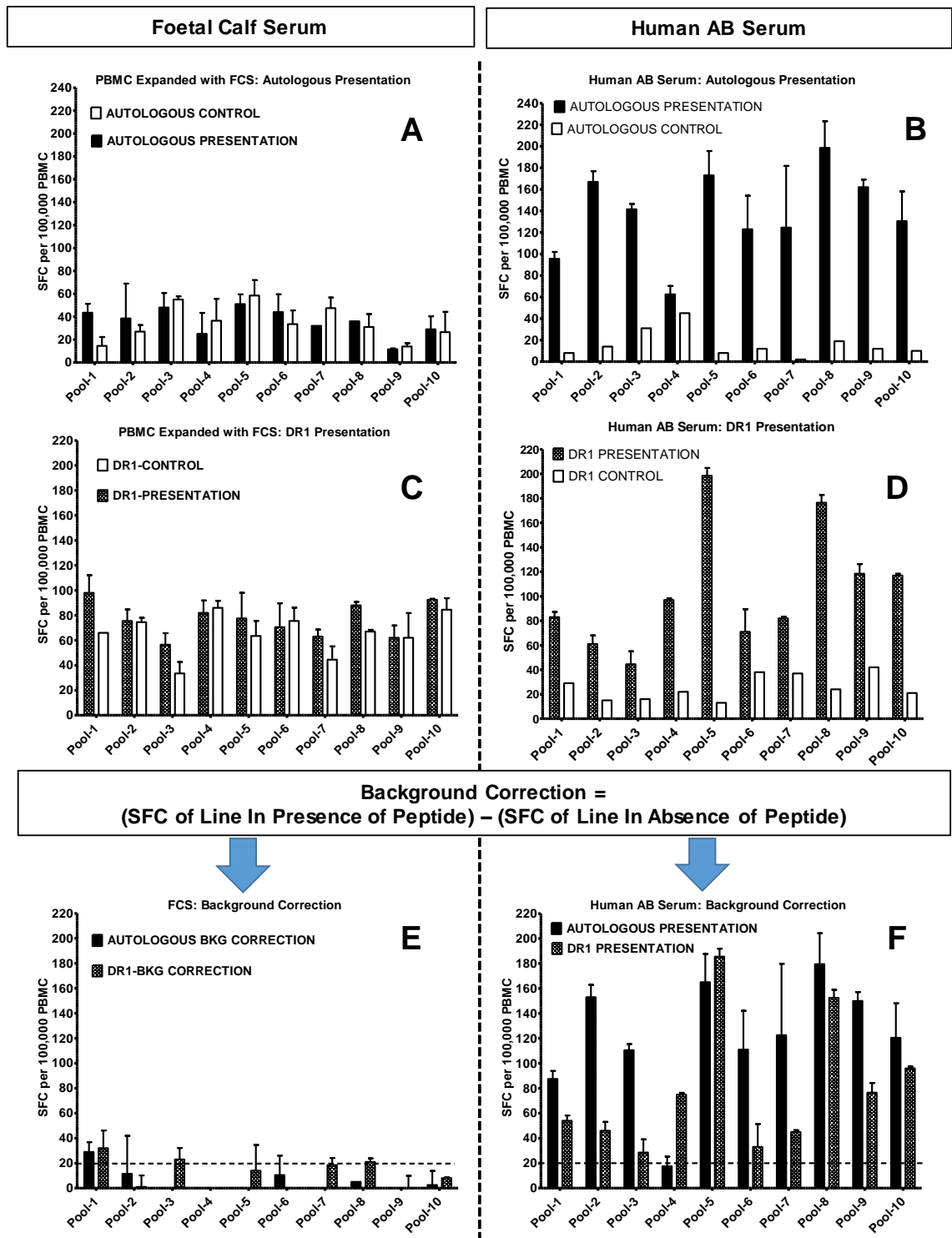


Figure 3.5: Comparison of Foetal Calf Serum (FCS) and Human AB Serum in expansion and ELISpot media. Autologous presentation ELISpot assays on peptide pool expanded lines using FCS and human serum are shown in (A) and (B) respectively. Background SFC values (white bars, negative control, no peptide) are side by side with the test (peptides added) SFC value (dark bars). DR1 presentation ELISpot assays using

FCS and human serum are shown in (C) and (D) respectively, in the same format as above. When the relevant background SFC values have been subtracted from each respective line SFC value, **resulting values above 20 SFC per 100,000 PBMC (represented by dashed line) are defined as positive**. This is applied to (A-D) to generate graphs shown in (E) and (F) for FCS and human serum respectively. Autologous presentation is shown side-by-side with DR1 presentation for comparison. Number of SFC for each well are background subtracted averaged over the number of repeats (mean with SD error bars, $n = 2$).

3.4.5 Distinguishing between Autologous and DR1 Presentation

Having established growth conditions with AB serum, the next step was detailed comparison of autologous and DR1 presentation on polyclonal lines. Here, we would expect autologous presentation to show a vast spread of responses mediated by multiple HLAs, with a reduction in responses when using DR1 presenting cells. Those peptide pools that elicited a response in both systems would warrant further investigation.

A key methodological consideration was the efficacy of washing. If DR1 presenting cells were inadequately washed after pulsing with peptide, then weakly or non-specifically bound peptide would be present in the ELISpot well. These peptides, not specific to DR1, would be free to dissociate and stimulate responses through autologous presentation. In such circumstances, the DR1 ELISpot would mirror the results pattern seen with autologous presentation. A clear difference between the techniques was necessary to identify peptide pools harbouring DR1-restricted responses.

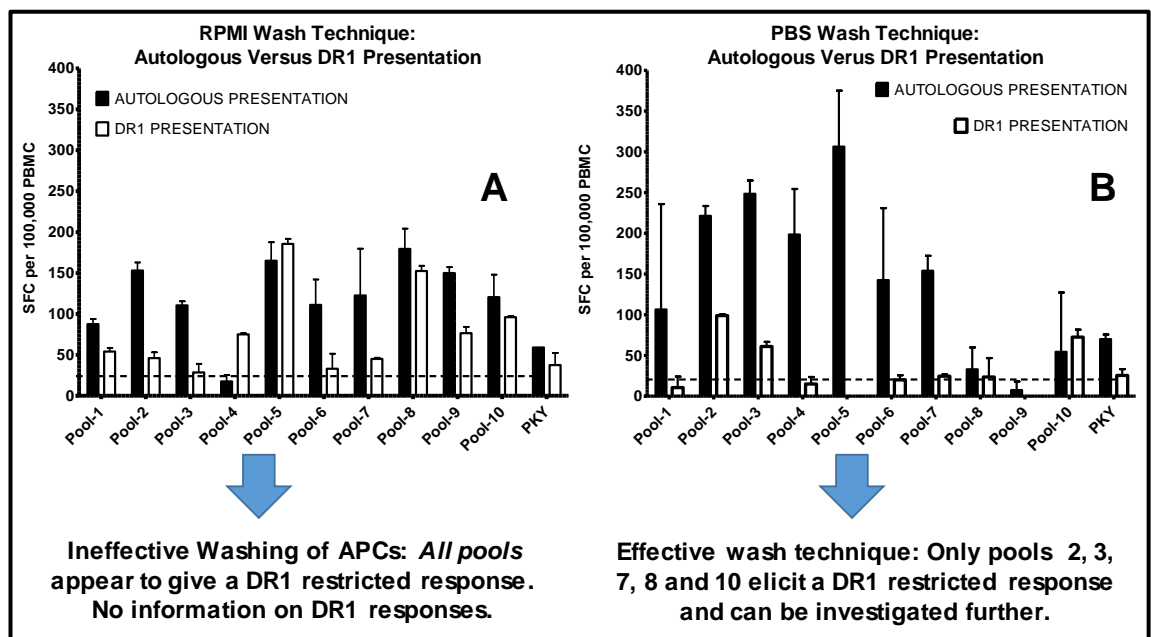


Figure 3.6: Using a PBS based wash Technique gives a clearer distinction between autologous and DR1 presentation. (A-B): Lines to M1 peptide pools were expanded and cells from the same line were divided and tested for responses on ELISpot using both autologous presentation (black bars) and DR1 presentation (white bars). For the

assay shown in (A), following peptide pulsing DR1.APCs were washed using standard RPMI media. For the assay in (B), DR1.APCs were washed in PBS (mean with SD error bars, n = 2).

The standard laboratory methodology for washing *peptide-pulsed APCs* in preparation for an ELISpot assay involves the use of RPMI media as wash buffer. Using RPMI, no consistent distinction was visible between the autologous and DR1 presentation methods, with all pools showing responses above the 20 SFC cut off, defined as a positive result (Fig. 3.6A). Under stringent PBS washing, DR1 presentation gave consistently lower responses than autologous presentation, with fewer pools eliciting a positive DR1 response (Fig 3.6B). This reduction in reactivity to each pool under DR1 presentation conditions was expected, given that not all peptides in each pool will be capable of binding to DR1, and from those that are, not all will constitute an epitope and elicit a response.

The same technique was further tested using two DR1 restricted T-cell clones: one specific for a viral epitope (DCD10 and HA₃₀₆₋₃₁₈ peptide) and the other for a cancer epitope (GD.D104 and 5T4_{xxx-xxx}, peptide sequence not disclosed)²¹⁴. This was a modification of the assay discussed in section 3.4.2 (Fig 3.3), in that wild type cells not expressing DR1 were pulsed alongside DR1.APCs, and the ability of each to activate their respective clones on ELISpot was measured (Fig 3.7).

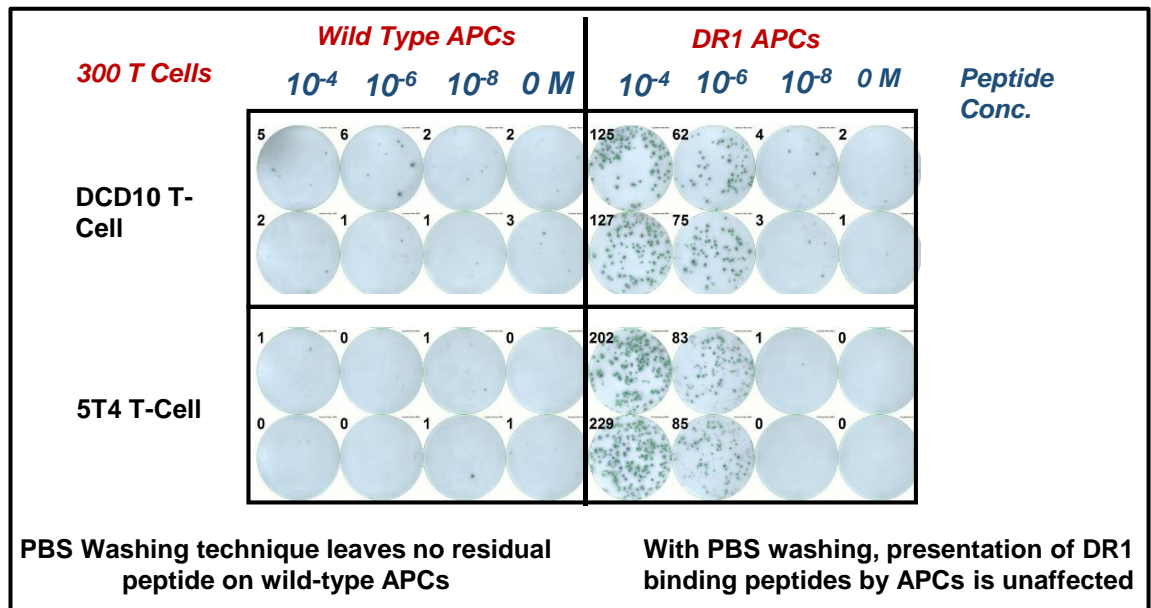


Figure 3.7 Assessment of presentation to DR1 restricted T-Cell clones after peptide pulsing by 174.WT and 174.DR1 APCs under stringent PBS wash conditions. Two clones were used, DCD10 specific to HA₃₀₆₋₃₁₈ and GD.D104 specific to 5T4_{xxx-xxx}. APCs were pulsed with respective peptides at three concentrations as shown, with relevant negative controls (0 M). Each repeat (n=2) is displayed in the ELISpot plate image, with number of spots as counted by the software in the top left corner of the well grid.

The PBS wash technique was able to reduce the IFN- γ response to wild type cells (DR1 negative cells) to nearly zero SFC, while maintaining DR1 restricted responses at 10^{-4} M and 10^{-6} M.

As an additional control, a DR1 negative donor was used to investigate how restrictive DR1 presentation was against a donor with no T-cell specificity to this HLA. In this case, autologous presentation yielded a normal response pattern against the M1 pools, while DR1 presentation elicited no responses above 20 SFC (Fig. 3.8). This was seen as the strongest indication of the technique's success in limiting responses to only those being presented by DR1, and preventing non-specific binding of peptide to pulsed

presenting cells. The assay was ready to be implemented in the larger screening of three proteins in two separate DR1+ donors.

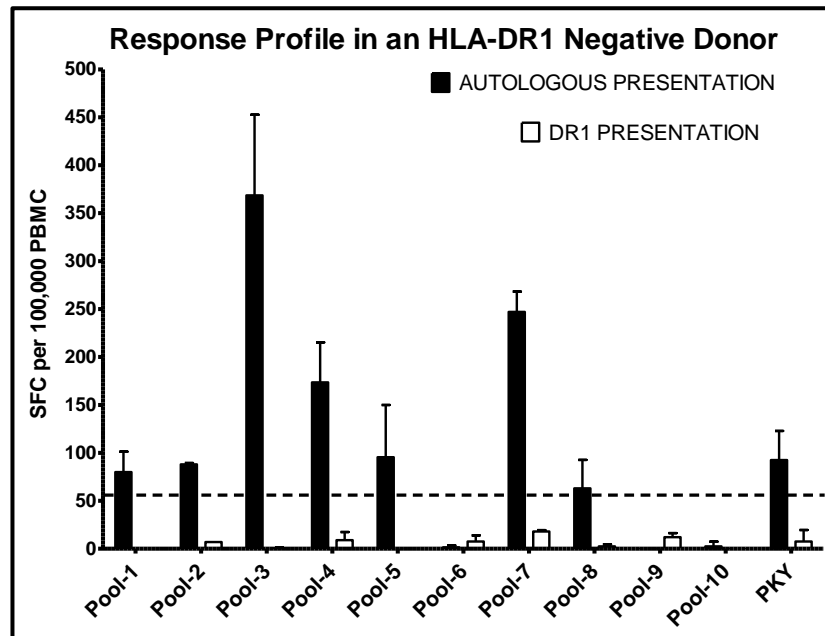


Figure 3.8 Comparison of autologous presentation and DR1 presentation in a DR1 Negative donor. Lines to M1 peptide pools were expanded and cells from the same line were divided and tested for responses on ELISpot using both autologous presentation (black bars) and DR1 presentation (white bars) in a DR1 negative donor (mean with SD error bars, $n = 2$).

3.5 Discussion

In this chapter, naked APCs were transduced with HLA-DR1 and demonstrated to efficiently present peptides to both CD4+ T-cell clones, and polyclonal lines derived from DR1+ donor PBMC. Methodological considerations such as culture serum and effective removal of non-specific peptide were taken into account to ensure that the protocol achieved the aim of identifying DR1-restricted responses.

As this methodology was the basis for epitope mapping of the conserved influenza proteins, certain parameters related to DR1 presentation on ELISpot would impact both the nature of the epitopes identified, and those that were potentially overlooked. These parameters are evaluated in this section based on literature and experimental observations, with conclusions taken into consideration for coming experiments.

3.5.1 Presentation of Peptide by DR1 in the Absence of HLA-DM

As shown in Figure 3.1 the naked APC lines have a deletion that includes HLA-DM in addition to other HLA genes. The role of HLA-DM (fully detailed in Chapter 1) is to catalyse the dissociation and binding of peptides to HLA class-II in the MIIC compartment of a cell. The outcome of its action is to favour presentation of high affinity peptides by HLA class-II, through the equilibrium-like conditions that allow the most kinetically stable pHLA conformations to predominate and be secreted to the cell surface¹⁶⁴. Therefore, peptides with a low affinity for class-II are less likely to be presented if processed in the MIIC compartment, i.e. when entering the cell as whole protein.

Yet, the addition of low affinity peptides as partially digested protein or as synthetic peptide fragments bypasses affinity-based selection. This is because short peptides or digested protein can be loaded onto class-II molecules in the recycling endosomal compartment^{162,163}, where selection is not mediated by HLA-DM and is thus

permissive of weak and low affinity binders. This work was elegantly carried out by the Unanue group^{235,236}.

Thus, transduced APCs lacking HLA-DM have pHLA at their surface that predominantly contains CLIP (the invariant chain)¹³⁵. Yet the addition of synthetic peptide at concentrations far in-excess of those found naturally, permits endocytosis and presentation through the recycling endosome compartment. Therefore, the absence of DM does not negatively impact peptide presentation on the T2.DR1 and 174.DR1 cells as displayed in the literature and experiments from this chapter.

3.5.2 A More Diverse DR1 Peptide Repertoire?

In the DR1.APC presentation system, the use of synthetic peptide (bypassing DM-mediated selection), and the abundance of low affinity CLIP-pHLA (for easy dissociation in the recycling endosome) may result in a more diverse DR1 peptide repertoire than under “natural” conditions. This is unavoidable when using synthetic peptide libraries and must be considered and addressed in later experiments.

The main considerations are whether the resulting T-cell responses are directed to “real” epitopes that are processed and presented naturally, and whether response magnitudes are comparable *in vivo*. The observation of a strong response to peptide on IFN- γ ELISpot does not always correlate with, or is not directly attributable to, *in vivo* responses to whole protein or virus. A clear example of this is the presence of cross-reactive T-cell memory populations to HIV epitopes in unexposed blood donors^{237,238}. Such work confirms that cross-reactive memory populations are capable of mounting detectable responses to epitopes, against which they were not initially primed.

To address this issue in subsequent chapters, the aim was to identify robust and reproducible responses shared between multiple donors. These criteria would decrease the probability of focusing on irrelevant responses that may not represent bona fide, processed epitopes.

3.5.3 Presence of A2

When first described, 174 and T2 cells were shown to express low levels of the HLA class-I allele HLA-A2 at 20-50 % of wild type expression levels²²⁶. Although lower, this does not diminish the ability of these cells to present peptides that can bind HLA-A2 and elicit non-DR1 responses on ELISpot. Consequently, A2+ DR1+ donors were avoided in the peptide pool stage of the screening process, as the presence of A2 responses would confound results at this stage and incur off target responses.

Yet, the very high prevalence of the HLA-A2 allele in European Caucasians means that the DR1+ A2+ genotype is common. Such donors were recruited at a later stage, upon completion of screening when individual epitopes were defined into shorter peptide sequences, and the likelihood of both a DR1 epitope and an A2 epitope within the same sequence was reduced.

3.5.4 Future Work

In this chapter, DR1 transduced APCs were able to isolate HLA-DR1 restricted CD4+ T-cell responses from polyclonal lines on IFN- γ ELISpot. This enabled early identification of peptides that contain a DR1-restricted epitope, and overcame the need for blocking antibodies or the use of matched and mismatched presenting cells. In this way, the process of screening large peptide libraries in subsequent chapters was simplified. Development of this method allowed advancement to more complex studies within the timeframe of the project.

4 Identification of HLA-DR1 Restricted Epitopes within the Internal Influenza Proteins

4.1 Abstract

The adaptive immune response to influenza-A is comprised of humoral and cell mediated immunity. Humoral effector immunity is directed at the external proteins haemagglutinin (HA) and neuraminidase (NA), while cell mediated immunity can additionally target the internal proteins. Relative to HA and NA these proteins are highly conserved, and individuals with strong CD8+ and CD4+ T-cell responses against them show reduced symptom severity and protection from pandemic strains.

CD4+ T-cells recognise peptide fragments presented on HLA class II molecules through their TCR. Few CD4+ T-cell influenza epitopes have been characterised in the literature and even fewer for HLA-DR1. In order to study the CD4+ T cell response to the conserved internal proteins Matrix (M1), Nucleoprotein (NP) and Polymerase Basic-1 (PB-1) in detail, peptide libraries of these three proteins were analysed using DR1.APCs on IFN- γ ELISpot using pooling matrices in two DR1+ donors. 20-30 amino acid regions associated with immunogenicity were identified, and the amino acid sequence of these regions analysed using an HLA binding algorithm to find the likely core sequence of the epitope.

Based on binding algorithm predictions, shorter peptides were tested for immunogenicity in four DR1+ donors. Crystal structures of two of these epitopes were generated and analysed in order to confirm the predicted anchor residues and gain insight into potential TCR contact residues. The epitopes that showed strong and consistent responses were taken forward in later chapters for further analysis using HLA-multimer staining and clonotypic dissection.

4.2 Introduction

Historically, the immunology of influenza has centred on the challenge of generating effective vaccination strategies and understanding the serological responses that ultimately prevent infection. This has led to a focus on the properties of antigenic shift and drift in the surface proteins haemagglutinin (HA) and neuraminidase (NA), which enable the virus to evade pre-existing antibody responses. The epidemiology associated with these proteins has been highly studied and much is known about the immunogenic regions and the protective nature of responses to them^{239–241}.

However, research into the cell mediated response is not limited to solvent exposed conformational epitopes on the virion surface. Instead, the ability of a T-cell to recognise short linear sequences derived from any of the viral proteins means that responses to the internal elements could have an equal importance. These elements are protected from antibody based selection pressure, and are central to viral structure and replicative function. When amino acid sequences of the eight viral proteins from different strains and subtypes are aligned, the internal elements show very high levels of conservation relative to HA and NA^{200,242}.

Having established that cell mediated immunity can respond to all viral proteins, recent studies have shown that a large percentage of T-cell responses are directed towards these elements^{234,242,243} in different challenge platforms and disease settings. Some of the major finds have shown that the magnitude of pre-existing responses to Matrix-1 (M1) and Nucleoprotein (NP) confer protection from severe disease when the antibody repertoire has been breached i.e. in seronegative patients^{48,213}. Polymerase Basic-1 (PB1) has also be identified as an important and highly conserved target of immune responses, although with less immunogenicity than M1 and NP^{213,233,234}.

The exact role of these responses, whether they are truly protective or not, is still the subject of debate, with studies suggesting they are markers of severe infection⁶⁸ and

others suggesting they inhibit seroconversion or the formation of a novel antibody response during infection with an unseen strain^{242,244}.

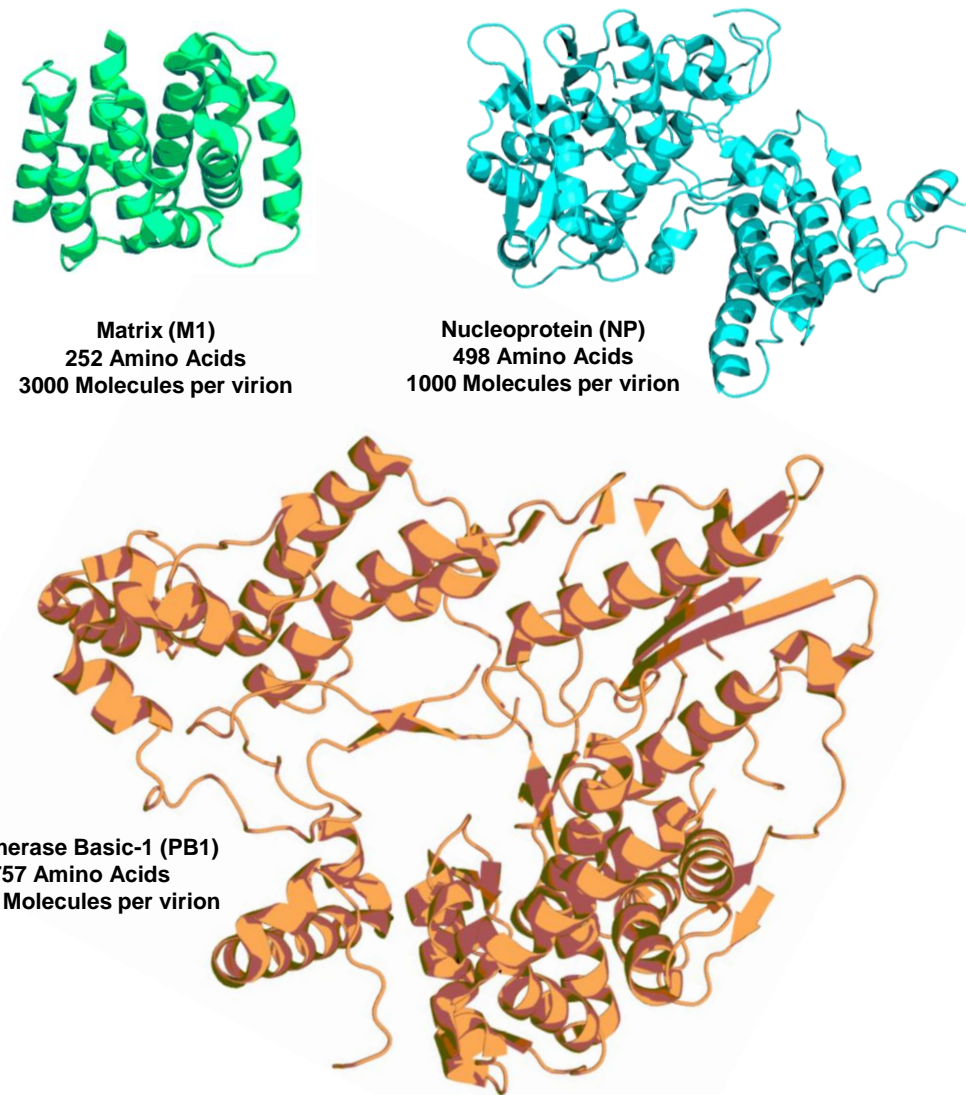


Figure 4.1 Schematic showing the structural representations of three internal influenza proteins. Matrix (1EA3²⁴⁵), Nucleoprotein (2Q06¹⁸⁴) and Polymerase Basic 1 (4WSB¹⁹⁰). Data on the number of molecules per virion²⁴⁶.

The common theme throughout the literature is that strong responses to these proteins exist, and that they are highly important in both protection and mediation of symptom severity. Although CD4 and CD8 T-cells are equally implicated, CD4+ T-cells play a more diverse and complex role in regulating the immune response. Through B-cell interactions they facilitate generation of an effective antibody response, they help prime and direct CD8 T-cells to sites of infection, suppress inflammation and have potential cytotoxic and innate-like functions during the antiviral state^{64,101,247,248}.

CD8+ T-cell responses to influenza have been relatively well studied. For example, epitopes like B35-LPF and A2-GIL have been characterised using crystallography²⁴⁹, biophysical and clonotypic analysis²⁵⁰, as well as being applied as diagnostic markers²²². However, because of the absence of available techniques and study platforms, CD4+ T-cell influenza responses have been less well characterised.

The majority of data on CD4+ T-cell mediated immunity to influenza is limited to protein-specific responses^{213,242}; only a few recent studies have performed detailed epitope characterisation^{233,251,252}. This is a major challenge if we want to understand which epitopes confer protection in multiple people with the same HLA allele, and then elucidate the structural and biophysical mechanisms behind this. By understanding the relationship between an epitope's sequence in the context of its host HLA, we can identify anchor and TCR contact residues that may be susceptible to mutation or potential therapeutic enhancement.

As our understanding of key HLA class-II epitopes and CD4+ T-cell responses progresses, new vaccination strategies, insights into epidemiology, and specific diagnostic markers may follow.

4.3 Aims

While the internal proteins of the influenza virus have been demonstrated as highly immunogenic for CD4+ T-cells, it is not known how many epitopes are present within these proteins that can generate robust and reproducible responses restricted to a single HLA allele. Furthermore, are these epitope-specific responses shared across the population in all individuals who possess the same HLA allele?

In order to address these questions, the aims of this chapter were to screen the internal influenza proteins M1, NP and PB1 using DR1.APCs and PBMC from two HLA-DR1+ donors on IFN- γ ELISpot. This would enable isolation of 20-30 amino acid immunogenic regions that contained DR1-restricted epitopes. These regions could be analysed using HLA-Binding algorithms in order to elucidate the binding registers which comprise the minimal epitopes.

The epitope sequences could be confirmed by testing shorter peptides, 13-16 amino acids in length, and X-ray crystallographic analysis of refolded pHLA proteins. Screening of short peptides against additional HLA-DR1+ donors would assess how common these epitope-specific responses are in the population and identify those epitopes which generated robust and reproducible immunogenicity.

Specific aims:

- 1) To screen peptide libraries of the internal influenza proteins Matrix-1 (M1), Nucleoprotein (NP) and Polymerase Basic-1 (PB1) using DR1.APCs in a pooling matrix format against PBMC from two HLA-DR1 donors.
- 2) To isolate the individual 20-30 amino acid regions that elicited IFN- γ responses from the peptide pools.
- 3) To analyse these regions using HLA-binding algorithms in order to elucidate the binding registers which comprise the minimal epitopes.

- 4) To test shorter peptides, 13-16 amino acids in length, in four HLA-DR1 donors to assess how common these responses were across the wider population.
- 5) To crystallise short peptides in HLA-DR1 and analyse structural information to validate binding algorithm predictions and identify the residues that mediate TCR recognition.

4.4 Results

4.4.1 Peptide Libraries and Pooling Matrix Design

Synthetic peptide libraries of the three internal proteins, Matrix-1 (M1), Nucleoprotein (NP) and Polymerase Basic-1 (PB1) were obtained commercially. Each library consisted of peptides usually twenty amino acids in length with an overlap of ten amino acids to ensure full coverage. For each internal protein, a unique pooling matrix was arranged, where each peptide was only present in two distinct pools (Fig. 4.2). This arrangement allowed pool specific responses to be cross-referenced, and thus individual peptides common to immunogenic pools to be identified. In this way, large numbers of peptides could be analysed for immunogenicity in a reduced number of assays.

M1	P-1	P-2	P-3	P-4	P-5
P-6	1	2	3	4	5
P-7	6	7	8	9	10
P-8	11	12	13	14	15
P-9	16	17	18	19	20
P-10	21	22	23	24	

NP	P-1	P-2	P-3	P-4	P-5	P-6	P-7
P-8	1	2	3	4	5	6	7
P-9	8	9	10	11	12	13	14
P-10	15	16	17	18	19	20	21
P-11	22	23	24	25	26	27	28
P-12	29	30	31	32	33	34	35
P-13	36	37	38	39			

PB1	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9
P-10	1	2	3	4	5	6	7	8	9
P-11	10	11	12	13	14	15	16	17	18
P-12	19	20	21	22	23	24	25	26	27
P-13	28	29	30	31	32	33	34	35	36
P-14	37	38	39	40	41	42	43	44	45
P-15	46	47	48	49	50	51	52	53	54
P-16	55	56	57	58	59	60	61	62	63
P-17	64	65	66	67	68	69	70	71	72
P-18	73	74							

Figure 4.2 Architecture of pooling matrices for M1, NP and PB-1. Pool number is in grey, while overlapping peptides are in white. Protein acronym is in the top left box of each matrix.

4.4.2 Overview of Screens in two HLA-DR1+ Donors

PBMC from two HLA-DR1+ donors were used to screen pooling matrices using the HLA-DR1 presentation and IFN- γ ELISpot methodology described in the previous chapter. The aim was identification of pool specific IFN- γ responses that were ***consistent across multiple replicates, and common to both donors***. In this way, the likelihood of finding robust and reproducible responses in the wider HLA-DR1 population was increased.

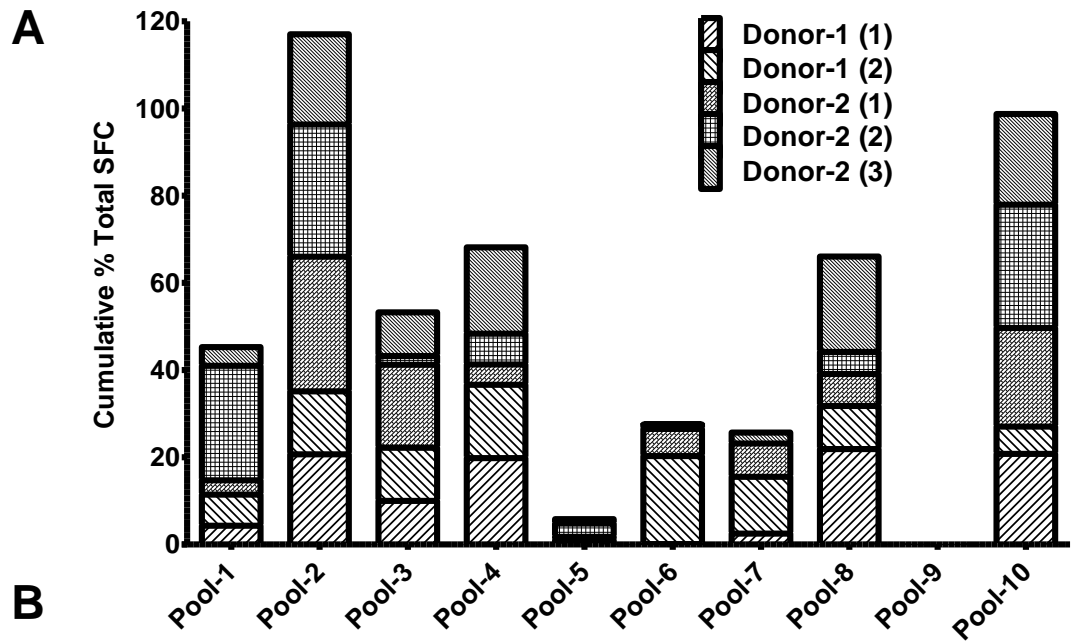
4.4.2.1 Matrix (M1)

The M1 protein is documented as being highly immunogenic to both CD4+ and CD8+ T-cells^{213,233,242}. For such an immunogenic protein, it was likely that multiple, potentially overlapping epitopes (presented by different HLAs) were concentrated across the peptide sequence. This was observed in the cumulative response pattern to peptide pools (Fig 4.3A) and in individual assays where eight out of ten pools showed some level of immunogenicity (Fig 4.3B).

Pool-2 elicited a positive response from both donors in every assay, while pool-3, pool-4, pool-8 and pool-10 showed responses present in both donors in at least one assay each. Weaker and inconsistent responses were observed to other pools. These may have been a result of assay variation, or responses private to only one donor.

Five pools indicated were taken forward for further analysis, with cross referencing of these pools isolating six peptides to be investigated individually (Fig 4.3C)

Matrix Normalised Cumulative Analysis



B

Donor (assay)	1	2	3	4	5	6	7	8	9	10
Donor-1 (1)										
Donor-1 (2)										
Donor-2 (1)										
Donor-2 (2)										
Donor-2 (3)										

↑ ↑ ↑ ↑ ↑

C

M1	P-1	P-2	P-3	P-4	P-5
P-6	1	2	3	4	5
P-7	6	7	8	9	10
P-8	11	12	13	14	15
P-9	16	17	18	19	20
P-10	21	22	23	24	

Figure 4.3 Summary of assays on M1 peptide pools.

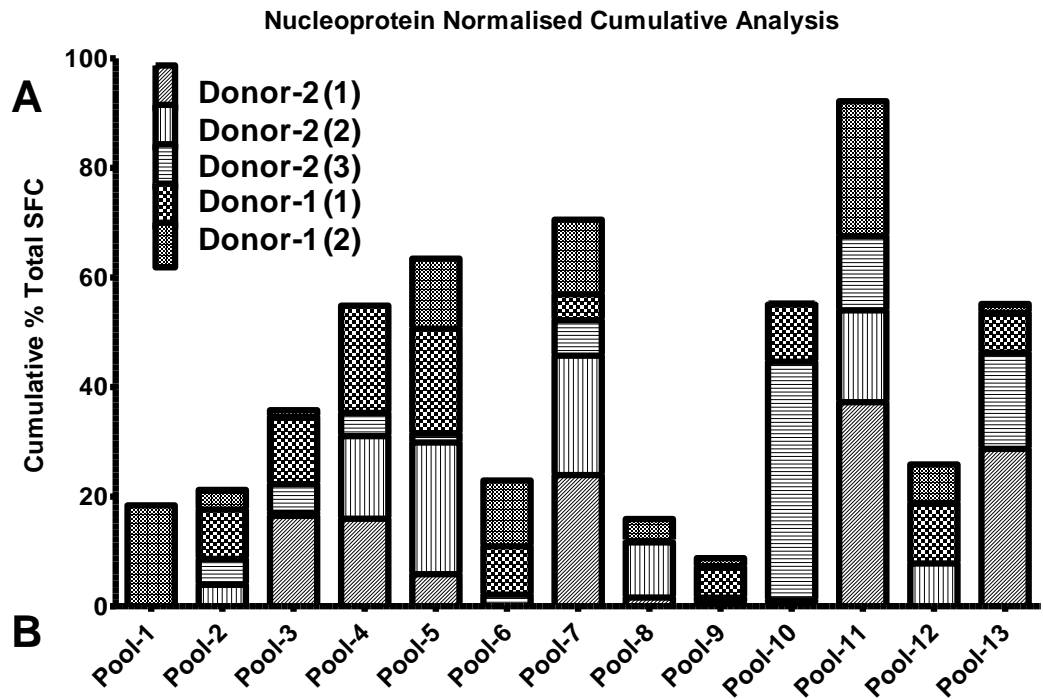
(A) Each pool specific SFC result is normalised by dividing by the total number of SFC across all peptide pools in that particular assay to give a percentage value. Percentage values are stacked for each pool to give a cumulative representation of the responses across multiple assays in two donors. (B) Representation of each assay, with the

response to a pool filled in green if the SFC number was a positive result (greater than 20 SFC per 100,000 PBMC). Arrows indicate the pools taken forward for further analysis. (C) Cross-referencing of positive pools indicating the peptides to be investigated individually (underlined in orange boxes).

4.4.2.2 Nucleoprotein

Like M1, the influenza nucleoprotein is highly immunogenic and multiple epitopes were likely present across the 402 amino acid sequence analysed. The response pattern to the thirteen peptide-pools showed a large degree of replicate variation (Fig 4.4). This inconsistency is exemplified across five separate assays, where 12 of 13 pools were able to elicit a response greater than 20 SFC in at least one replicate (Fig 4.4B).

Amidst this large degree of assay variation, consistent and common responses were present. Pool-11, pool-7 and pool-5 were seen in both donors and generated the highest cumulative responses (Fig.4.4A). Shared responses were observed to pool-3, pool-4 and pool-13, these were positive in three of five assays. Pool-10 exhibited a large response in the cumulative analysis, yet this was largely attributed to a single assay, and neither donor saw this pool more than once, so it was not taken forward (Fig 4.4C).



Donor (assay)	1	2	3	4	5	6	7	8	9	10	11	12	13
Donor-1 (1)		█		█		█	█		█	█		█	█
Donor-1 (2)	█				█	█	█				█		
Donor-2 (1)			█	█			█				█		█
Donor-2 (2)				█	█		█				█		
Donor-2 (3)			█				█			█	█		█

↑ ↑ ↑ ↑ ↑

C

NP	P-1	P-2	P-3	P-4	P-5	P-6	P-7
P-8	1	2	3	4	5	6	7
P-9	8	9	10	11	12	13	14
P-10	15	16	17	18	19	20	21
P-11	22	23	24	25	26	27	28
P-12	29	30	31	32	33	34	35
P-13	36	37	38	39			

Figure 4.4 Summary of assays on NP peptide pools. Format and presentation is the same as in Fig 4.3. **(A)** Cumulative normalised analysis of all responses. **(B)** Representation of pools that elicited a positive result in each assay. **(C)** Cross-referencing of pooling matrix.

4.4.2.3 Polymerase Basic-1 (PB1)

PB1 is the largest of the three proteins analysed, the least abundant in the virion and the least immunogenic. Consequently, for a given length of sequence fewer epitopes were expected relative to M1 and NP. This was observed as a much clearer response pattern with fewer transient and inconsistent responses, despite a larger number of peptides in each pool.

Cumulative analysis highlighted six immunogenic pools, with minimal reactivity towards others (Fig 4.5A). Examination of individual assay results (Fig 4.5B) revealed that these highlighted responses were not common to each donor, with only pool-2, pool-5 and pool-13 shared. Pool-7 and pool-14 showing strong and consistent responses in a single donor. Pool-12 appeared large in cumulative analysis but was dominated by a single assay and was therefore discounted.

Pool-2, pool-5 and pool-13 fitted the criteria for further investigation, while pool-7 and pool-14 did not, yet were highly consistent in their respective responders. Given the lower overall immunogenicity and small number of peptides to be investigated for this protein, these pools were taken forward in the hope that any epitopes found may be of interest in later studies with additional donors.

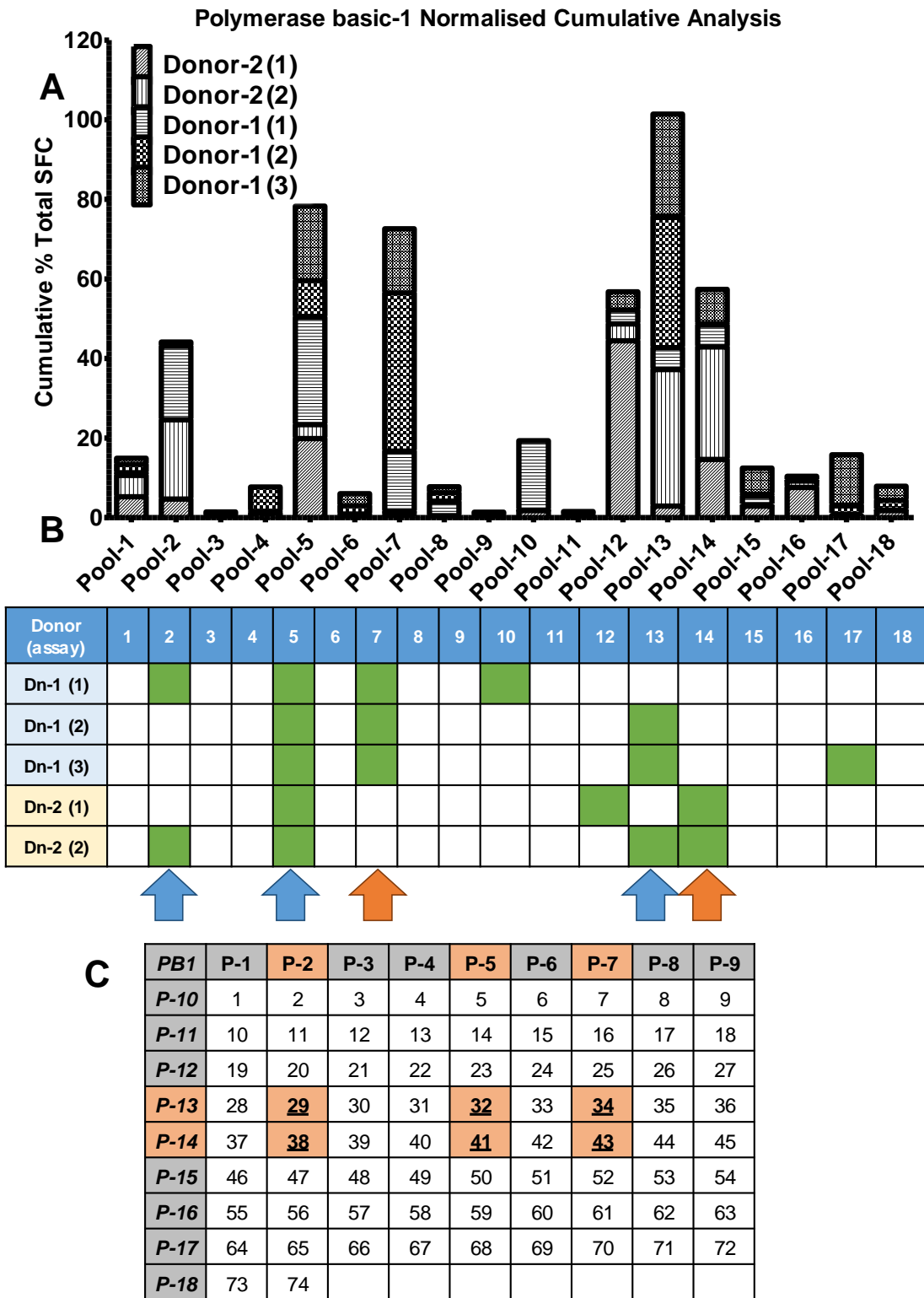


Figure 4.5 Summary of assays on PB-1 peptide pools. Format and presentation is the same as in Fig 4.3. Arrows indicate the pools taken forward for further analysis. Blue arrows indicate the pools which were clearly positive in both donors. Orange arrows indicate pools positive for only one donor.

4.4.3 Identification of Immunogenic Peptides

In order to identify individual peptides, the pool matrices were cross-referenced based on the responses determined in the previous section. Individual peptides were tested by re-stimulation of a line ***expanded against the relevant parent pool***. Each line was confirmed as responsive to the expansion pool, before it was restimulated with the ***relevant individual peptide***. This method ensured that potential intra-pool interactions during the expansion phase, for example competition for HLA binding between pool constituents, were maintained; and testing was not performed on cultures that were biased to produce one response to a single peptide.

In addition, peptides ***not identified as immunogenic by the pooling matrices***, were included in each set of assays. These control peptides and standard negative controls (no peptide) ensured responses were specific, and not the result of an over-stimulated culture, in which non- or partially-specific responses to all peptides are observed (often with high background levels relative to other lines). All IFN- γ ELISpot assays were performed with HLA-DR1 APCs.

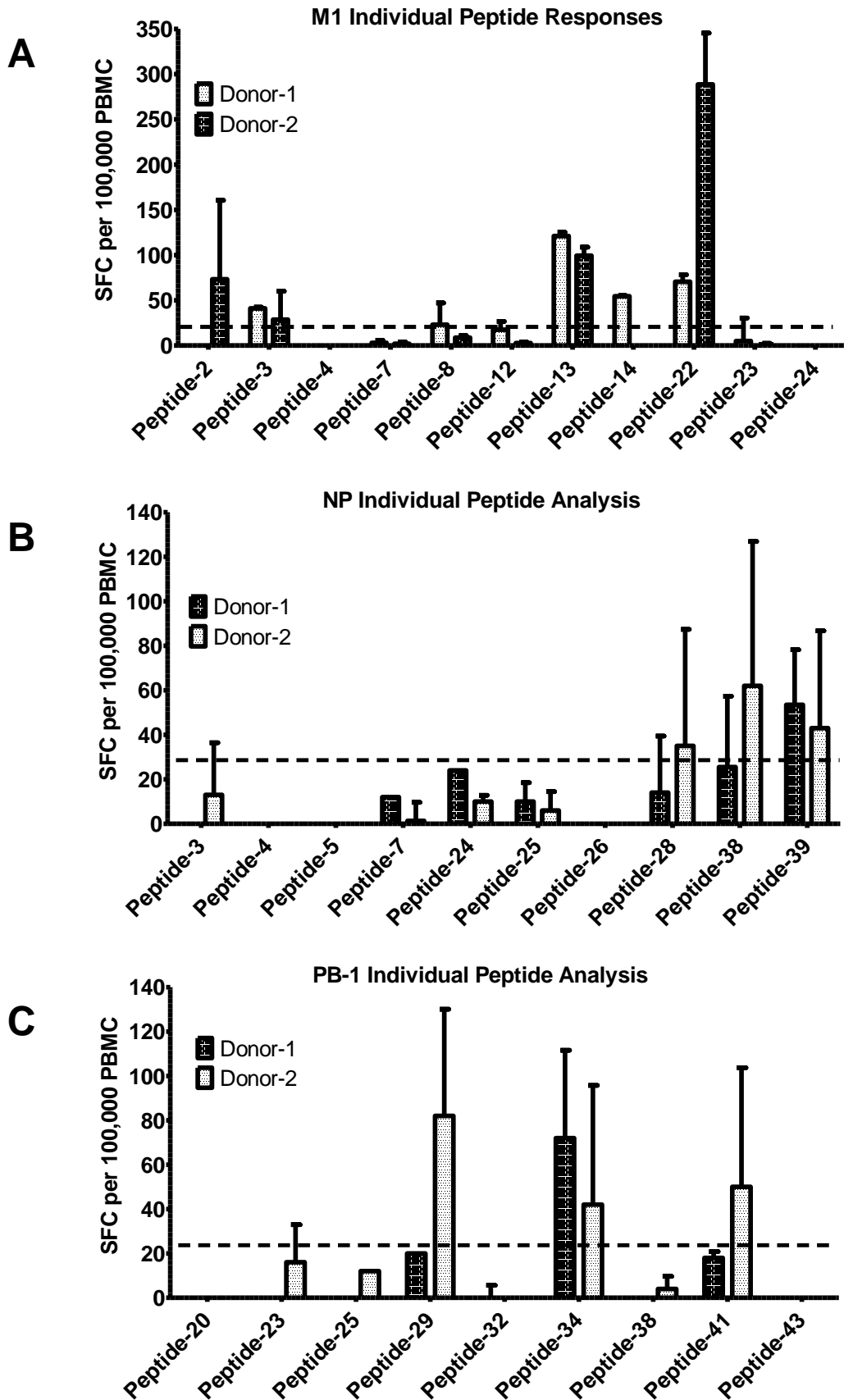


Figure 4.6 Individual peptide analyses of regions identified from pool assays on HLA-DR1 IFN- γ ELISpot. (A) Matrix-1, (B) Nucleoprotein, (C) Polymerase Basic-1.

Figure 4.6 Individual peptide analyses of regions identified from pool assays on HLA-DR1 IFN- γ ELISpot. (A) Matrix-1, (B) Nucleoprotein, (C) Polymerase Basic-1.

For each assay, a line cultured against a parent pool, and shown to be reactive to that pool on IFN- γ ELISpot, was then retested with specific individual peptides from that pool. Due to limited numbers of PBMC and a broad range of testing that occurred, not all peptides were tested equal numbers of times in each donor (mean with SD error bars, donor 1: n = 2, donor-2: n = 3).

4.4.3.1 Matrix (M1)

Testing of peptides identified through cross-referencing of M1 pools revealed peptide-22, and the overlapping peptide-13 and peptide-14 to elicit responses in both donors (Fig 4.6A). These individual peptides accounted for the observed reactivity of the M1 pools.

Although the inclusion of negative control peptides ensured that responses were not aberrant occurrences, these peptides resulted in the discovery of responses not explicitly observed in the pooling data. For example, for pool-2: peptide-12 and peptide-22 were specifically tested, with either peptide-7 or peptide-2 included as a negative control. Here peptide-2 gave a positive and repeatable response in donor-2. As this was used as one of multiple of negative control peptides, it was not extensively tested in donor-1 (the overlapping peptide-3 was), so responses in this donor were not observed.

Analysis of the immune epitope database (IEDB) revealed an HLA-DR1 epitope within peptide-2 (and partially in its overlapping partner peptide-3), which had been highly characterised in early influenza literature²⁵³, and may have been the potential mediator of this response. It was therefore carried forward under the assumption it may be generating a consistent sub-dominant or private response, which could be of interest in later investigations.

4.4.3.2 Nucleoprotein (NP)

Testing of individual nucleoprotein peptides revealed the overlapping peptide-38 and peptide-39 to elicit the strongest responses in both donors (Fig 4.6B). Peptide-28 elicited consistent but weaker responses. The overlapping peptide-24 and peptide-25 had very weak responses close to the significance cut off, with values of 10-20 SFC seen in multiple repeats in both donors. This region was taken forward for further analysis, but it was likely to contain a subdominant epitope with weak reactivity.

These identified peptides did not fully account for the observed responses to the NP pools (Fig 4.4); as pool-5 was consistently recognised in both donors, yet testing of peptide-5 (negative control) and peptide-26 revealed no responses. This suggested the pool contained immunogenic peptides that were not flagged-up by a relevant cross-referencing pool, or that responses were missed through insufficient testing during the individual assays. This pool was later investigated using binding algorithms in section 4.4.4 in order to elucidate the peptide to which the consistent responses could have been directed.

4.4.3.3 Polymerase Basic-1 (PB1)

Three peptides were shown to be immunogenic from PB1 (Fig 4.6C), of which peptide-34 gave consistent and positive responses in each donor. Peptide-41 showed strong reactivity in donor-2 but was slightly below the significance level in donor-1. A similar trend was observed for peptide-29. These three peptides appeared to account for the pattern of pool reactivity, especially for peptide-41/pool-14 where the responses are predominantly through donor-2 (Fig 4.5B). Although this may be a private response, the consistency and lack of other immunogenic peptides meant that it was taken forward for sequence analysis and testing in multiple HLA-DR1+ donors.

4.4.3.4 Summary of Identified Regions

Protein	Peptide Number	Peptide	Sequence Position
M1	2	VLSIVPSGPLKAEIAQRLED	11-30
	3	KAEIAQRLEDVFAGKNTDLE	21-40
	Combined	VLSIVPSGPLKAEIAQRLEDVFAGKNTDLE	11-40
M1	13	AGALASCMGLIYNRMGAVTT	121-140
	14	IYNRMGAVTTEVAFGLVCAT	131-150
	Combined	AGALASCMGLIYNRMGAVTTEVAFGLVCAT	121-150
M1	22	QMVQAMRTIGTHPSSSAGLK	210-230
NP	28	GIDPFRLQLNSQVFLIRPN	271-290
NP	38	RASAGQISVQPTFSVQRNLPFER	371-393
	39	QPTFSVQRNLPFERATIMAAFTG	379-402
	Combined	RASAGQISVQPTFSVQRNLPFERATIMAAFTG	371-402
NP	24	ARSALILRGSVAHKSLPAC	231-250
	25	VAHKSLPACVYGLAVASGY	241-260
	Combined	ARSALILRGSVAHKSLPACVYGLAVASGY	221-260
PB1	29	NVVRKMMTNSQDTELSFTIT	284-203
PB1	34	NVLSIAPIMFSNKMARLGKG	335-354
PB1	41	PGMMGMFNMLSTVLGVSIL	405-424

Table 4.1 Summary of identified regions found by pools and individual analysis.

Parent protein, corresponding peptide numbers, amino acid sequence and the position of those residues within the whole protein are listed.

4.4.4 Defining Epitopes within Long Sequences

Immunogenic regions 32 to 20 amino acids in length from each of the three proteins were identified following the screening process (Table 4.1). The observed immunogenicity could have arisen from single or multiple epitopes within these sequences, presented by HLA-DR1 independent of proteolytic cleavage. Each long peptide will theoretically be able to adopt multiple binding registers when presented in the open-ended HLA class II groove. Within a peptide, the binding register or “core” is defined by the P1, P4, P6 and P9 residues which anchor into the HLA binding pockets. Each binding register represents an epitope with the potential to stimulate a distinct T-cell population (for more see introduction section on class II processing and presentation).

The challenge of decoding which of multiple potential registers or “cores” result in the observed immunogenicity is the reason why many studies define long sequences but do not progress further^{243,252}. Those that do progress require “chop-down” or sub libraries of shorter peptides combined with further immuno- or binding-assays to provide conclusive information²³³. These methods are comprehensive but time consuming, and produce the clearest result when using a single T-cell clone instead of a polyclonal line.

It is also unclear how informative using reduced length peptides are when screening the HLA-Class II system. Residues beyond the core, i.e. flanking regions, may have a stabilising effect²⁵⁴, with the optimal peptide length for a T-cell assay suggested as 18-20 amino acids²⁵⁵. Shorter peptides may activate cognate T-cells less efficiently, due to weaker HLA binding or a greater disposition to proteolytic cleavage of the core nonamer. Experimentally distinguishing between an important flanking residue and an anchor residue may be inconclusive depending on the sequence in question.

Based on these limitations and the timeframe of the PhD project, obtaining chop-down libraries and T-cell clones for nine distinct peptide sequences in two donors was not feasible. Instead, an alternative strategy was undertaken.

4.4.4.1 Using HLA-Binding to Identify Epitopes

There are many factors which influence immunodominance, or the magnitude of a response to an epitope, such as processing²⁵⁶, antigen abundance²⁵⁷ and the stability of the pHLA complex²⁵⁸. Work in cancer has demonstrated that epitopes which exhibit strong HLA binding generally elicit stronger cognate T-cell responses than epitopes which bind weakly²⁵⁹. *In vivo*, when antigen processing and presentation occurs via the MHC compartment in the presence of HLA-DM, the natural processing machinery exerts a selection preference for epitopes that form the most stable pHLA class II complexes^{141,260}. Although this high affinity selection process is bypassed when using molar excess of synthetic peptide, it will have been undoubtedly relevant during priming of cognate memory populations that have been restimulated and identified in this study.

Thus, the working hypothesis was that within the regions identified, the immunogenic epitopes are comprised of ***registers that bind HLA-DR1 with highest affinity***. The binding strength of different registers within a region could be estimated using prediction algorithms, and the preferred registers tested using peptides between 13-14 amino acids in length.

4.4.4.2 Outputs of Binding Data

HLA class-II crystal structures and information on eluted peptides using mass spectrometry have provided a wealth of data on the amino acid sequences capable of binding to specific HLA proteins and the characteristics of peptide anchor residues and HLA binding pockets. This information forms the basis of HLA binding algorithms, which predict the affinity of ***each*** potential binding register in a given sequence based on the physical characteristics of each amino acid.

The algorithm used here was NetPanMHCII 3.1²⁶¹⁻²⁶⁴. The 32-20 amino acid sequences were processed, and predictions analysed. The algorithm outputs a sequence of specified length (e.g. 13-mer peptides), identifies the 9-amino acid binding

“core” and gives a predicted affinity. The results are ranked in order of strongest to weakest binders, and each core is displayed only once, in the most favoured register position (Fig 4.7A).

A key factor of the program is that it takes into account flanking sequences and chemical properties **beyond the core** into the overall affinity prediction. Therefore, when comparing a specific core within different output lengths (e.g. 11-mers and 15-mers), the size and amino acid composition of the flanking region will change. The resulting affinity differences may impact the ranking order of potential cores within a given input sequence (Fig 4.7B).

Generally, the ranking order of potential cores is independent of output length, but this can still affect a number of examples (see following section). Taking this into consideration, summarised outputs (Table 4.2) were derived from the consensus predictions of all possible binding 11-mers, 13-mers and 15-mers. This ensured binding registers were not estimated from only the anchor residues within each core, but included potential flanking contributions as well.

Overlapping Peptides Identified in Screens

M1 Peptide-13 AGALASCMGLIYNRMGAVTT
M1 Peptide-14 IYNRMGAVTTEVAFGLVCAT

Sequence Processed By Algorithm

AGALASCMGLIYNRMGAVTTEVAFGLVCAT

Peptide length 13

Threshold for Strong binding peptides (%Rank) 2%

Threshold for Weak binding peptides (%Rank) 10%

Allele: DRB1_0101

Seq	Allele	Peptide	Identity	Pos	Core	Core_Rel	1-log50k(aff)	Affinity(nM)
9	DRB1_0101	LIYNRMGAVTTEV	Sequence	2	YNRMGAVTT	0.915	0.817	7.26
6	DRB1_0101	CMGLIYNRMGAVT	Sequence	3	LIYNRMGAV	0.455	0.655	41.67
5	DRB1_0101	SCMGLIYNRMGAV	Sequence	2	MGLIYNRMG	0.400	0.629	55.29
3	DRB1_0101	LASCMGLIYNRMG	Sequence	3	CMGLIYNRM	0.585	0.558	119.59
14	DRB1_0101	MGAVTTEVAFGLV	Sequence	3	VTTEVAFGL	0.785	0.475	293.87
0	DRB1_0101	AGALASCMGLIYN	Sequence	1	GALASCMGL	0.355	0.409	601.75
1	DRB1_0101	GALASCMGLIYNR	Sequence	2	LASCMGLIY	0.380	0.389	741.11
12	DRB1_0101	NRMGAVTTEVAFG	Sequence	2	MGAVTTEVA	0.640	0.370	911.87

Number of strong binders: 1 Number of weak binders: 1

Figure 4.7A Example of the output of NetPanMHCII 3.1. Two overlapping peptides were identified from the previous section. The algorithm processes the sequence by analysis of all possible 13-mer binding registers with respect to HLA-DR1 (DRB1*0101). Registers are ranked based on predicted binding affinity (nM) strongest to weakest. The top sequences capable of binding are displayed. For each possible binding register the “core” 9 amino acids defined by peptide anchor residues at P1, P4, P6 and P9 is shown. The output is set so that **each core is displayed once** (see materials and methods).

Output length Can Impact Predicted Affinities								
# Peptide length 11			Output Length = 11 Amino Acids					
# Threshold for Strong binding peptides (%Rank) 2%								
# Threshold for Weak binding peptides (%Rank) 10%								
# Allele: DRB1_0101								
Seq	Allele	Peptide	Identity	Pos	Core	Core_Rel	1-log50k(aff)	Affinity(nM)
8	DRB1_0101	PLKAEIAQRLE	Sequence	1	LKAEIAQRL	0.805	0.564	111.75
0	DRB1_0101	VLSIVPSGPLK	Sequence	1	LSIVPSGPL	0.765	0.562	114.50
1	DRB1_0101	LSIVPSGPLKA	Sequence	2	IVPSGPLKA	0.550	0.528	165.23
# Peptide length 13			Output Length = 13 Amino Acids					
# Threshold for Strong binding peptides (%Rank) 2%								
# Threshold for Weak binding peptides (%Rank) 10%								
# Allele: DRB1_0101								
Seq	Allele	Peptide	Identity	Pos	Core	Core_Rel	1-log50k(aff)	Affinity(nM)
0	DRB1_0101	VLSIVPSGPLKAE	Sequence	1	LSIVPSGPL	0.495	0.666	37.30
1	DRB1_0101	LSIVPSGPLKAEI	Sequence	2	IVPSGPLKA	0.825	0.632	53.74
6	DRB1_0101	SGPLKAEIAQRLE	Sequence	3	LKAEIAQRL	0.820	0.621	60.59
# Peptide length 15			Output Length = 15 Amino Acids					
# Threshold for Strong binding peptides (%Rank) 2%								
# Threshold for Weak binding peptides (%Rank) 10%								
# Allele: DRB1_0101								
Seq	Allele	Peptide	Identity	Pos	Core	Core_Rel	1-log50k(aff)	Affinity(nM)
0	DRB1_0101	VLSIVPSGPLKAEIA	Sequence	3	IVPSGPLKA	0.740	0.704	24.65
6	DRB1_0101	SGPLKAEIAQRLEDV	Sequence	3	LKAEIAQRL	0.800	0.629	55.17
15	DRB1_0101	QRLEDVFAGKNTDLE	Sequence	5	VFAGKNTDL	0.550	0.372	897.30

Figure 4.7B Example with M1 peptide-2 and peptide-3 of how the output length can impact the predicted binding affinities and hence ranking of each potential core within a sequence.

The same M1 sequence (VLSIVPSGPLKAEIAQRLEDVFAGKNTDLE) was processed based on three different output lengths: 11-mer, 13-mer and 15-mer. The resulting top binding cores and predicted affinities of the housing register are shown. The relative ranking position of each core with respect to one another changes. There is no overall top ranked consensus across three output lengths, so in this example two cores (LSIVPSGPL and LKAEIAQRL) must be taken for further investigation.

Protein	Peptide Number	Peptide	Sequence Position	NetMHCIIpan 3.1 Prediction		Reduced Length Peptide for Testing
				Core-1	Core-2	
M1	2	VLSIVPSGPLKAEIAQRLED	11-30	LSIVPSGPL	LKAEIAQRL	SGPLKAEIAQRLED*
	3	KAEIAQRLEDVFAGKNTDLE	21-40	LKAEIAQRL	N/A	
	Combined	VLSIVPSGPLKAEIAQRLEDVFAGKNTDLE	11-40	LSIVPSGPL	LKAEIAQRL	
M1	13	AGALASCMGLIYNRMGAVTT	121-140	YNRMGAVTT	LIYNRMGAV	GLIYNRMGAVTTEV**
	14	IYNRMGAVTTEVAFGLVCAT	131-150	YNRMGAVTT	LIYNRMGAV	
	Combined	AGALASCMGLIYNRMGAVTTEVAFGLVCAT	121-150	YNRMGAVTT	LIYNRMGAV	
M1	22	QMVQAMRTIGTHPSSSAGLK	210-230	VQAMRTIGT	MRTIGTHPS	
NP	28	GIDPFRLQNSQVFLIRPN	271-290	FRLQNSQV	N/A	DPFRLQNSQVFS
NP	38	RASAGQISVQPTFSVQRNLPFER	371-393	FSVQRNLPF	N/A	LPFERATIMAAFT PTFSVQRNLPFER
	39	QPTFSVQRNLPFERATIMAAFTG	379-402	FERATIMAA	FSVQRNLPF	
	Combined	RASAGQISVQPTFSVQRNLPFERATIMAAFTG	371-402	FERATIMAA	FSVQRNLPF	
NP	24	ARSALILRGsvAHKSCLPAC	231-250	LRGsvAHKS	ALILRGsvA	
	25	VAHKSCLPACVYGLAVASGY	241-260	YGLAVASGY	CVYGLAVAS	
	Combined	ARSALILRGsvAHKSCLPACVYGLAVASGY	221-260	No Consensus		
PB1	29	NVVRKMMTNSQDTELSFTIT	284-203	VRKMMTNSQ	N/A	NVVRKMMTNSQDT
PB1	34	NVLSIAPIMFSNKMARLGKG	335-354	MFSNKMARL	VLSIAPIMF	IMFSNKMARLGKG
PB1	41	PGMMMGMFNMLSTVLGVSIL	405-424	FNMLSTVLG	N/A	GMFNMLSTVLGVS

Table 4.2 Summary of immunogenic regions and the corresponding binding predictions by NetPanMHCII 3.1. The reduced length peptides chosen for further testing are based on the predictions and are displayed in the final column. Single amino acids highlighted in red, are those missing from one of the two parent peptide sequences while the remaining residues are present in both (8 of 9 residues found in both parent peptides, red is the one missing).

*The predicted core in this region agreed with a 14-amino acid sequence SGPLKAEIAQRLED detailed in published work²⁵³. **The predicted core of the 14-amino acid sequence of GLIYNRMGAVTTEV agreed with published work finding this to be an HLA-DR1 epitope²³³.

4.4.4.3 Analysis of Binding Predictions

For certain immunogenic regions, the top predicted cores were unanimous and gave estimated binding affinities less than 20 nM (for simplicity all affinities are based on 13-mer predictions), suggesting strong binding to HLA-DR1. The strongest predicted binders were in the NP peptide-28 (core: FRLLQNSQV, 5 nM) and M1 peptide-13/14 (core: YNRMGAVTT, 7 nM) and PB1 peptide-41 (core: FNMLSTVLG, 10 nM). These three regions had an aromatic residue at position P1, known to favour binding in that pocket²⁶⁵, and hydrophobic residues at P4. The anchor residues at P6 and P9 were uncharged but varied in size and polarity.

The M1 peptide-22 (core: VQAMRTIGT, 17 nM) also showed a single core with predicted binding highly favoured over other cores in the same sequence. Instead of an aromatic at P1 a bulky aliphatic residue, valine, was present. Replicating motifs in the strong binders detailed above, the residue at P4 was methionine, with P6 and P9 both threonine.

Within this analysis a number of complex, overlapping regions of potential immunogenicity were observed. For M1 peptide-2/3 three different cores of similar affinity were predicted (LSIVPSGPL, 37 nM; IVPSGPLKA 54 nM; LKAEIAQRL, 61 nM), each with an aliphatic leucine or isoleucine at the first anchor position. In this case, only one

core (LKAEIAQRL) displayed any relative overlap. The other cores did not overlap with peptide-3 by more than three residues.

LKAEIAQRL was complete in peptide-2, but contained 8 amino acids in peptide-3, missing only the first anchor. Loss of the first anchor has been shown in other studies to maintain some level of response to an otherwise complete peptide²³³, and may have explained the consistent but weak responses to peptide-3.

Examination of the IEDB and early literature showed the sequence containing this core (LKAEIAQRL) to have been characterised as a HLA-DR1 epitope²⁵³. The sequence defined in the literature was selected for further investigations (SGPLKAEIAQRLED). The absence of this information would have necessitated full investigation of each core, exemplifying the challenge certain immunogenic regions may pose.

NP peptide-38 and peptide-39 were the final two components of the NP peptide library and as a result were longer, at 23 amino acids each, in order to accommodate the full c-terminal sequence of the protein. They overlapped by 14 amino acids, and within this overlapping region a core (FSVQRNLPF) was predicted at 49 nM affinity. Yet peptide-39 contained a non-overlapping core with an affinity of 28 nM (FERATIMAA). Both predictions contained the aromatic phenylalanine as the P1 first anchor.

As a result, reduced peptides corresponding to each core were chosen for testing. It was hoped that a strongly dominant epitope would be characterised, given the magnitude of responses observed to these two peptides and their parent pools.

A similar strategy was applied to PB1 peptide-34 where two cores (MFSNKMARL and VLSIAPIMF) of similar affinity (44 nM) were observed and two corresponding reduced 13-mer peptides chosen for further analysis. Neither core appeared overly favourable, with a methionine or a valine at the first anchor then proline or methionine at position 6.

The most difficult analysis occurred with the overlapping NP peptide-24 and peptide-25. These peptides gave responses close to, or below the significance cut off in individual assays (Fig 4.6C), and were carried forward to look for any obvious binding

register. Four binding cores, three of similar affinity (22-29 nM) and one weaker (78 nM) were present, none of which fully overlapped both peptides, and each showing variation depending on output length. This complexity, coupled with the relatively weak SFC numbers observed during screens meant that this region was not investigated further.

From these predictions, peptides 13-14 amino acids in length were obtained at high synthetic purity (>90 %). A slightly longer sequence containing the core VQAMRTIGT was ordered (Table 4.3) to include 5 amino acids at the N-terminus (instead of 2-3) permitting crystallographic investigation of the N-terminal flanking region. The reduced sequences of M1 peptide-2 and peptide-13/14, contained 3 amino acids to conform with sequences used in literature studies.

4.4.5 Confirmation of Epitope Immunogenicity

Nine peptides of reduced length were taken forward for testing (Table 4.3). These peptides were ordered commercially at 90 % purity (HPLC) in order to reduce the risk of impurities and unwanted chemical modifications shown to impact immunogenicity²⁶⁶. Testing was expanded to include two additional HLA-DR1+ donors, and was performed on lines cultured against a single peptide only.

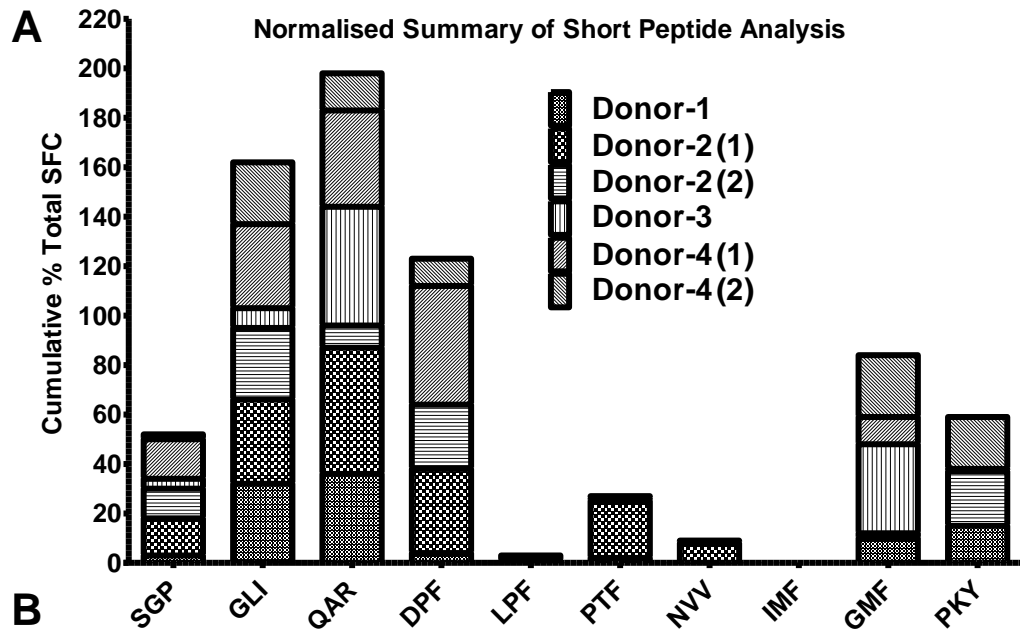
Protein	Short Peptide	Sequence Position
M1	SGPLKAEIAQRLED	17-30
	GLIYNRMGAVTTEV	129-142
	QARQMVQAMRTIGTHP	208-222
NP	DPFRLLQNSQVFS	273-285
	LPFERATIMAAFT	388-401
	PTFSVQRNLPFER	380-392
PB-1	NVVRKMMTNSQDT	284-297
	IMFSNKMARLGKG	342-354
	GMFNMLSTVLGVS	410-422

Table 4.3 Short peptides for further testing.

Reduced peptides showed a specific pattern of immunogenicity when compared across four donors (Fig 4.8). The peptides GLI and QAR gave strong and consistent responses in each individual and assay. SGP, GMF and DPF showed a reduced response strength in terms of relative SFC but were recognised in three of the four donors. These five peptides were carried forward for further

analysis.

The remaining peptides LPF, NVV, PTF and IMF elicited minimal responses below the significance level when tested (Fig 4.8B). IMF represented one of two predicted cores in PB1 peptide-34; this elicited no detectable responses in two donors, despite positive responses to the parent 20-mer in the side-by-side assays.



Donor (Assay)	Matrix			Nucleoprotein			Polymerase Basic-1			HA
	SGP	GLI	QAR	DPF	LPF	PTF	NVV	IMF	GMF	PKY
Donor-1				Orange	Dark Grey			Dark Grey		
Donor-2	(1)				Dark Grey					Dark Grey
	(2)						Dark Grey	Dark Grey	Orange	
Donor-3	Orange				Dark Grey			Dark Grey		
Donor-4	(1)				Dark Grey					Dark Grey
	(2)						Dark Grey	Dark Grey		

Figure 4.8. Summary analysis of IFN- γ DR1 ELISpot testing on lines cultured with short peptides across 4 HLA-DR1+ donors. (A) Cumulative analysis on normalised SFC, where each response was dividing by the total number of SFC across all peptide tested in that particular assay to give a percentage value. Percentage values are stacked to give a cumulative representation of the responses across multiple assays in four donors.

(B) Representation of each assay, with the response to a specific peptide filled in green if the SFC number was a positive result (greater than 20 SFC per 100,000 PBMC). Boxes are orange if a response was borderline (of two replicates one was just above the significance level and one was just below but the mean was below 20 SFC). White indicates no response. Dark grey indicates not tested.

Follow-up experiments, revision of data and analysis of the literature shed light on why some of these peptides failed to replicate the reactivity seen to their parent peptide. Misidentification of the core was only specific to one example, other possible reasons related to the HLA-DR1 presentation methodology and precursor frequencies in the sample blood. For case-by-case explanation of why some peptides may have failed to elicit responses, see discussion section 4.5.4.

These results identify some of the epitopes that may be providing protection from Influenza in the HLA-DR1+ population. Later assays using a larger starting number of PBMC would overcome some methodological issues such as precursor frequency and begin to explore the prevalence and genetic make-up of these responses across the population.

4.4.6 Confirming Algorithm Predictions Using Crystal Structures

To confirm that the chosen sequences bound to HLA-DR1 as predicted by the binding algorithms in the previous section, several reduced length peptides were refolded to produce class-II monomer for crystallisation. Crystal trays were set up using the TOPS screen²¹⁵, an array of HEPES and Tris buffered conditions optimised for crystallisation of HLA-class I protein.

Crystals were obtained for QAR and SGP from concentrations of 8.00 mg/ml and 8.23 mg/ml respectively (Fig. 4.9). QAR crystals from condition A11 diffracted to 1.64 - 1.70 Å resolution and SGP from condition F9 to 2.66 – 3.96 Å. The data sets of highest resolution were solved to generate structures (Table 4.4) that were used to explore the orientation and features of each peptide within the class-II binding groove.

These proteins were of interest, as neither had a typical aromatic anchor at the P1 position in the first HLA binding pocket, being leucine for SGP and valine for QAR. Structural analysis would be able to identify if these were the true anchors, and ensure that the peptide was not forming disordered conformations or unpredicted interactions with the HLA. Of the other reduced peptides which were refolded only GLI and DPF produced crystals which diffracted to < 4 Å, but these failed to produce data sets that could be solved to any satisfactory level.

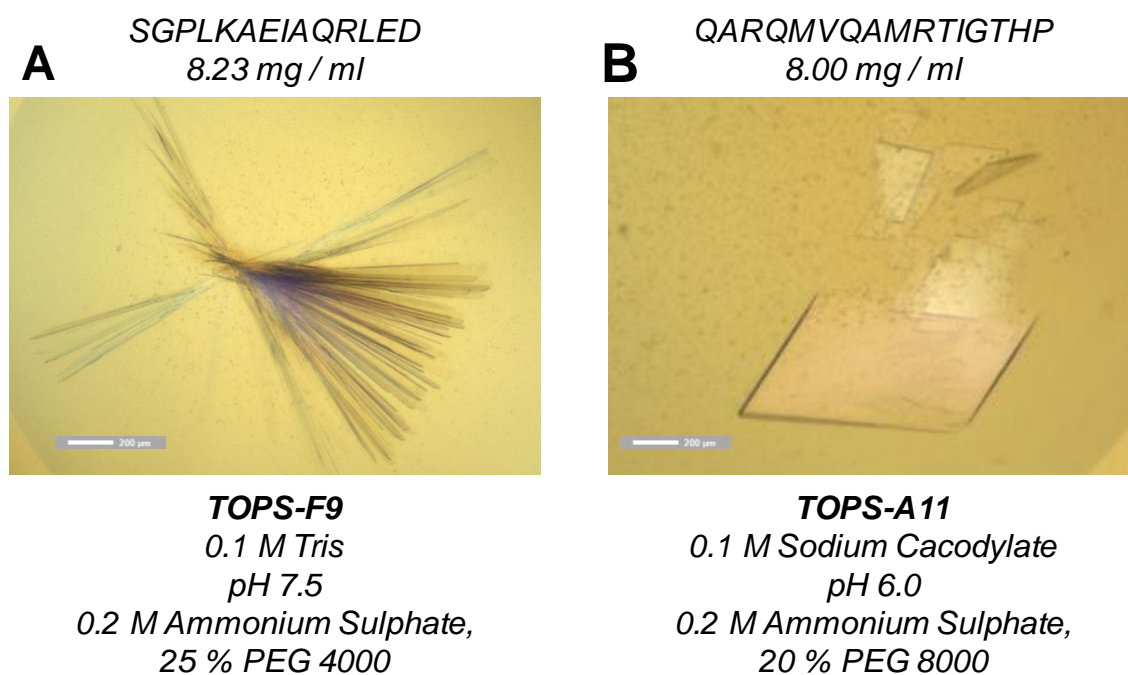


Figure 4.9. Images and details of the successful crystallisation conditions for (A) SGP and (B) QAR. TOPS-F9 and A11 correspond to positions on the 96 well screen, with exact contents detailed below. White scale bar corresponds to 200 µm.

	SGP	QAR
Resolution	2.66	1.64
Completeness	98.9%	99.8%
Space Group	P 1 21 1	P 21 21 2
R Value	0.200	0.197
Free R Value	0.261	0.238
Free R Value Test Set Size	4.8%	5.0%
Free R Value Test Set Count	1267	2819
Mean B Value	42.95	39.43
Number of Reflections	25022	54085
RMS Deviations From Ideal Values		
Bond Lengths	0.016	0.018
Bond Angles	2.220	1.997

Table 4.4. Crystallographic Data Table. SGP details in the left column and QAR in the right.

4.4.6.1 Peptide Anchor Residues

The features of the peptide which facilitate binding and p-HLA complex stability are the backbone (main chain) contacts and four buried residues that sit in binding pockets that “anchor” into the HLA. Polar contacts between the backbone and side chains of the binding groove hold the main chain in highly rigid conformation as observed by B-Factor analysis (Fig. 4.10) that is a consistent feature of class-II structures^{127,267}.

Anchors agreed with the predicted cores assigned in section 4.4.4 by net pan MHC II. The first binding pocket of HLA-DR1 is highly hydrophobic and accommodates large aliphatic or aromatic residues. This was occupied by leucine and valine for SGP and QAR respectively, each residue had a low B-factor (dark blue Fig. 4.10) and was buried deep within the HLA (Fig. 4.11).

The P4 residues occupying the second binding pocket had greater B-factors (Fig. 4.10) and were not buried as deeply within the HLA (Fig. 4.11C, 4.11G). This pocket has been shown to bind acidic side chains in other crystal structures²⁶⁷. The acidic side chain P4Glu of SGP formed polar contacts with DR1 β Gln70 and DR1 β Arg71 (Fig. 4.12A), while QAR-P4Met, a large hydrophobic residue was buried closer to the hydrophobic surface of the pocket (coloured orange Fig. 4.11G) and formed no polar contacts (Fig. 4.12B).

At P6 the situation was reversed, with SGP-P6Ala forming no polar contacts (Fig. 4.12A), and QAR-P6Thr forming polar contacts with DR1 α Asp66 and DR1 α Asn62 (Fig. 4.12B). The P9 pocket is hydrophobic with SGP-P9Leu deeply buried (Fig. 4.12A), but the comparatively smaller polar QAR-P9Thr-OH forming a polar contact with DR1 α Asn69 at the pocket entrance, with QAR-P9Thr-CH3 buried towards the hydrophobic pocket (Fig. 4.12B).

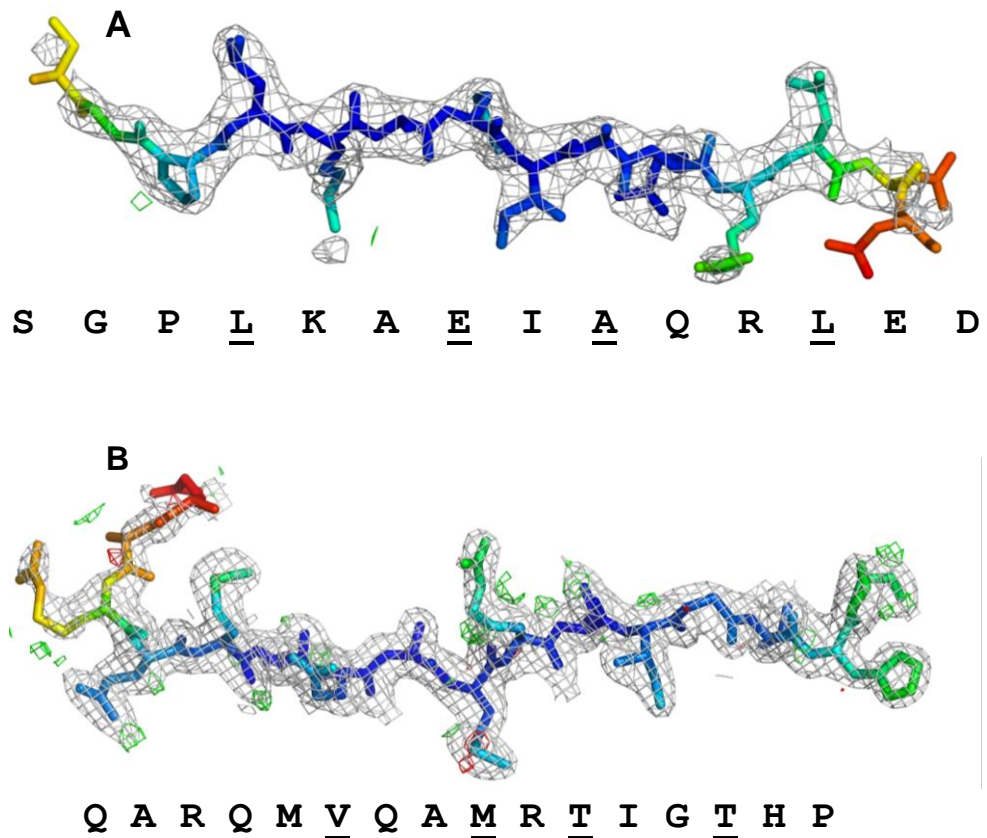


Figure 4.10. *B-factor and electron density representation of each peptide at 0.3 sigma. (A) SGP peptide shown at 2.66Å. (B) QAR shown at 1.66Å. Anchor residues are underlined. B-factor colouring spectrum represents lowest B-factors in dark blue, and increasingly high B-factors as lighter blue, green, yellow, orange and red. This represents the degree of order based on elastic scattering.*

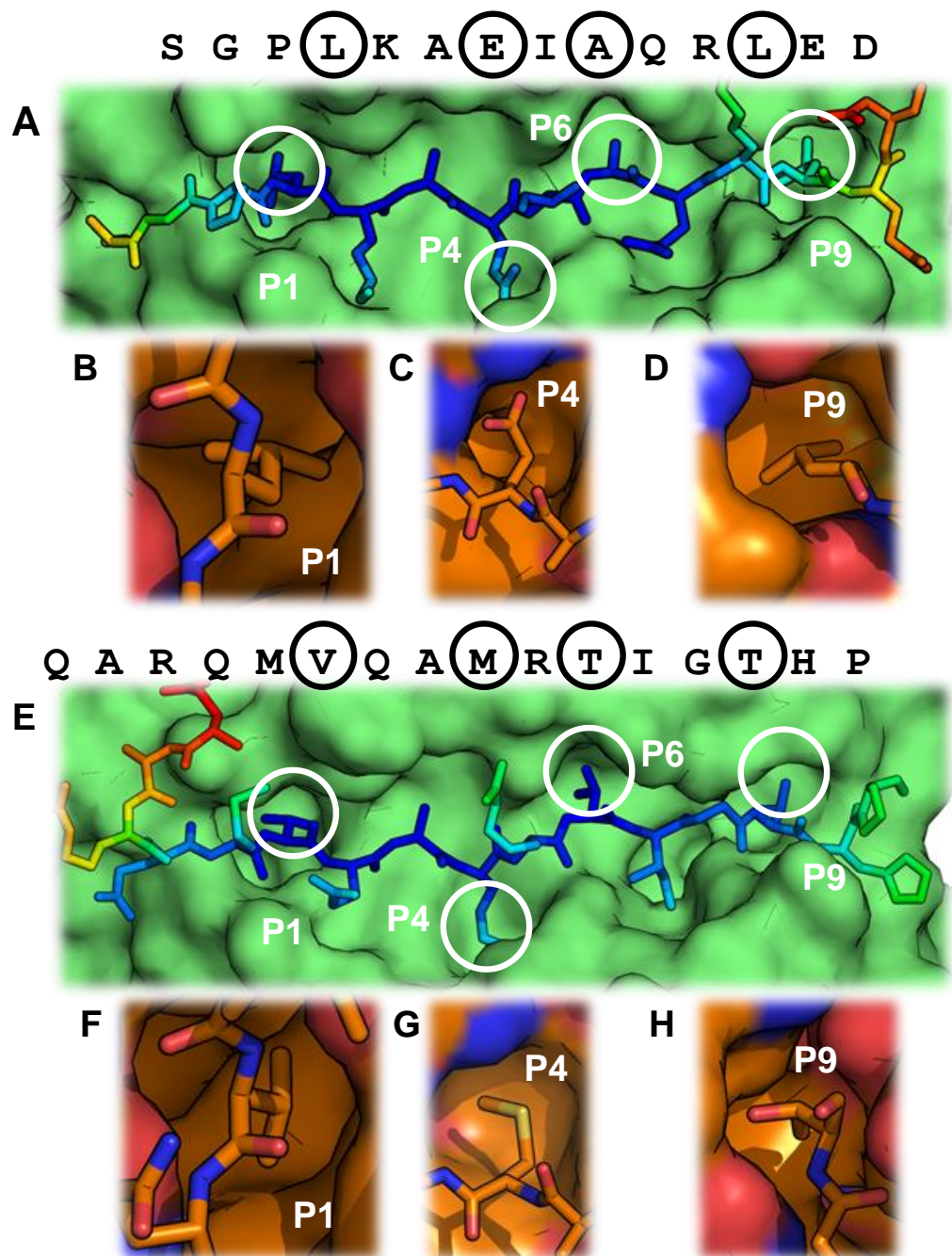


Figure 4.11. Identification of peptide anchor residues and orientation within the binding pockets. Binding grooves of **(A)** SGP and **(E)** QAR, with HLA surface in green and peptide sticks coloured by B-factor. Each binding pocket and corresponding anchor residue is circled white. To better show the binding pockets for P1, P4 and P9 they are shown below each groove picture (SGP – **B**, **C**, **D** and QAR- **F**, **G**, **H**), both peptide and HLA are coloured by element (carbon orange, nitrogen blue, oxygen red, sulphur yellow). This displays the likely hydrophobic interactions (orange) as well as potential polar or charged contacts (red/blue).

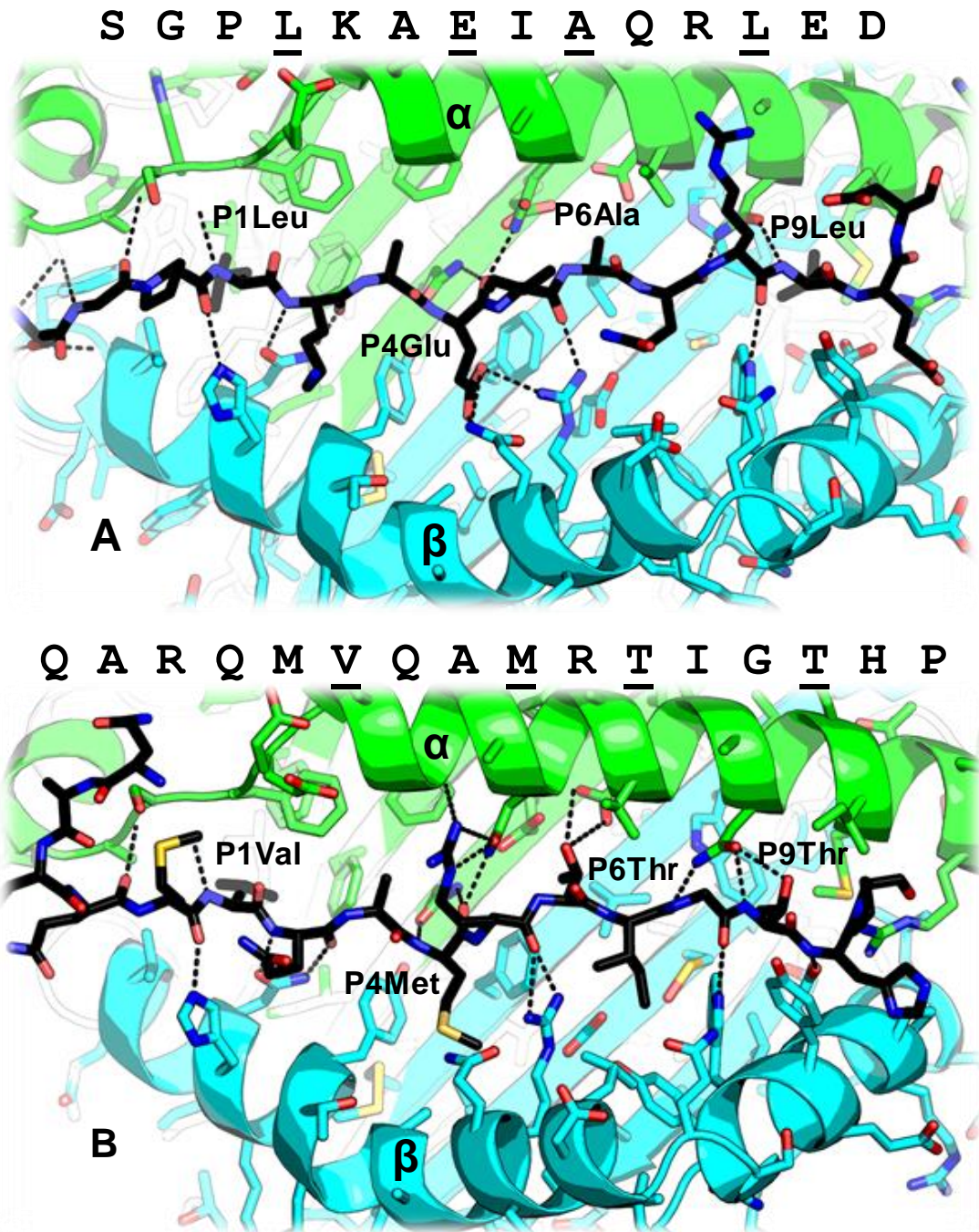


Figure 4.12. Polar interactions between the peptide and side chains of the binding groove. The HLA groove is displayed by secondary structure representation for (A) SGP and (B) QAR, with side chains shown. α -chain in green, β -chain in blue. Peptide is displayed in stick format with a black carbon backbone. Side chain and backbone sticks are coloured by element (carbon orange, nitrogen blue, oxygen red, sulphur yellow). Polar interactions between peptide backbone and side chain with HLA groove side chains are represented by black dashed lines.

4.4.6.2 Defining the Epitope: TCR contact residues?

The literature definitions of an HLA class-II epitope are very broad, with any length of peptide capable of stimulating a CD4+ T-cell response defined as an epitope. At the molecular level, solvent exposed amino acids at, or close to, the conventional docking site of the TCR could help define what the T-cell “sees” and hence, the epitope.

For HLA-DR1 these TCR contact residues are P2, P3, P5, P7 and P8 but may include N- and C-terminal flanking residues beyond the core, specifically P-1 as detailed in several complex structures^{167,219}. The orientation of these contact residues was analysed for both QAR and SGP (Fig. 4.13). Inside the core (P1-P9), residues SGP-P2Lys, SGP-P8Arg and QAR-P5Arg are charged and thus capable of forming salt bridges with incoming TCR chains. Polar glutamines at SGP-P7 and QAR-P2 have hydrogen bonding potential, and aliphatic isoleucines at SGP-P5 and QAR-P7 may contribute to hydrophobic interactions.

Beyond the core, at C-terminal positions P10 and P11, charged residues were present in each peptide that may bind the TCR β -chain. At the N-terminal position P-1 of SGP, proline presented a hydrophobic surface for interaction, while serine and glycine showed high B factors and may not be oriented toward the TCR interface.

The N-terminal flank of QAR consists of five residues, QARQM, and presents a more complex morphology (Fig. 4.13C, 4.13D). Based on B-factor analysis, the terminal P-5Gln and P-4Ala were disordered and fell over the DR1 α -chain (Fig. 4.13D), while the P-3Arg was oriented such that the side chain guanidinium group pointed out, away from the core (Fig. 4.13C). The P-2Gln side chain lay over the DR1 β -chain, while P-1Met chain pointed up, and was the most likely residue to form hydrophobic interactions with the TCR.

The observation that the N-terminal flank of QAR loops back over the DR1 α -chain, instead of pointing away from the likely TCR interface is of interest, as the flanking residues may be poised to make TCR interactions in this conformation. Whether this occurs when in complex with the TCR requires further investigation.

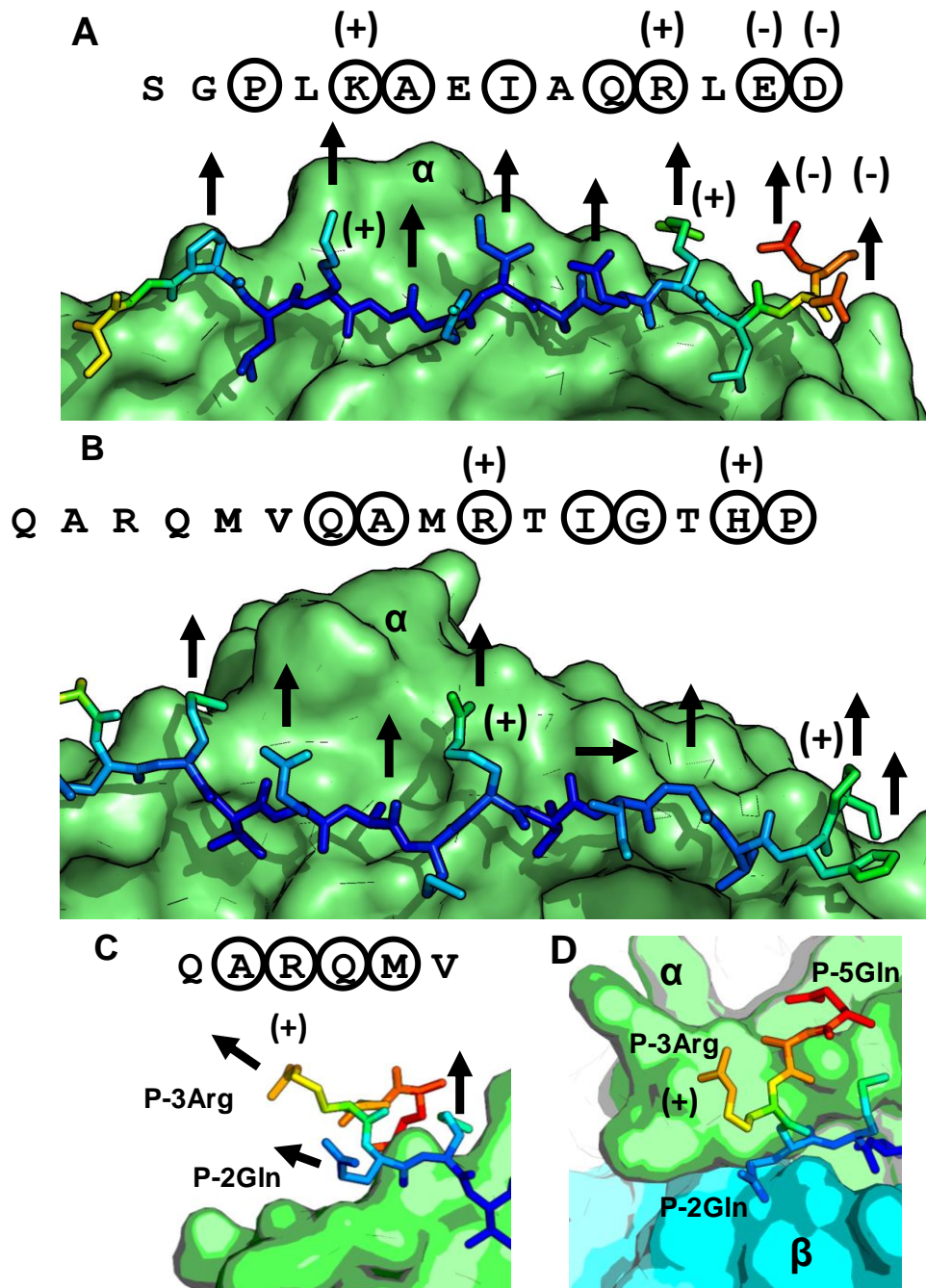


Figure 4.13. Orientation of potential TCR contact residues and QAR N-terminal flanking region. For (A) SGP and (B) QAR, peptide is coloured by B-factor in stick format, with the HLA α -chain in green behind. Charged side chains are labelled positive or negative. Arrows point in the side chain direction for each potential contact position (P-1, 2, 3, 5, 7, 8, 19, 11). The N-terminal peptide flanking residues are shown for QAR from the side (C) and top-down (D).

4.5 Discussion

Due to variation of influenza surface proteins (HA and NA), the investigation of its most conserved proteins at the epitope level may offer insight into the nature of the immune response. CD4+ T-cells sit at the heart of the cell mediated response, yet the epitopes that facilitate their action have not been studied in detail.

In this chapter, several immunogenic regions of the internal proteins of Influenza were identified initially in two HLA-DR1+ donors. These regions were analysed using HLA-binding algorithms and the predicted core epitopes tested in four DR1+ donors with highly pure peptides of a reduced length. Five consistently immunogenic epitopes were found: three from M1, one from NP and one from PB1. Crystallographic analysis confirmed binding algorithm predictions for two epitopes and indicated the likely solvent exposed residues that mediate T-cell recognition.

Although five HLA-DR1 epitopes were defined, it is likely that a complete picture of the immune response mediated by this allele has not been revealed. Especially for the larger proteins NP and PB-1, where the single epitopes identified account for a fraction of the total response to screening pools.

The reasons for this are related to the process of epitope mapping, which is highly complex and is affected by multiple factors that can confound or halt each stage of investigation. These include: the variable nature of the immune response, the overall immunogenicity of a protein, concentration of epitopes across a sequence, presentation by different HLA alleles, donor responses to culture conditions, the use of synthetic peptides and the screening methodology itself.

These factors are assessed in the following discussion in order to account for the identified, misidentified and potentially overlooked DR1 epitopes within these proteins.

4.5.1 The Precursor Frequency and Sensitivity of Cognate T Cells

Assuming that an epitope is presented by HLA-DR1, and that a cognate T-cell population exists in the peripheral blood of a donor, the limiting factors in identification of a response are the precursor frequencies of cognate cells and their sensitivity to antigen.

In these experiments, every T-cell line was set up with a starting number of 600,000 PBMC. This means if a T-cell is present in peripheral blood at a frequency higher than 1 in 600,000, it is likely that it will be detected in every assay. If the frequency is less than 1 in 600,000 then the probability of detection decreases, for example, to one positive result in every two assays.

During epitope mapping, the precursor frequency of a population is balanced by its sensitivity to antigen, which can be defined as the propensity to expand in culture and the magnitude of its cytokine response during an immunoassay. A T-cell may be present at a high precursor frequency, yet not proliferate with high efficiency or elicit strong IFN- γ responses on ELISpot. For these less active T-cells, it could require a greater number of precursors to reach consistent and positive levels of detection above background.

Precursor frequency and T-cell sensitivity may be related, but these factors are mainly attributed to the most recent exposure to antigen (during influenza infection or vaccination) and the biophysical characteristics of the TCR and pHLA interaction, respectively. Other factors such as pHLA stability and antigen presentation may play a role, but are unlikely to limit detection, due to the excess of synthetic peptide used in culture and immunoassay conditions.

Ultimately, the epitopes found in the chapter were chosen for their ability to elicit strong and consistent responses, with M1 short peptides GLI and QAR followed by NP short peptide DPF best satisfying these criteria. Epitopes such as M1-SGP (weak but consistent IFN- γ responses) and PB1-GMF (strong but inconsistent IFN- γ responses) may have been on the threshold of detection based on their balance of precursor frequency and sensitivity.

Responses below or close to this threshold, in either donor, could partially account for the transient and inconsistent response patterns seen towards the peptide pools. Such responses could have been detected in only one or two assays, and would have been discounted due to lack of consistency. Such thresholds and parameters have given rise to the terms immunodominant and sub-dominant when attempting to classify epitopes based on their tendency to elicit a response of a defined magnitude.

4.5.2 Culturing with a Peptide Pool versus an Individual Peptide

In the previous section, the term “propensity to expand in culture” was deliberately used. This has relevance when appraising the response to a peptide pool or an individual peptide. In culture, the addition of cytokines such as IL-2 and growth factors are intended to drive T-cell expansion in response to antigen, yet if multiple antigens are present i.e. multiple peptides in a peptide pool, the cytokine balance or milieu may be influenced by the dominant responses.

For example, if a peptide, which elicits a strong Th1 response, is present, this may drive expansion of clones to other peptides through helper cytokines that would have been insensitive under normal culture conditions. Conversely, if Th17 or Th2 or T-regulatory responses were stimulated, suppressive or different phenotypes, which do not favour expansion of IFN- γ producing cells may have arisen from a culture.

The second factor when culturing with a pool, is whether peptide competition for HLAs exist during the activation and expansion process? I.e. do peptides that bind strongly outcompete others that bind to the same HLA, and inhibit responses to weak binders? This question has been explicitly tested by the Kwok group in tetramer guided epitope mapping studies^{268,269}. Lines were cultured to a mixed pool of immunodominant and subdominant peptides and the resulting tetramer stains compared to lines cultured with individual peptides²⁶⁸. Counterintuitively, these experiments showed that the mixed pools favoured responses to all peptides, including subdominant, and suggested that competition for the same HLA during the expansion process was not a limiting factor.

Their explanation of the increased response to subdominant peptides was that **greater cytokine release, or “help,”** contributed to the expansion of weaker populations. Not suggested by them, is whether a stabilising effect on that particular HLA occurs, thus ensuring that other weakly binding epitopes are more likely to be presented. The greater abundance of the complementary HLA at the cell surface due to stabilisation with a strong binder would increase the probability of loading and presentation of other peptides via the recycling endosome (see class II presentation introduction).

What this means for pools and the methodology used in this chapter is that the detection of a response to a subdominant (low sensitivity or low precursor frequency) epitope will be highly influenced by the binding characteristics and responses of **other peptides in the same pool**. Therefore, a response may be detected in one peptide pool, while in another with different constituents, it may not.

This applies explicitly in the case of M1 peptide-2, where a response was detected from the highly immunogenic M1 pool-2, which contained a dominant epitope (VQAMRTIGT). Yet the corresponding parent pool-6 showed minimal HLA-DR1 responses.

Peptide-2 was detected by chance as it was used as a negative control. Many other subdominant responses may have been present that were missed due to these dynamic interactions. Whether such subdominant responses in culture (associated with low proliferative capacity and low antigen sensitivity), are also subdominant *in vivo* and hence of lesser clinical relevance is an important future consideration.

4.5.3 Using Algorithms to Define Epitopes

The use of algorithms was based on the premise that the immunogenic epitopes are comprised of **registers that bind HLA-DR1 with highest affinity**. The strategy was undertaken for practical reasons and it appears successful, with strong responses to peptides that had predicted affinities less than 20 nM (Fig 4.8). The SGP 14-mer peptide has a predicted affinity of 67 nM, and showed consistent but relatively reduced response strength in terms of SFC.

In the misidentified regions, or pool responses that were unaccounted for, did the strategy of epitope prediction fail? This is highly relevant to epitopes such as SGP, where a consistent and detectable response is present but the binding affinity of predicted cores was not sufficiently skewed towards a single register. Analysis of the process detailed in this chapter suggest this only applies to M1-SGP itself, and NP peptides-24/25 where the potential predictions coupled with weak responses deterred from further investigation.

In the cases of NP peptide-38/39 and PB1-34 two potential cores were identified in each, and tested where possible. These regions are specifically discussed in section 4.5.5, but their misidentification or lack of responses did not result from algorithm failure, instead methodological and practical issues were to blame.

Following the converse argument, were strongly binding epitopes identified through algorithm predictions ***missed by the pool screening methodology?*** Indeed, input of each entire protein sequence into the binding algorithm, and analysis of the top 20 hits per sequence, finds all 5 epitopes that were detected and characterised here ranking highly. The cores from peptides M1-GLI and NP-DPF are the top ranked in their respective proteins, while PB1-GMF core is 5th highest and M1-QAR 8th.

Yet, strong binding cores are not the sole criteria for an epitope, as shown with NP-FSV and NP-FER which were not able to elicit consistent responses. Where peptides bind an HLA with very high affinity but elicit no IFN- γ response is an interesting area of investigation. Whether they are simply not processed *in vivo* (and therefore have no memory populations) or are not Th1 epitopes i.e. Th17 (IL-17 producers) or regulatory (IL-10) is beyond the scope of this investigation. It is also possible that they are simply presented through another HLA allele that outcompetes HLA-DR1 during processing, something directly observed here (see section 4.5.5).

4.5.4 Epitope Missed within Highly Immunogenic Peptide Pools

Following a thorough review of binding algorithm and summarised assay data, some potential epitopes were highlighted that could have explained responses to pools in which immunogenic peptides ***were not identified.***

4.5.4.1 Nucleoprotein Pool-11

NP peptide pool-11 was extremely immunogenic in both donors, yet only peptide-28 was identified. Analysis of binding predictions for the whole NP protein shows peptide-23 to contain the second strongest binding core. This peptide was never tested as its corresponding parent pool gave consistently weak responses.

All three cores from the discontinued investigation of peptide-24 and peptide-25 feature in the top 10 from NP and potentially warrant further investigation.

4.5.4.2 Polymerase Basic-1: Pool-5, Pool-12 and Pool-13

During the analysis of PB1, a major source of confusion were the strong responses to pool-5 and pool-13, coupled with the inconsistent but strong response to pool-12. Analysis of the top ten binding cores within the whole PB1 protein revealed eight to be contained in the immunogenic pools-5, 12 or 13, from which only two peptides were identified (FNMLSTVLG and VLSIAPIMF) and carried forward for further investigation.

Peptide-32 was tested in two assays but did not yield any response, while all others (peptide-50, peptide-25, peptide-22 and peptide-33) were not tested as the relevant cross-reactive pool was not sufficiently active, potentially due to the factors discussed.

Had epitope prediction been coupled with assay results from the beginning, it is possible that even more HLA-DR1 responses could have been identified within NP and PB1. Such lessons will be relevant in returning to this study and in future epitope mapping projects.

4.5.5 Analysis of Three Misidentified Epitopes

It is likely that some epitopes associated with HLA-DR1, particularly sub-dominant or transient responses were missed. Had the goal been to fully map these protein for every possibly response then greater effort would have been applied in this endeavour. Instead, the goal was identification of the ***shared and consistent responses*** - those that may be mediating protection and limiting symptom severity.

Those regions that were initially identified as fitting the above criteria, but did not fully mature are discussed in this final section. These regions are still of interest, as they serve as specific case studies of the complications encountered during epitope mapping, and it is worth accounting for their incomplete progress.

4.5.5.1 NP-38 & NP-38: FSVQRNLPF & FERATIMAA

The key to these peptides relies on a discussion from the previous chapter, which states that the HLA-DR1 presentation methodology could present a more diverse peptide repertoire than that presented *in vivo*, due to the use of synthetic peptide and the recycling endosomal pathway. The NP peptides 38/39 contained two predicted cores (FSVQRNLPF and FERATIMAA), both of which were tested and gave minimal and inconsistent responses via DR1 presentation. This was puzzling due to the strong responses to full-length peptides and parent pools (Fig. 4.3.A and 4.5.B) and the absence of other strong binders within this region.

Further analysis was carried out by repeating the assays using normal autologous presentation side-by-side with DR1 presentation (data not shown). This experiment and subsequent analysis of the literature clarified the responses patterns observed. For both short peptides, autologous presentation but not DR1 presentation, yielded positive results in two donors tested, and a strong response (greater than 200 SFC) in each donor. Analysis of recent literature showed each peptide contains epitopes presented through another HLA allele.

FSVQRNLPF has been characterised as a DR12 epitope²⁷⁰, the other DR allele of donor-2 who displayed the strongest response. FERATIMAA is a well-documented B*35/B*07 epitope that elicits strong CD8+ T-cell responses across the population²⁵⁰. The explanation for why these two peptides were consistently detected on an HLA-DR1 screening platform is the primary concern. It appears to be explained by their length of 23 amino acids, and the fact they both contain strong DR1 binding cores. It is possible that each peptide was loaded during the pulsing process, highly stable to PBS washing of APCs, before addition to the overnight ELISpot assay, during which cleavage and exchange to other donor HLA antigens could take place.

4.5.5.2 PB1 Peptide-29: VRKMMTNSQ

The only case where an inconsistent response from the pool screens was investigated further was that of PB1 peptide-29. This was due to the lack of other

immunogenic regions within this protein. Of the two cross-referenced pools, the strongest responses were to PB1 pool-13, which contained another more immunogenic epitope within peptide-34. Responses to the other parent pool, PB1 pool-2 were detected in 2 of 5 assays (one per donor), but when detected, the response magnitudes were strong (~100 SFC), suggesting a low precursor frequency of highly sensitive clones.

Subsequent investigation yielded a reduced peptide PB1-NVV that gave negative responses during testing. The reduced peptide chosen was based on a unanimous prediction from the parent peptide, although its binding affinity was the weakest of all those tested. This again could be an example of where a 20-mer is more effective at eliciting a response than a 13-mer, and hence when an epitope is on the ***threshold of detection***, this makes a measurable difference.

Initial pooling interactions, weak HLA binding and a low precursor frequency likely account for the poor response to this reduced peptide. If the assays were repeated, lines could be set up with a greater number of starting PBMC to overcome low precursor frequency, and both DR1 and autologous presentation could be tested side-by-side to account for weak HLA binding, or the potential that this epitope was presented predominantly through another HLA.

4.5.5.3 PB1-34: MFSNKMARL & VLSIAPIMF

This was the final peptide for which a strong response was found to the parent pools and the individual peptide demonstrated no reactivity. From PB1 peptide-34 two cores were present in the same sequence, both predicted as binding with 40 nM affinity. Only one peptide (IMFSNKMARLGKG) was obtained and tested. Results show no detectable response in two donors and further testing was not carried out. Testing of the second core will be carried out in future investigations.

4.5.6 Future Work

The identified epitopes from the internal proteins matrix, nucleoprotein and polymerase basic-1 that elicited consistent responses in four HLA-DR1+ donors were taken forward for further analysis. Having established that each epitope was able to stimulate a PBMC line of IFN- γ producing CD4+ T-cells in four donors, the next step was to probe the magnitude and specificity of these responses at the cellular level. This would permit clonotypic analysis of relevant populations and a better understanding of the cells that ultimately mediate immunity to the conserved elements of the influenza virus.

5 Epitope-Specific CD4+ T-cells from HLA-DR1 Donors Exhibit Shared Cellular and Genetic Characteristics

5.1 Abstract

In the context of an acute infection, analysis of epitope-specific CD4+ T-cell populations and their constituent T-cell receptor (TCR) sequences can help resolve the cellular and genetic characteristics that mediate immunity and protection.

Using five epitopes derived from conserved influenza proteins identified in the preceding chapter, analysis of CD4+ T-cell populations in five HLA-DR1 donors was performed. HLA-multimer staining enabled stratification of epitopes based on the magnitude of their cellular responses, while TCR sequences of epitope-specific populations were obtained using next generation sequencing with the α -chain data fully analysed. TRAV gene usage and CDR3 α diversity were compared between donors, while CDR3 α motifs and amino acid frequencies were compared for each TRAV gene (and CDR3 length) in order to dissect the underlying genetic and physical architecture.

Focused TRAV gene usage was observed in response to two epitopes, with broader usage observed in response to four other epitopes. Epitopes with broader TRAV gene usage appeared to show greater response magnitudes and reproducibility in all donors. TCR CDR3 α amino acid motifs were identified with strong conservation of physical properties such as charge and polarity at key positions. Additionally, several “public” TCR amino acid sequences were identified. The described work contributes to further understanding of CD4+ T-cell immunity to conserved influenza proteins, revealing unexpected sharing of response characteristics across multiple HLA-DR1 donors.

5.2 Introduction

The use of fluorochrome-conjugated HLA-multimer technology allows the direct enumeration of epitope-specific T-cell populations by flow cytometry. This is useful when probing polyclonal lines expanded from PBMC, where cells of interest can be analysed relative to their parent population of CD8+ or CD4+ T-cells. The mean fluorescence intensity (MFI) of HLA-multimer binding cells can indicate the relative avidities of the overall population or sub-populations that respond to an antigen. The relative magnitude and MFI of an epitope-specific population may be linked to the efficacy and strength of the resulting T-cell response and thus relate to protection from disease.

Subsequent clonotypic analysis can dissect the genetic composition of antigen-specific populations based on their constituent TCR sequences. This is of interest when analysing the diversity and V(D)J gene usage of TCRs that respond to a particular epitope. In response to immunodominant HLA-class I influenza epitopes, examples of both highly focused^{221,271} and diverse repertoires²⁵⁰ have been found in multiple individuals. Limited data are available for influenza-specific CD4+ T-cells, with clonotypic analysis restricted to HA₃₀₇₋₃₁₉ in three HLA-DR1 donors^{168,272}.

When looking across human and mice populations, sharing of TCR sequences at the amino acid and nucleotide level is observed in response to important autoimmune^{273,274} and viral epitopes^{249,275}. This sharing, termed “publicity,” is thought to arise from an increased prevalence of certain TCRs in the naïve repertoire due to convergent recombination^{276,277}. TCRs from this enriched pool may have an increased probability of selection in response to immunodominant epitopes following repeated exposure²⁷⁸ and may be important indicators of the epitope-specific responses which confer protection.

The study of public responses to the universal HLA-A2 restricted M1₅₈₋₆₆ epitope has spanned two decades of research^{221,279} and provided a wealth of information on CD8+ T-cell repertoire dynamics^{278,280,281} and structural insights^{249,250}. No parallel exists for CD4+ T-cell responses to influenza or other highly transmissible viral pathogens.

Those studies which have compared epitope-specific CD4+ T-cell populations have mainly focused on single epitopes²⁷² in chronic viral or autoimmune settings^{271,273,282–284}. It is not known how acute infection shapes the responding CD4+ T-cell repertoires, and whether strong comparisons with CD8+ T-cell mediated immunity exist. Acute infections may not exhibit the same gene usage patterns seen in response to either self or persistent antigens.

Given that peptides are presented by HLA class-II in a flat, extended conformation⁵ with a natural variation in flanking length^{169,170}, it is hypothesised that responding TCR repertoires may exhibit greater diversity and wider gene usage than class-I responsive repertoires²⁸⁵. Testing of this hypothesis requires several distinct epitopes and multiple donors to begin to encompass the many potential responses that could arise.

In this chapter, comparison of epitope-specific CD4+ T-cell responses in multiple individuals identified shared cellular and genetic patterns that mediate immunity in the HLA-DR1+ population. Similar cellular responses, biased TRAV gene usage, conserved CDR3 α amino acid residues and public TCR sequences were all observed. These observations helped guide further structural analysis in order to elucidate the molecular mechanisms of TCR repertoire recognition, and raised further questions concerning epitope-specific immunity.

5.3 Aims

When using HLA-multimer staining to analyse epitope-specific cell populations, it is important to establish that identified cells are functionally relevant. The first aim of this chapter was to confirm that multimer staining correlated with effector function by comparison of IFN- γ ELISpot and flow cytometry data from the same cell populations expanded against conserved influenza epitopes.

The second aim was to assess the extent to which five different HLA-DR1+ donors were able to respond to the same conserved epitopes. Specifically, were all donors able to respond, and were the responses comparable in terms of %CD4+ T-cells and MFI?

The final aim was to investigate the TCR sequences that mediate epitope-specificity through clonotypic analysis. Examination of TRAV gene usage, CDR3 length and amino acid composition would provide insight into how conserved the mechanisms of TCR recognition are across individuals who share a common HLA allele.

Specific aims:

- 1) To compare the techniques of IFN- γ ELISpot and HLA-multimer staining on the same lines expanded against conserved influenza epitopes to see if correlations exist.
- 2) To assess the epitope-specific populations in five HLA-DR1 donors using HLA-multimers and flow cytometry.
- 3) To analyse the clonotypic information from epitope-specific CD4+ T-cells in five donors through fluorescence activated cell sorting and next generation sequencing.

5.4 Results

5.4.1 Flow Cytometry Analysis of Epitope-Specific CD4+ Populations

In order to analyse epitope-specific populations on flow cytometry, biotin tagged pHLA-monomers were generated using recombinantly expressed HLA-DR1 α - and β -chains, and synthetic peptide. Six batches of monomer comprising three matrix peptides (SGP, GLI, QAR), one nucleoprotein (DPF), one polymerase basic-1 (GMF) and one control HA peptide (PKY) were successfully produced using the same high purity peptides tested in the previous chapter.

Small quantities of monomers (2-3 μ g) were multimerised as needed using a PE-conjugated dextran backbone, and used to stain expanded polyclonal lines with an optimised protocol²¹⁷. HLA-multimer populations were quantified within CD3+/Live/CD4+/CD8- populations by percentage of the parent CD4+ population and the mean fluorescence intensity (MFI) of the multimer binding cells. Irrelevant HLA-DR1 multimers and fluorescence minus one (FMO) were used to set each HLA-multimer gate.

5.4.1.1 Comparison of HLA-Multimer Staining and IFN- γ ELISpot Data

In the previous chapter, each of the six peptides was able to elicit strong and consistent IFN- γ ELISpot responses restricted to HLA-DR1. Side-by-side analysis of ELISpot responses with flow cytometry data determined any correlation between the techniques and hence indicated the extent to which multimer binding cells exhibited T_H1 functionality.

PBMC from two HLA-DR1 donors were expanded to each peptide; IFN- γ ELISpot was performed on day-12 followed by HLA-multimer staining of the same lines on day-14 (Fig. 5.1).

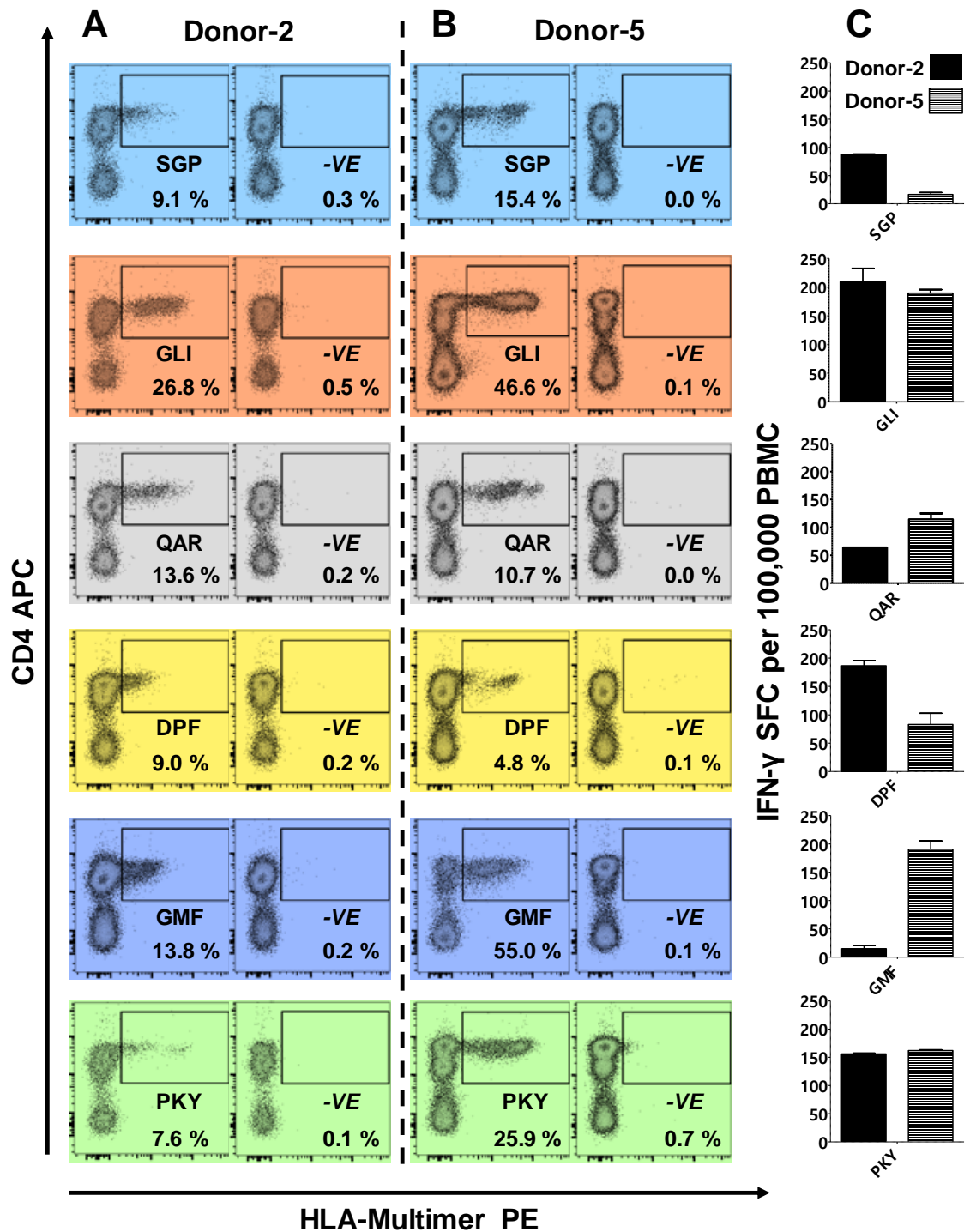


Figure 5.1. Comparison of HLA-Multimer staining and IFN- γ ELISpot responses in two HLA-DR1 donors. Donor numbers correspond to experiments detailed in the previous chapter. HLA-multimer stains are shown alongside irrelevant HLA Class-II multimer negative controls for donor-2 (A) and donor-5 (B) with % of CD4+ T-cells shown for each gate. Data for each epitope is shown as a colour-coded row. (C) IFN- γ ELISpot data for each donor and epitope is displayed as SFC per 100,000 PBMC with

background (negative control) subtracted, donor-2 in black, donor-5 in hatched bars (mean with SD error bars, n = 2). FMO controls and gating strategy in appendix section 8.2.

HLA-multimer+ populations were observed in response to each epitope, with stronger shifts and greater numbers in donor-5 (Fig. 5.1B). Donor-2 had slightly lower shifts (in terms of MFI) but as fewer cells were obtained for analysis, the staining was weaker (Fig. 5.1A). When compared to IFN- γ ELISpot results (Fig. 5.1C), the epitopes that showed consistent staining and > 100 SFC in both donors were GLI and PKY, while QAR and DPF each showed consistent staining and > 50 SFC (a positive response is > 20 SFC per 100,000 starting PBMC).

The epitopes SGP and GMF showed donor-specific inconsistencies between the numbers of cells stained on flow cytometry and the corresponding IFN- γ responses. It is probable that this results from a discrepancy in the ELISpot immunoassay, or limited activity in the cell line. HLA-multimer staining may offer a more reliable quantification of epitope-specific responses, especially when quantified with respect to parent populations.

An attempt to correlate both sets of data was carried out using regression analysis (Fig. 5.2). This yielded limited success, all correlations were positive, yet only one (donor-5 IFN- γ SFC vs %CD4+) gave a significant R-squared value greater than 0.70 ($p = 0.006$). This may be the result of ELISpot variation, with more advanced immunoassay techniques such as intracellular cytokine staining showing greater correlations. Ultimately, these epitope-specific populations do show functionality and are worth investigating further.

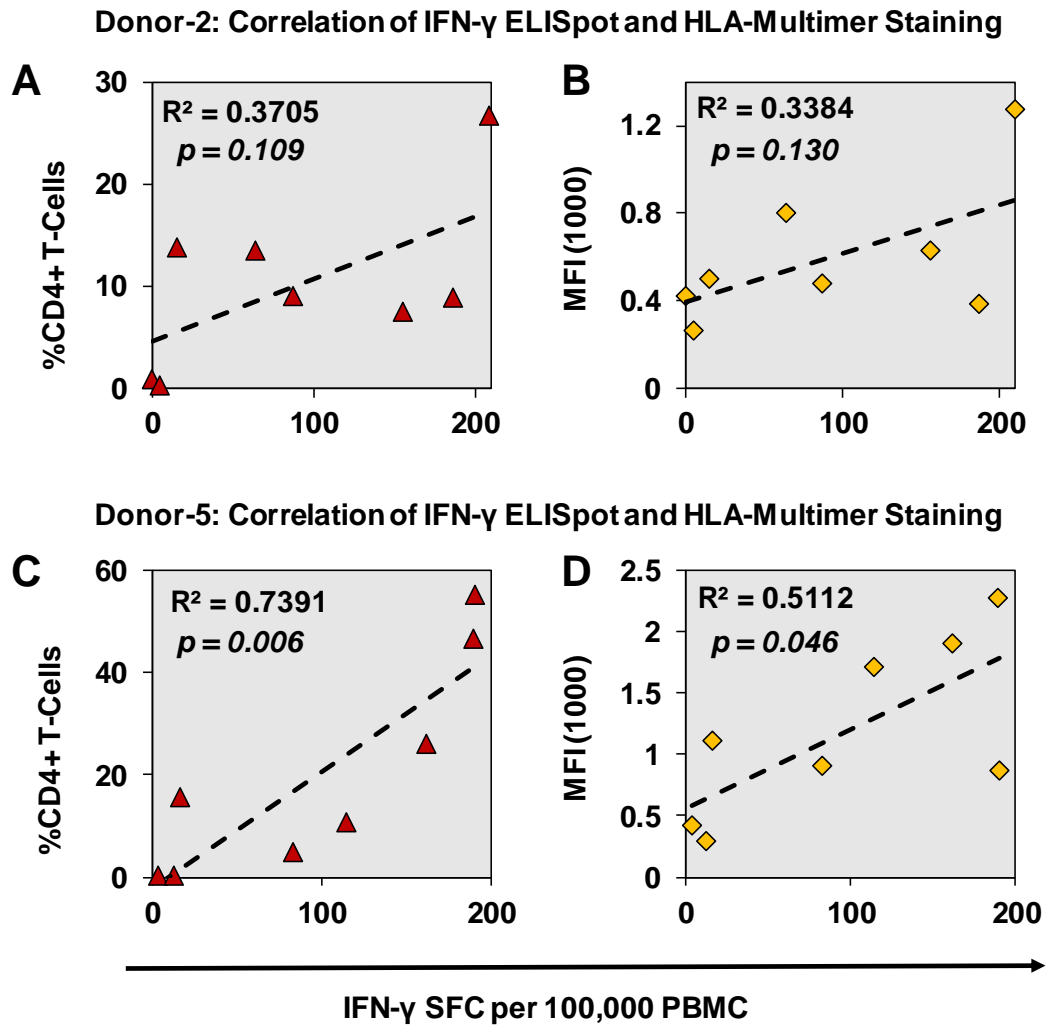


Figure 5.2. Regression analysis of flow cytometry statistics (%CD4+ and MFI) with IFN- γ ELISpot data (SFC). Analysis is shown for each donor separately, with linear trend lines fitted and R-squared and p values shown on each plot. Donor-2 IFN- γ versus %CD4+ (A) and MFI (B) Donor-5 IFN- γ versus %CD4+ (C) and MFI (D).

5.4.1.2 Analysis of Epitope-Specific Populations in Five DR1 Donors

Given that two donors exhibited comparable responses to each epitope tested, additional multimer staining were performed in three more HLA-DR1 donors. Four of these donors had been used to confirm the immunogenicity of the epitopes characterised in the previous chapter, while the responses in another were examined for the first time.

The first question driving this investigation was whether all HLA-DR1 donors tested had epitope-specific T-cells that could be expanded and detected from one million starting PBMC? Following this, were the T-cell response magnitudes to each epitope comparable between donors, or highly donor specific?

Staining was performed in each donor at day 12-14 following expansion (Fig. 5.3A). Multimer analysis gates were set using FMO and an irrelevant HLA class-II multimer (Fig. 5.3B), the analysis was repeated twice using PBMC samples taken at different time points in donor-3 and donor-5 to investigate reproducibility (data not shown). Epitope-specific responses were detectable in all donors, with the exception of GMF in donor-1 and DPF in donor-4.

Donor-4 appeared to have preferential expansion of CD19+ cells after 14 days in culture and as a result had fewer numbers of CD3+ cells, evident visually when comparing stains. Inspection of the plots show variation in magnitude and the fluorescence shifts of epitope-specific populations. GLI, PKY and QAR show larger populations and more consistency between donors than SGP, DPF and GMF.

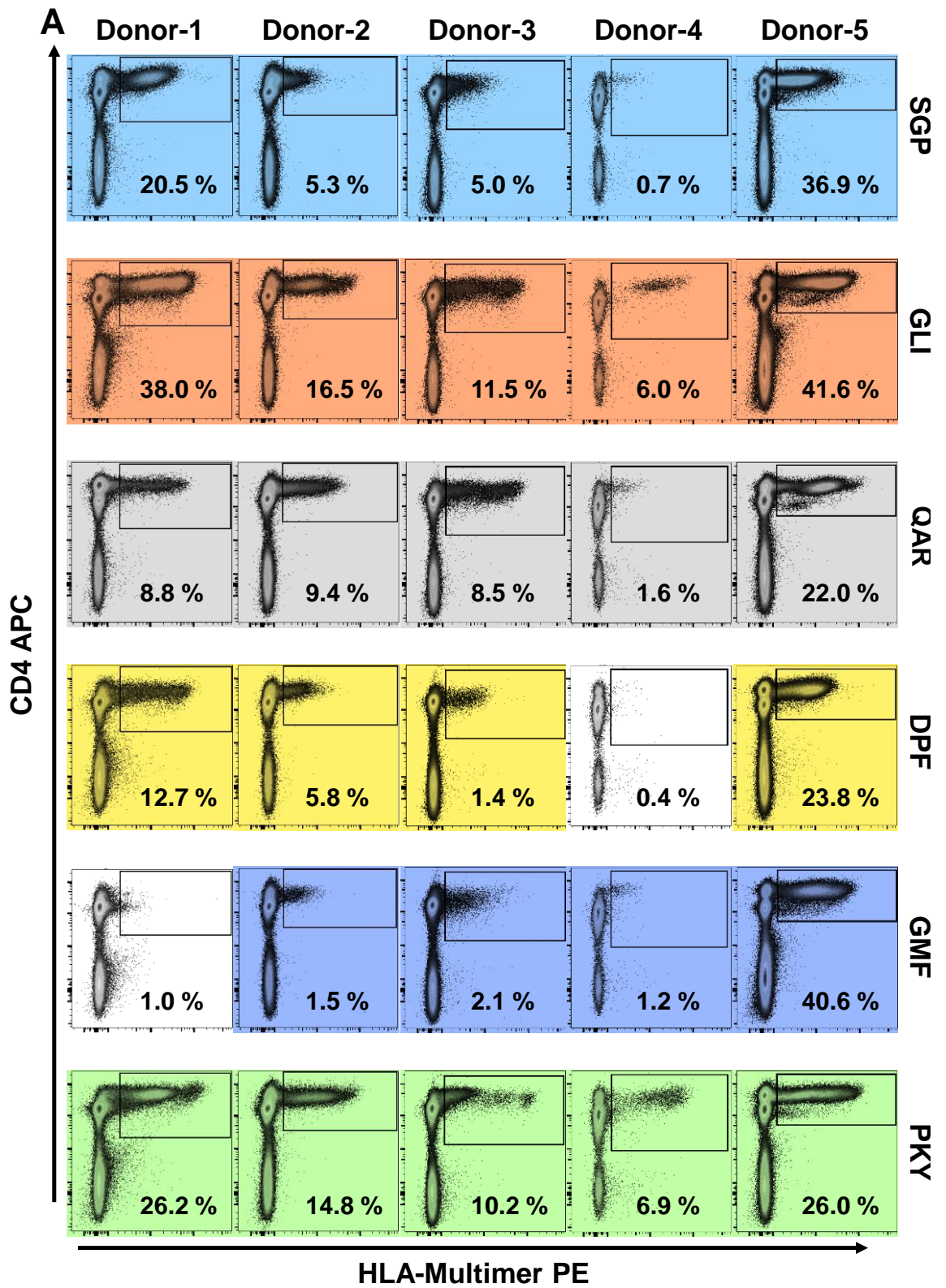


Figure 5.3A. Comparison of epitope-specific stains in 5 HLA-DR1 donors.

Continued on next page.

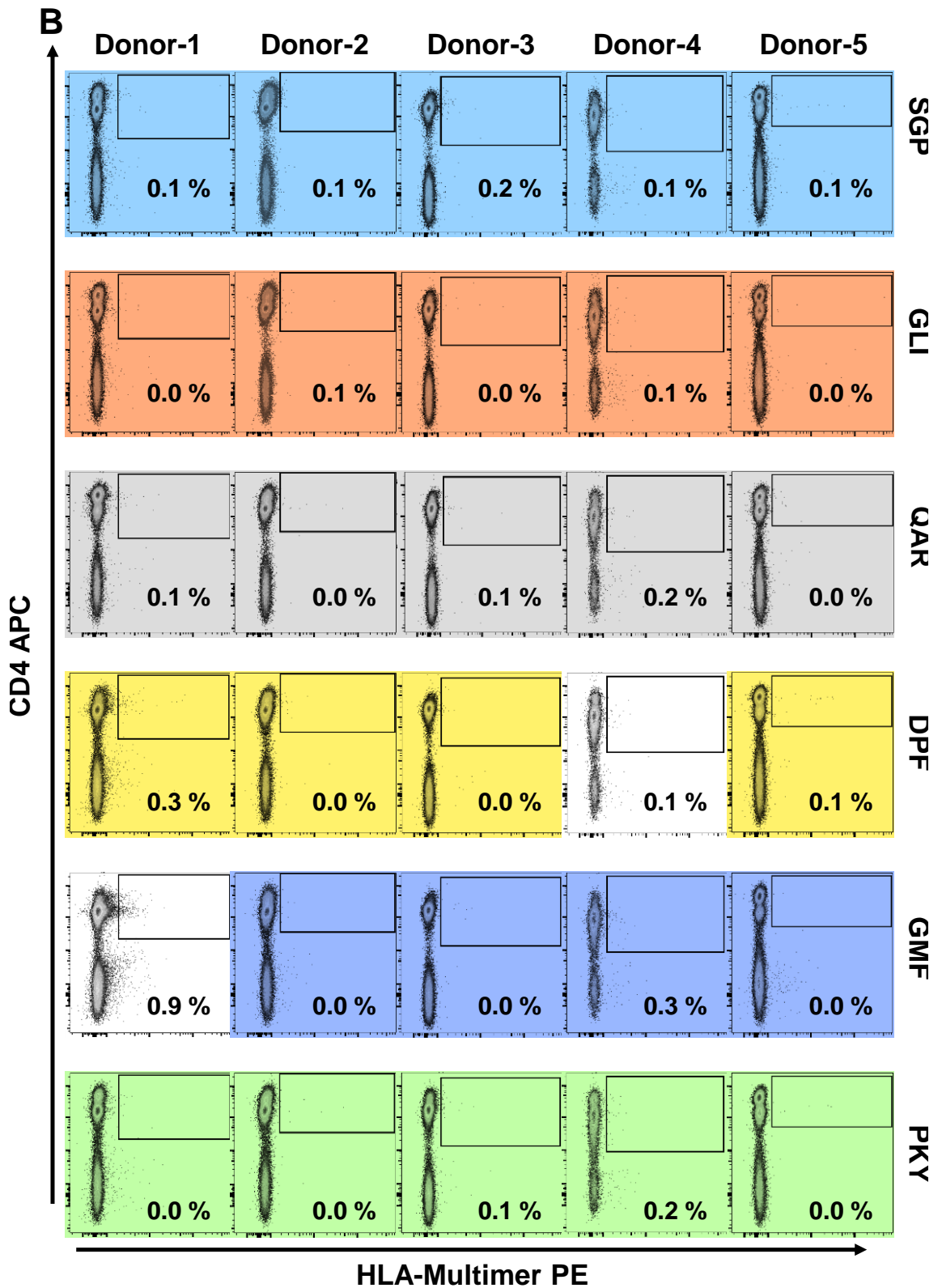


Figure 5.3. Comparison of epitope-specific stains in 5 HLA-DR1 donors.

Epitope-specific data (A) and corresponding lines stained with irrelevant HLA class-II multimers (B) are presented in colour-coded rows, and donor specific data in columns.

Gate % of parent CD4+ T-cells are detailed on each plot. Where stains are coloured

white (donor-4, DPF and donor-1, GMF) this represents insignificantly stained populations when compared to the irrelevant control and FMO. Gates were set by irrelevant and FMO controls.

When normalised data (Fig. 5.4A, 5.4B) and mean values were compared (Fig 5.4C, 5.4D), the matrix epitope GLI showed the highest %CD4+ and MFI values across five donors. The control HA-PKY epitope exhibited a similar magnitude to GLI but showed less consistency between donors. This suggests that GLI and PKY consistently elicit the strongest responses with respect to other HLA-DR1 epitopes, potentially a result of increased precursor frequency or an increased avidity of the cognate TCR population (see discussion section 5.5).

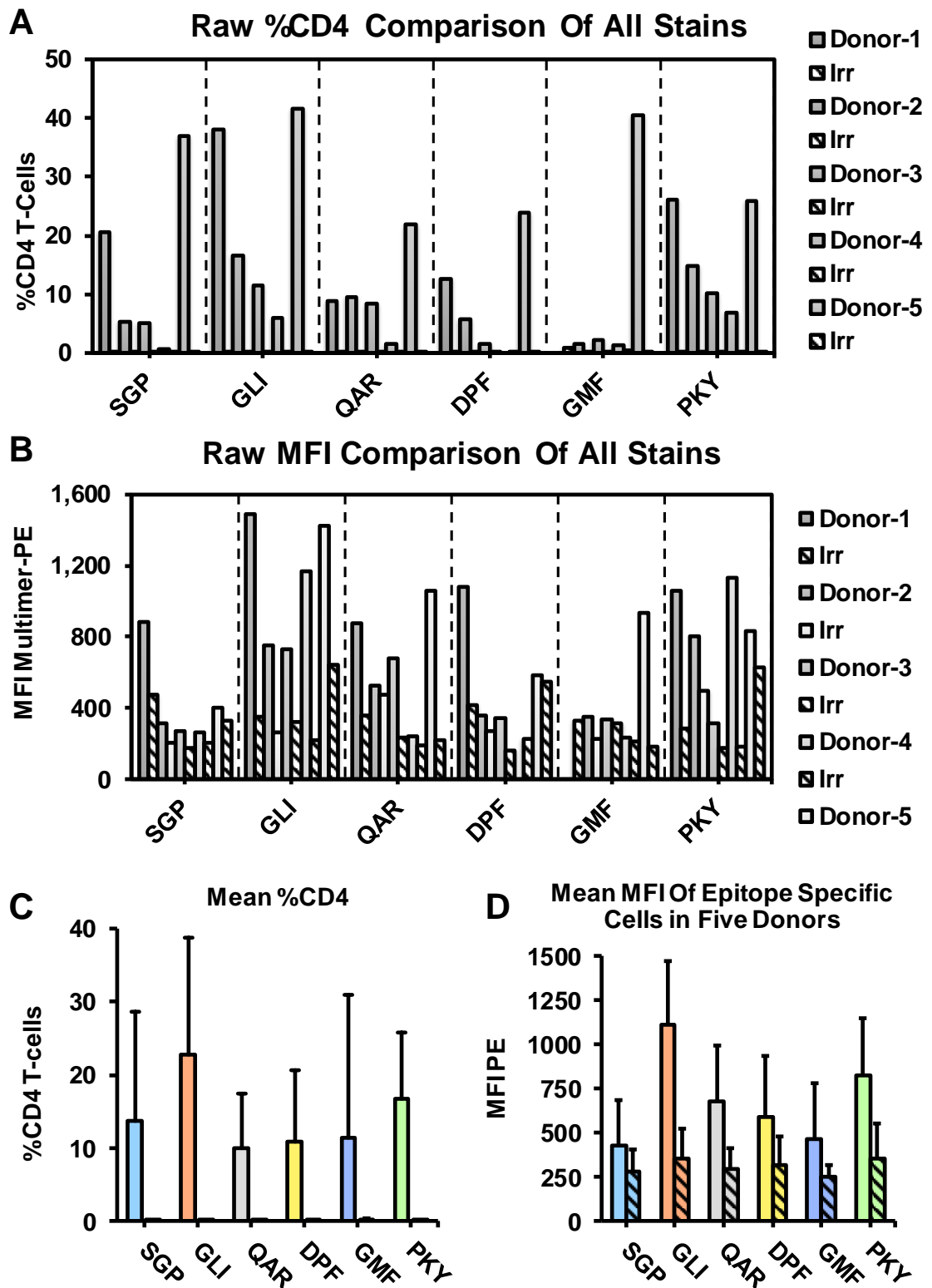


Figure 5.4. Comparison of epitope-specific HLA-multimer staining in five HLA-DR1 donors. Negative controls, irrelevant HLA class-II multimer staining show in hatched bars on all plots. **(A)** %CD4+ T-cells which comprise HLA-multimer+ CD4+ T-cells. **(B)** Raw MFI of HLA-multimer+ CD4+ T-cells per donor and epitope. **(C)** Mean %CD4+ value

across five donors per epitope (n = 5, SD error bars). (D) Mean MFI value across all donors per epitope (n = 5, SD error bars).

5.4.2 TCR α -Chain Analysis of Epitope-Specific CD4+ T-cells

In order to understand the genetic basis of these CD4+ T-cell responses, clonotypic analysis of epitope-specific populations was carried out. Where possible, HLA-multimer positive CD4+ T-cells from the stains in the previous section (Fig. 5.4) were sorted by FACS. Messenger RNA was extracted and used for cDNA synthesis, which was amplified by two rounds of PCR. The PCR product was analysed using next generation sequencing, yielding detailed information about the TCR α -chain usage for each epitope.

Samples were obtained for GLI and PKY epitopes in all five donors, while QAR, DPF and SGP samples were obtained for all except donor-4. Samples for GMF were obtained for donor-2, 3 and 5.

A threshold of 50 reads per sequence was set; any clonotypes with fewer than 50 reads were eliminated. This value was based on the presence of low-frequency TCR sequences in multiple samples from the same donor, i.e. they did not show epitope specificity. The biological nature of these TCRs, cross reactive or cellular contamination, are yet to be determined. A small number of such sequences were present below 50 reads, hence the threshold value (see example in appendix section 8.3).

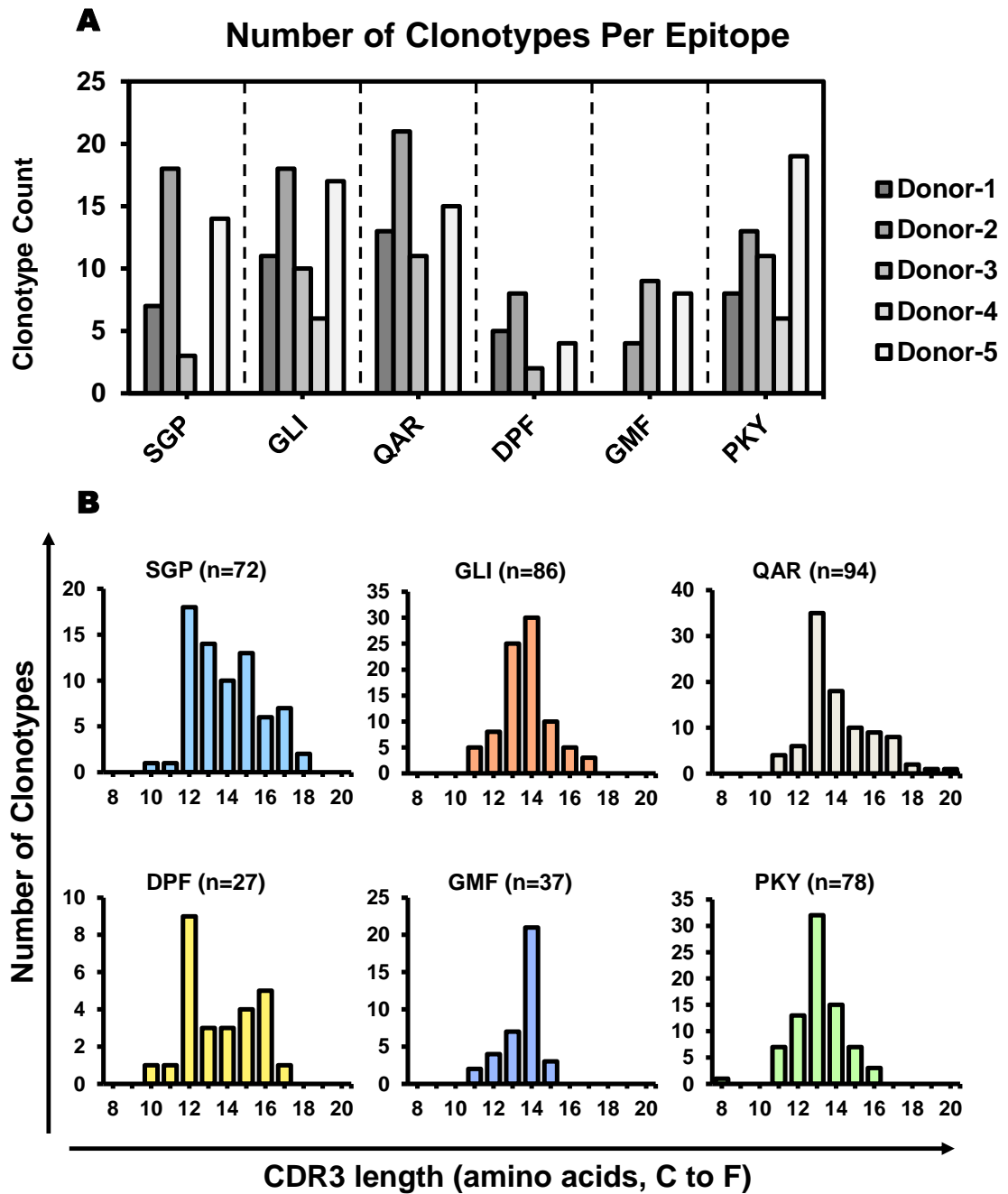


Figure 5.5. Summary of TCR α -chain data in five donors. (A) Number of clonotypes per epitope from each donor. The results of one sample per donor are shown. (B) Distribution of CDR3 TCR α -chain lengths (number of amino acids) across all samples in all donors, specific to each epitope.

Between two and twenty-one distinct α -chain clonotypes were detected in each sample (Fig. 5.5A). Fewer clonotypes comprised the response to DPF and GMF, while other epitopes showed a large range depending on the donor.

CDR3 α lengths in response to each epitope from eight to twenty amino acids when measured from the primary cysteine to the terminal phenylalanine. The majority of samples centred around 12-16 amino acids, with epitope-specific patterns emerging (Fig. 5.5B).

GLI and PKY showed resemblance to a normal distribution, with modal values at 14 and 13 amino acids respectively. Other epitopes showed slightly skewed plots with QAR favouring 13 amino acids or longer, and GMF peaking at 14 amino acids with a skew towards shorter sequences. SGP and DPF had bimodal distributions; both had a mode of 12, followed by lower use of 13-17 amino acid lengths.

These length distributions may reflect the optimal number of CDR3 amino acids necessary for good contacts with cognate pHLA surfaces. However, it is one of many factors contributing to the receptor ligand interface; others such as the germline encoded CDR1 and CDR2 sequences may exert a much greater and more obvious impact.

5.4.3 Epitope-Specific TRAV Gene Usage

When investigating epitope-specific responses, analysis of V gene usage in multiple individuals can identify genetic bias that is shared across the population^{249,279}. Recent human studies of CD4+ T-cells in celiac disease^{271,284} and HIV²⁸⁶ have found repertoires exhibiting highly restricted use of certain TRAV and TRBV genes, as well as TRAJ and TRBJ. These may represent important interactions between the germline-encoded CDR1 and CDR2 amino acids and the pHLA surface that facilitate binding¹⁶⁷.

In order to analyse the responses, TRAV gene usage was normalised for each donor (as a fraction of the total number of epitope-specific clonotypes) and cumulative results presented for each epitope (Fig. 5.6).

Two epitopes, QAR (number of clonotypes, $n = 60$, 4 donors) and GMF ($n = 21$, 3 donors), each exhibited highly biased use of a single gene (Fig. 5.6A, 5.6B). QAR responses using TRAV38-2/DV8 had an average frequency of 61 % in four donors (range 54 -73 %). GMF responses using TRAV2 averaged 72 % in three donors (range 67-75 %), representing the highest gene usage seen across all epitopes. The average frequencies were greater than 50 %, but much less than some examples in CD8 T-cell responses where average frequencies greater than 90 % have been observed for specific V genes in response to influenza epitopes²⁵⁰. Comparative data for other CD4+ responses is not available, having been carried out on small populations of clones²⁷¹ or using qPCR based sequencing methodologies²⁸⁶.

Usage to the remaining epitopes showed less bias and lower average frequencies in response to highlighted genes. SGP ($n = 42$, 4 donors) showed a broader use of three TRAV genes (Fig. 5.6C). TRAV23DV6 was the top hit in three of four donors with an average of 36 % (range 0-57 %), there was limited use of TRAV13-2 (average 26 %, range 6-43 %) and TRAV5 (average 15 %, range 0-33 %).

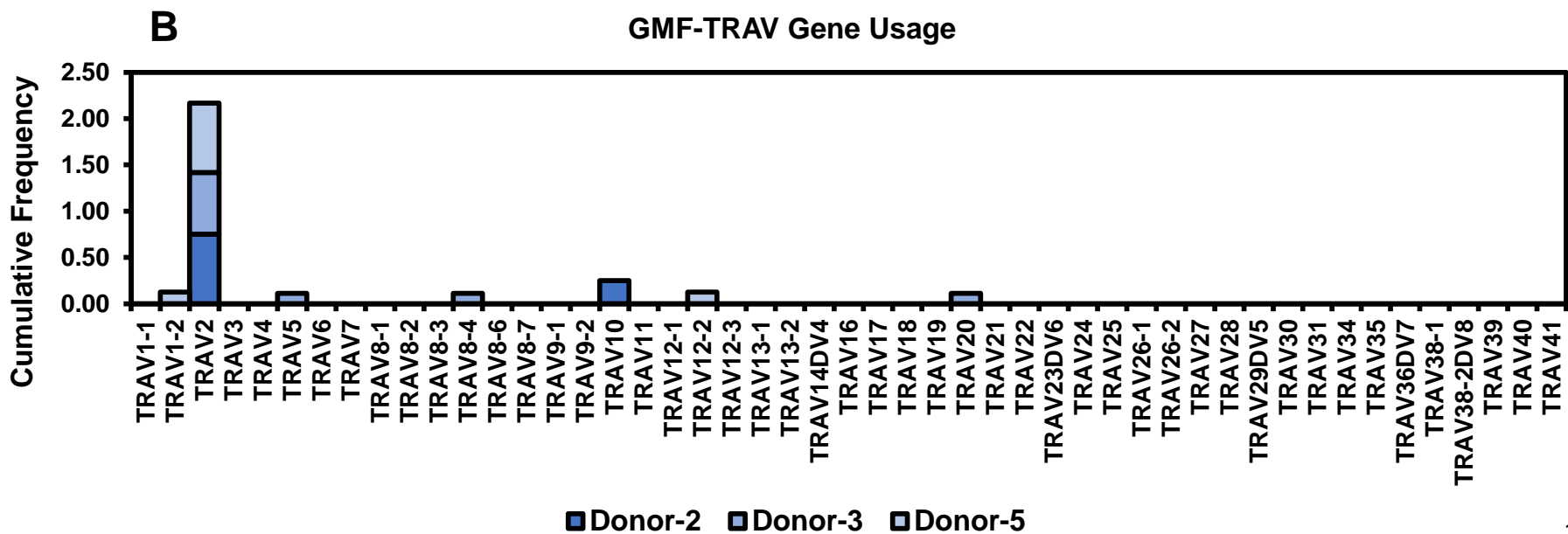
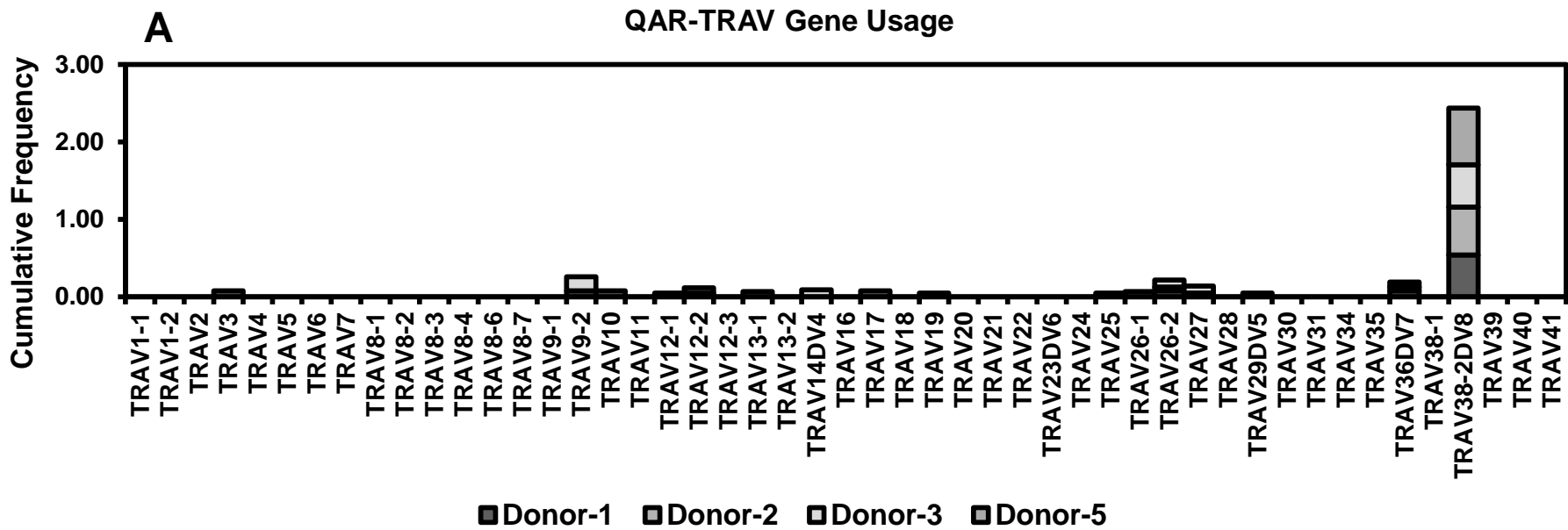
Analysis of DPF ($n=19$, 4 donors) was skewed by low numbers of clonotypes (Fig. 6D), and showed no apparent usage pattern common to more than two donors.

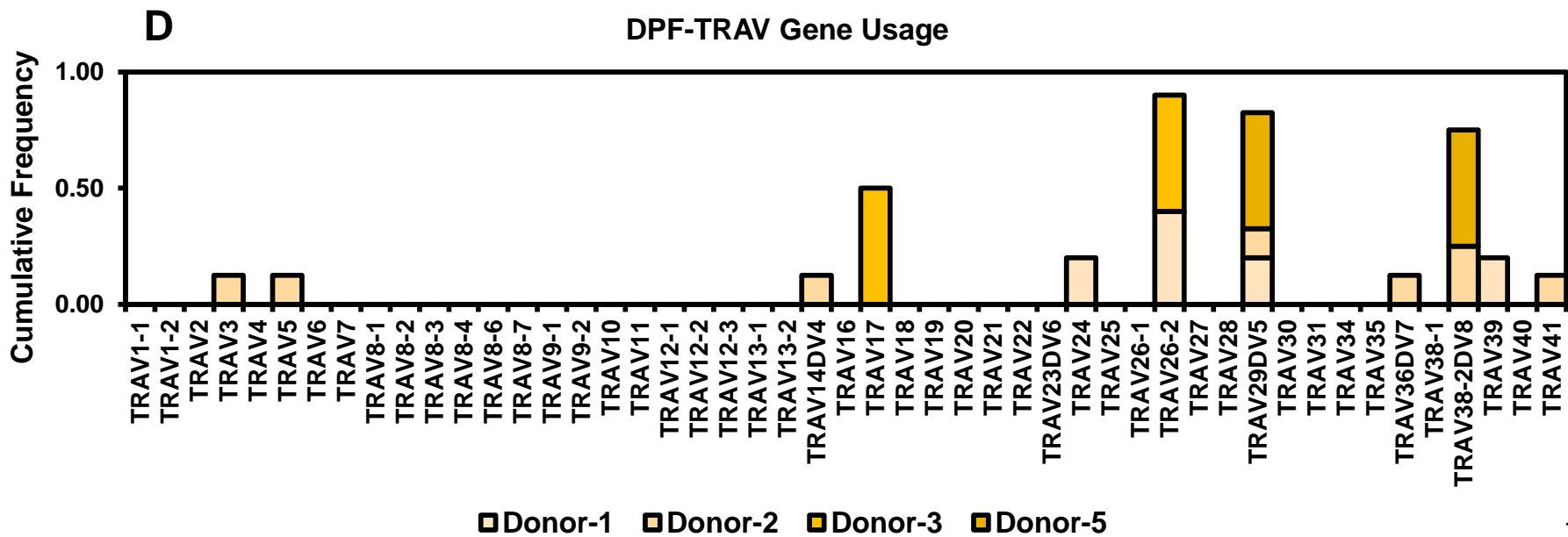
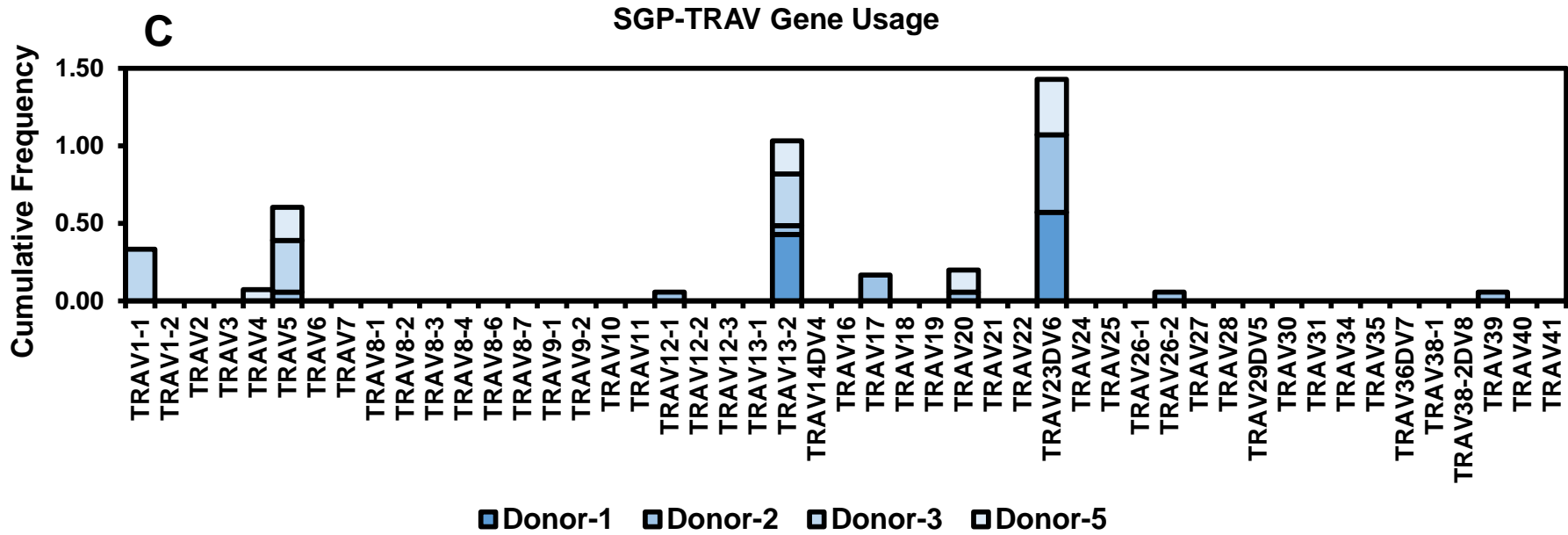
The final two epitopes GLI ($n = 62$) and PKY ($n = 57$), generated the strongest responses across all donors (section 5.4.1). Analysis of TRAV usage for both epitopes

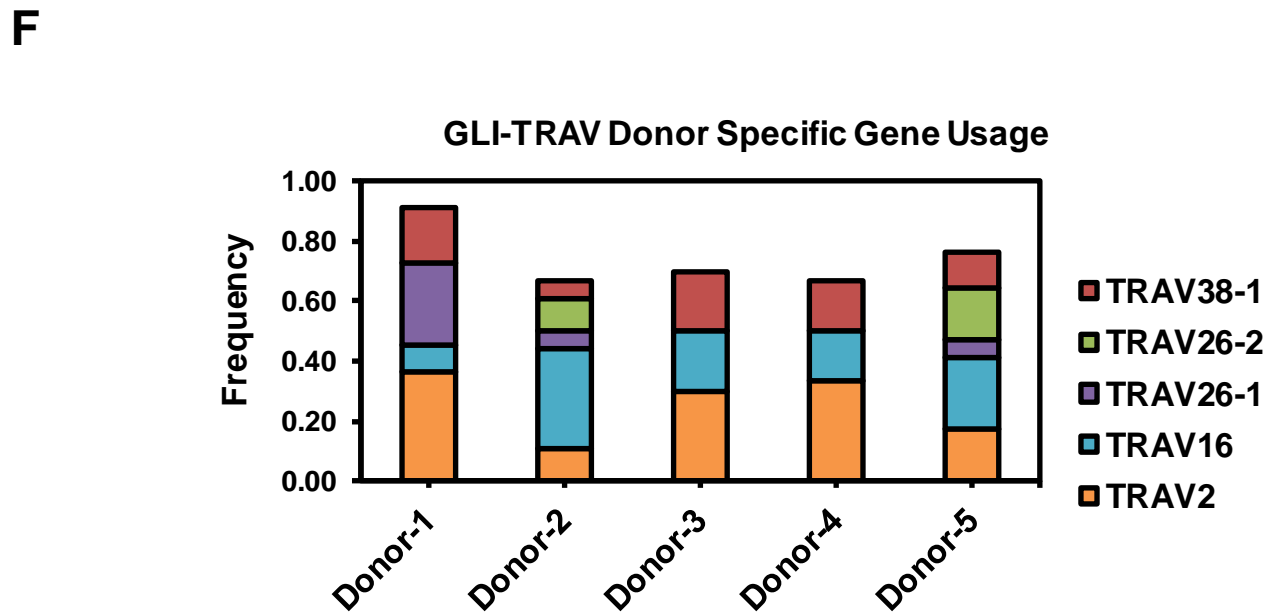
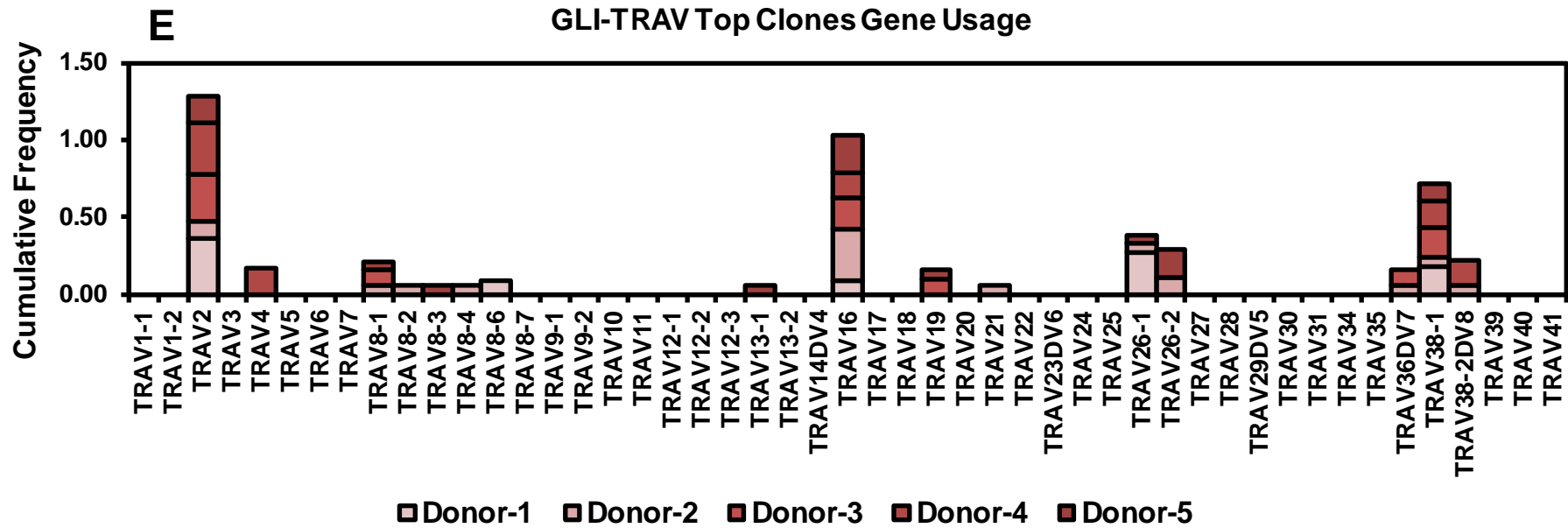
in five donors showed the broadest gene usage of all epitopes tested. GLI TRAV usage (Fig. 5.6E, 5.6F) was focused in three genes, TRAV2 (average 26 %, range 11-36 %), TRAV16 (average 21 %, range 9-33 %) and TRAV38-1 (average 14 %, range 6-20 %), with usage specific to two donors displayed in six additional genes.

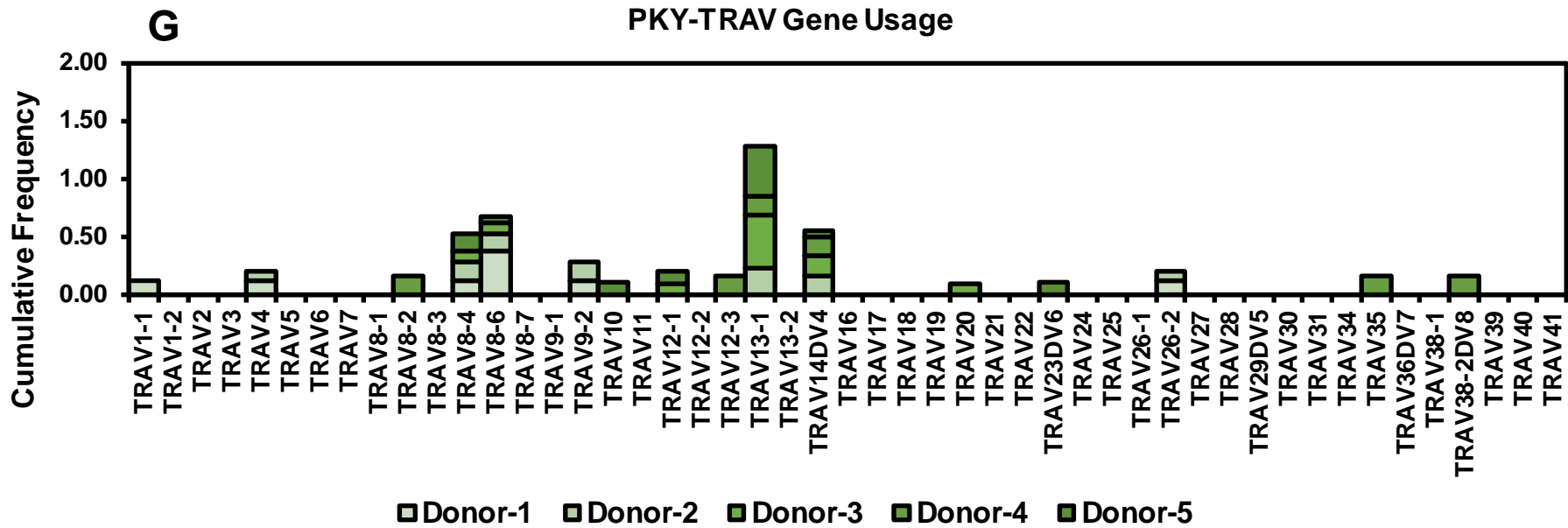
In response to PKY (Fig. 6G, 6H), TRAV8-4 (average 11 %, range 0- 16 %), TRAV8-6 (average 13 %, range 0-38 %), TRAV13-1 (average 25 %, range 0-45 %) and TRAV14DV4 (average 11 %, range 0-18 %) were each used by four of five donors tested.

These relatively broader usage patterns may suggest GLI and PKY are less limited in the selection of TRAV genes capable of forming strong structural interactions through their germline regions. Although speculative, it could explain the stronger responses observed, as there would be a larger theoretical pool of CD4+ T-cells with the potential to respond to these ligands.









H

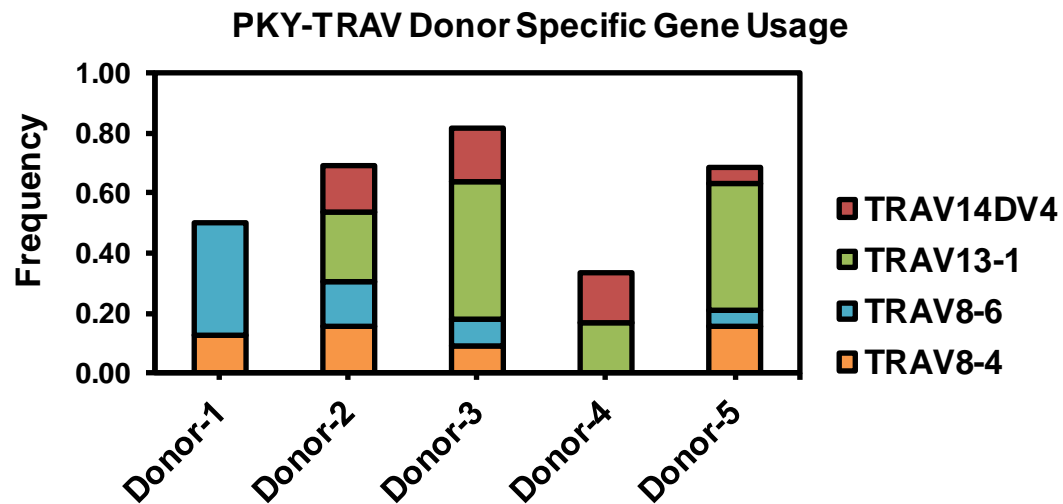


Figure 5.6. Analysis of epitope-specific TRAV gene usage in multiple donors.

Donor specific frequencies of each gene with respect to the total number of clonotypes per epitope were calculated. Frequencies for each gene were plotted as a cumulative analysis of the responses from all donors to each epitope: QAR (A), GMF (B), SGP (C), DPF (D), GLI (E), PKY (G)

Breakdown of usage patterns by donor are shown for the epitopes GLI (F) and PKY (H) as a cumulative analysis of dominant TRAV genes.

5.4.4 TCR α CDR3 Diversity

Analysis of an epitope-specific CD4⁺ T-cell response at the level of CDR3 sequences reveals the clonotypic diversity of a sample. In this analysis, epitope-specific TCR α -chain CDR3 amino acid sequences were presented for each donor (Fig. 5.7). The Shannon index (H') was calculated for every sample, with lower H' values corresponding to lower biological diversity. The mean H' value and standard deviation across all donors for each epitope was calculated.

QAR exhibited the highest levels of diversity across four donors, with a mean H' of 1.77 (SD 0.52) and pie charts containing multiple small slices in each donor. Donor-3 had a highly dominant TCR at greater than 75 % frequency, while in all other donors the dominant sequence occupied between 25-50 % of the total. Based on this data, and section 5.4.3 (Fig. 5.6A), the response to QAR was likely mediated by a large pool of TRAV38-2 clonotypes with the capacity to compete for the same pHLA.

PKY showed similar levels of diversity across five donors (mean 1.65, SD 0.37), although a greater number of large slices (> 20 % of the total, less than 50 %) and fewer small slices were observed. The prevalence of large slices may indicate that two to five dominant, high avidity, clones are competing for PKY in each donor; this may account for the large MFI values seen in the corresponding flow cytometry data (Fig. 5.3).

In contrast, GLI showed responses dominated by single clones in five donors (greater than 50 %) with multiple smaller slices in the remaining 50 % that contribute to relatively diverse responses (mean 1.41, SD 0.35). The mean H' value for SGP was similar, but the variation between donors was the highest of all epitopes (SD 0.76), with little consistency in H' values or pie chart architecture. This suggests the nature of the response to SGP was specific to each donor, with no similar patterns observed across the population (at this same size).

GMF had a lower mean H' value, highly consistent across three donors (SD 0.18), with three to five clonotypes in equilibrium (slices > 10 %) across each sample, and few small slices.

DPF exhibited the least response diversity of all epitopes in four donors (mean 0.72). The pie charts were dominated by large slices, in some cases greater than 75 %, which indicate fewer clonotypes contributing to each response (Fig. 5.3). The CDR3 structural requirements necessary to bind HLA-DR1 presented DPF at sufficient avidity may be very narrow and therefore it contrasts strongly with the highly diverse responses seen in other epitope samples.

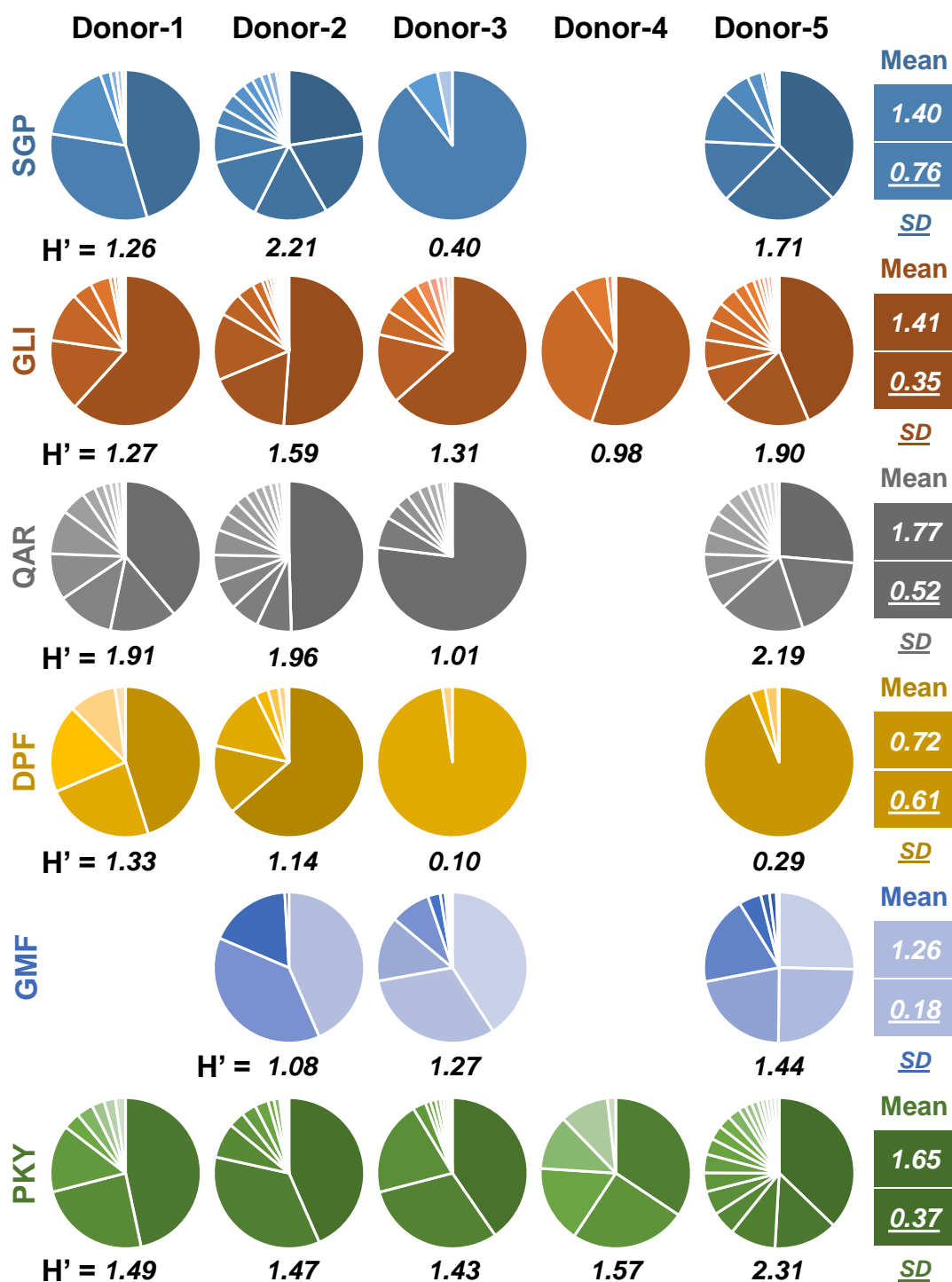


Figure 5.7. Clonotypic diversity in response to each epitope across all donor samples. Pie charts were compiled using TCR sequence frequency in each donor, with the corresponding Shannon diversity index calculated from the same data shown below each chart. Mean H' values and standard deviation are calculated for each epitope.

5.4.5 Public CDR3 Amino Acid Sequences

Given the strength of CD4+ T-cell responses and focused gene usage in response to the characterised epitopes, it was of interest to search for public TCR sequences that were shared in multiple donors. Few public sequences (by CDR3 amino acids) have been characterised in response to HLA-Class II epitopes^{286,287}, and none in response to the conserved internal proteins of influenza.

Analysis of TCR α -chain data across five donors showed fifteen shared sequences (Table 5.1). Thirteen of which were shared between two donors in response to the same epitope, while two, (PKY-CAASFSDGQKLLF and QAR-CAYLTGTASKLTF) were shared between three donors. Several sequences were identical the amino acid level, but differed at the nucleotide level (red letters in table 5.1), an indicator of convergent recombination in the generation of publicity^{276,277}.

These public sequences were rare, but highly important, as they represent TCR mediated immunity that is conserved between individuals to the codon level. Preferential expansion of these clonotypes during infection is likely to have resulted in a precursor frequency that permits consistent detection in multiple HLA-DR1 individuals. These TCR sequences may be highly protective and therefore have been driven out under viral selection pressure across the population.

Donor	CDR3	Genetic Seq	V Gene	J gene
SGP	123456	1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6		
1	CAAS S RIYNQGGKLI F	TGTGCAGCAAGCT CCCGG ATTTATAACCAGGGAGGAAAGCTTATCTTC	TRAV23DV6	TRAJ23
2	CAAS S RIYNQGGKLI F	TGTGCAGCAAGC AGCCG CATTTATAACCAGGGAGGAAAGCTTATCTTC	TRAV23DV6	TRAJ23
	1234	1 2 3 4 5 6 7 8 9 0 1 2 3		
1	CA A TRRGADGLTF	TGTGCAG CAACC AGGAGAGGTGCTGACGGACTCACCTTT	TRAV23DV6	TRAJ45
2	CA A TRRGADGLTF	TGTGCAG CCACA AGGAGAGGTGCTGACGGACTCACCTTT	TRAV23DV6	TRAJ45
GLIYN	123456	1 2 3 4 5 6 7 8 9 0 1 2 3 4		
2	CAF M RYNAGNMLTF	TGTGCTTTCATG AGG TATAATGCAGGCAACATGCTCACCTTT	TRAV38-1	TRAJ39
3	CAF M RYNAGNMLTF	TGTGCTTTCATG AGA TATAATGCAGGCAACATGCTCACCTTT	TRAV38-1	TRAJ39
	123456	1 2 3 4 5 6 7 8 9 0 1 2 3 4		
1	CAL R EANTGNQFYF	TGTGCTCTA AGG GAGGCTAACACCCGGTAACCAGTTCATTTTT	TRAV16	TRAJ49
5	CAL R EANTGNQFYF	TGTGCTCTA AGA GAGCGCAACACCCGGTAACCAGTTCATTTTT	TRAV16	TRAJ49
QAR	1234	1 2 3 4 5 6 7 8 9 0 1 2 3		
1	CAY L TGTASKLTF	TGTGCTTAT ATT ACCGGCACTGCCAGTAAACTCACCTTT	TRAV38-2DV8	TRAJ44
2	CAY L TGTASKLTF	TGTGCTTAT ATA ACCGGCACTGCCAGTAAACTCACCTTT	TRAV38-2DV8	TRAJ44
	12345	1 2 3 4 5 6 7 8 9 0 1 2 3		
1	CAY L AGTAS K LTF	TGTGCTTATTTA G CCGGCACTGCCAGTAAACTCACCTTT	TRAV38-2DV8	TRAJ44
3	CAY L AGTAS K LTF	TGTGCTTATTTA GCA GGCACTGCCAGTAAACTCACCTTT	TRAV38-2DV8	TRAJ44
	1234	1 2 3 4 5 6 7 8 9 0 1 2 3		
2	CAY L TGTASKLTF	TGTGCT TATTT GACCGGCACTGCCAGTAAACTCACCTTT	TRAV38-2DV8	TRAJ44
	CAY L TGTASKLTF	TGTGCT TATCTT ACCGGCACTGCCAGTAAACTCACCTTT	TRAV38-2DV8	TRAJ44
	CAY L TGTASKLTF	TGTGCT TACTTA ACCGGCACTGCCAGTAAACTCACCTTT	TRAV38-2DV8	TRAJ44
	CAY L TGTASKLTF	TGTGCT TACCTT ACCGGCACTGCCAGTAAACTCACCTTT	TRAV38-2DV8	TRAJ44
3	CAY L TGTASKLTF	TGTGCT TATTTA ACCGGCACTGCCAGTAAACTCACCTTT	TRAV38-2DV8	TRAJ44
	CAY L TGTASKLTF	TGTGCT TATCTG ACCGGCACTGCCAGTAAACTCACCTTT	TRAV38-2DV8	TRAJ44
	CAY L TGTASKLTF	TGTGCT TACCTC ACCGGCACTGCCAGTAAACTCACCTTT	TRAV38-2DV8	TRAJ44
5.1	CAY L TGTASKLTF	TGTGCT TACCTA ACCGGCACTGCCAGTAAACTCACCTTT	TRAV38-2DV8	TRAJ44
DPF	12345	1 2 3 4 5 6 7 8 9 0 1 2		
2	CAY R STGNQFYF	TGTGCTTAT AGATCA ACCGGTAACCAGTTCATTTTT	TRAV38-2DV8	TRAJ49
5	CAY R STGNQFYF	TGTGCTTAT AGGAGC ACCGGTAACCAGTTCATTTTT	TRAV38-2DV8	TRAJ49
GMF	123456	1 2 3 4 5 6 7 8 9 0 1 2 3 4		
3	CAVE E GSSASKIIF	TGTGCTGTGGAG GAGGGA AGCAGTGCCTCCAAGATAATCTTT	TRAV2	TRAJ3
5	CAVE E GSSASKIIF	TGTGCTGTGGAG GAAGGC AGCAGTGCCTCCAAGATAATCTTT	TRAV2	TRAJ3
3	CAVEGDNAGNMLTF	TGTGCTGTGGAGGGGATAATGCAGGCAACATGCTCACCTTT	TRAV2	TRAJ39
5	CAVEGDNAGNMLTF	TGTGCTGTGGAGGGGATAATGCAGGCAACATGCTCACCTTT	TRAV2	TRAJ39
PKY				
2	CAASFSDGQKLLF	TGTGCAGCAAGTTTTTCAGATGGCCAGAAGCTGCTCTTT	TRAV13-1	TRAJ16
3	CAASFSDGQKLLF	TGTGCAGCAAGTTTTTCAGATGGCCAGAAGCTGCTCTTT	TRAV13-1	TRAJ16
5	CAASFSDGQKLLF	TGTGCAGCAAGTTTTTCAGATGGCCAGAAGCTGCTCTTT	TRAV13-1	TRAJ16
1	CALSNDYKLSF	TGTGCTCTGAGTAACGACTACAAGCTCAGCTTT	TRAV9-2	TRAJ20
2	CALSNDYKLSF	TGTGCTCTGAGTAACGACTACAAGCTCAGCTTT	TRAV9-2	TRAJ20
2	CAMSATDSWGKLQF	TGTGCAATGAGTGCAACTGACAGCTGGGGGAAATGTCAGTTT	TRAV14DV4	TRAJ24
3	CAMSATDSWGKLQF	TGTGCAATGAGTGCAACTGACAGCTGGGGGAAATGTCAGTTT	TRAV14DV4	TRAJ24
	12345	1 2 3 4 5 6 7 8 9 0 1 2 3 4		
2	CAMS P TDSWGKLQF	TGTGCAATGAGT CCA ACTGACAGCTGGGGGAAATGTCAGTTT	TRAV14DV4	TRAJ24
3	CAMS P TDSWGKLQF	TGTGCAATGAGT CCT ACTGACAGCTGGGGGAAATGTCAGTTT	TRAV14DV4	TRAJ24
	123	1 2 3 4 5 6 7 8 9 0 1 2 3		
3	CV V YTG T ASKLTF	TGTGT GTT TATACCGGCACTGCCAGTAAACTCACCTTT	TRAV12-1	TRAJ44
5	CV V YTG T ASKLTF	TGTGT GTA TATACCGGCACTGCCAGTAAACTCACCTTT	TRAV12-1	TRAJ44

Table 5.1. Shared TCR α CDR3 amino acid and nucleotide sequences in response to each epitope. Nucleotide differences are highlighted in red. Sequences identical at the genetic level are coloured blue. Epitope and donor number are listed in the far-left column, followed by CDR3 amino acid sequence, genetic sequence, then V and J gene (left to right).

5.4.6 Analysis of Amino Acid Motifs in Response to Each Epitope

Theoretically, certain common trends should exist in the physical characteristics of the CDR3 regions that mediate epitope-specific responses; for example, conservation of charged or polar residues in the centre of the CDR3 region that facilitate binding. Motif based analysis is a common tool used in studies on epitope-specific TCR populations.

In order to visualise the amino acid usage of CDR3 regions, sequence motif software was used (WebLOGO) and amino acids were coloured based on polarity, charge, aromaticity and conformation (proline). The clearest approach to visualising patterns of amino acid usage involved stratifying CDR3 sequences based on their parent TRAV genes and length (Fig. 5.8). In this way, key features remained visible and were not lost amidst the detail of multiple unrelated genes and length variations.

Inspection of motifs appear to show conserved physical properties in response to some epitopes. For example, in response to SGP (Fig. 5.8A) the majority of motifs favour a basic residue, five amino acids from the starting cysteine. This residue was followed by a predominance of polar uncharged residues close to the c-terminus.

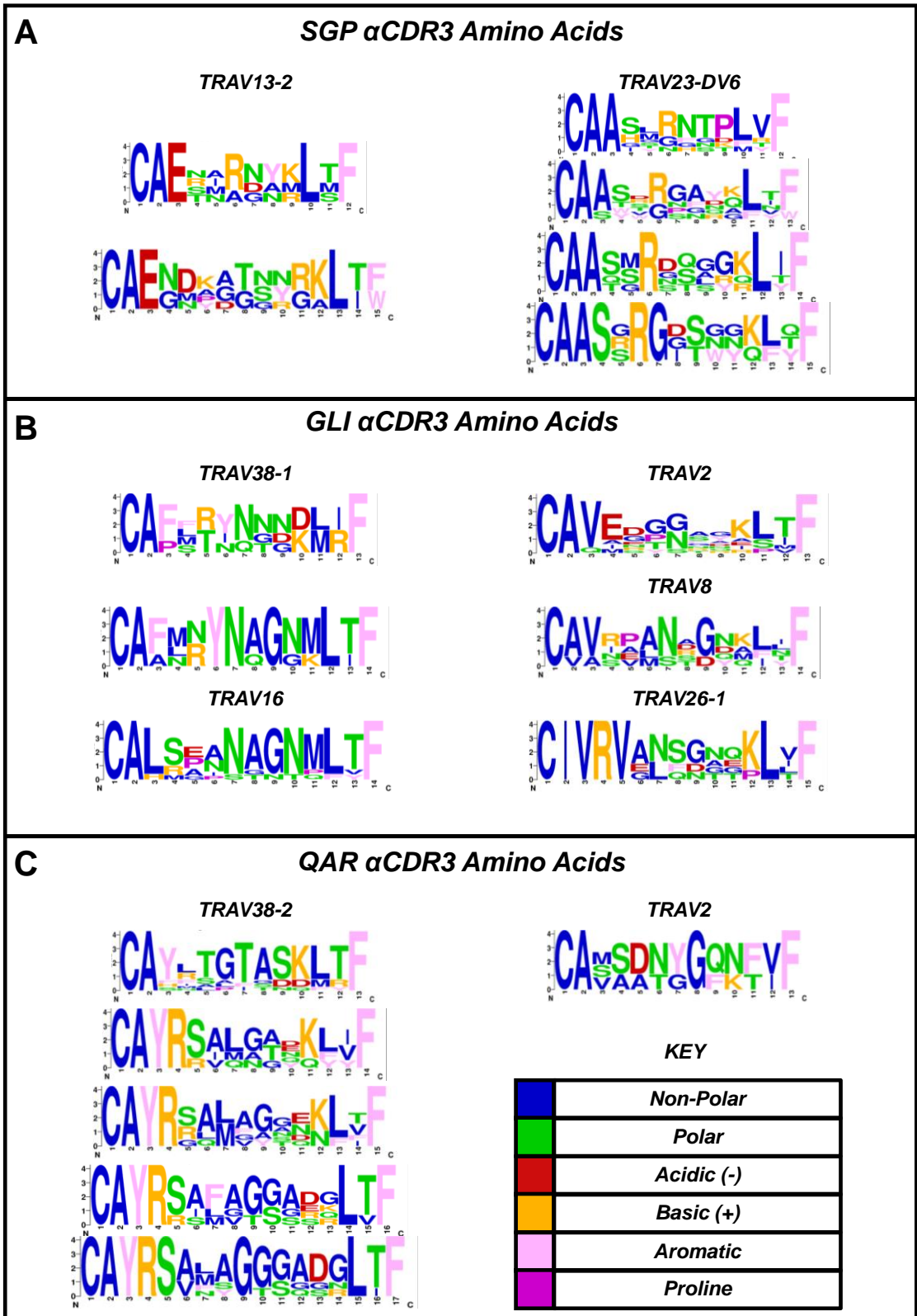
Motifs in response to GLI favoured two or more polar-uncharged residues positioned centrally and close to the c-terminus in four of the five TRAV genes analysed (Fig. 5.8B). No pattern was apparent at N-terminal residues.

When looking at TRAV38-2 motif responses to QAR (Fig. 5.8C), the N and C termini appeared highly conserved, while the central amino acids were charge neutral, predominantly non-polar with a low frequency use of polar residues. There was a complete absence of charged residues in the central region in all lengths, suggesting a highly hydrophobic interface between the TCR CDR3 α and the pHLA. The only other motif in response to QAR, TRAV2, showed an acidic residue at C-terminal position five, but the remaining positions conform to the uncharged pattern observed before.

The small number of TCRs detected in response to DPF (Fig. 5.8D) and GMF (Fig. 5.8E) meant that only one motif was created for each epitope, and it was hard to read into any significant usages. Four motifs were created for PKY (Fig. 5.8F), these

favour central acidic residues spread across positions 5-9 in each 11-mer, 12-mer and 13-mer CDR3 length motifs.

The patterns within observed motifs in response to SGP, GLI, QAR and PKY were striking and appeared highly unique to each epitope. Identification of patterns and physical characteristics at specific CDR3 positions form a base for modelling interactions between the pHLA and its cognate repertoire. If congruent physical characteristics can be found in the relevant HLA class-II crystals structures, then explanations of gene usage bias and amino acid motifs can be explored.



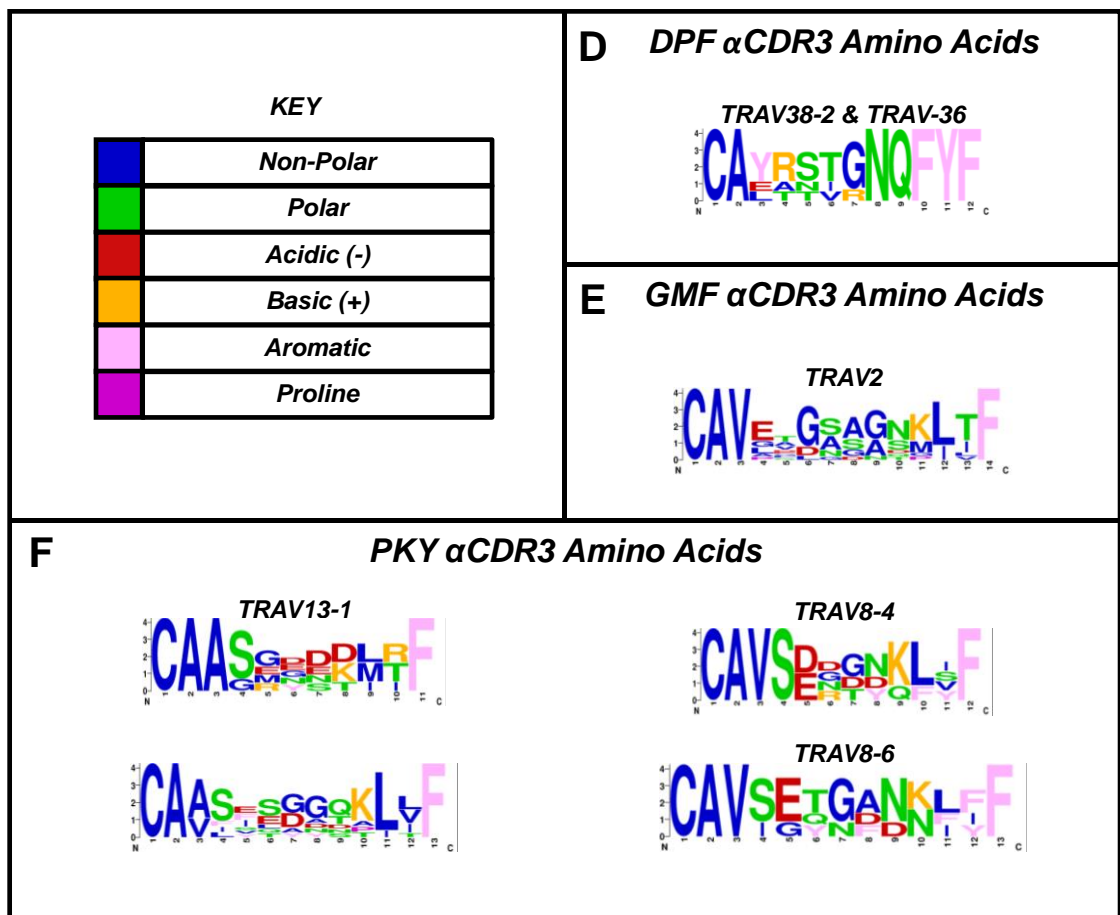


Figure 5.8. Amino acid motif analysis of α chain CDR3 regions in all sequences obtained in response to each epitope. A minimum of three TCR sequences are represented in each plot. Representations are proportional in terms of length of CDR3, which is detailed on the x-axis of each plot. Y-axis is the size proportion (bits) in height. The specific TRAV gene is detailed above each set of plots. (A) SGP (B) GLI (C) QAR (D) DPF (E) GMF and (F) PKY.

5.5 Discussion

In this chapter, the CD4⁺ T-cell responses to five conserved internal influenza epitopes and one external haemagglutinin epitope were compared by flow cytometry across five donors. Two epitopes, M1-GLI and HA-PKY, exhibited consistent and strong responses in each donor, while the remaining epitopes showed variability and donor specific patterns. Where possible, the TCR- α clonotypic repertoires were dissected to investigate the gene usage and sequences that mediated responses to each epitope.

Highly focused TRAV gene usage was observed in responses to M1-QAR and PB1-GMF, with other epitopes eliciting broader usage across multiple donors. TCR sequence diversity in response to each epitope varied, with highest diversity seen in response to M1-QAR and the lowest to NP-DPF. Several of the epitope-specific α -chains exhibited distinct CDR3 motifs involving conserved use of charged, polar or hydrophobic amino acids at single or multiple positions across the loop. Several public CDR3 sequences, identical at the amino acid and/or nucleotide level, were shared in up to three of the five donors.

The observations detailed here raise some important topics of discussion on the cellular and genetic features that underlie CD4⁺ T-cell mediated immunity.

5.5.1 Response Hierarchies and Protective Capacity

The epitopes HA-PKY and M1-GLI were detailed as eliciting the best responses relative to other epitopes based on the %CD4⁺ and MFI values (Fig. 5.4). From the six epitopes tested, these two exhibited the largest and most consistent response patterns in five donors. This confirmed earlier IFN- γ ELISpot responses (section 5.4.1 and chapter 4) and suggested these peptides were stimulating a greater number of precursors with high avidity TCRs and strong proliferative profiles. Yet the question remains as to whether the hierarchy we observe in the *in vitro* setting, especially when using HLA multimer reagents, relates to the magnitude of *in vivo* responses and protective capacity.

Methodological explanations for these strong responses may involve the half-life of synthetic peptides in culture, or the corresponding quality of the HLA-multimer reagent

used. The persistence of peptide in culture, increased stability of pHLA complexes and resulting pHLA multimer reagents may combine to improve expansion, avidity and flow cytometry staining, thus creating a response hierarchy less representative of the true *in vivo* response characteristics and effectiveness of protection.

Work in mice with HLA class-I epitopes and LCMV has demonstrated that the relationship between protective capacity of an epitope and its cognate T-cell population is complex^{257,288}. Epitope-mediated protection did not correlate with the number of functionally active cognate T-cells, but was instead related to the minimum amount of antigenic peptide necessary for recognition i.e. the sensitivity.

Therefore, peptide titrations and *ex vivo* analysis may complement the response hierarchies observed here to provide a more accurate picture of the protective capacity of epitope-specific populations.

5.5.2 Explaining Public/Shared Clonotypic Features

Three notable findings in this chapter concern the shared properties of TCR- α chains that facilitate recognition of epitopes between multiple donors. Specifically, highly focused TRAV gene usage, the presence of public TCR sequences and the conservation of certain amino acids across CDR3 motifs. The explanations of all three phenomena are intrinsically related and likely to have an underlying structural basis. This will be explored in detail in the following chapter using TCR-pHLA complex data, but will be discussed briefly here in relation to the observed findings.

5.5.2.1 Focused TRAV Gene Usage

TRAV gene usage is determined by both the germline-encoded amino acids that contact the pHLA (CDR1 and CDR2) and the CDR3 sequences capable of being generated. The CDR1 and CDR2 loops form the majority of HLA specific contacts and may place limitations on those TRAV genes that can recognise a specific allele⁵.

The α -chain CDR1 loop has been shown in multiple TCR-pHLA complex structures^{219,289} to form interactions with both the N-terminus of the peptide and the HLA class-II binding groove. Therefore, certain germline-encoded CDR1 residues may form

highly favourable peptide *and* HLA contacts that result in focused TRAV usage in response to an epitope. This may explain the striking patterns associated with QAR and GMF, where single TRAV genes dominated each response.

In response to other epitopes such as GLI, DPF and PKY where TRAV usage is broader, the structural requirements imposed by α chain CDR1 and CDR2 interactions may be less stringent. This could result from a greater conformational diversity in TCR docking orientation (crossing angle)⁵, or beta chain interactions which exert a greater energetic contribution to binding and permit flexible α -chain conformations.

5.5.2.2 Public CDR3 Sequences

Although CDR3 sequences are not germline encoded, N-terminal (e.g. TRAV9-2) and C-terminal (e.g. TRAJ20) germline elements (e.g. **CALS** and **SNDYKLSF**) can remain unaltered if few nucleotide additions or deletions have taken place during recombination (e.g. **CALSNDYKLSF**, public PKY sequence detailed in Table. 5.1).

Similar CDR3 sequences arising from minimal somatic alteration of V(D)J elements are likely to be present in multiple individuals across the population (convergent recombination). If they can facilitate recognition of epitopes relevant in common diseases and infections, they will constitute public T-cell responses^{276,277}.

Where epitope-specific TRAV gene usage is highly focused due to CDR1 and CDR2 structural constraints, the probability of publicity may also be increased. This is because fewer genes, and therefore a narrower pool CDR3 sequences, are available to facilitate recognition. In this way, structural limitations can increase the likelihood of public TCRs.

This may explain the abundance of shared TCR sequences in response to QAR (Table. 1), where gene usage was nearly exclusive to TRAV38-2DV8. All three public CDR3 amino acid sequences are distinct at the nucleotide level, and relatively short at thirteen amino acids; both features have been highlighted as evidence of convergent recombination^{276,277}.

Public sequences in response to other epitopes where gene usage was broader also show a predominance of short sequences twelve to fourteen amino acids in length. Only one longer public sequence, sixteen residues in length, was identified.

5.5.2.3 Conserved CDR3 Amino Acid Usage

In the absence of identical sequences, patterns were observed within CDR3 motifs that were highly consistent across TRAV genes, donors and CDR3 lengths. Two examples are the presence of a non-germline arginine at CDR3 α 6 in TRAV23 responses to SGP (Fig. 5.8A), and a central asparagine at CDR3 α 7 in multiple TRAV genes specific to GLI (Fig. 5.8B). Many more examples were present where single or several amino acids with identical physical properties were positionally conserved.

The primary explanations are probably structural, where specific side chain properties are necessary to form CDR3 α interactions with the unique pHLA surface formed by each epitope within HLA-DR1. Each surface presents a patchwork of charged, polar and hydrophobic regions that preferentially accommodates certain CDR3 amino acids in combination with CDR1 and CDR2 determined binding orientations.

If a TCR has the correct TRAV gene (and β chain), but does not have certain positionally conserved residues necessary for binding, recognition will not occur and this is evident across the epitope-specific motifs.

5.5.3 Future Work

In this chapter, epitope-specific CD4⁺ T-cells from five HLA-DR1 donors exhibited similarities and striking patterns at the cellular and genetic level. These response characteristics demonstrated a degree of sharing in the mechanisms which underlie CD4⁺ T-cell recognition across the population. Further investigations aimed to explain these response characteristics by analysis at the molecular level.

In the subsequent chapter, a specific recognition mechanism was studied directly. Two different TCRs, each in complex with DR1-PKY, were compared structurally in order to visualise how their CDR loops interacted with both the peptide and HLA. Corresponding clonotypic sequence information and complex structural data were

aligned in order to rationalise features such as gene usage bias and CDR3 amino acid motifs. In this way, the structural basis of observed genetic characteristics observed in this chapter were better understood, and this work helped inform future investigations into CD4+ T-cell repertoires and antigen recognition.

Future work, to complement data presented here, includes analysis of TCR β -chain sequence information, *ex vivo* HLA-multimer staining and *ex vivo* IFN- γ ELISpot assays to assess the magnitude of these responses in the absence of culture. As part of a longitudinal investigation into the importance conserved epitopes in influenza infection, their cellular and genetic analysis pre- and post-vaccination, or post-infection, could be highly informative.

6 Recognition of DR1-PKY by Two Distinct TRAV8-4 TCRs is Mediated by Residues that are Conserved across PKY-Specific TCR α Repertoires

6.1 Abstract

X-Ray crystallography can be used to generate protein structure information at atomic resolution. This enables detailed analysis of secondary and tertiary structure, solvent-exposed and buried surface area as well as the location of amino acid side chains that are available to interact with other molecules. This information can answer mechanistic questions concerning binding partners, enzyme active sites and conformational flexibility.

In order to investigate the binding of TCRs specific to the HA₃₀₆₋₁₈ peptide “PKY” and explain the corresponding α -chain sequence information in the previous chapter, two TCR-pHLA ternary complex structures were compared. The structures were analysed in order to understand the contribution of each CDR loop in binding both the peptide and the HLA.

This allowed rationalisation of gene usage and CDR3 motifs present in cognate TCR α repertoires from five HLA-DR1 donors. Specific CDR1 α germline-encoded contacts and CDR3 α rearranged contacts were identified as key structural features in pHLA recognition by the epitope-specific T-cells of multiple donors. This analysis provided a blueprint by which other CD4+ T-cell repertoire information could be analysed following the generation of relevant structural complexes.

6.2 Introduction

Analysis of protein structures at atomic resolution can identify molecular mediators of receptor-ligand interactions that result in biological phenotypes. X-ray crystallography relies on the successful diffraction of X-rays by a protein crystal to produce diffraction patterns. The mathematical attributes of these patterns can be decoded by Fourier transform to produce an electron density map of the repeating unit constituting the crystal lattice.

The polypeptide chain is fitted to the density map in order to agree with observed data and build the protein structure. At high resolution ($<3 \text{ \AA}$) detailed orientation of amino acid side chains and peptide backbone can be observed. These can help understand how sub units of a protein or binding partners can come together through non-covalent interactions, and identify the critical residues needed for their formation.

When analysing immune receptors, such as the TCR and pHLA molecules, several questions can be answered through interrogation of structural data. Structures of the pHLA alone, such as those presented in chapter 4, provide information on the anchor and solvent-exposed residues that constitute an epitope. Many such class-II structures have been published^{127,267}. Yet less than twenty human pHLA class-II molecules in complex with a cognate TCR are available at present²⁸⁹, and only five for HLA-DR1, from which only one is a viral epitope²¹⁹.

6.2.1 Features of the TCR-pHLA Class II Protein Complex

A conceptual challenge of immunology is understanding how the TCR is capable of distinguishing between a vast pool of structurally similar HLA molecules in order to facilitate a T-cell response to very specific peptides. Several theories on the mechanisms of T-cell activation^{38,290}, the necessity of TCR cross-reactivity^{211,212,291}, the evolutionary origin of the TCR, the contributions of class-II flanking regions²⁸⁵, the role of self antigen⁹, the thymic development of regulatory T-cells^{22,292} and the relevance of non-canonical binding have arisen around the subject of TCR recognition^{23,27,293}.

Structural information is a primary source of information to help understand some of these questions. When analysing a TCR-pHLA complex, several features provide indications of how the epitope is recognised with high levels of specificity.

The crossing angle is defined as the binding orientation of the TCR over the pHLA molecule⁵, and roughly described as the angle between the peptide and a line passing between the α and β chain of the TCR. This parameter, when calculated in the same manner, allows comparison of docking geometries relative to the peptide backbone and indicates the extent of peptide or HLA interactions. Following this direct identification of TCR to pHLA contacts, the TCR “footprint,” can elucidate the binding regions of each CDR loop. For both class-I and class-II systems, germline CDR1 and CDR2 amino acids have been estimated to account for 75 % of the interface contacts²³, and are thought to account for HLA-specific recognition. The CDR2 makes contacts exclusively to the HLA, while the CDR1 loop sits closer to the peptide and forms additional epitope-specific interactions in several cases²⁸⁹. The CDR3 loop predominantly contacts the peptide, forming strong, discriminatory interactions that facilitate epitope-specific recognition. Exceptions to this “canonical” binding modes exist, and are the focus recent research^{31,32}.

6.2.2 Rationalising Repertoire Information using Structural Analysis

Through identification of the peptide and HLA specific contacts made by germline and somatic TCR residues, structural information can further our understanding of the immune response to specific epitopes. A number of recent studies have compared complex structural data to TCR repertoire information in order to explain gene usage bias and CDR3 sequence patterns^{250,294}.

This work has predominantly been carried out in HLA class-I systems, while the few class-II studies have used TCR sequences derived from clones^{167,219,284,295} or β -chain information solely²⁷². Large scale epitope-specific repertoire analysis and structural alignment with HLA class-II has not been undertaken. The difficulty in obtaining paired sequence information followed by crystallising the TCR-pHLA in complex, means that

structural information will always be the limiting factor relative to the abundance of epitope-specific TCR sequence information. For this reason, a minimal number of crystal structures serve as the model around which much clonotypic information is understood.

6.2.3 PKY as a Model System

In previous chapters, PKY was used as a positive control due to its ability to generate a consistent HLA-DR1 specific immune response. It has historically been used to investigate several aspects of HLA class-II antigen processing and presentation as well as the response to influenza haemagglutinin²⁹⁶. It is termed “universal” due to its ability to bind to multiple HLA-DR alleles and elicit strong immune responses^{168,219}. Abundant CD4+ T-cell response information exists in response to this epitope, and it remains the closest class-II system available to A2-GIL (M1₅₈₋₆₆)²⁴⁹ in terms of available structural^{166,219} and clonotypic^{219,272} information related to influenza responses at the population level. This availability of a published TCR-pHLA-DR1 crystal structure²¹⁹ makes PKY the ideal candidate to begin structural and repertoire data comparisons.

6.3 Aims

Structural analysis can identify the interactions between a TCR and cognate pHLA that are central to recognition. These interactions may be highly conserved across diverse TCR repertoires that are specific to the same epitope.

The aim of this chapter was to compare two distinct $\alpha\beta$ TCRs, both utilising TRAV8-4 but differing in their CDR3 α sequence and paired β -chain, in order to assess the extent of these conserved interactions in the recognition of DR1-PKY. The identified structural features and key residues could be aligned with clonotypic sequence information to understand whether such features were conserved across PKY-specific repertoires or were unique to specific TRAV genes and sequences.

Specific aims:

- 1) To compare two TCR HLA-DR1 crystal complex structures to identify similarities and differences in the features that mediate recognition of the HA₃₀₆₋₃₁₈ epitope (PKYVKQNTLKLAT).
- 2) To align the identified structural interactions with TRAV gene usage bias and CDR3 α motifs obtained from PKY specific repertoire data in five HLA-DR1 donors.

6.4 Results

6.4.1 Comparison of TCR-pHLA Complex Structures

In order to understand the interactions between two different TCRs and the same pHLA, inspection of two complex structures using the PKY peptide (HA₃₀₆₋₃₁₈) presented by HLA-DR1 was carried out. The first structure was a published complex with the HA1.7 TCR²¹⁹ (Fig. 6.1) and utilised a flexible octapeptide linker between the peptide and TCR to increase the probability of trimer formation (TCR-pHLA). The HA1.7 TCR was derived from a CD4+ T-cell clone specific to the PKY peptide which also cross-reacts with the same peptide presented by HLA-DR4 for which another complex structure has been published.

The second structure was solved in our group (refolding of pHLA was performed by the author, the TCR was refolded by Aaron Wall and structure solution was performed by Dr David Cole) and remains unpublished. It consists of PKY bound by HLA-DR1 and complexed with the F11 TCR at 1.9 Å resolution (Fig. 6.2). The F11 TCR was sequenced from a PKY specific clone derived from Donor-1, and both chains appeared in their clonotypic sequence information at low frequency, as well as in data from the same donor in published studies¹⁶⁸.

The affinities of HA1.7 and F11 for DR1-PKY have been determined by BIAcore analysis¹⁶⁸ as 50 µM and 26.7 µM respectively (25 °C). These are much weaker than affinities determined for class-I viral epitopes²⁹⁷. When compared to the published affinities of the public JM22 TCR (TRAV27, TRBV19) for A2-GIL (M1₅₆₋₆₄) of 6 µM²⁹⁸ and 1.79 µM²⁵⁰, these interactions are between four to twenty-five times weaker (depending on values used). This is consistent with weaker average TCR affinities for available class-II systems than those for class-I²⁹⁷ and points to fundamental structural differences that could be explored further.

Gene	HA1.7	F11
TRAV	TRAV8-4	TRAV8-4
TRAJ	TRAJ48	TRAJ30
TRBV	TRBV28	TRBV24-1
TRBJ	TRBJ1-2	TRBJ1-2
Sequence		
CDR1 α	SSVPPY	SSVPPY
CDR2 α	Y TSAATLV	Y TSAATLV
CDR3 α	CAVSE <u>SPFGNEKLT</u> F <i>15-mer</i>	CAVSE <u>QDDKIIF</u> F <i>12-mer</i>
CDR1 β	MDHEN	KGHDR
CDR2 β	SYDVKM	SFDVKD
CDR3 β	CASSSTGLPYGYTF <i>14-mer</i>	CATSDSYGYTF <i>12-mer</i>

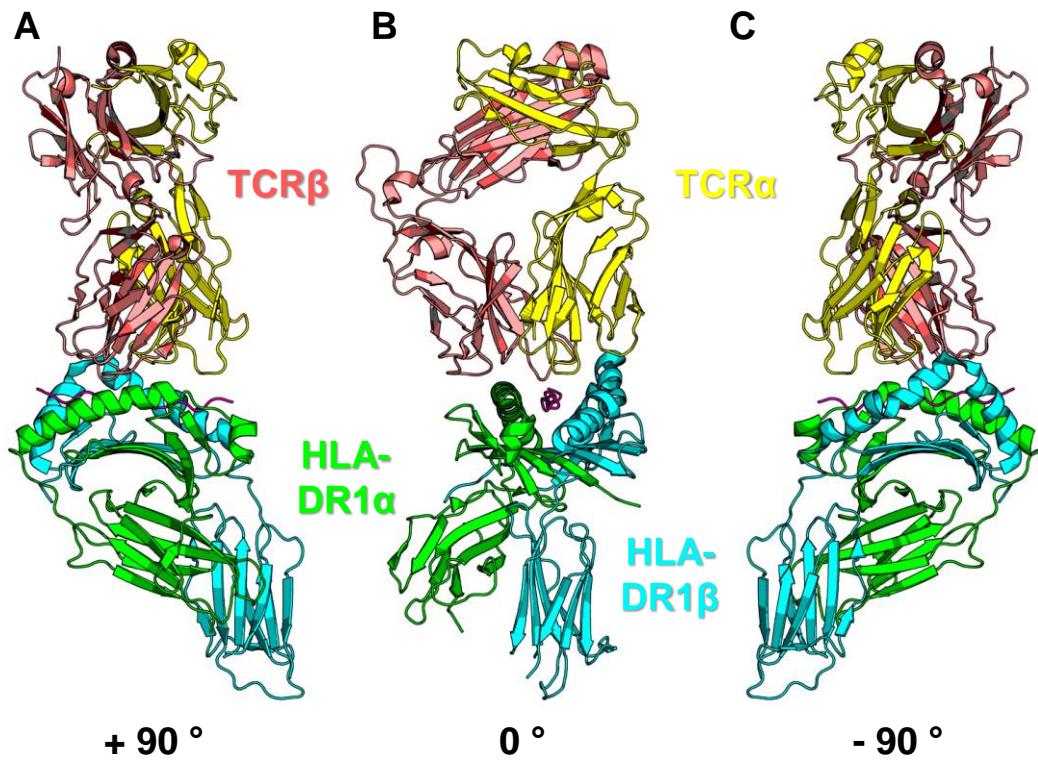
Table 6.1. Genetic and amino acid composition of HA1.7 and F11 TCRs. Where text is in red, the same gene and amino acids are present in each TCR sequence.

Each TCR utilises the same TRAV gene (TRAV8-4) but have different TRBV genes and distinct CDR3 α sequences of differing lengths (Table 6.1). When repertoires show bias towards a limited number of TRAV genes, as observed in this study, is this because they contact the pHLA in a highly conserved manner, despite having distinct CDR3 α lengths, amino acid composition and β -chain pairing?

Each structure shows TCR binding in a similar orientation (Fig. 6.1, 6.2), with the TCR α -chain over the N-terminus of the peptide and the DR1 β -groove helix, while the TCR β -chain sits over the C-terminus and the DR1 α -groove helix.

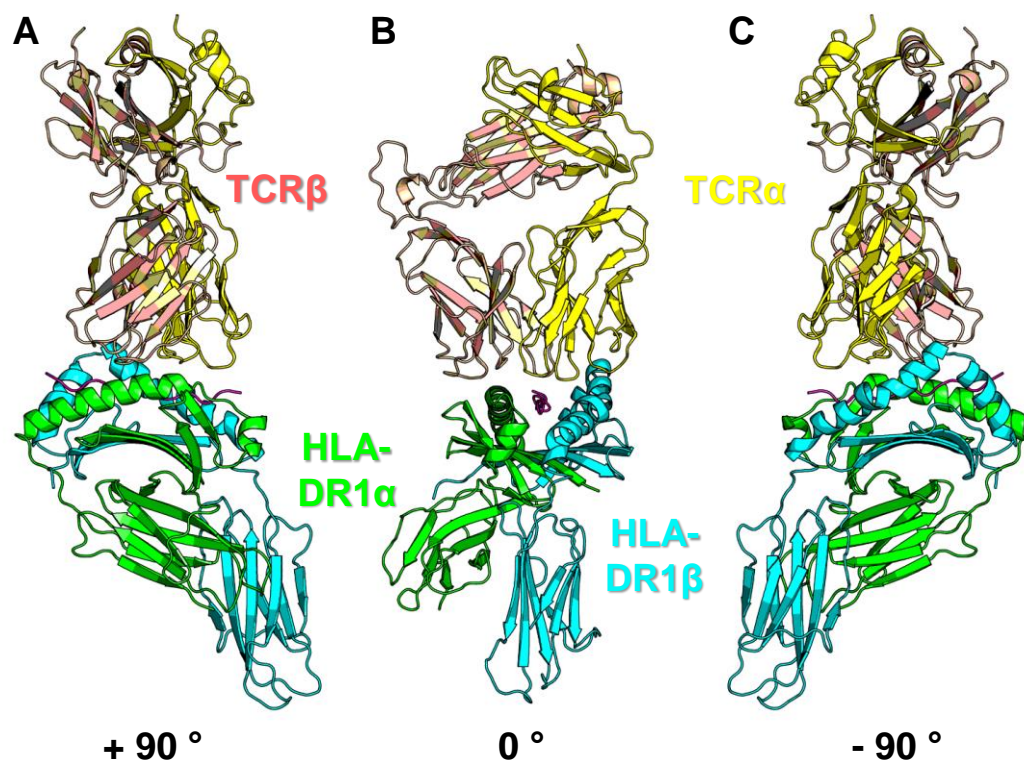
The orientation of the TCR over the pHLA is quantified by the “crossing angle”, a parameter with a high level of variability across many complex structures in both class-I (37-90°) and class-II (37-115°) systems²⁸⁹ (Fig. 6.1B, 6.2B). Between HA1.7 and F11, angles of 47.22° and 47.43° respectively were within 0.5 degrees, demonstrating a highly conserved binding orientation despite the presence of distinct β -chains.

When secondary structural features were aligned, both pHLA components show high levels of overlap (Fig. 6.3). HLA similarity was most apparent, while TCR chain alignment showed greater variation, specifically in the CDR3 loops, which exhibited minimal overlap due to their distinct lengths. Polar contacts between the peptide backbone and HLA were highly conserved (Fig. 6.4), reflecting the significant rigidity and conservation in epitope presentation. Many of the interactions documented in published HLA class-II structural data¹²⁷ and those determined in this project (chapter 4) were present, in addition to peptide-specific side chain contacts made with the HLA.



HA1.7 Complex	Resolution	2.60	RMS Deviations From Ideal Values	
	Completeness	99.6	Bond Lengths	0.008
	Space Group	C 1 2 1	Bond Angles	1.500
	R Value	0.221		
	Free R Value	0.255		
	Free R Value Test Set Size	5.0%		
	Free R Value Test Set Count	1871		
	Mean B Value	38.7		
	Number of Reflections	37122	Crossing Angle	47.22

Figure 6.1. HA1.7-DR1-PKY complex structure secondary structural representation and data table. Structural data taken from PDB file 1fyt published by Hennecke et al¹⁹. Three orientations are shown: (B) 0° facing down the binding groove from the N-terminus, (A) +90 facing the DR1 α-chain, (C) -90 facing the DR1 β-chain. Structures are represented by secondary structure cartoon analysis.



F11 Complex	Resolution	1.91	RMS Deviations From Ideal Values	
	Completeness	99.9	Bond Lengths	0.019
	Space Group	P 2 ₁ 2 ₁ 2	Bond Angles	1.944
	R Value	0.206		
	Free R Value	0.240		
	Free R Value Test Set Size	4.9%		
	Free R Value Test Set Count	4762		
	Mean B Value	55.4		
	Number of Reflections	91698	Crossing Angle	47.43

Figure 6.2. F11-DR1-PKY complex structure secondary structural representation and data table. Three orientations are shown: **(B)** 0° facing down the binding groove from the N-terminus, **(A)** +90 facing the DR1 α-chain, **(C)** -90 facing the DR1 β-chain.

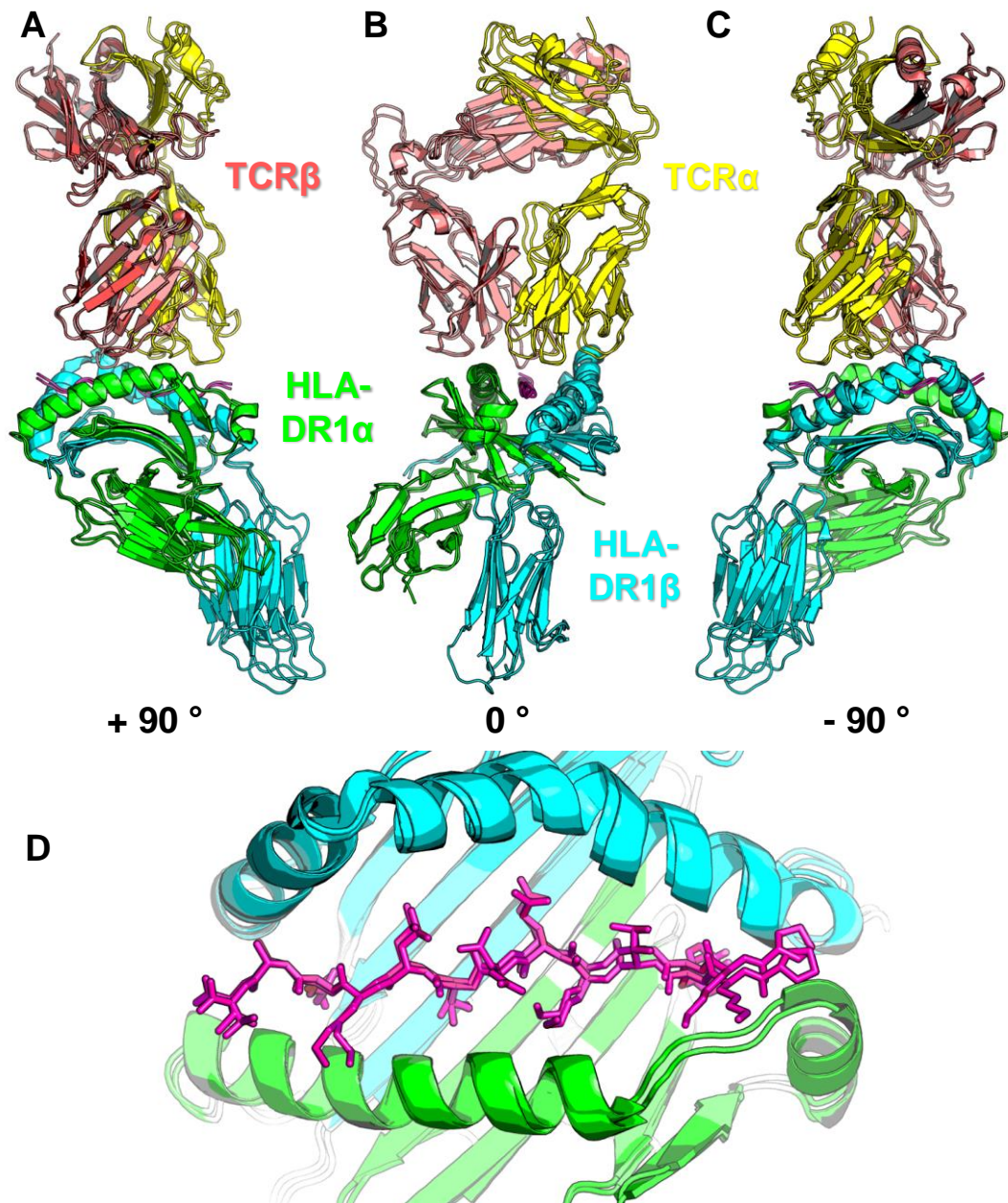


Figure 6.3. Structural comparison of F11 and HA.17 complexes by secondary structural overlay alignment. (A, B, C) Three complex orientations are shown rotated in the z plane. (D) View of the peptide overlay in stick representation within the binding groove. Overlay alignment was carried out by the align function in PyMol viewer, giving the closest overlap of the majority of each structure.

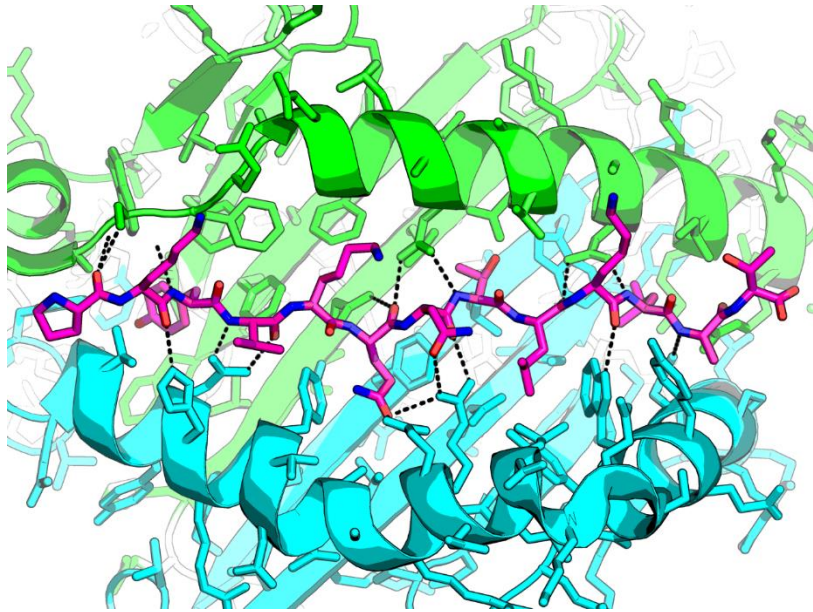
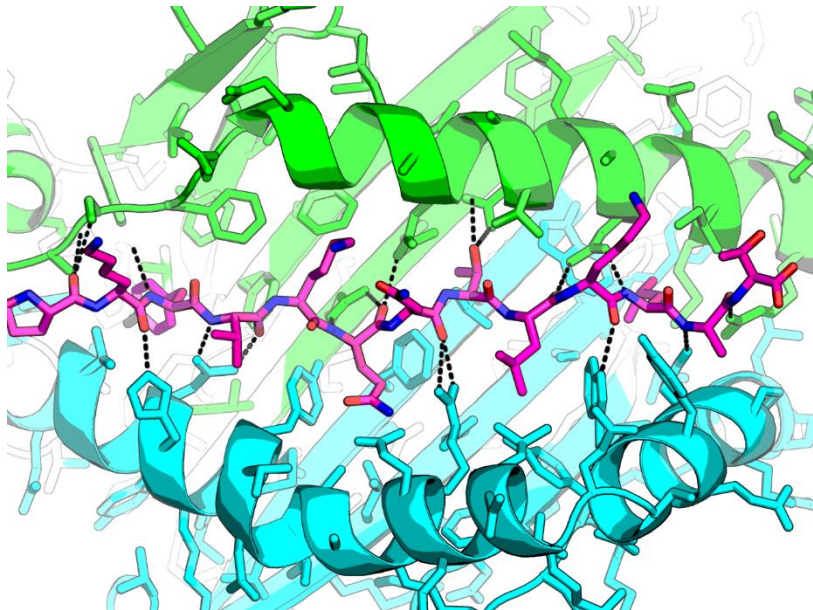
A**HA1.7 Complex****B****F11 Complex****P K Y V K Q N I L K L A T**

Figure 6.4. *Stabilising polar peptide contacts with the HLA groove for (A) HA1.7 complex and (B) F11. Polar contacts are represented by black dashed lines between the backbone and side chains of the peptide, and the side chains of amino acids that constitute the HLA binding groove.*

6.4.2 TCR-pHLA Interface

The amino acid interactions that govern T-cell to epitope recognition occur at the contact interface between the pHLA and the TCR. The interface is formed by non-covalent interactions such as salt bridges, hydrogen bonds and van der Waals (vdWs) forces, which contribute to form an energetically favourable binding event. By analysing the nature of these interactions in a TCR-pHLA complex structure, CDR amino acid properties essential for either peptide or HLA binding can be identified.

Inspection of 4 Å contact footprints (Fig. 6.5A, 6.5B)) showed that the majority of peptide residues were contacted by either the CDR3 α (blue) or CDR3 β (turquoise) loop over the central region of the HLA. CDR2 α (green) and CDR2 β (pink) loops exclusively contacted the HLA on the outside of the groove, while the CDR1 α (red) and CDR1 β (yellow) loops contacted the inside of the groove, within 4 Å of both peptide and HLA.

The total number of atom to atom contacts within 4 Å was similar for each complex, with 112 for HA1.7 and 102 for F11. The binding affinities determined at 50 μ M and 26.7 μ M for HA1.7 and F11 respectively indicate that number of atom to atom contacts is not the sole determinant of affinity.

These contact footprints, and overlays of the CDR backbone loops (Fig. 6.5C, 6.5D), confirm the “canonical” binding mode of both TCRs: with CDR3s predominantly contacting the peptide, CDR2s exclusively contacting the HLA, and the CDR1s in proximity to both components. CDR1 and CDR2 loops appeared highly conserved in orientation above the contact region, while CDR3 loops exhibited variation due to the difference in length.

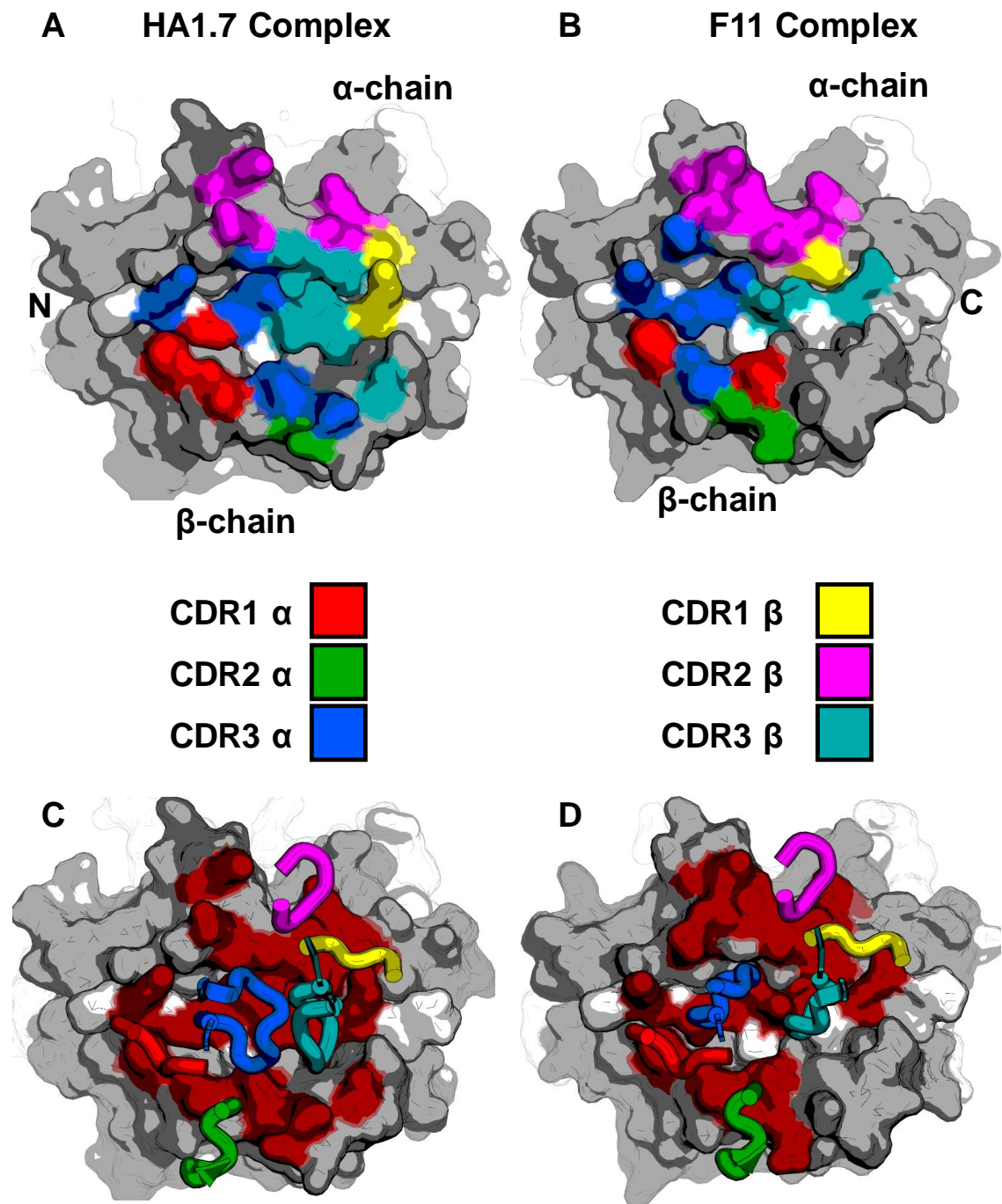


Figure 6.5. TCR Contact Footprint representation over pHLA surface plot. Surface of the HLA that does not bind TCR is shown in grey, peptide in white, amino acids contacted by the TCR are coloured according to the CDR loop (see central key) for **(A) HA1.7** and **(B) F11** complexes. Colours are shown at the solvent exposed surface and represent regions that are within 4.00 Å distance of the TCR.

(C) HA1.7 and **(D) F11.** Direct overlay of each TCR CDR loop (backbone represented as a coloured tube) above the contact footprint in uniform red colouring. The entire CDR

loops according to IMGT definition are represented. See later figures for analysis of individual amino acid contributions.

6.4.3 Understanding Repertoires with Structural Information

In the previous chapter, the analysis focused on the gene usage and CDR3 motifs of the TCR α -chain repertoires in response to specific epitopes. Observed in response to PKY was a biased usage of four TRAV genes (TRAV8-4, TRAV8-6, TRAV13-1, TRAV14) and a prevalence of acidic residues in the CDR3 α motifs across a range of lengths. Here, we investigate these interactions and provide a structural explanation for the epitope-specific bias and patterns.

6.4.3.1 CDR1 α Interactions

As seen from contact footprints and CDR loop overlays, the CDR1 α loop sat close to the inner side of the DR1 β -chain groove helix and the N-terminal peptide residues, with the CDR2 α loop directly over the DR1 β -chain. Both the HA1.7 and F11 TCRs utilised CDR1 α residues CDR1 α Val28, CDR1 α Pro29 and CDR1 α Tyr31 at the contact interface (Fig. 6.6 & Table 6.2). CDR1 α Val28 and CDR1 α Pro29 made conserved vdWs interactions with DR1 β His81 in both complexes (Fig. 6.6B, 6.6D). Additionally, CDR1 α Val28 was involved in a peptide contact with P-1Lys and P2Val in both complexes (Fig. 6.6A, 6.6C). CDR1 α Tyr31 was employed slightly differently by the two TCRs, making vdWs contacts with DR1 β Gln70 in the F11 structure (Fig. 6.6D), and vdWs contacts with DR1 β Thr77 in the HA1.7 structure (Fig. 6.6B).

Interestingly, CDR1 α Val28 was present in two of the genes selected (TRAV8-4 and TRAV8-6) (Fig. 6.6E). The other two genes not encoding CDR1 α Val28 (TRAV13-1 and TRAV14DV4), contained Ala or Pro in this position, both small weakly hydrophobic residues that would be congenial with the binding mode employed by CDR1 α Val28. CDR1 α Pro29 was only selected in TRAV8-4, with CDR1 α Ser27 being present in the other V-genes observed. However, Ser has similar properties in terms of size and charge to Pro, so could be structurally compatible during CDR1 loop ligation with DR1-PKY.

CDR1 α Tyr31 and CDR1 α Ser27 were conserved in all of the genes identified in PKY specific repertoires (Fig. 6.6E). Further genetic investigation demonstrated that only TRAV8-4, TRAV8-6, TRAV13-1 and TRAV14DV4 contained a motif that included

Val/Ala/Pro28, Pro/Ser29, Tyr31 and Ser27. Thus, the structural analysis suggested how the V-genes detected during recognition of DR1-PKY have been uniquely selected to make important contacts with the HLA and peptide. These observations help to illuminate the structural basis of conserved selection of specific TCR α V-genes in response to DR1-PKY.

Whether these CDR1 α residues are essential, or dispensable, for TCR binding to DR1-PKY would require further structural or mutational analysis in order to account for their contribution to the gene usage pattern.

6.4.3.2 CDR2 α Interactions with DR1 β

CDR2 α interacted exclusively with the DR1 β -chain (Fig. 6.7 & Table 6.3). The contacts were mediated by CDR2 α Ser51 and CDR2 α Ala52 vdWs interactions in both structures, as well as an additional vdWs contact between CDR2 α Th50 and DR1 β Gln70 in the F11 complex (Fig. 6.7B).

The CDR2 α contacts were fewer in number (Table 6.3) and contained no polar interactions when compared to CDR1 α . When compared to the other genes that mediate the PKY specific response, only Ser51 appeared conserved (Fig. 6.7C). This may point to a lower importance of CDR2 α in recognition of DR1-PKY, and suggests these interactions are not critical for TCR binding.

NOTE FOR ALL FIGURES: Where the amino acids are represented in stick format (the peptide/side chains/CDRs in this case), black corresponds to carbon atoms, red to oxygen atoms (which may carry a negative dipole or charge) and blue to nitrogen atoms (which may carry a positive dipole or charge).

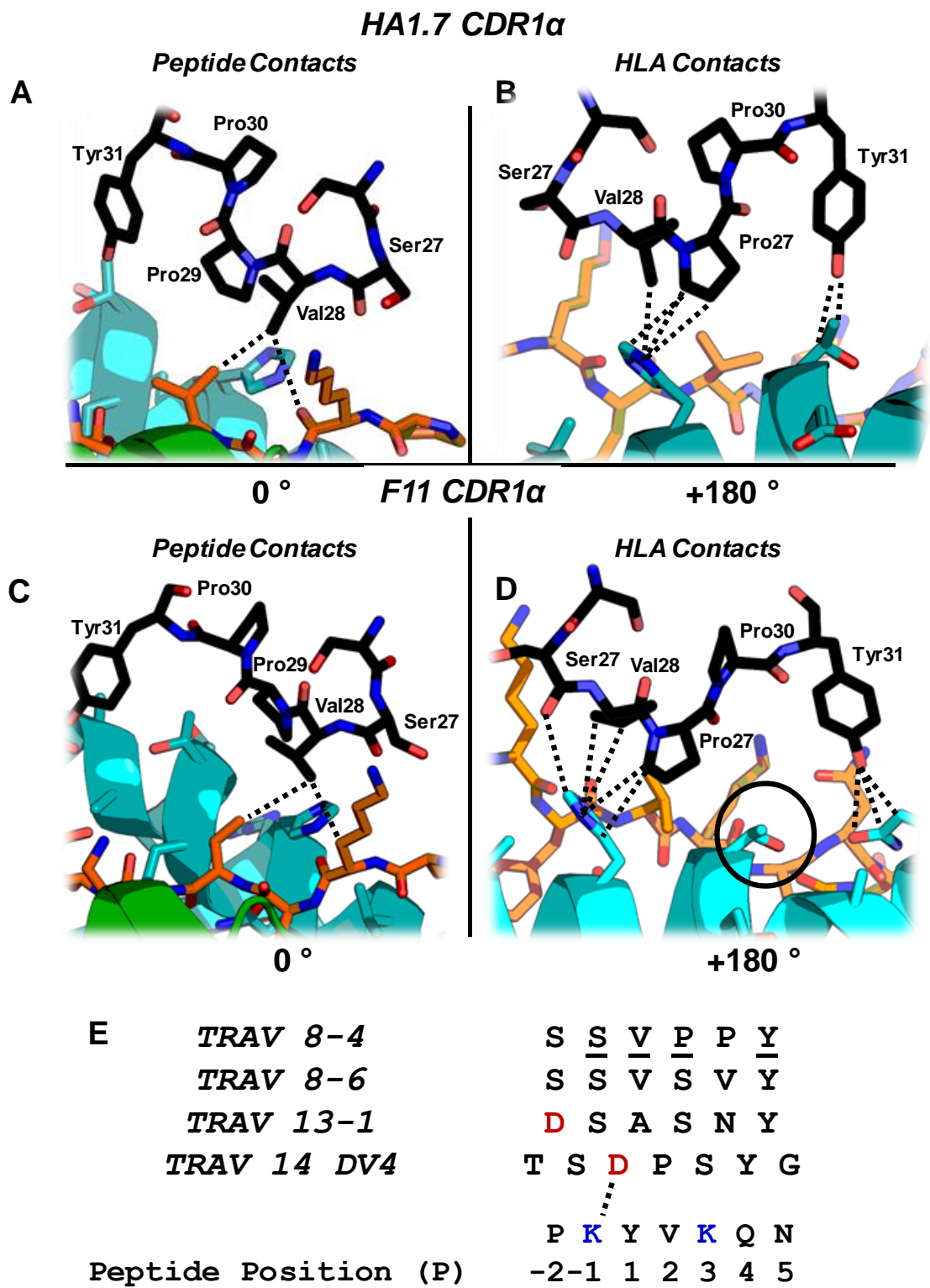


Figure 6.6 CDR1 α Contacts with both peptide and HLA.

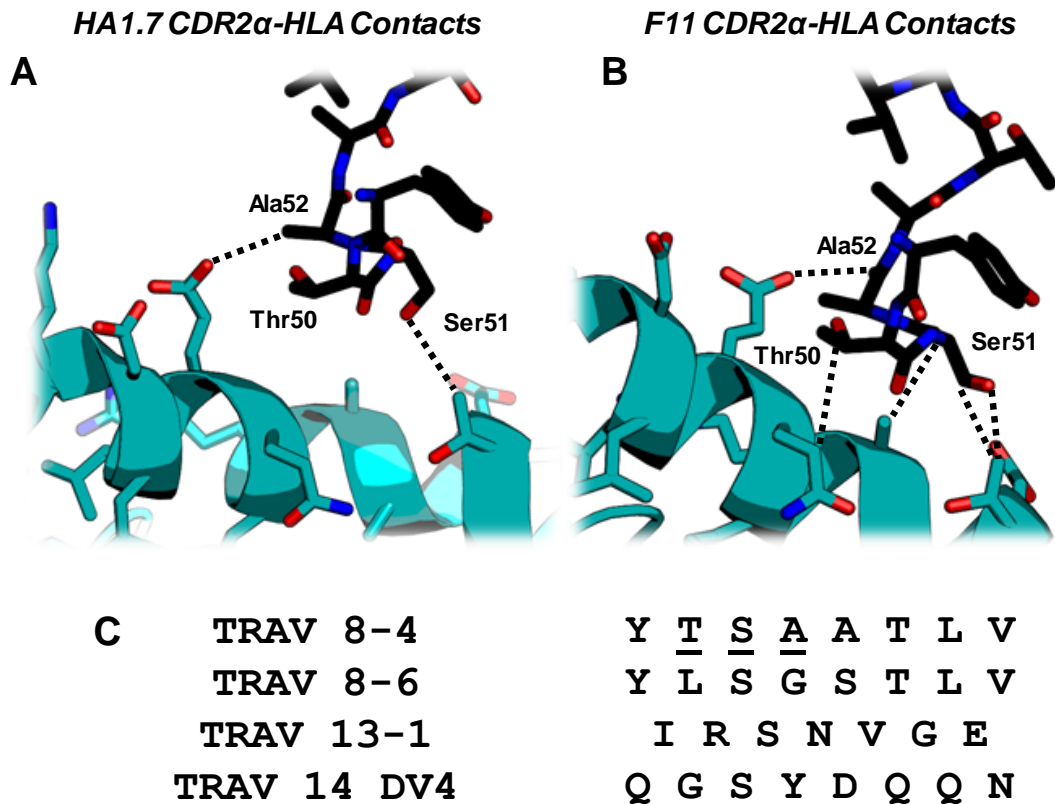
Figure legend on following page.

HA1.7 TCR Contacts							
TCR CDR1 α			Peptide/HLA Contact				
NO.	AA	ATOM	Contact	NO.	AA	ATOM	DIST
28	VAL	CG2[C]:	DR1- β	81	HIS	NE2[N]	3.84
28	VAL	CG2[C]:	Peptide	307	LYS	O [O]	4.00
			Peptide	309	VAL	CG2[C]	3.19
29	PRO	CG [C]:	DR1- β	81	HIS	ND1[N]	3.80
			DR1- β	81	HIS	ND1[N]	3.76
			DR1- β	81	HIS	CE1[C]	3.61
			DR1- β	81	HIS	NE2[N]	3.89
31	TYR	OH [O]:	DR1- β	77	THR	CB [C]	3.73
			DR1- β	77	THR	CG2[C]	3.93

F11 TCR							
TCR CDR1 α			Peptide/HLA Contact				
NO.	AA	ATOM	Contact	NO.	AA	ATOM	DIST
27	SER	O [O]	DR1- β	81	HIS	CE1[C]	3.93
28	VAL	CB [C]	DR1- β	81	HIS	NE2[N]	3.95
28	VAL	CG1[C]	DR1- β	81	HIS	NE2[N]	3.86
			Peptide	2	LYS	CB [C]	3.91
			Peptide	4	VAL	CG1[C]	3.98
29	PRO	CD [C]	DR1- β	81	HIS	CG [C]	4.00
			DR1- β	81	HIS	ND1[N]	3.71
31	TYR	OH [O]	DR1- β	70	GLN	CG [C]	3.65
			DR1- β	70	GLN	CD [C]	3.41
			DR1- β	70	GLN	NE2[N]	3.83
			DR1- β	70	GLN	OE1[O]	3.54

Figure 6.6 & Table 6.2. CDR1 α Contacts with both peptide and HLA. HA1.7 CDR1 α (black chain) contacts with (A) peptide (orange chain) and (B) HLA (blue chain) with side chains within 4 Å shown by dashed lines. Corresponding F11 CDR1 contacts with (C) peptide and D. HLA. E. CDR1 sequences of the top genes identified in clonotyping data, with acidic residues coloured in red.

Table 6.2. Corresponding Contact details of all atoms within 4.00 Å. Between the CDR1 α loop and the peptide (shaded in grey) or the HLA.



HA1.7 TCR Contacts							
TCR CDR2 α			Peptide/HLA Contact				
NO.	AA	ATOM	Contact	NO.	AA	ATOM	DIST
51	SER	OG [O]:	DR1- β	77	THR	CG2[C]	3.91
52	ALA	CB [C]:	DR1- β	69	GLU	OE2[O]	3.48

F11 TCR							
TCR CDR2 α			Peptide/HLA Contact				
NO.	AA	ATOM	Contact	NO.	AA	ATOM	DIST
50	THR	O [O]	DR1- β	70	GLN	CG [C]	3.84
51	SER	C [C]	DR1- β	73	ALA	CB [C]	3.90
51	SER	O [O]	DR1- β	73	ALA	CB [C]	3.71
51	SER	CB [C]	DR1- β	77	THR	CG2[C]	3.30
51	SER	OG [O]	DR1- β	77	THR	CG2[C]	3.34
52	ALA	O [O]	DR1- β	69	GLU	OE2[O]	3.46

Figure 6.7 & Table 6.3. CDR2 α contacts with the peptide and HLA.

(A) HA1.7 CDR2 α (black chain) contacts with HLA (blue chain) within 4 Å shown by black dashed lines. (B) Corresponding F11 contacts. (C) CDR2 α sequences of the top genes identified in clonotyping data. Underlined amino acids are those mediating contact in the above images.

Table 6.3. Corresponding Contact details of all atoms within 4.00 Å.

6.4.3.3 CDR3 α Interactions and Repertoire Motifs

The hypervariable CDR3 α and CDR3 β loops account for the majority of TCR to peptide interactions. The variability arising from V(D)J recombination engenders the host with a highly diverse pool of TCR sequences capable of recognising vast numbers of potential pathogenic epitopes. Where multiple distinct TCR sequences see the same epitope, conserved amino acid features or physical properties that mediate the interaction are likely to be present. This phenomenon of epitope-specific sequence “motifs” or conserved positional properties has been observed in CD8+ and CD4+ T-cell repertoires, a limited number of which have corresponding structural complex data^{281,294}.

Analysis of the TRAV8-4 HA1.7 and F11 CDR3 α interactions display a similar pattern (Fig. 6.8A, 6.8B) despite a difference in length of three amino acids, and greater overall TCR footprint of the longer CDR3 loop (Fig. 6.5A). Both CDR3 α -chains utilised the side chains of three residues to interact with the peptide, two of which were acidic (CDR3 α Glu94 and CDR3 α Glu102 in HA1.7; CDR3 α Glu94 and CDR3 α Asp97 in F11) and formed highly favourable salt bridges with corresponding basic residues in the peptide (Lys P-1 and P3). The third interaction was weaker and distinct to each TCR: CDR3 α Phe97 to P5-Asn in HA1.7, and in F11, CDR3 α Gln95 to P2Val.

These interactions account for the predominance of acidic residues spread across the motifs derived from TCR α -chain repertoire information in the four dominant TRAV genes (Fig. 6.8C). A degree of flexibility in the CDR α loops and the opportunity to form two salt bridges with either P-1-Lys or P3-Lys could allow for positional variation in the occurrence of acidic residues across the motif plot.

The prevalence of acidic residues in response to PKY has been observed in other studies on CD4+ T-cell clones specific to DR1-PKY and DR4-PKY, and was initially predicted based on the PKY amino acid sequence. Yet this is the first example where large scale CD4+ T-cell repertoire data and motif comparison to more than one structure has been undertaken. The defining message is that for the same or similar TRAV gene with distinct CDR3 sequences and β -chain pairing, the germline CDR1 α and CDR2 β interactions with pHLA and TCR orientation are conserved. The conserved orientation is

likely to be favourable and facilitates hypervariable CDR3 α sequences to interact with peptide via conserved amino acid motifs or positional properties. This means that amidst the vast potential of immune diversity, the interactions that determine epitope-specific CD4+ T-cell recognition are focused and consistent in repertoires derived from multiple donors.

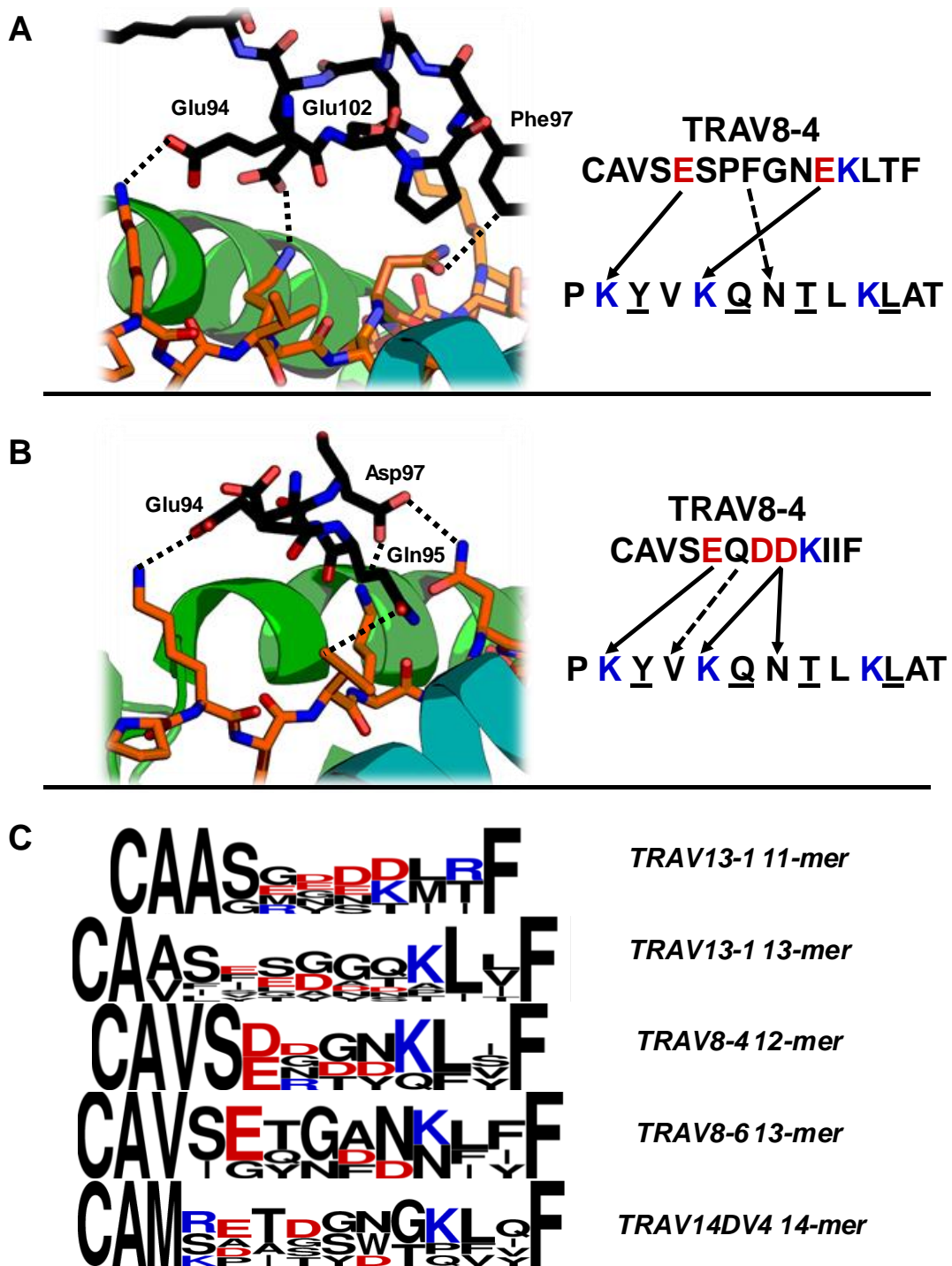


Figure 6.8. CDR3 α loop contacts with the peptide and corresponding acidic enriched motifs from repertoire data.

(A) HA1.7 CDR3 α (black chain) to peptide (orange chain) contacts. (B) F11 CDR3 α to peptide contacts. (C) CDR3 α Motif clonotyping data for different gene/length combinations, acidic residues shown in red and basic residues in blue.

HA1.7 TCR Contacts							
TCR CDR3 α			Peptide/HLA Contact				
NO.	AA	ATOM	Contact	NO.	AA	ATOM	DIST
94	GLU	OE1[O]:	Peptide	307	LYS	CE [C]	3.56
94	GLU	OE2[O]:	Peptide	307	LYS	CE [C]	3.86
94	GLU	CD [C]:	Peptide	307	LYS	NZ [N]	3.41
94	GLU	OE1[O]:	Peptide	307	LYS	NZ [N]	3.27
94	GLU	OE2[O]:	Peptide	307	LYS	NZ [N]	2.83
96	PRO	CB [C]:	DR1- β	70	GLN	CG [C]	3.34
			DR1- β	70	GLN	NE2[N]	3.89
96	PRO	CG [C]:	DR1- β	70	GLN	NE2[N]	3.65
97	PHE	CE2[C]:	DR1- β	66	ASP	O [O]	3.67
			DR1- β	70	GLN	CB [C]	3.77
97	PHE	CZ [C]:	DR1- β	70	GLN	CB [C]	3.36
97	PHE	CE1[C]:	DR1- β	70	GLN	CG [C]	3.86
97	PHE	CZ [C]:	DR1- β	70	GLN	CG [C]	3.55
97	PHE	CE1[C]:	Peptide	312	ASN	OD1[O]	3.61
			Peptide	312	ASN	ND2[N]	3.97
102	GLU	OE1[O]:	DR1- α	58	GLY	CA [C]	3.72
102	GLU	CD [C]:	DR1- α	61	ALA	CB [C]	3.94
102	GLU	OE2[O]:	DR1- α	61	ALA	CB [C]	3.17
102	GLU	CD [C]:	Peptide	310	LYS	NZ [N]	3.42
102	GLU	OE2[O]:	Peptide	310	LYS	NZ [N]	3.21
102	GLU	OE1[O]:	Peptide	310	LYS	NZ [N]	3.55

Table 6.4A. Corresponding Contact details of all atoms within 4.00 Å between HA1.7 CDR3 α and peptide.

F11 TCR							
TCR CDR3 α			Peptide/HLA Contact				
NO.	AA	ATOM	Contact	NO.	AA	ATOM	DIST
94	GLU	OE1[O]	Peptide	2	LYS	CD [C]	3.88
			Peptide	2	LYS	CE [C]	3.96
			Peptide	2	LYS	NZ [N]	2.93
94	GLU	CD [C]	Peptide	2	LYS	NZ [N]	3.90
95	GLN	OE1[O]	DR1- β	77	THR	C [C]	3.99
95	GLN	NE2[N]	DR1- β	77	THR	O [O]	3.71
95	GLN	CD [C]	DR1- β	77	THR	O [O]	3.73
95	GLN	OE1[O]	DR1- β	77	THR	O [O]	2.97
95	GLN	NE2[N]	Peptide	4	VAL	CB [C]	3.66
95	GLN	CB [C]	Peptide	4	VAL	CG1[C]	3.74
95	GLN	CD [C]	Peptide	4	VAL	CG1[C]	3.92
95	GLN	NE2[N]	Peptide	4	VAL	CG2[C]	3.57
96	ASP	CB [C]	DR1- α	55	GLU	OE2[O]	3.69
96	ASP	CG [C]	DR1- α	55	GLU	OE2[O]	3.97
96	ASP	OD2[O]	DR1- α	55	GLU	OE2[O]	3.35
97	ASP	OD2[O]	Peptide	5	LYS	CE [C]	3.30
97	ASP	CG [C]	Peptide	5	LYS	NZ [N]	3.57
97	ASP	OD2[O]	Peptide	5	LYS	NZ [N]	2.72
			Peptide	7	ASN	CG [C]	3.94
97	ASP	CG [C]	Peptide	7	ASN	ND2[N]	3.91
97	ASP	OD1[O]	Peptide	7	ASN	ND2[N]	3.26
97	ASP	OD2[O]	Peptide	7	ASN	ND2[N]	3.74
			Peptide	7	ASN	OD1[O]	3.43
98	LYS	CE [C]	DR1- α	55	GLU	OE2[O]	3.60
98	LYS	NZ [N]	DR1- α	55	GLU	OE2[O]	3.91

Table 6.4B. Corresponding Contact details of all atoms within 4.00 Å between F11 CDR3 α and peptide.

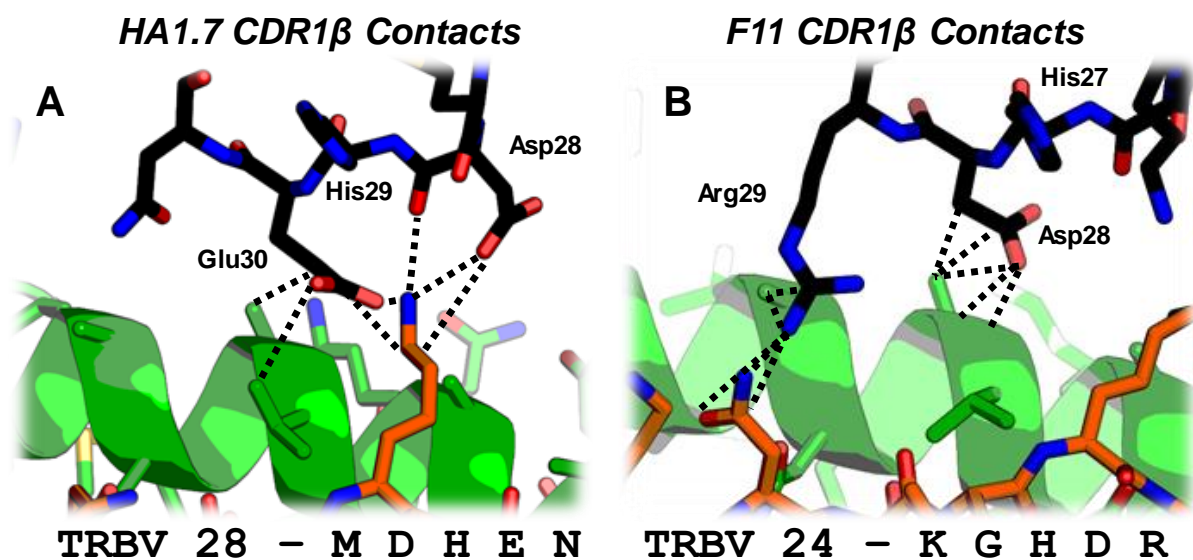
6.4.4 Conserved Interactions with the TCR β -chains and DR1-PKY

Although clonotypic data are not currently available for TCR β -chain deployment against DR1-PKY, structural analysis of the HA1.7 and F11 revealed a number of conserved interactions between β -chain residues and DR1-PKY, despite unique TRBV genes and distinct CDR3 loops (Table. 6.1).

Both TCRs contacted HLA residues DR1 α Ala64 and DR1 α Val65 using negatively charged amino acids in their CDR1 β loops (CDR1 β Glu30 for HA1.7, Fig. 6.9A or CDR1 β Asp29 for F11, Fig. 6.9B). HA1.7 contacted the peptide P7Lys through multiple salt bridge interactions (CDR1 β Asp28, CDR1 β His29 and CDR1 β Glu30, Fig. 6.9A), while F11 instead formed a polar contact with P5Asn and CDR1 β Arg29 (Fig. 6.9B).

Similarly, both TCRs maintained a network of HLA interactions with residues DR1 α Gln57, DR1 α Ala61, DR1 α Ala64, DR1 α Lys67 and DR1 α Lys39 through their CDR2 β loops (Fig. 6.10). Either forming salt bridges/hydrogen bonds through polar amino acids (CDR2 β Asp51, CDR2 β Lys55 and CDR2 β Glu56 for HA1.7, or CDR2 β Asp50 and CDR2 β Asn55 for F11) or vdWs contacts through aromatic amino acids (CDR2 β Tyr50 for HA1.7, or CDR2 β Phe48 for F11).

The CDR3 loops did not share the same degree of binding mode similarity, with the CDR3 β loop in HA1.7 making three polar contacts between CDR3 β Ser96, CDR3 β Thr97 and CDR3 β Gly98 and peptide P8Lys, P5Asn and P6Thr residues respectively (Fig. 6.11A). In comparison, the CDR3 loop in F11 made a number of contacts with peptide residues P7Asn, P6Thr and P8Lys mainly through a single residue CDR3 β Glu94 (Fig. 6.11B). It will be interesting to see if the majority of TCRs in response to DR1-PKY have similar CDR β residues, as was observed in the α -chain analysis. This will be investigated in future work by generating clonotyping data for the β -chain usage.



HA1.7 TCR Contacts							
TCR CDR1 β			Peptide/HLA Contact				
NO.	AA	ATOM	Contact	NO.	AA	ATOM	DIST
28	ASP	OD1[O]:	Peptide	315	LYS	CE [C]	3.98
28	ASP	O [O]:	Peptide	315	LYS	NZ [N]	2.90
28	ASP	OD1[O]:	Peptide	315	LYS	NZ [N]	3.07
30	GLU	CG [C]:	DR1- α	64	ALA	O [O]	3.93
			DR1- α	64	ALA	CB [C]	3.87
30	GLU	OE1[O]:	DR1- α	65	VAL	CG2[C]	3.84
30	GLU	OE2[O]:	DR1- α	68	ALA	CB [C]	3.72
30	GLU	CD [C]:	Peptide	315	LYS	CE [C]	3.71
30	GLU	OE1[O]:	Peptide	315	LYS	CE [C]	3.66
30	GLU	OE2[O]:	Peptide	315	LYS	CE [C]	2.97
30	GLU	CD [C]:	Peptide	315	LYS	NZ [N]	3.24
30	GLU	OE1[O]:	Peptide	315	LYS	NZ [N]	3.12
30	GLU	OE2[O]:	Peptide	315	LYS	NZ [N]	2.66

F11 TCR							
TCR CDR1 β			Peptide/HLA Contact				
NO.	AA	ATOM	Contact	NO.	AA	ATOM	DIST
28	ASP	OD2[O]	DR1- α	64	ALA	C [C]	3.88
28	ASP	OD2[O]	DR1- α	64	ALA	O [O]	3.91
28	ASP	CB [C]	DR1- α	64	ALA	CB [C]	3.79
28	ASP	CG [C]	DR1- α	64	ALA	CB [C]	3.65
28	ASP	OD1[O]	DR1- α	64	ALA	CB [C]	3.85
28	ASP	OD2[O]	DR1- α	65	VAL	N [N]	3.92
28	ASP	OD2[O]	DR1- α	65	VAL	CA [C]	3.98
28	ASP	OD2[O]	DR1- β	69	GLU	CB [C]	3.93
28	ASP	OD2[O]	DR1- β	69	GLU	CG [C]	3.78
29	ARG	CZ [C]	DR1- α	61	ALA	CB [C]	3.86
29	ARG	NH2[N]	DR1- α	61	ALA	CB [C]	3.73
29	ARG	NH2[N]	Peptide	7	ASN	CG [C]	3.58
			Peptide	7	ASN	ND2[N]	3.35
			Peptide	7	ASN	OD1[O]	3.77

Figure 6.9. & Table 6.5. CDR1 β contacts with peptide and HLA. Format is identical to previous figures, CDR1 α sequence below each image. (A) HA1.7 and (B) F11.

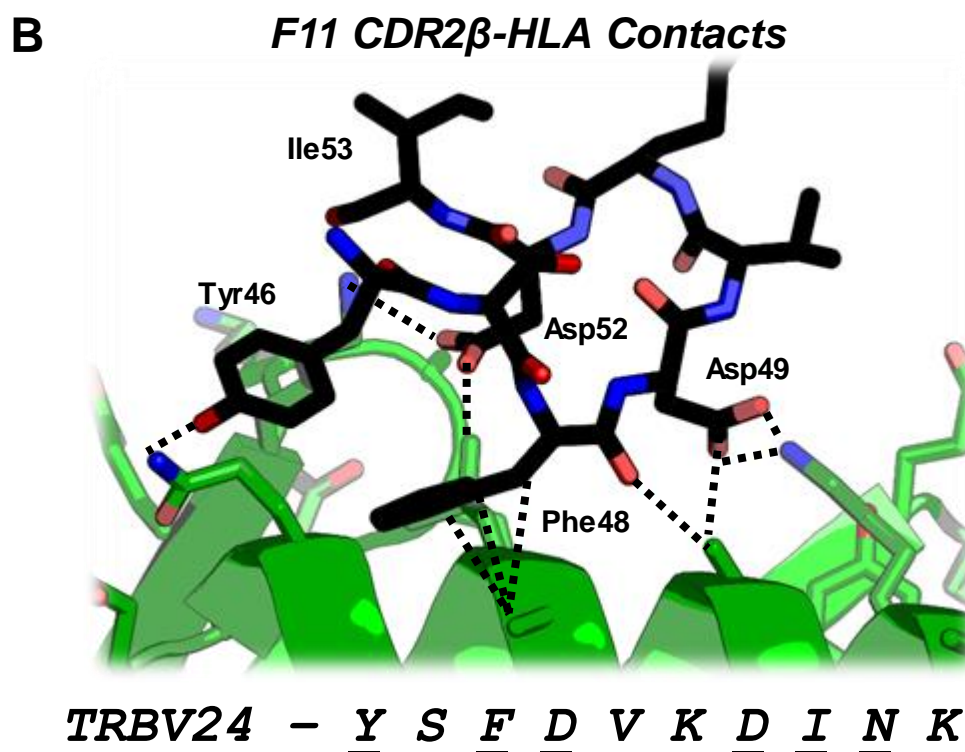
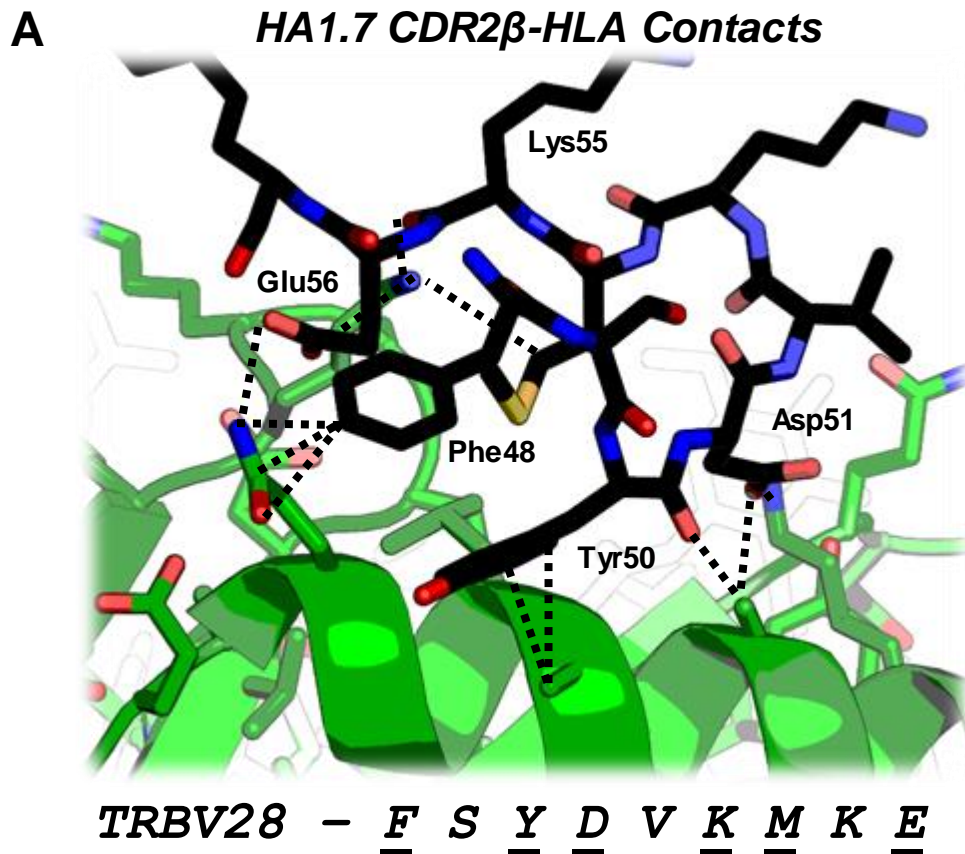


Figure 6.10. CDR2 β contacts with the HLA. (A) HA1.7 (B) F11. CDR2 loop sequence is shown below each image. Contact table is located in the appendix due to size.

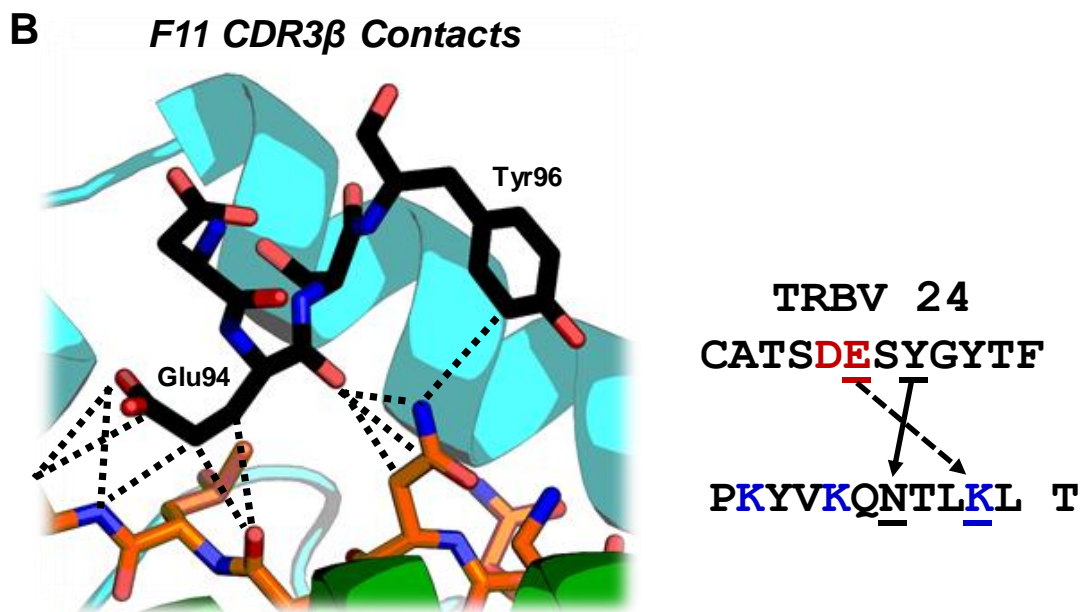
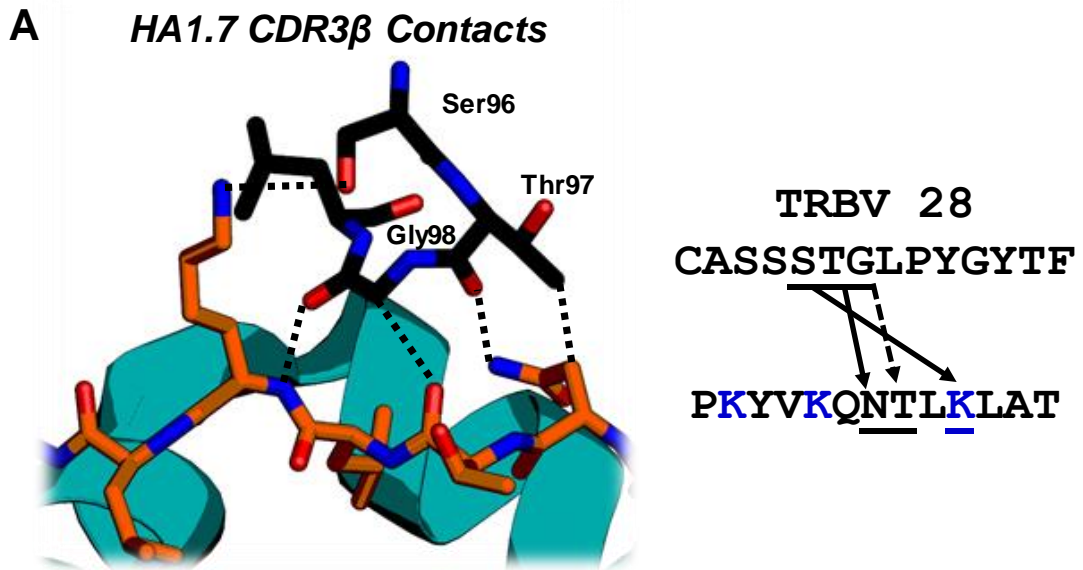


Figure 6.11. CDR3 β contacts with the peptide. (A) HA1.7 contacts with the sequences shown to the right. Amino acids involved in contact are shown underlined. Acidic residues in red and basic in blue. **(B)** F11 contacts, format as previous.

HA1.7 TCR Contacts							
TCR CDR3 β			Peptide/HLA Contact				
NO.	AA	ATOM	Contact	NO.	AA	ATOM	DIST
96	SER	OG [O]:	Peptide	315	LYS	NZ [N]	3.95
97	THR	OG1[O]:	DR1- α	61	ALA	CB [C]	3.42
97	THR	CG2[C]:	DR1- α	62	ASN	OD1[O]	3.69
97	THR	CG2[C]:	Peptide	310	LYS	NZ [N]	3.79
97	THR	O [O]:	Peptide	312	ASN	CB [C]	3.54
97	THR	CG2[C]:	Peptide	312	ASN	CB [C]	3.49
97	THR	O [O]:	Peptide	312	ASN	CG [C]	3.64
			Peptide	312	ASN	ND2[N]	2.89
			Peptide	313	THR	O [O]	3.24
97	THR	C [C]:	Peptide	313	THR	O [O]	3.72
98	GLY	N [N]:	DR1- α	65	VAL	CG2[C]	3.78
98	GLY	N [N]:	Peptide	313	THR	O [O]	3.71
98	GLY	CA [C]:	Peptide	313	THR	O [O]	3.13
			Peptide	314	LEU	CA [C]	3.96
98	GLY	O [O]:	Peptide	314	LEU	CA [C]	3.93
			Peptide	314	LEU	C [C]	3.98
98	GLY	CA [C]:	Peptide	315	LYS	N [N]	3.62
98	GLY	C [C]:	Peptide	315	LYS	N [N]	3.77
98	GLY	O [O]:	Peptide	315	LYS	N [N]	3.08
			Peptide	315	LYS	CA [C]	3.93
			Peptide	315	LYS	CB [C]	3.60

F11 TCR							
TCR CDR3 β			Peptide/HLA Contact				
NO.	AA	ATOM	Contact	NO.	AA	ATOM	DIST
94	GLU	O [O]	Peptide	7	ASN	CB [C]	3.18
			Peptide	7	ASN	CG [C]	3.37
94	GLU	C [C]	Peptide	7	ASN	ND2[N]	3.59
94	GLU	O [O]	Peptide	7	ASN	ND2[N]	2.65
94	GLU	CG [C]	Peptide	8	THR	O [O]	3.37
94	GLU	CB [C]	Peptide	8	THR	O [O]	3.31
94	GLU	CG [C]	Peptide	10	LYS	N [N]	3.95
94	GLU	OE1[O]	Peptide	10	LYS	N [N]	3.98
94	GLU	CD [C]	Peptide	10	LYS	CB [C]	3.91
94	GLU	OE1[O]	Peptide	10	LYS	CB [C]	3.78
96	TYR	CE1[C]	Peptide	7	ASN	ND2[N]	3.52

Table 6.6. CDR3 β contacts with peptide and HLA within 4.00 Å.

6.4.5 Summary

The responses to PKY were the second most diverse at the clonotypic level, but the least TRAV gene usage biased of the six epitopes studied. By comparing repertoire data with crystallographic information, we can identify the amino acid interactions that govern recognition and offer an explanation for TRAV and TRBV gene usage. If representative complex structures can be obtained for the other epitopes such as GMF, SGP and QAR, which did not display the same diversity and gene usage, then structural breakdown of their repertoire information can take place.

Further investigation will lead to a greater understanding of CD4+ T-cell responses in general, but will specifically shed light on the mechanistic basis by which these important T-cell repertoires can mediate protection from highly relevant acute infections like Influenza. This understanding can inform new research and become the basis for novel vaccination strategies that analyse protection at the level of the HLA and responding clonotypic T-cell repertoires.

6.5 Discussion

In this chapter, TCR-pHLA complex structures of the HA1.7 and F11 TCRs were compared in order to understand how conserved the TCR α -chain interactions were between the same TRAV gene in combination with distinct CDR3 α sequences and paired with different TRBV chains.

As expected, CDR2 α interactions were exclusively with the HLA, while CDR1 α residues side chains were within 4 Å contact distance of both the peptide and HLA, with the potential to form van der Waals and polar contacts. These interactions correlated with conserved CDR1 α sequence residues seen in the four TRAV genes that dominated clonotypes specific to PKY and partially explain the biased usage.

CDR3 α mediated contacts were consistent in nature regardless of loop length or overall amino acid composition. The identification of two salt bridges used by both HA1.7 and F11 TCRs explained the conserved presence of aspartic or glutamic acid residues in sequence motifs of the top TRAV genes identified by repertoire analysis. These strong interactions would be energetically favourable and contribute significantly to PKY peptide recognition.

The implications on further comparison of structural data and TCR repertoire information are positive, suggesting that observed CDR amino acid and gene usage are easily identified and extrapolated from TCR-pHLA complex data. Similar analysis of the other class-II epitopes identified in this study, especially those which exhibited highly biased gene usage (QAR and GMF) may identify the interactions that result in overwhelming preference of a single TRAV gene and striking CDR3 motif patterns (Fig. 6.12).

Could one or two complex structures with relevant TRAV genes explain these striking gene usage and CDR3 motif patterns?

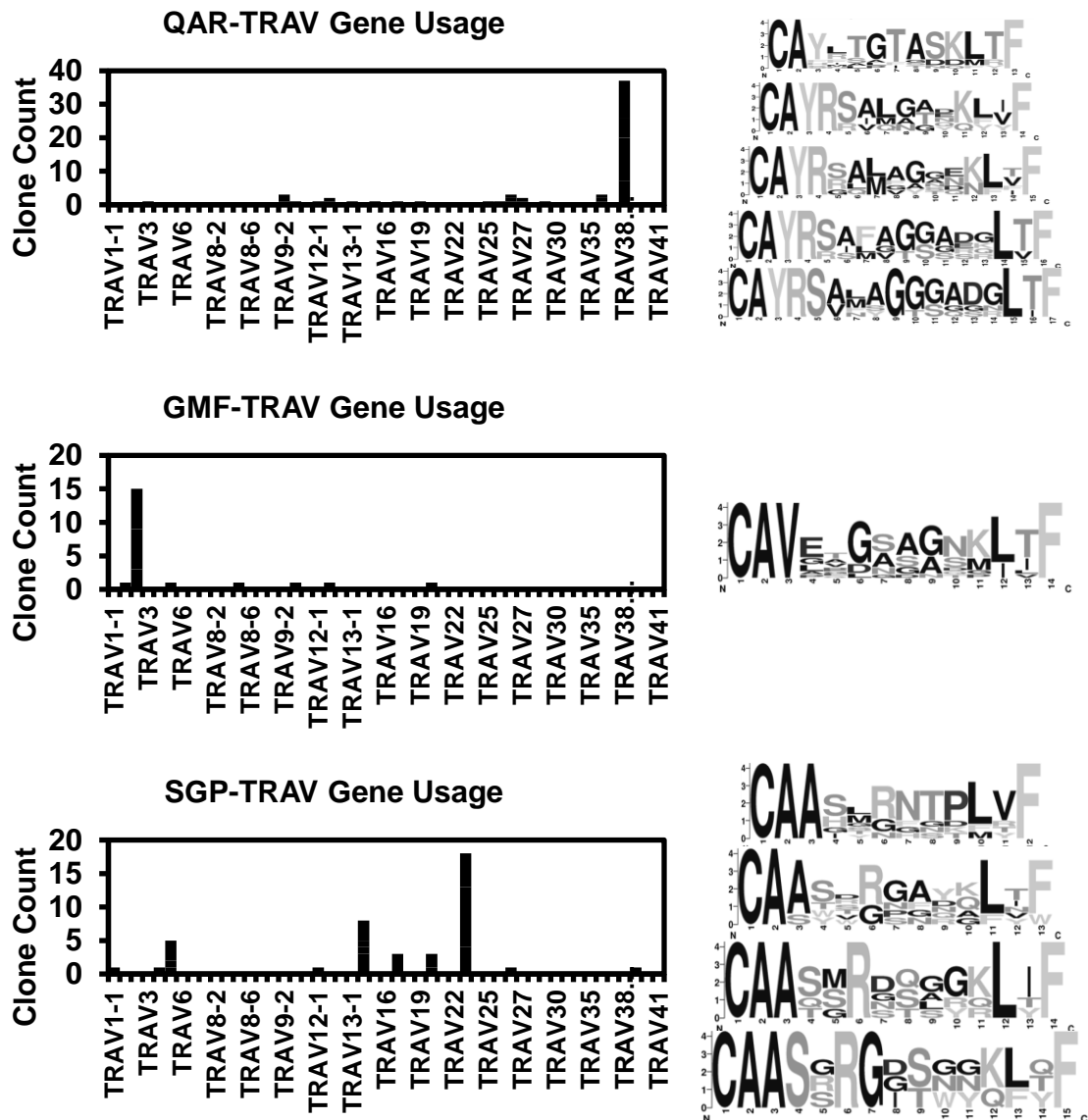


Figure 6.12. Examples of Gene usage and CDR3 motifs from the previous chapter, which may be understood by the generation and analysis of relevant TCR-pHLA complex structures. This involves acquisition of relevant paired TCR $\alpha\beta$ chain information and formation of a complex during crystallisation suitable of high resolution analysis.

6.5.1 Implications Beyond TRAV8-4 and TRAV8-6

Work in this chapter suggests that one or two complex structures may be sufficient to elucidate the mechanistic basis of highly conserved repertoire patterns related to dominant genes. The relevance of structural information on one TRAV gene when applied to different TRAV genes specific to the same epitope is a more difficult challenge.

In response to PKY, four TRAV genes were favoured in the repertoires of five HLA-DR1 donors. Of these, TRAV8-4 and TRAV8-6 are highly similar based on sequence comparison, while TRAV13-1 and TRAV14DV4 are relatively different (Table 6.7). The similarity means that the structural data analysed is most relevant to TRAV8-4 (F11 and HA1.7) and TRAV8-6 repertoires, as these TCRs are likely to form similar germline contacts and orientation.

Given the flexibility and variability of TCR CDR3 loops, peptide interactions in TRAV8-4 structures may be just as relevant to TRAV13-1 and TRAV14DV4 repertoire motifs, i.e. hypervariable regions are impossible to fully group by nature, regardless of their parent genes. Understanding of epitope-specific gene usage potentially mediated by germline encoded CDR1 peptide and HLA contacts requires structural information to support predictions. This has been the basis of a number of recent studies in the CD4+ T-cell responses to celiac epitopes at the clonal level^{167,284,295}.

Germline	CDR1	CDR2	CDR3
TRAV8-4	CNYSSSV.....PPY LFW	YTSA..ATLV	SDAAEYF CAVS
TRAV8-6	CNYSSSV.....SVY LFW	YLSG..STLV	SDTAEYF CAVS
TRAV13-1	CTYSDSA.....SNY FPW	IRSN...VGE	EDSAVYF CAAS
TRAV14/DV4	CTYDTSDP.....SYG LFW	QGSY..DQQN	GDSAMYF CAMRE

Table 6.7. Variation of germline sequences from TRAV8-4 (crystal structure) around the CDR loops of TRAV genes used to respond to PKY in five donors. TRAV8-4 is coloured in red, and conserved residues in other TRAV genes are coloured red to display sequence variations from the structural data.

6.5.2 Conservation of Salt Bridges over Weaker Contacts

When looking at extensive repertoire data, particularly CDR3 motifs, the extent to which weak interactions are conserved, and thus understood by limited structural information is not clear. Highly favoured salt bridge interactions are likely to be conserved and exhibit similar orientations when present in epitope-specific repertoires. Yet whether weaker non-covalent interactions necessarily exhibit strong positional rigidity relative to the peptide, or HLA, is harder to infer.

If a CDR3 α sequence and cognate peptide contain complementary salt bridging charged residues (for example in Fig. 6.8C), then the likelihood of an interaction is high and very energetically favourable. Hypothesising such complementary interactions from repertoire motifs may be valid. Yet as interactions get weaker, from non-polar hydrogen bonds, to hydrophobic then weak van der Waals, positional preferences are less defined (no clear positive to negative direction) and these may not be driving overall TCR recognition. Therefore, hypotheses based on weak interactions are less likely to be valid.

Certain repertoire motifs and amino acid conservations may require highly relevant structural models to be understood, due to the many potential energetic drivers behind the TCR-pHLA binding interaction. Where strong complementary forces are present, repertoire dynamics may be understood by a general or loosely related example.

6.5.3 Future Work

In addition to obtaining the corresponding TCR β -chain repertoire information, the next step is obtaining representative TCR-pHLA complex structures for the other epitopes identified in this project. Their analysis will enable understanding of corresponding TRAV and TRBV repertoire usage and patterns. Progress in such investigations is limited by the difficulty in generating complex structural data.

Structural analysis and understanding TCR repertoires has relevance in many areas of T-cell immunology. When key residues, in either the epitope or the TCR, that mediate recognition are identified, these can be altered in order to enhance the corresponding immune response for T-cell therapy and novel vaccination.

The ultimate goal is the accumulation of sufficient repertoire and corresponding structural information to accurately model the effects of changing TCR or epitope residues in immunotherapy. The binding algorithms used in previous chapters relied on existing information to attain accurate predictive capacity of a binary system. Modelling of ternary interaction systems such as the TCR, peptide and HLA is much more complex.

If the effects of amino acid or gene substitution on TCR-pHLA binding affinity could be accurately predicted, it would be a powerful tool for immunological research and therapeutic development. Given that the TCR recognition detailed in this chapter showed strong elements of conservation, modelling and predictive tools may be feasible, and within reach in coming years.

7 Discussion

In 2012, Wilkinson et al²¹³ found that CD4+ T-cell responses specific to conserved influenza proteins correlated with heterosubtypic protection against pandemic flu. In this project, several epitopes that mediate these responses in the HLA-DR1 population were identified and used to uncover the genetic and molecular basis of populations that provide protection.

In chapter 3, transduction of naked APCs with HLA-DR1 and their implementation on ELISpot assay was achieved. These DR1-APCs were used to successfully identify DR1-restricted responses to the conserved internal proteins in chapter 4. Structures of two of these epitopes were obtained and their analysis confirmed predictions from HLA binding algorithms and provided information on the likely amino acids that were contacted by cognate TCRs.

In chapter 5, these epitopes enabled further investigation of the CD4 immune response to the conserved proteins through HLA-multimer staining and clonotypic analysis. Strong epitope-specific responses were mediated by narrow TRAV gene usage, and conservation of amino acid properties across CDR3 sequences as well as some rare public clonotypes. The underlying nature of these responses implied a uniformity in the mechanisms of class-II mediated immunity to specific viral epitopes.

In chapter 6, Crystallographic analysis of DR1-PKY in complex with two distinct TCRs provided an insight into how clonotypic repertoire data could be aligned with structural analysis to rationalise gene usage and amino acid motifs. The implications of these results on current knowledge of CD4+ T-cell mediated immunity, potential vaccine relevance and application to other challenges will be discussed here.

7.1 The Basis of Heterosubtypic Protection?

In this study, the importance of CD4+ T-cells specific to conserved epitopes of the Influenza virus was exemplified by the detection of robust and reproducible responses in five healthy donors. Whether a result of infection or vaccination, these populations may have played a critical role in generation of an antibody response, as

well as priming of CD8+ populations to eradicate the virus⁶¹, in addition to the provision of innate and cytotoxic functions that have been documented as protective^{64,101}.

As the antigenic targets of these cells are highly conserved, it is likely they provide this protection regardless of challenge strain and thus may contribute to heterosubtypic immunity^{48,213}. Nearly every adult is expected to have encountered influenza at one time in their life, most likely during childhood²⁹⁹. Why some individuals are asymptomatic and others hospitalised is a mystery³⁰⁰, but it is likely to be related to important characteristics of their immune response.

Heterosubtypic immunity is known to correlate with the absence of an antibody response³⁰¹ and the presence of a CD8+ T-cell response to the conserved internal proteins⁴⁸. The recent observation that CD4+ T-cells also correlate with cross-protective immunity has cemented their positive role in our understanding of influenza infection²¹³. However, work by Andrea Sant suggests that CD4+ T-cells specific to internal proteins, as opposed to external proteins, *do not aid seroconversion* of the antibody repertoire to novel strains^{302,303} and are instead a barrier to some important aspects of immunity.

This is because T-cells specific to external proteins were poised toward a “T_{FH}-like” phenotype which best facilitates the generation of a neutralising antibody response, while those specific to NP were not. Instead NP specific cells were associated with a T_H1 (IFN- γ , CXCR3+) phenotype, which does not directly contribute to antibody production²⁴². In order to both clear virus and provide an immediate barrier to infection, this neutralising antibody response is essential.

In these papers, the authors state that conserved responses hinder CD4+ T-cell help towards the generation of a novel antibody response to shifted or drifted strains, and cites the negative impact of memory responses on the expansion of a naïve repertoire. Instead, DiPiazza et al propose the administration of vaccines that do not contain these internal elements and thus drive seroconversion unassisted by conserved responses²⁴². This is the basis of many “split vaccines”³⁰⁴, while live attenuated or inactivated vaccines still contain internal proteins (but not in a standard or quantified amount).

The Sant hypothesis clearly states the goal of a universal vaccine is to facilitate seroconversion to novel Influenza-A strains. The CD8+/CD4+ work^{48,213} and studies on S-Flu^{305,306} demand a different criterion, namely the reduction in symptom severity and prevention of host to host transmission respectively. Given that novel strains with the potential to cause pandemic may arise from poorly monitored sources, the sequences of upcoming HA proteins cannot be predicted with complete certainty.

Therefore, only the internal proteins offer the chance to provide heterosubtypic protection against all possible future strains. Antibody focused vaccine design always requires some foresight or predictive capacity. If it were possible to identify the coming pandemic and seasonal threats with complete certainty, then the Sant approach would be valid.

7.2 Implications for Vaccine Design

Downstream clinical applications that build on the work of this project may exploit responses to conserved class-II epitopes in order to establish heterosubtypic immunity. Whether CD4+ T-cell immunity can be preferentially boosted or induced, relies on the success of vaccines that explicitly target T-cells.

Bona fide T-cell vaccines involving modified peptides³⁰⁷, recombinant protein, or MVA^{222,308} and adenoviral vectors vaccines^{309,310} are relatively novel technologies, with several systems that are yet to complete clinical trials³⁰⁴. These vaccines diverge from the traditional aims of inducing strong antibody responses, and instead target CD8+ and CD4+ T-cell responses to provide cross-protective immunity.

In Influenza research, the novel S-Flu vaccine limits expression of the HA protein by the removal of its signal sequence and successfully elicits protection in pig and ferret models^{305,306}. Although this is not an exclusive T-cell vaccine, it lessens the focus on an HA specific antibody response and has been demonstrated to provide robust protection through T-cell mediated immunity.

If CD4+ T-cells can be successfully targeted, there is scope for even more advanced vaccination models that can alter the responding repertoire to class-II epitopes

and potentially created a better form of immunity²⁸⁵. This involves modifying amino acids outside of the core of an epitope, and vaccinating against the modified peptide in order to subtly alter the expanded T-cell pool which recognises the original sequence¹⁶⁸.

The preliminary work for such an investigation, i.e. identification of HLA-restricted epitopes, understanding of HLA binding and assessment of wild type responses, has been described in this project. What is unknown is what kind of repertoire we would want to create in order to best provide protective immunity.

7.3 Broad or Narrow Repertoires, What's Best for Protection?

Questions on the kind of T-cell repertoire which provides the best possible protective capacity from disease are likely to become increasingly important as the prevalence of epitope-specific repertoire data increases. Simply, is a highly focused repertoire composed of similar TCRs in terms of gene usage and CDR3 sequence more effective at eliminating pathogen than a broader, more genetically/biochemically diverse, TCR repertoire?

This has been answered, in part, in HIV infection where highly skewed repertoires containing public TCR sequences and conserved CDR3 motifs in multiple patients are thought to play an important role in controlling chronic infection²⁸⁶. A focused CD4+ T-cell repertoire, like that observed in response to some epitopes detailed here, appears to be beneficial.

In this project, skewing of TRAV gene usage in response to three epitopes was striking, and in some cases this was linked to highly conserved CDR3 amino acid usage. The remaining three epitopes exhibited slightly broader gene usage and still showed CDR3 conservation.

In terms of clonal diversity (pie charts, chapter 5) some responses were dominated by one or two TCR sequences, while others had much more diverse hierarchies. Therefore, in response to influenza, we can extrapolate that we have a spectrum of breadth and diversity depending on the epitope and donor in question. Where, on this spectrum, do we want our TCR repertoires to be?

Is it possible that focused repertoires are more “immunologically efficient,” requiring a smaller pool of clones to respond to, and ultimately control infection? A larger pool of clones, with a variation in TCR binding affinity and overall efficacy, may show some levels of redundancy, ultimately detracting from the responses of more effective clones.

Do we have an exhaustible supply of immune resources, i.e. B-cells, CD8+ T-cells and APCs, and therefore we want to limit their use to the most effective action only? Certainly, this would be the case in the immunocompromised, i.e. HIV with a reduced CD4+ compartment, but to the healthy population, would artificial focusing of the repertoire hold potentially catastrophic consequences?

With the wealth of pathogenic diversity, does some pre-existing, albeit less efficient, cross-reactive, or antigenically sinful response provide some level of protection? Does the initial response, no matter the efficiency, give us time to nullify a new threat before the more efficient clones eventually take over? A broad repertoire may be capable of tolerating the many potential mutations that leads to viral evolution, and thus will not be compromised in response to novel strains.

If it is possible to significantly alter repertoire breadth by intelligent vaccine design, then this will warrant serious consideration of both the coverage and gaps such changes may leave.

7.4 Application of Findings to Other Challenges

This project has dealt with CD4+ T-cell responses that were most likely forged during acute viral infection and possibly vaccination. Do the observations on response magnitudes and repertoire characteristic detailed here, have relevance in other disease settings where CD4+ T-cells may play a protective or pathogenic role?

What do the responses to self, chronic or allergic antigens look like in terms of the cellular magnitude and repertoire architecture? Chronic disease may shape the diversity and number of responding clones in a very unique manner. In the presence of

persistent and/or self-antigen, factors including T-cell exhaustion, the impact of regulatory T-cells and an absence of high-affinity TCRs due to thymic selection will complicate the response.

If the methodology of this project was applied in the cancer, autoimmune or chronic viral setting then the resulting cellular and repertoire information would be severely limited. This is because expanded epitope-specific cells may not show high avidity for self or persistent antigen, they may be exhausted and thus fail to expand or produce cytokine, as well as have more complex cytokine profiles than simply T_H1 IFN- γ . Specifically with CD4⁺ T-cells, the repertoire isolated may contain a mixture of inflammatory and regulatory clones, both of which will have very different roles in pathology.

As a result, any number of potential scenarios can be envisaged and it is unlikely that the data will be easily aligned with a true understanding of the disease in question. Instead, when dissecting the roles of CD4⁺ T-cells in cancer or chronic infection methods, hypotheses, and questions will have need to carefully considered and will likely be much more advanced than those presented in this project.

7.5 Future Work

Following the completion of β -chain analysis and isolation of paired $\alpha\beta$ -TCR sequence information for structural complex analysis, some wider questions could be posed of the T-cell repertoires documented in this project.

To confirm their protective nature, further investigation would involve looking at *ex vivo* responses pre- and post- vaccination, with a vaccine that includes the internal proteins (i.e. not a split vaccine). If these responses are significantly raised, it would indicate their importance in active immunity to processed virus.

The use of these epitopes as diagnostic markers of CD4⁺ T-cell populations in ongoing T-cell vaccine trials could be an excellent clinical application, moving away from the focus on A2-GIL responses as the surrogate marker of cell mediated immunity.

Finally, the next generation of vaccines may involve subtle alteration of epitopes in order to induce highly protective changes to induced immunity at the repertoire level¹⁶⁸. This differs from heteroclitic peptide vaccinations in that the “core” of the epitope remains unchanged, while its surroundings, the flanking residues, are altered to produce the desired immunological effect.

Although highly ambitious, such work may be the future of both vaccination against infectious disease and cancer. We can hope to tailor our pre-existing T-cell populations to induce highly focused protection with minimal off-target or antigenically sinful effects.

8 Appendix

8.1 Details of Peptides

FluA PB1

1. MDVNPTLLFLKVPQAQNAIST
2. KVPAQNAISTTFPYTGDPFY
3. TFPYTGDPYSHGTGTGYTM
4. SHGTGTGYTMDTVNRTHQYS
5. DTVNRTHQYSEKGRWTTNTE
6. EKGRWTTNTEGTAPQLNPID
7. TGAPQLNPIDGPLPEDNEPS
8. PLPEDNEPSGYAQTDCVLEA
9. YAQTDCVLEAMAFLEESHYPG
10. MAFLEESHYPGIFENSCIETM
11. IFENSCIETMEVVQQTRVDK
12. EVVQQTRVDKLTQGRQTYDW
13. LTQGRQTYDWTLLNRNQPAAT
14. TLLNRNQPAATALANTIEVFR
15. ALANTIEVFRSNGLTANESG
16. SNGLTANESGRLIDFLKDVM
17. RLIDFLKDVME SMKKEEMGI
18. ESMKKEEMGITTHFQRKRRV
19. TTHFQRKRRVRDNMTKKMIT
20. RDNMTKKMITQRTIGKKKQR
21. QRTIGKKKQRLNKR SYLIRA
22. LNKRSYLIRALTLNMTKDA
23. LTLNMTKDAERGKLRRAI
24. ERGKLRRAIATPGMQIRGF
25. ATPGMQIRGFVYFVETLARS
26. VYFVETLARSICEKLEQSGL
27. ICEKLEQSGLPVGGNEKKAK
28. VGGNEKKAKLANVVRKMMTN
29. NVVRKMMTNSQDTELSFTIT
30. QDTELSFTITGDNTKWNENQ

31 . GDNTKWENQNPFRMFLAMIT
32 . NPRMFLAMITYMTRNQPEWF
33 . YMTRNQPEWFRNVLSIAPIM
34 . NVLSIAPIMFSNKMARLGKG
35 . SNKMARLGKGYMFESKSMKL
36 . YMFESKSMKLRTQI PAEMLA
37 . RTQIPAEMLASIDLKYNDS
38 . SIDLKYFNDRKKEKIRP
39 . TRKKEKIRPLLIETASLS
40 . LLIEGTASLSPGMMMGFMN
41 . PGMMMGFMNMLSTVLGVSIL
42 . LSTVLGVSILNLGQKRYTKT
43 . NLGQKRYTKTTYWWDGLQSS
44 . TYWWDGLQSSDDFALIVNAP
45 . DDFALIVNAPNHEGIQAGVD
46 . NHEGIQAGVDRFYRTCKLLG
47 . RFYRTCKLLGINMSKKKSYI
48 . INMSKKKSYINRTGTFFETS
49 . NRTGTFFETSFFYRYGFVAN
50 . FFYRYGFVANFSMELPSFGV
51 . FSMELPSFGVSGINESADMS
52 . SGINESADMSIGVTVIKNNM
53 . IGVTVIKNNMINNDLGPATA
54 . INNDLGPATAQMALQLFIKD
55 . QMALQLFIKDYRYTYRCHRG
56 . YRYTYRCHRGDTQIQTRRSF
57 . DTQIQTRRSFEIKKLWEQTR
58 . EIKKLWEQTRSKAGLLVSDG
59 . SKAGLLVSDGGPNLYNIRNL
60 . GPNLYNIRNLHIPEVCLKWE
61 . HIPEVCLKWELMDEDYQGR
62 . LMDEDYQGRLCNPLNPFVSH
63 . CNPLNPFVSHKEIESMNAV
64 . KEIESMNAVMPAHGPAKN

65 . MMPAHGPAKNMEYDAVATTH
66 . MEYDAVATTHSWIPKRNRSI
67 . SWIPKRNRSILNTSQRGVLE
68 . LNTSQRGVLEDEQMYQRCCN
69 . DEQMYQRCCNLFKFFPSSS
70 . LFEKFFPSSSYRRPVGISSM
71 . YRRPVGISSMVEAMVSRARI
72 . VEAMVSRARIDARIDFESGR
73 . DARIDFESGRIKKEEFTEIM
74 . EEFTEIMKICSTIEELRRQK

74 peptides

FluA Nucleoprotein

1. GRMVGIGRIFYIQMCTELKL
2. YIQMCTELKLTDYEGRLIQN
3. TDYEGRLIQNSITIERMVL
4. SITIERMVLSAFDERRNRYL
5. AFDERRNRYLEEHP
6. EEHP
7. KKTGGPIYRRRDGKVVRELI
8. RDGKVVRELILYDKEEIRRI
9. LYDKEEIRRIWRQANNGEDA
10. WRQANNGEDATAGLTHLMIW
11. TAGLTHLMIWHSNLNDATYQ
12. HSNLNDATYQRTRALVRTGM
13. RTRALVRTGMDPRMCSLMQG
14. DPRMCSLMQGSTLPRRSGAA
15. STLPRRSGAAGAAVKGVGTM
16. GAAVKGVGTMVMEIIRMIKR
17. VMEIIRMIKRGINDRNFWRG
18. GINDRNFWRGNGRRTRVAY
19. ENGRTRVAYERMCNILKGG
20. ERMCNILKGGFQTAQRAMM
21. FQTAQRAMMDQVRESRNP
22. DQVRESRNPNAEIEDLIFL
23. NAEIEDLIFLARSALILRGS
24. ARSALILRGSVAHKSCLPAC
25. VAHKSCLPACVYGLAVASGY
26. VYGLAVASGYDFEREGYSLV
27. DFEREGYSLVGIDPFRLQ
28. GIDPFRLQNSQVFSLIRPN
29. SQVFSLIRPNENPAHKSQ
30. ENPAHKSQLVWMACHSAAFE
31. WMACHSAAFEDLRVSSFIRG
32. DLRVSSFIRGTRVVPRGQLS

33. TRVVPRGQLSTRGVQIASNE
34. TRGVQIASNENMEAMDSNTL
35. NMEAMDSNTLELRSRYWAIR
36. ELRSRYWAIRTRSGGNTNQQ
37. TRSGGNTNQQRASAGQISVQ
38. RASAGQISVQPTFSVQRNLPFER
39. QPTFSVQRNLPFERATIMAAFTG

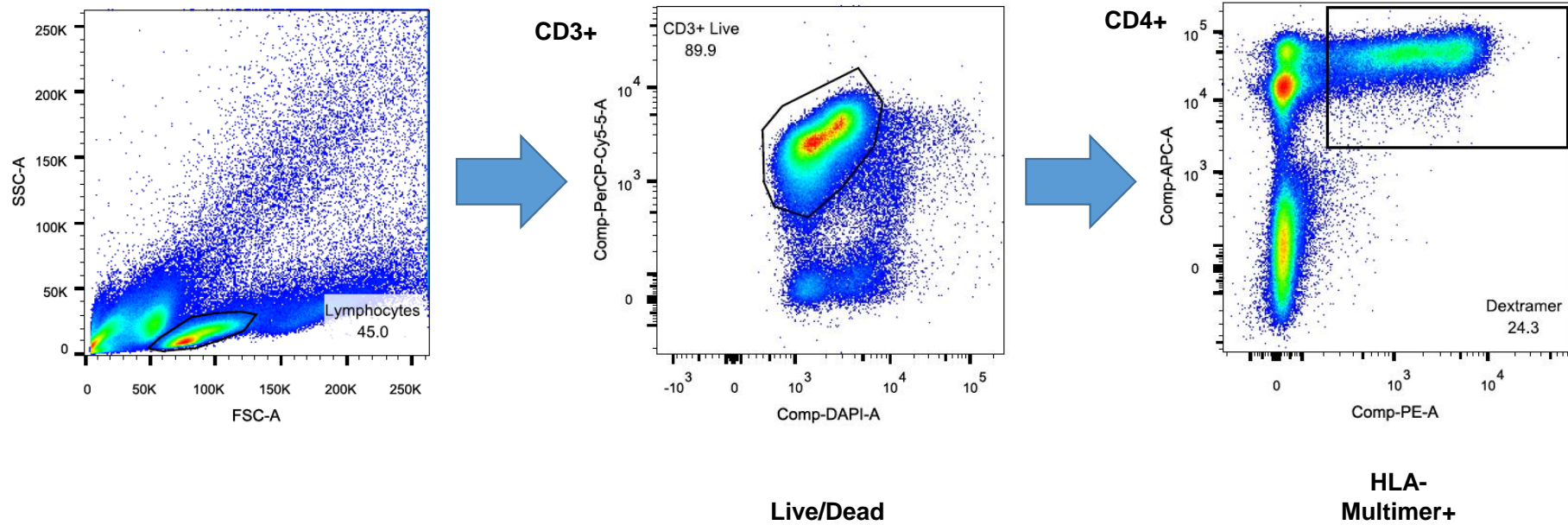
39 peptides

FluA Matirx1

1. MSLLTEVETYVLSIVPSGPL
2. VLSIVPSGPLKAEIAQRLED
3. KAEIAQRLEDVFAGKNTDLE
4. VFAGKNTDLEVLMEWLKTRP
5. VLMEWLKTRPILSPLTKGIL
6. ILSPLTKGILGFVFTLTVPS
7. GFVFTLTVPSERGLQRRRFV
8. ERGLQRRRFVQNALNGNDP
9. QNALNGNDPNNDKAVKLY
10. NNMDKAVKLYRKLKREITFH
11. RKLKREITFHGAKEIALSYS
12. GAKEIALSYSAGALASCMGL
13. AGALASCMGLIYNRMGAVTT
14. IYNRMGAVTTEVAFGLVCAT
15. EVAFGLVCATCEQIADSQHR
16. CEQIADSQHRSHRQMVTTN
17. SHRQMVTTNPLIRHENRMV
18. PLIRHENRMVLASTTAKAME
19. LASTTAKAMEQMAGSSEQAA
20. QMAGSSEQAAEAMDIASQAR
21. EAMDIASQARQMVQAMRTIG
22. QMVQAMRTIGTHPSSSAGLK
23. THPSSSAGLKDDLLENLQAY
24. DDLLENLQAYQKRMGVQMRFK

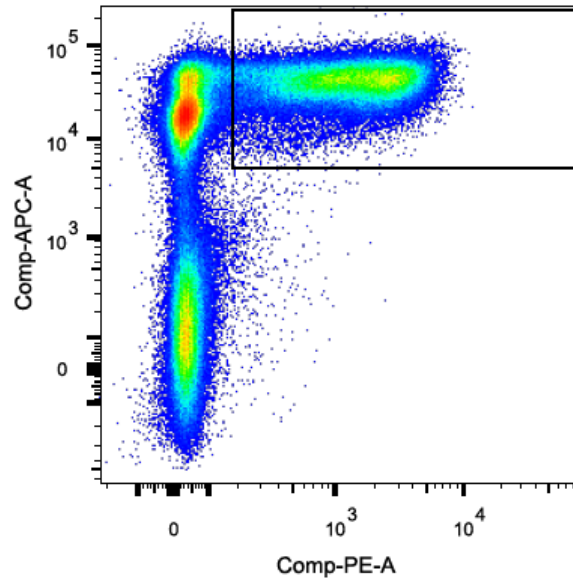
24 peptides

8.2 Flow Cytometry & Sorting Data

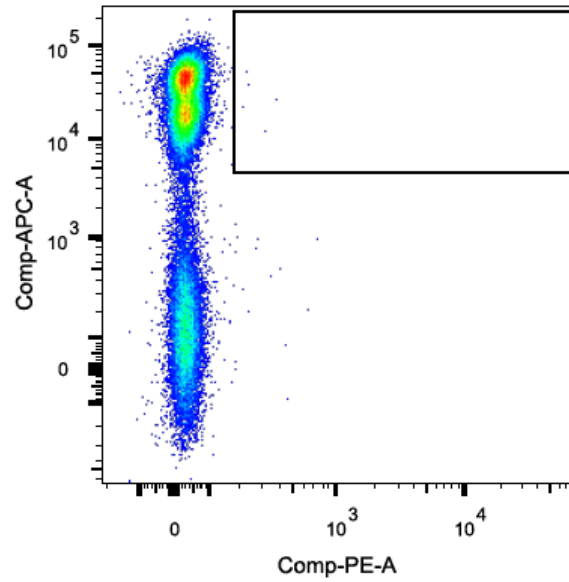


Gating strategy used for analysis of flow cytometry data in chapter 5.

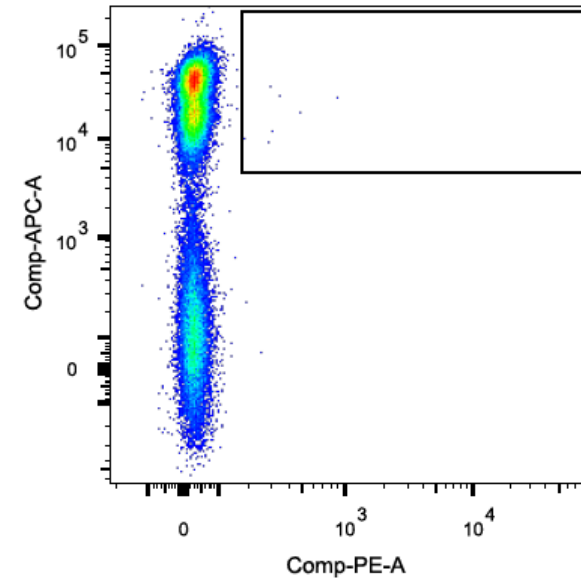
Test- Cognate HLA-Multimer



Irrelevant class-II HLA-multimer

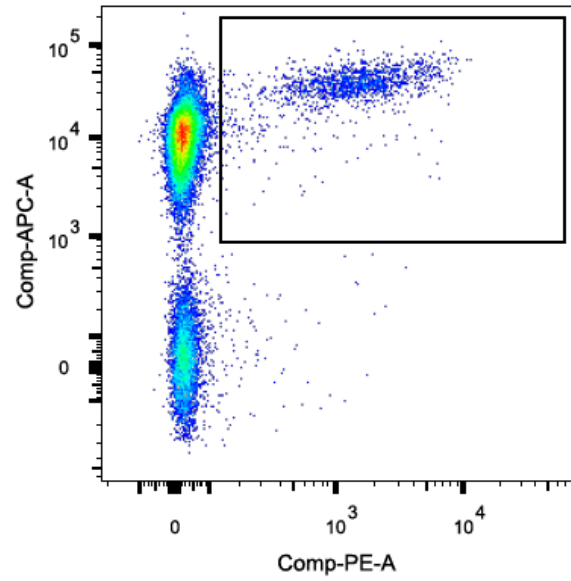


FMO

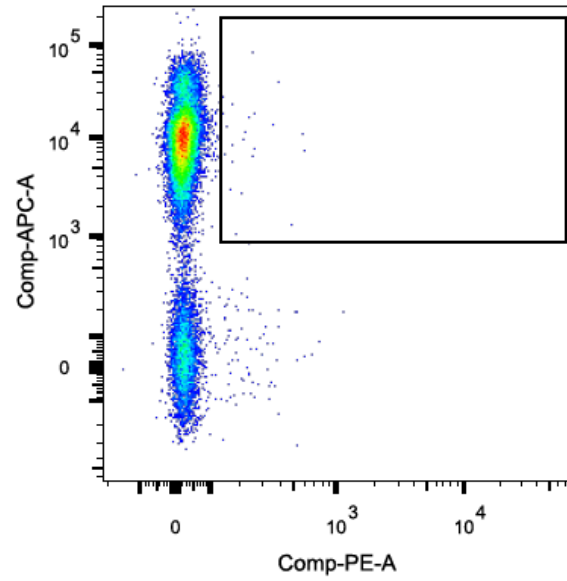


Example 1 of HLA-multimer staining and corresponding irrelevant and FMO controls

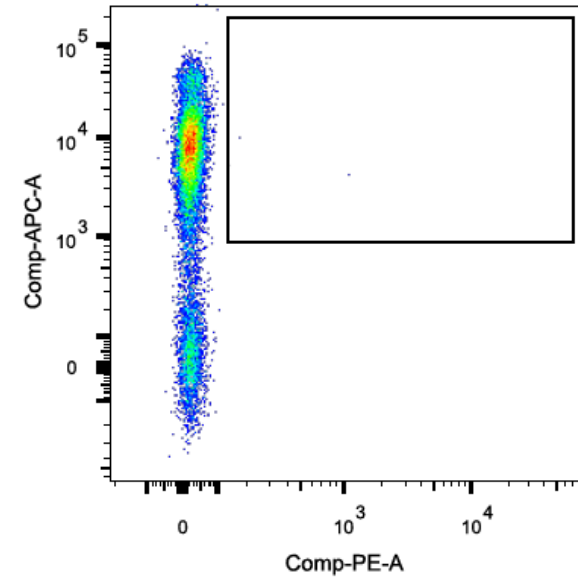
Test- Cognate HLA-Multimer



Irrelevant class-II HLA-multimer

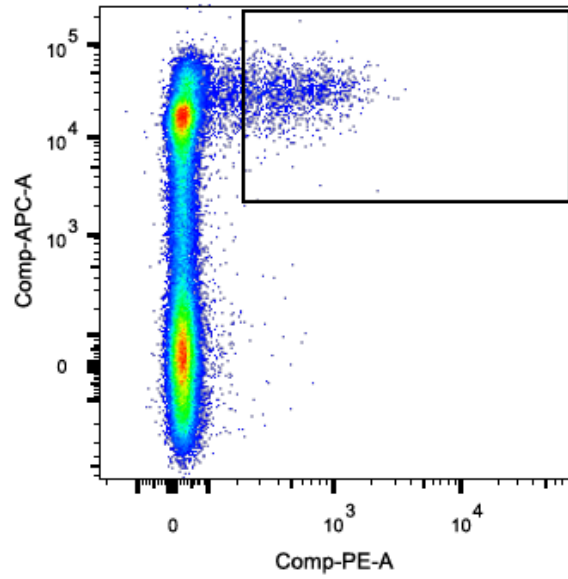


FMO

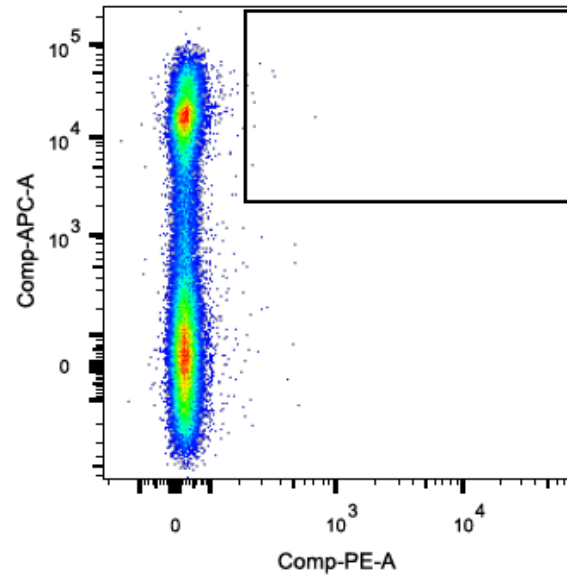


Example 2 of HLA-multimer staining and corresponding irrelevant and FMO controls

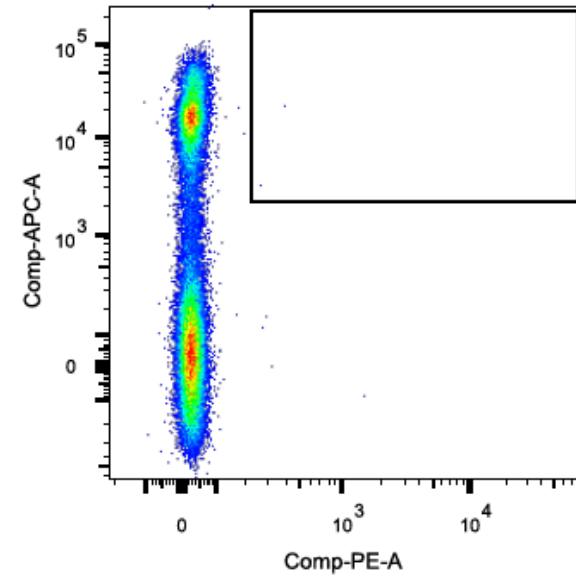
**Test- Cognate HLA-
Multimer**



**Irrelevant class-II
HLA-multimer**



FMO



Example 3 of HLA-multimer staining and corresponding irrelevant and FMO controls

Donor	Bleed (Date)	SGPLK	GLIYN	QARQMV	DPF	GMF	PKY
3	1 (June-2016)	594	1,400	1,570	236	870	589
	2 (July-2016)	2,132	9,515	9,436	1,291	1,584	3,433
1	1 (June-2016)	22,245	37,404	6,355	8,567	0 (n=2)	24,804
2	1 (March-2016)	1,900	15,000	10,758	2,078	1,000	16,859
	2 (Dec-2015)	352		364	10	27	
4	1 (June-2016)	58	1645	68	19	82	2,600
5	1 (March-2016)	40,309	78,000	40,000	47,324	56,249	39,039
	2 (June-2016)	47,441	76,053	18,403	12,782	18,228	28,396

Number of cells isolated from cell sorting and taken forward for clonotypic analysis Boxes in green were analysed, yellow ere not.

8.3 Application of the 50 read threshold cut off

DONOR-2					
SGP	GLIYN	QAR	DPF	GMF	PKY
CAENMKDTGRRALTF	CALQGGSEKLVF	CAYLTGTASKLTF	CA YRSTGNQFYF	CAVEDGSDGQKLLF	CALSNDYKLSF
CAVDMKGVVWYFNKIFY	CAVEDTNSGYALNF	CAYLTGTASKLTF	CAESMEAAGNKLTF	CVVSAGEGRDDKIIF	CLVGDMSNGYSTLTF
CAASMRHSTLTF	CALFTGGGNKLTf	CAYRRALFGNEKLTf	CAMREGPGTGGFKTIF	CAVPLGAAAGNKLTF	CAMSA TDSWGKLOF
CATDAGSWGKLOF	CALSEVNNNAGNMLTF	CAYRSSMAGGADGLTF	CAVRDPDYGNRLAF	CAVEKSQKLLF	CAMSP TDSWGKLOF
CAESMHGQKLLF	CALSEANA GNMLTF	CASLTGTASKLTF	CAVQNYGGA TNKLIF	CAVLLGSSASKIIF	CAASFSDGQKLLF
CVVSGFGNVLHC	CVVSEANSGYALNF	CAHFTGTASKLTF	CALNTGNQFYF	CAVEGDNA GNKLTF	CAASGYSTLTF
CILRDVRRNTGTASKLTF	CALSARNSGNTPLVF	CAYRSAISDDMRF	CA YRSVRNQFYF	CAGGGSNY QLIW	CAVSGQSGQGNLIF
CAASWRPGYALNF	CAVSLLTSYDKVIF	CAGAAAAGNNRKLW	CAARVFGNEKLTf	CAVPLGAAAGNMLTF	CAVSDNDYKLSF
CATDARRYGGA TNKLIF	CAVYNTNAGKSTF	CAVGGYSTLTF	CALSDFHGNQGGKLIIF	CAV TTSYGNRLAF	CILLFGNEKLTf
CAASSRYNQGGKLIIF	CAVIGTGRRALTF	CAAISQGGSEKLVF	CA YRSARSQGNLIF	CAVRRWSNTGKLIIF	CAVSGNGANNLFF
CAASV RGN YQLIW	CILLFGNEKLTf	CAVHYNNNDMRF	CAVYTGANSKLTf	CAMREGPGTGGFKTIF	CAVSESTDKLIIF
CAAIYNFNKIFY	CALSPANSGNTPLVF	CATTNSGYALNF	CATDGEGGGADGLTF		CALVASGGSYIPTF
CAAKSNA GGTSYGKLTf	CIVRVGLQGAQKLVF	CAVSDNYGQNFVF	CILRDWNYGGSQGNLIF		CAVFLAGTALIF
CATDEGSWGKLOF	CILKSAAGGTSYGKLTf	CAYLTGTASKLTF	CAVEQNTGFQKLVF		CALSEGTSYDKVIF
CAASMRDSSYKLIIF	CAV PNFNGNEKLTf	CVVNSNTGGFKTIF	CAMRERN TDKLIIF		CAVSEPGANNLFF
CAVEGGA TNKLIF	CAVPLGGTSYGKLTf	CAYLTGTASKLTF	CIVRAKGSYNQGGKLIIF		CAMSGTDSWGKLOF
CAAHLRNTPLVF	CAFMRYNAGNMLTF	CILRATSDYKLSF	CATDKNTGKLIIF		CAL E VSNFNGNEKLTf
CAASSRV TGGGNKLTf	CA YRSVSGGADGLTF	CA YRSAMAGNQFYF	CAVDMRGGADGLTF		CAVERQGAQKLVF
CAAMFRGSRLTF	CAFMPKPPWGTDKLIIF	CA YRSAMAYGQNFVF	CAVNEVSDGQKLLF		CAVTRFSDGQKLLF
CAGEVSRGNQFYF	CAVRPLTSSGRLTF	CALTANTDKLIIF	CYEASHSVNTGTASKLTF		CAATGYSTLTF
CAASKGTSYGKLTf	CALSGSGNQFYF	CAYRRIQGAQKLVF			CAVSDSGNSGYALNF
CALSFRDSTGLRLYF	CAVEAGGGNKLTF	CILKSAAGGTSYGKLTf			CAETSRSRLTF
CAASPPYGNRLAF	CALWVGTNKLIF	CAWTPVGSQGNLIF			CAV VNSGGSNYKLTf
CAPPGNNDMRF	CALRDGYNKLIF	CARLTGTASKLTF			CAMSSGYSSAKIIF
CAASWRHSTLTF	CAARGQSGQGNLIF	CA YRSALNSYKLIIF			CAAYSGGYNKLIF
CLVGSADYSSNTGKLIIF	CAVLPGTYKYIF	CA YRSALGTDKLIIF			CAVSETGANNLFF
CASNSGNTPLVF	CAVEGITQGGSEKLVF	CAGARDSSYKLIIF			CATERAWVTGGGNKLTf
CIVRVGGNKLVF	CYEASHSVNTGTASKLTF	CAAVNDYKLSF			CALTRRDSWGKLOF
CAMRERGGGGADGLTF	CAV KSYGQNFVF	CA YRSGFATGNQFYF			CILKSAAGGTSYGKLTf
CAVERSTGGFKTIF	CIVRVGLQGAQKLVF	CAYIAIQGAQKLVF			CAFSYSSASKIIF
	CAVERSTGGFKTIF	CAYRSAFTGNQFYF			CAVSVFDSWGKLOF
	CIVRVSYSGGADGLTF	CAYRSALGGQLTf			CA YHMEYGNKLVF
	CAVSGRDGGA TNKLIF	CAYRKIMGGADGLTF			CAVKDSGTYKYIF
	CAASRNSGGSNYKLTf	CAVRPLTSSGRLTF			CAVNVPPGNQFYF
	CAVQADSWGKLOF	CAASMRGGNKLTF			CAMSIYNQGGKLIIF
	CAVGMNSGYSTLTF	CALGTLQGAQKLVF			CAVFSGGYNKLIF
	CALQGGGNKLTf	CAGHPDYKLSF			CAVYSGGYQKVTf
	CAVNARGTGGFKTIF	CAYRRALF			CILSDLISFNKLTf
		CAEGGNRLAF			CVVSDSVSGGYNKLIF
		CA YRSAMAGGADGLTF			CAFMINRQTGANLFF
		CAVSGRDGGA TNKLIF			CALSEAMDSNYQLIW
		CAASRNSGGSNYKLTf			CAVNTGGGNKLTf
		CIVRVGLQGAQKLVF			CAVPGGVNTDKLIIF
		CVVSDNYGQNFVF			CLVGDWNNNARLMF
		CTISNFGNEKLTf			CAVSDGGNEKLTf
		CAVEITGKLIIF			CLVKGSAKIIF
					CALDFMDSNYQLIW
					CAASGGGSNYKLTf
					CAVKGSGGSYIPTF
					CAVSGTSYGKLTf
					CAVSRGSWGKLOF
					CAASYNTDKLIIF
					CALSQGGKLIIF
					CAVSGRDGGA TNKLIF
					CAASRNSGGSNYKLTf
					CALSPQGYNTDKLIIF
					CALERDSGYSTLTF

Example of Donor-2 clonotypic sequences in the absence of any read cut off value.

Where a sequence is highlighted in blue or grey, this means that it is present multiple times in the across the sequences in the spreadsheet. As each column represents an epitope-specific population, there should not be overlap between populations. Yet due to the inclusion of low frequency samples below 50 reads, many sequences are present that overlap multiple samples.

See next page.

DONOR-2					
SGP	GLIYN	QAR	DPF	GMF	PKY
CAENMKDTGRRALTF	CALQGGSEKLVF	CAYLTGTASKLTF	CAYRSTGNQFYF	CAVEDGSDGQKLLF	CALSNDYKLSF
CAVDMKGVWYNFNKIFYF	CAVEDTNSGYALNF	CAYLTGTASKLTF	CAESMEAAGNKLTF	CVVSAGEGRDDKIIF	CLVGDMINSGYSTLTF
CAASMRHSTLTF	CALFTGGGNKLT	CAYRRALFGNEKLT	CAMREGPGTGGFKTIF	CAVPLGAAAGNKLT	CAMSA TDSWGKLOF
CATDAGSWGKLOF	CALSEVNNNAGNMLTF	CAYRSSMAGGADGLTF	CAVRDPDYGNRLAF	CAVEKSQKLLF	CAMSP TDSWGKLOF
CAESMHGQKLLF	CALSEANA GNMLTF	CASLTGTASKLTF	CAVQNYGGATNKLIF		CAASFSDGQKLLF
CVVSGFGNVLHC	CVVSEANSGYALNF	CAHFTGTASKLTF	CAL TNTGNQFYF		CAASGYSTLTF
CILRDVRRNTGTASKLTF	CALSARNSGNTPLVF	CAYRSAISDDMRF	CAYRSVRNQFYF		CAVSGQGSQGNLIF
CAASWRPGYALNF	CAVSDLLTSYDKVIF	CAGAAAAGNQRKLIW	CAARVFGNEKLT		CAVSDNDYKLSF
CATDARRYGGATNKLIF	CAVYNTNAGKSTF	CAVGGYSTLTF			CILLFGNEKLT
CAASSRIYNOGGKLI	CAVIGTGRRALTF	CAAISQGGSEKLVF			CAVSGNGANLFF
CAASVRGNYQLIW	CILLFGNEKLT	CAVHYNNNDMRF			CAVSESTDKLI
CAAIYNFNKIFYF	CALSPANSGNTPLVF	CATTNSGYALNF			CALVASGGSYIPTF
CAAASNAAGGTSYGKLT	CIVRVGLQGAQKLVF	CAVSDNYGQNFVF			CAVFLAGTALIF
CATDEGSWGKLOF	CILKSAGGTSYGKLT	CAYLTGTASKLTF			
CAASMRDSSYKLI	CAVPNFGNEKLT	CVVNSNTGGFKTIF			
CAVEGGATNKLIF	CAVPLGGTSYGKLT	CAYLTGTASKLTF			
CAHLRNTPLVF	CAFMRYNAGNMLTF	CILRATSDYKLSF			
CAASSRV TGGGNKLT	CAYRSVSGGADGLTF	CAYRSAMAGNQFYF			
		CAYRSAMAYGQNFVF			
		CALTANTDKLI			
		CAYRRIQGAQKLVF			

All low frequency sequences below 50 reads were removed. Only two sequences are present that overlap between epitope-specific populations. This justifies the use of a 50 read cut off to clean up data and remove the presence of promiscuous sequences, the biological explanation of which is yet to be determined.

8.4 CDR2 β Contact Tables

HA1.7 TCR Contacts							
TCR CDR2 β			Peptide/HLA Contact				
NO.	AA	ATOM	Contact	NO.	AA	ATOM	DIST
48	PHE	CZ [C]:	DR1- α	57	GLN	CD [C]	3.82
			DR1- α	57	GLN	OE1[O]	3.62
			DR1- α	57	GLN	NE2[N]	3.75
50	TYR	CE1[C]:	DR1- α	57	GLN	O [O]	3.92
50	TYR	CD1[C]:	DR1- α	57	GLN	O [O]	3.45
50	TYR	CE1[C]:	DR1- α	57	GLN	CB [C]	3.95
50	TYR	CG [C]:	DR1- α	61	ALA	CB [C]	3.91
50	TYR	CD1[C]:	DR1- α	61	ALA	CB [C]	3.72
50	TYR	O [O]:	DR1- α	64	ALA	CB [C]	3.36
51	ASP	OD1[O]:	DR1- α	64	ALA	CA [C]	3.81
51	ASP	OD2[O]:	DR1- α	64	ALA	CB [C]	3.78
51	ASP	CB [C]:	DR1- α	64	ALA	CB [C]	3.59
51	ASP	CG [C]:	DR1- α	64	ALA	CB [C]	3.40
51	ASP	OD1[O]:	DR1- α	64	ALA	CB [C]	3.57
			DR1- α	67	LYS	CD [C]	3.44
			DR1- α	67	LYS	CE [C]	3.33
51	ASP	CG [C]:	DR1- α	67	LYS	NZ [N]	3.82
51	ASP	OD1[O]:	DR1- α	67	LYS	NZ [N]	2.67
54	MET	CG [C]:	DR1- α	39	LYS	NZ [N]	3.75
55	LYS	O [O]:	DR1- α	39	LYS	CE [C]	3.97
55	LYS	C [C]:	DR1- α	39	LYS	NZ [N]	3.63
55	LYS	O [O]:	DR1- α	39	LYS	NZ [N]	2.61
55	LYS	N [N]:	DR1- α	39	LYS	NZ [N]	3.74
56	GLU	OE2[O]:	DR1- α	39	LYS	CG [C]	3.99
56	GLU	CD [C]:	DR1- α	39	LYS	CD [C]	3.99
56	GLU	OE2[O]:	DR1- α	39	LYS	CD [C]	2.84
			DR1- α	39	LYS	CE [C]	3.63
56	GLU	CG [C]:	DR1- α	39	LYS	NZ [N]	3.79
56	GLU	CD [C]:	DR1- α	39	LYS	NZ [N]	3.76
56	GLU	OE2[O]:	DR1- α	39	LYS	NZ [N]	3.22
56	GLU	CD [C]:	DR1- α	57	GLN	CG [C]	3.97
56	GLU	CG [C]:	DR1- α	57	GLN	NE2[N]	3.98
56	GLU	CD [C]:	DR1- α	57	GLN	NE2[N]	3.52
56	GLU	OE1[O]:	DR1- α	57	GLN	NE2[N]	3.09

F11 TCR							
TCR CDR2 β			Peptide/HLA Contact				
NO.	AA	ATOM	Contact	NO.	AA	ATOM	DIST
46	TYR	OH [O]	DR1- α	57	GLN	CD [C]	3.95
			DR1- α	57	GLN	NE2[N]	2.98
48	PHE	CD2[C]	DR1- α	57	GLN	O [O]	3.36
48	PHE	CE2[C]	DR1- α	57	GLN	O [O]	3.57
48	PHE	O [O]	DR1- α	60	LEU	O [O]	4.00
48	PHE	CB [C]	DR1- α	61	ALA	N [N]	3.81
48	PHE	CD2[C]	DR1- α	61	ALA	N [N]	3.99
48	PHE	CB [C]	DR1- α	61	ALA	CA [C]	3.71
			DR1- α	61	ALA	CB [C]	3.83
48	PHE	CG [C]	DR1- α	61	ALA	CB [C]	3.68
48	PHE	CD2[C]	DR1- α	61	ALA	CB [C]	3.80
48	PHE	O [O]	DR1- α	64	ALA	CB [C]	3.13
49	ASP	OD2[O]	DR1- α	64	ALA	CA [C]	3.87
49	ASP	CG [C]	DR1- α	64	ALA	CB [C]	3.61
49	ASP	OD1[O]	DR1- α	64	ALA	CB [C]	3.93
49	ASP	OD2[O]	DR1- α	64	ALA	CB [C]	3.67
49	ASP	CG [C]	DR1- α	67	LYS	CE [C]	3.99
49	ASP	OD1[O]	DR1- α	67	LYS	CE [C]	3.90
49	ASP	OD2[O]	DR1- α	67	LYS	CE [C]	3.25
49	ASP	CG [C]	DR1- α	67	LYS	NZ [N]	3.53
49	ASP	OD1[O]	DR1- α	67	LYS	NZ [N]	3.77
49	ASP	OD2[O]	DR1- α	67	LYS	NZ [N]	2.61
52	ASP	CG [C]	DR1- α	39	LYS	CE [C]	3.98
52	ASP	OD1[O]	DR1- α	39	LYS	CE [C]	2.76
			DR1- α	39	LYS	NZ [N]	3.04
			DR1- α	60	LEU	CD2[C]	3.57
53	ILE	O [O]	DR1- α	39	LYS	CE [C]	3.99
53	ILE	N [N]	DR1- α	39	LYS	NZ [N]	3.78
53	ILE	C [C]	DR1- α	39	LYS	NZ [N]	3.63
53	ILE	O [O]	DR1- α	39	LYS	NZ [N]	2.66
54	ASN	OD1[O]	DR1- α	39	LYS	CD [C]	3.61
			DR1- α	39	LYS	CE [C]	3.25
54	ASN	CG [C]	DR1- α	39	LYS	NZ [N]	3.90
54	ASN	OD1[O]	DR1- α	39	LYS	NZ [N]	3.24
			DR1- α	57	GLN	CG [C]	3.50

9 Bibliography

1. Medzhitov, R. & Janeway, C. A. Decoding the Patterns of Self and Nonself by the Innate Immune System. *Science (80-.)*. **296**, 298–300 (2002).
2. Kringelum, J. V., Nielsen, M., Padkjær, S. B. & Lund, O. Structural analysis of B-cell epitopes in antibody:protein complexes. *Mol. Immunol.* **53**, 24–34 (2013).
3. Stave, J. W. & Lindpaintner, K. Antibody and Antigen Contact Residues Define Epitope and Paratope Size and Structure. *J. Immunol.* **191**, 1428–1435 (2013).
4. Neefjes, J., Jongsmma, M. L., Paul, P. & Bakke, O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol* **11**, 823–36 (2011).
5. Rudolph, M. M. G., Stanfield, R. L. R. & Wilson, I. a. How TCRs bind MHCs, peptides, and coreceptors. *Annu. Rev. Immunol.* **24**, 419–466 (2006).
6. Davis, M. M. & Bjorkman, P. J. T-cell antigen receptor genes and T-cell recognition. *Nature* **334**, 395–402 (1988).
7. Schatz, D. G. & Swanson, P. C. V(D)J recombination: mechanisms of initiation. *Annu. Rev. Genet.* **45**, 167–202 (2011).
8. Ma, Y., Pannicke, U., Schwarz, K. & Lieber, M. R. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108**, 781–94 (2002).
9. Persaud, S. P., Parker, C. R., Lo, W.-L., Weber, K. S. & Allen, P. M. Intrinsic CD4(+) T cell sensitivity and response to a pathogen are set and sustained by avidity for thymic and peripheral complexes of self peptide and MHC. *Nat. Immunol.* **15**, 1–11 (2014).
10. Morris, G. P. & Allen, P. M. How the TCR balances sensitivity and specificity for the recognition of self and pathogens. *Nat. Immunol.* **13**, 121–128 (2012).
11. Fry, A., Jones, L., Kruisbeek, A. & Matis, L. Thymic requirement for clonal deletion during T cell development. *Science (80-.)*. **246**, 1044–1046 (1989).

12. Alam, S. M. *et al.* T-cell-receptor affinity and thymocyte positive selection. *Nature* **381**, 616–620 (1996).
13. Fu, G. *et al.* Themis sets the signal threshold for positive and negative selection in T-cell development. *Nature* **504**, 441–445 (2013).
14. Anderson, M. S. *et al.* The Cellular Mechanism of Aire Control of T Cell Tolerance. *Immunity* **23**, 227–239 (2005).
15. Liston, A., Lesage, S., Wilson, J., Peltonen, L. & Goodnow, C. C. Aire regulates negative selection of organ-specific T cells. *Nat. Immunol.* **4**, 350–354 (2003).
16. Fontenot, J. D., Gavin, M. A. & Rudensky, A. Y. Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. *Nat. Immunol.* **4**, 330–336 (2003).
17. Gavin, M. A. *et al.* Foxp3-dependent programme of regulatory T-cell differentiation. *Nature* **445**, 771–775 (2007).
18. Hori, S., Nomura, T. & Sakaguchi, S. Control of Regulatory T Cell Development by the Transcription Factor Foxp3. *Science (80-.)*. **299**, 1057–1061 (2003).
19. Maloy, K. J. & Powrie, F. Regulatory T cells in the control of immune pathology. *Nat. Immunol.* **2**, 816–822 (2001).
20. Moran, A. E. & Hogquist, K. A. T-cell receptor affinity in thymic development. *Immunology* **135**, 261–267 (2012).
21. Kitagawa, Y. *et al.* Guidance of regulatory T cell development by Satb1-dependent super-enhancer establishment. *Nat. Immunol.* **18**, 173–183 (2016).
22. Feng, Y. *et al.* A mechanism for expansion of regulatory T-cell repertoire and its role in self-tolerance. *Nature* **528**, 132–136 (2015).
23. Garcia, K. C. Reconciling views on T cell receptor germline bias for MHC. *Trends in Immunology* **33**, 429–436 (2012).
24. Feng, D., Bond, C. J., Ely, L. K., Maynard, J. & Garcia, K. C. Structural evidence for a germline-encoded T cell receptor–major histocompatibility complex interaction ‘codon’. *Nat. Immunol.* **8**, 975–983 (2007).
25. Colf, L. A. *et al.* How a Single T Cell Receptor Recognizes Both Self and Foreign MHC. *Cell* **129**, 135–146 (2007).

26. Jones, L. L., Colf, L. A., Stone, J. D., Garcia, K. C. & Kranz, D. M. Distinct CDR3 conformations in TCRs determine the level of cross-reactivity for diverse antigens, but not the docking orientation. *J. Immunol.* **181**, 6255–64 (2008).
27. Van Laethem, F. *et al.* Deletion of CD4 and CD8 coreceptors permits generation of alphabetaT cells that recognize antigens independently of the MHC. *Immunity* **27**, 735–50 (2007).
28. Van Laethem, F. *et al.* MHC restriction is imposed on a diverse TCR repertoire by CD4 and CD8 coreceptors during thymic selection. *Trends Immunol.* **33**, 437–41 (2012).
29. Holland, S. J. *et al.* PNAS Plus: The T-cell receptor is not hardwired to engage MHC ligands. *Proc. Natl. Acad. Sci.* **109**, (2012).
30. Tikhonova, A. N. *et al.* $\alpha\beta$ T cell receptors that do not undergo major histocompatibility complex-specific thymic selection possess antibody-like recognition specificities. *Immunity* **36**, 79–91 (2012).
31. Gras, S. *et al.* Reversed T Cell Receptor Docking on a Major Histocompatibility Class I Complex Limits Involvement in the Immune Response. *Immunity* **45**, 749–760 (2016).
32. Beringer, D. X. *et al.* T cell receptor reversed polarity recognition of a self-antigen major histocompatibility complex. *Nat. Immunol.* **16**, (2015).
33. Sigalov, A., Aivazian, D. & Stern, L. Homooligomerization of the Cytoplasmic Domain of the T Cell Receptor α Chain and of Other Proteins Containing the Immunoreceptor Tyrosine-Based Activation Motif. *Biochemistry* **43**, 2049–2061 (2004).
34. Roh, K.-H., Lillemeier, B. F., Wang, F. & Davis, M. M. The coreceptor CD4 is expressed in distinct nanoclusters and does not colocalize with T-cell receptor and active protein tyrosine kinase p56lck. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E1604–13 (2015).
35. Lin, R. S., Rodriguez, C., Veillette, A. & Lodish, H. F. Zinc is essential for binding of p56(lck) to CD4 and CD8alpha. *J. Biol. Chem.* **273**, 32878–82 (1998).

36. Kim, P. W., Sun, Z.-Y. J., Blacklow, S. C., Wagner, G. & Eck, M. J. A zinc clasp structure tethers Lck to T cell coreceptors CD4 and CD8. *Science* **301**, 1725–8 (2003).
37. Artyomov, M. N., Lis, M., Devadas, S., Davis, M. M. & Chakraborty, A. K. CD4 and CD8 binding to MHC molecules primarily acts to enhance Lck delivery. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16916–16921 (2010).
38. van der Merwe, P. A. & Dushek, O. Mechanisms for T cell receptor triggering. *Nat. Rev. Immunol.* **11**, 47–55 (2011).
39. Li, Y., Yin, Y. & Mariuzza, R. a. Structural and biophysical insights into the role of CD4 and CD8 in T cell activation. *Front. Immunol.* **4**, 206 (2013).
40. Bridgeman, J. S., Sewell, A. K., Miles, J. J., Price, D. a & Cole, D. K. Structural and biophysical determinants of $\alpha\beta$ T-cell antigen recognition. *Immunology* **135**, 9–18 (2012).
41. Valitutti, S., Müller, S. & Cella, M. Serial triggering of many T-cell receptors by a few peptide MHC complexes. *Nature* (1995).
42. Choudhuri, K., Wiseman, D., Brown, M. H., Gould, K. & van der Merwe, P. A. T-cell receptor triggering is critically dependent on the dimensions of its peptide-MHC ligand. *Nature* **436**, 578–82 (2005).
43. Grakoui, A. The Immunological Synapse: A Molecular Machine Controlling T Cell Activation. *Science (80-.)*. **285**, 221–227 (1999).
44. Bromley, S. K. *et al.* The immunological synapse and CD28-CD80 interactions. *Nat. Immunol.* **2**, 1159–1166 (2001).
45. Wülfing, C. *et al.* Costimulation and endogenous MHC ligands contribute to T cell recognition. *Nat. Immunol.* **3**, 42–47 (2002).
46. McMichael, A. J. & Rowland-Jones, S. L. Cellular immune responses to HIV. *Nature* **410**, 980–7. (2001).
47. Kägi, D., Ledermann, B., Bürki, K., Zinkernagel, R. M. & Hengartner, H. Molecular mechanisms of lymphocyte-mediated cytotoxicity and their role in immunological protection and pathogenesis in vivo. *Annu. Rev. Immunol.* **14**, 207–32 (1996).

48. Sridhar, S. *et al.* Cellular immune correlates of protection against symptomatic pandemic influenza. *Nat. Med.* **19**, 1305–12 (2013).
49. Price, D. a *et al.* Public clonotype usage identifies protective Gag-specific CD8+ T cell responses in SIV infection. *J. Exp. Med.* **206**, 923–36 (2009).
50. Berthoud, T. K. *et al.* Potent CD8+ T-Cell Immunogenicity in Humans of a Novel Heterosubtypic Influenza A Vaccine, MVA-NP+M1. *Clin. Infect. Dis.* **52**, 1–7 (2011).
51. Kvistborg, P. *et al.* TIL therapy broadens the tumor-reactive CD8(+) T cell compartment in melanoma patients. *Oncoimmunology* **1**, 409–418 (2012).
52. Phan, G. Q. *et al.* Cancer regression and autoimmunity induced by cytotoxic T lymphocyte-associated antigen 4 blockade in patients with metastatic melanoma. *Proc. Natl. Acad. Sci.* **100**, 8372–8377 (2003).
53. Kawakami, Y. *et al.* Identification of new melanoma epitopes on melanosomal proteins recognized by tumor infiltrating T lymphocytes restricted by HLA-A1, -A2, and -A3 alleles. *J. Immunol.* **161**, 6985–92 (1998).
54. Godfrey, D. I., Uldrich, A. P., McCluskey, J., Rossjohn, J. & Moody, D. B. The burgeoning family of unconventional T cells. *Nat. Immunol.* **16**, 1114–1123 (2015).
55. Van Rhijn, I., Godfrey, D. I., Rossjohn, J. & Moody, D. B. Lipid and small-molecule display by CD1 and MR1. *Nat. Rev. Immunol.* **15**, 643–654 (2015).
56. Iezzi, G., Scotet, E., Scheidegger, D. & Lanzavecchia, A. The interplay between the duration of TCR and cytokine signaling determines T cell polarization. *Eur. J. Immunol.* **29**, 4092–101 (1999).
57. Gerner, M. Y., Torabi-Parizi, P. & Germain, R. N. Strategically Localized Dendritic Cells Promote Rapid T Cell Responses to Lymph-Borne Particulate Antigens. *Immunity* **42**, 172–185 (2015).
58. Eickhoff, S. *et al.* Robust Anti-viral Immunity Requires Multiple Distinct T Cell-Dendritic Cell Interactions. *Cell* **162**, 1322–1337 (2015).
59. Honda, T. *et al.* Tuning of Antigen Sensitivity by T Cell Receptor-Dependent Negative Feedback Controls T Cell Effector Function in Inflamed Tissues.

- Immunity* **40**, 235–247 (2014).
60. van Panhuys, N., Klauschen, F. & Germain, R. N. T-Cell-Receptor-Dependent Signal Intensity Dominantly Controls CD4⁺ T Cell Polarization In Vivo. *Immunity* **41**, 63–74 (2014).
 61. Smith, C. M. *et al.* Cognate CD4(+) T cell licensing of dendritic cells in CD8(+) T cell immunity. *Nat. Immunol.* **5**, 1143–1148 (2004).
 62. Gallegos, A. M., Pamer, E. G. & Glickman, M. S. Delayed protection by ESAT-6-specific effector CD4⁺ T cells after airborne *M. tuberculosis* infection. *J. Exp. Med.* **205**, 2359–68 (2008).
 63. Laidlaw, B. J. *et al.* CD4(+) T Cell Help Guides Formation of CD103(+) Lung-Resident Memory CD8(+) T Cells during Influenza Viral Infection. *Immunity* **41**, 633–645 (2014).
 64. Strutt, T. M. *et al.* Memory CD4⁺ T cells induce innate responses independently of pathogen. *Nat. Med.* **16**, 558–564, 1p following 564 (2010).
 65. Street, N. E. & Mosmann, T. R. Functional diversity of T lymphocytes due to secretion of different cytokine patterns. *FASEB J.* **5**, 171–7 (1991).
 66. Damjanovic, D. *et al.* Immunopathology in influenza virus infection: uncoupling the friend from foe. *Clin. Immunol.* **144**, 57–69 (2012).
 67. Penaloza-MacMaster, P. *et al.* Vaccine-elicited CD4 T cells induce immunopathology after chronic LCMV infection. *Science (80-.).* **347**, 278–282 (2015).
 68. Zhao, Y. *et al.* High levels of virus-specific CD4⁺ T cells predict severe pandemic influenza A virus infection. *Am. J. Respir. Crit. Care Med.* **186**, 1292–7 (2012).
 69. Breitfeld, D. *et al.* Follicular B helper T cells express CXC chemokine receptor 5, localize to B cell follicles, and support immunoglobulin production. *J. Exp. Med.* **192**, 1545–52 (2000).
 70. Schaerli, P. *et al.* CXC chemokine receptor 5 expression defines follicular homing T cells with B cell helper function. *J. Exp. Med.* **192**, 1553–62 (2000).
 71. Reinhardt, R. L., Liang, H.-E. & Locksley, R. M. Cytokine-secreting follicular T

- cells shape the antibody repertoire. *Nat. Immunol.* **10**, 385–93 (2009).
72. Hale, J. S. & Ahmed, R. Memory T follicular helper CD4 T cells. *Front. Immunol.* **6**, 1–9 (2015).
73. Streeck, H., D'Souza, M. P., Littman, D. R. & Crotty, S. Harnessing CD4⁺ T cell responses in HIV vaccine development. *Nat. Med.* **19**, 143–9 (2013).
74. Alam, S., Knowlden, Z. a G., Sangster, M. Y. & Sant, A. J. CD4 T cell help is limiting and selective during the primary B cell response to influenza infection. *J. Virol.* **88**, 314–324 (2013).
75. Martins, K. A. O. *et al.* Adjuvant-enhanced CD4 T Cell Responses are Critical to Durable Vaccine Immunity. *EBIOM* **3**, 67–78 (2016).
76. Craft, J. E. Follicular helper T cells in immunity and systemic autoimmunity. *Nat. Rev. Rheumatol.* **8**, 337–47 (2012).
77. Chen, F. *et al.* An essential role for TH2-type responses in limiting acute tissue damage during experimental helminth infection. *Nat. Med.* **18**, 260–266 (2012).
78. Perrigoue, J. G. *et al.* MHC class II–dependent basophil–CD4⁺ T cell interactions promote TH2 cytokine–dependent immunity. *Nat. Immunol.* **10**, 697–705 (2009).
79. Kuperman, D. A. *et al.* Direct effects of interleukin-13 on epithelial cells cause airway hyperreactivity and mucus overproduction in asthma TL - 8. *Nat. Med.* **8** **VN-re**, 885–889 (2002).
80. Mionnet, C. *et al.* CX3CR1 is required for airway inflammation by promoting T helper cell survival and maintenance in inflamed lung. *Nat. Med.* **16**, 1305–1312 (2010).
81. Robinson, D. S. *et al.* Predominant T H2 -like Bronchoalveolar T-Lymphocyte Population in Atopic Asthma. *N. Engl. J. Med.* **326**, 298–304 (1992).
82. Mangan, P. R. *et al.* Transforming growth factor-beta induces development of the T(H)17 lineage. *Nature* **441**, 231–4 (2006).
83. Ivanov, I. I. *et al.* The Orphan Nuclear Receptor ROR γ t Directs the Differentiation Program of Proinflammatory IL-17⁺ T Helper Cells. *Cell* **126**, 1121–1133 (2006).
84. Acosta-Rodriguez, E. V *et al.* Surface phenotype and antigenic specificity of

- human interleukin 17-producing T helper memory cells. *Nat. Immunol.* **8**, 639–646 (2007).
85. Annunziato, F. *et al.* Phenotypic and functional features of human Th17 cells. *J. Exp. Med.* **204**, (2007).
86. Gagliani, N. *et al.* Th17 cells transdifferentiate into regulatory T cells during resolution of inflammation. *Nature* **523**, 221–5 (2015).
87. Weaver, C. T., Harrington, L. E., Mangan, P. R., Gavrieli, M. & Murphy, K. M. Th17: An Effector CD4 T Cell Lineage with Regulatory T Cell Ties. *Immunity* **24**, 677–688 (2006).
88. Huber, S. *et al.* Th17 Cells Express Interleukin-10 Receptor and Are Controlled by Foxp3⁻ and Foxp3⁺ Regulatory CD4⁺ T Cells in an Interleukin-10-Dependent Manner. *Immunity* **34**, 554–565 (2011).
89. Withers, D. R. *et al.* Transient inhibition of ROR- γ t therapeutically limits intestinal inflammation by reducing TH17 cells and preserving group 3 innate lymphoid cells. *Nat. Med.* **22**, 319–323 (2016).
90. Punkenburg, E. *et al.* Batf-dependent Th17 cells critically regulate IL-23 driven colitis-associated colon cancer. *Gut* **65**, 1139–1150 (2016).
91. Deng, Z. *et al.* Enterobacteria-secreted particles induce production of exosome-like S1P-containing particles by intestinal epithelium to drive Th17-mediated tumorigenesis. *Nat. Commun.* **6**, 6956 (2015).
92. Geis, A. L. *et al.* Regulatory T-cell Response to Enterotoxigenic *Bacteroides fragilis* Colonization Triggers IL17-Dependent Colon Carcinogenesis. *Cancer Discov.* **5**, (2015).
93. Jones, G. W. *et al.* Interleukin-27 inhibits ectopic lymphoid-like structure development in early inflammatory arthritis. *J. Exp. Med.* **212**, 1793–1802 (2015).
94. Jones, G. W. *et al.* Loss of CD4⁺ T Cell IL-6R Expression during Inflammation Underlines a Role for IL-6 *Trans* Signaling in the Local Maintenance of Th17 Cells. *J. Immunol.* **184**, 2130–2139 (2010).
95. Komatsu, N. *et al.* Pathogenic conversion of Foxp3⁺ T cells into TH17 cells in

- autoimmune arthritis. *Nat. Med.* **20**, 62–68 (2013).
96. Iwashiro, M., Peterson, K., Messer, R. J., Stromnes, I. M. & Hasenkrug, K. J. CD4(+) T cells and gamma interferon in the long-term control of persistent friend retrovirus infection. *J. Virol.* **75**, 52–60 (2001).
 97. Brien, J. D., Uhrlaub, J. L. & Nikolich-Zugich, J. West Nile virus-specific CD4 T cells exhibit direct antiviral cytokine secretion and cytotoxicity and are sufficient for antiviral protection. *J. Immunol.* **181**, 8568–75 (2008).
 98. Qui, H. Z. *et al.* CD134 plus CD137 dual costimulation induces Eomesodermin in CD4 T cells to program cytotoxic Th1 differentiation. *J. Immunol.* **187**, 3555–64 (2011).
 99. Curran, M. A. *et al.* Systemic 4-1BB activation induces a novel T cell phenotype driven by high expression of Eomesodermin. *J. Exp. Med.* **210**, 743–55 (2013).
 100. Haabeth, O. A. W. *et al.* How Do CD4(+) T Cells Detect and Eliminate Tumor Cells That Either Lack or Express MHC Class II Molecules? *Front. Immunol.* **5**, 174 (2014).
 101. Brown, D. M., Lee, S., Garcia-Hernandez, M. D. L. L. & Swain, S. L. Multifunctional CD4 cells expressing gamma interferon and perforin mediate protection against lethal influenza virus infection. *J. Virol.* **86**, 6792–803 (2012).
 102. Takahashi, T. *et al.* Immunologic self-tolerance maintained by CD25+CD4+ naturally anergic and suppressive T cells: induction of autoimmune disease by breaking their anergic/suppressive state. *Int. Immunol.* **10**, 1969–80 (1998).
 103. Suri-Payer, E., Amar, A. Z., Thornton, A. M. & Shevach, E. M. CD4+CD25+ T cells inhibit both the induction and effector function of autoreactive T cells and represent a unique lineage of immunoregulatory cells. *J. Immunol.* **160**, 1212–8 (1998).
 104. Chen, W. *et al.* Conversion of Peripheral CD4+CD25– Naive T Cells to CD4+CD25+ Regulatory T Cells by TGF- β Induction of Transcription Factor Foxp3. *J. Exp. Med.* **198**, (2003).
 105. Fantini, M. C. *et al.* Cutting Edge: TGF- β Induces a Regulatory Phenotype in CD4+CD25– T Cells through Foxp3 Induction and Down-Regulation of Smad7. *J.*

- Immunol.* **172**, (2004).
106. Chinen, T. *et al.* An essential role for the IL-2 receptor in Treg cell function. *Nat. Immunol.* **17**, 1322–1333 (2016).
 107. Holt, M. P., Punkosdy, G. A., Glass, D. D. & Shevach, E. M. TCR Signaling and CD28/CTLA-4 Signaling Cooperatively Modulate T Regulatory Cell Homeostasis. *J. Immunol.* **198**, 1503–1511 (2017).
 108. Gondek, D. C., Lu, L.-F., Quezada, S. A., Sakaguchi, S. & Noelle, R. J. Cutting edge: contact-mediated suppression by CD4⁺CD25⁺ regulatory cells involves a granzyme B-dependent, perforin-independent mechanism. *J. Immunol.* **174**, 1783–6 (2005).
 109. Chaudhry, A. *et al.* Interleukin-10 Signaling in Regulatory T Cells Is Required for Suppression of Th17 Cell-Mediated Inflammation. *Immunity* **34**, 566–578 (2011).
 110. Betts, G. *et al.* Suppression of tumour-specific CD4⁺ T cells by regulatory T cells is associated with progression of human colorectal cancer. *Gut* **61**, 1163–71 (2012).
 111. Savage, P. a, Malchow, S. & Leventhal, D. S. Basic principles of tumor-associated regulatory T cell biology. *Trends Immunol.* **34**, 33–40 (2013).
 112. Scurr, M. *et al.* Escalating regulation of 5T4-specific IFN- γ (+) CD4(+) T cells distinguishes colorectal cancer patients from healthy controls and provides a target for in vivo therapy. *Cancer Immunol. Res.* **1**, 416–426 (2013).
 113. Sakaguchi, S., Miyara, M., Costantino, C. M. & Hafler, D. a. FOXP3⁺ regulatory T cells in the human immune system. *Nat. Rev. Immunol.* **10**, 490–500 (2010).
 114. Curiel, T. J. *et al.* Specific recruitment of regulatory T cells in ovarian carcinoma fosters immune privilege and predicts reduced survival. *Nat. Med.* **10**, 942–9 (2004).
 115. Veldhoen, M. *et al.* Transforming growth factor- β ‘reprograms’ the differentiation of T helper 2 cells and promotes an interleukin 9–producing subset. *Nat. Immunol.* **9**, 1341–1346 (2008).
 116. Dardalhon, V. *et al.* IL-4 inhibits TGF- β -induced Foxp3⁺ T cells and, together with

- TGF- β , generates IL-9+ IL-10+ Foxp3- effector T cells. *Nat. Immunol.* **9**, 1347–1355 (2008).
117. Eyerich, S. *et al.* Th22 cells represent a distinct human T cell subset involved in epidermal immunity and remodeling. *J. Clin. Invest.* **119**, 3573–3585 (2009).
 118. Carrier, Y., Yuan, J., Kuchroo, V. K. & Weiner, H. L. Th3 Cells in Peripheral Tolerance. I. Induction of Foxp3-Positive Regulatory T Cells by Th3 Cells Derived from TGF- β T Cell-Transgenic Mice. *J. Immunol.* **178**, 179–185 (2007).
 119. Addey, C. *et al.* Functional plasticity of antigen-specific regulatory T cells in context of tumor. *J. Immunol.* **186**, 4557–64 (2011).
 120. Yamane, H. & Paul, W. Memory CD4+ T cells: fate determination, positive feedback and plasticity. *Cell. Mol. Life Sci.* **69**, 1577–1583 (2012).
 121. Zhou, L., Chong, M. M. W. & Littman, D. R. Plasticity of CD4+ T Cell Lineage Differentiation. *Immunity* **30**, 646–655 (2009).
 122. Bjorkman, P. J. *et al.* Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329**, 506–512 (1987).
 123. Fremont, D. H., Matsumura, M., Stura, E. a, Peterson, P. a & Wilson, I. a. Crystal structures of two viral peptides in complex with murine MHC class I H-2Kb. *Science* **257**, 919–927 (1992).
 124. Stern, L., Brown, J., Jardetzky, T. & Gorga, J. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. (1994).
 125. Tynan, F. E. *et al.* T cell receptor recognition of a ‘super-bulged’ major histocompatibility complex class I-bound peptide. *Nat. Immunol.* **6**, 1114–22 (2005).
 126. Ljunggren, H.-G. *et al.* Empty MHC class I molecules come out in the cold. *Nature* **346**, 476–480 (1990).
 127. Painter, C. A. & Stern, L. J. Conformational variation in structures of classical and non-classical MHCII proteins and functional implications. *Immunol. Rev.* **250**, 144–157 (2012).
 128. Srinivasan, M., Domanico, S. Z., Kaumaya, P. T. P. & Pierce, S. K. Peptides of 23

- residues or greater are required to stimulate a high affinity class II restricted T cell response. *Eur. J. Immunol.* **23**, 1011–1016 (1993).
129. Wang, D. *et al.* HLA class II antigens and T lymphocytes in human nasal epithelial cells. Modulation of the HLA class II gene transcripts by gamma interferon. *Clin. Exp. Allergy* **27**, 306–14 (1997).
130. Donia, M. *et al.* Aberrant Expression of MHC Class II in Melanoma Attracts Inflammatory Tumor-Specific CD4⁺ T- Cells, Which Dampen CD8⁺ T-cell Antitumor Reactivity. *Cancer Res.* **75**, 3747–3759 (2015).
131. Sconocchia, G. *et al.* HLA class II antigen expression in colorectal carcinoma tumors as a favorable prognostic marker. *Neoplasia* **16**, 31–42 (2014).
132. Adamopoulou, E. *et al.* Exploring the MHC-peptide matrix of central tolerance in the human thymus. *Nat. Commun.* **4**, 2039 (2013).
133. Rudensky, A. Y. *et al.* Intracellular assembly and transport of endogenous peptide-MHC class II complexes. *Immunity* **1**, 585–94 (1994).
134. Kleijmeer, M. J., Oorschot, V. M. & Geuze, H. J. Human resident langerhans cells display a lysosomal compartment enriched in MHC class II. *J. Invest. Dermatol.* **103**, 516–23 (1994).
135. Riberdy, J. M., Avva, R. R., Geuze, H. J. & Cresswell, P. Transport and intracellular distribution of MHC class II molecules and associated invariant chain in normal and antigen-processing mutant cell lines. *J. Cell Biol.* **125**, 1225–37 (1994).
136. Denzin, L. K. & Cresswell, P. HLA-DM induces clip dissociation from MHC class II dimers and facilitates peptide loading. *Cell* **82**, 155–165 (1995).
137. Roche, P. a & Furuta, K. The ins and outs of MHC class II-mediated antigen processing and presentation. *Nat. Rev. Immunol.* **15**, 203–216 (2015).
138. Allan, E. R. O. *et al.* NADPH oxidase modifies patterns of MHC class II-restricted epitopic repertoires through redox control of antigen processing. *J. Immunol.* **192**, 4989–5001 (2014).
139. Clement, C. C. *et al.* The Dendritic Cell MHC II Peptidome Derives from a Variety

- of Processing Pathways and Includes Peptides with a Broad Spectrum of HLA-DM Sensitivity. *J. Biol. Chem.* **291**, jbc.M115.655738 (2016).
140. Hornell, T. M. C. *et al.* Human Dendritic Cell Expression of HLA-DO Is Subset Specific and Regulated by Maturation. *J. Immunol.* **176**, 3536–3547 (2006).
 141. Kropshofer, H. *et al.* Editing of the HLA-DR-peptide repertoire by HLA-DM. *EMBO J.* **15**, 6144–54 (1996).
 142. Eisenlohr, L. C. & Hackett, C. J. Class II major histocompatibility complex-restricted T cells specific for a virion structural protein that do not recognize exogenous influenza virus. Evidence that presentation of labile T cell determinants is favored by endogenous antigen synthesis. *J. Exp. Med.* **169**, 921–31 (1989).
 143. Mukherjee, P. *et al.* Efficient presentation of both cytosolic and endogenous transmembrane protein antigens on MHC class II is dependent on cytoplasmic proteolysis. *J. Immunol.* **167**, 2632–41 (2001).
 144. Lich, J. D., Elliott, J. F. & Blum, J. S. Cytoplasmic processing is a prerequisite for presentation of an endogenous antigen by major histocompatibility complex class II proteins. *J. Exp. Med.* **191**, 1513–24 (2000).
 145. Eisenlohr, L. C. Alternative generation of MHC class II-restricted epitopes: not so exceptional? *Mol. Immunol.* **55**, 169–71 (2013).
 146. Miller, M. a, Ganesan, A. P. V & Eisenlohr, L. C. Toward a Network Model of MHC Class II-Restricted Antigen Processing. *Front. Immunol.* **4**, 464 (2013).
 147. Tewari, M. K., Sinnathamby, G., Rajagopal, D. & Eisenlohr, L. C. A cytosolic pathway for MHC class II-restricted antigen processing that is proteasome and TAP dependent. *Nat. Immunol.* **6**, 287–294 (2005).
 148. Miller, M. a, Ganesan, A. P. V, Luckashenak, N., Mendonca, M. & Eisenlohr, L. C. Endogenous antigen processing drives the primary CD4+ T cell response to influenza. *Nat. Med.* **21**, 1216–1222 (2015).
 149. Bevan, M. J. Cross-priming for a secondary cytotoxic response to minor H antigens with H-2 congenic cells which do not cross-react in the cytotoxic assay. *J. Exp. Med.* **143**, 1283–8 (1976).

150. Joffre, O. P., Segura, E., Savina, A. & Amigorena, S. Cross-presentation by dendritic cells. *Nat. Rev. Immunol.* **12**, 557–69 (2012).
151. Denzin, L. K., Robbins, N. F., Carboy-Newcomb, C. & Cresswell, P. Assembly and intracellular transport of HLA-DM and correction of the class II antigen-processing defect in T2 cells. *Immunity* **1**, 595–606 (1994).
152. Sanderson, F. *et al.* Accumulation of HLA-DM, a regulator of antigen presentation, in MHC class II compartments. *Science* **266**, 1566–9 (1994).
153. Morris, P. An essential role for HLA-DM in antigen presentation by class II major histocompatibility molecules. *Nature* **368**, 551–4 (1994).
154. Sloan, V. S. *et al.* Mediation by HLA-DM of dissociation of peptides from HLA-DR. *Nature* **375**, 802–806 (1995).
155. van Ham, S. M. *et al.* HLA-DO is a negative modulator of HLA-DM-mediated MHC class II peptide loading. *Curr. Biol.* **7**, 950–7 (1997).
156. Denzin, L. K., Sant'Angelo, D. B., Hammond, C., Surman, M. J. & Cresswell, P. Negative regulation by HLA-DO of MHC class II-restricted antigen processing. *Science* **278**, 106–9 (1997).
157. Mosyak, L., Zaller, D. M. & Wiley, D. C. The structure of HLA-DM, the peptide exchange catalyst that loads antigen onto class II MHC molecules during antigen presentation. *Immunity* **9**, 377–83 (1998).
158. Nicholson, M. J. *et al.* Small molecules that enhance the catalytic efficiency of HLA-DM. *J. Immunol.* **176**, 4208–20 (2006).
159. Pos, W. *et al.* Crystal structure of the HLA-DM-HLA-DR1 complex defines mechanisms for rapid peptide selection. *Cell* **151**, 1557–68 (2012).
160. Guce, A. I. *et al.* HLA-DO acts as a substrate mimic to inhibit HLA-DM by a competitive mechanism. *Nat. Struct. Mol. Biol.* **20**, 90–8 (2013).
161. Kämper, N. *et al.* γ -Interferon-regulated chaperone governs human lymphocyte antigen class II expression. *FASEB J.* **26**, 104–16 (2012).
162. Pinet, V., Vergelli, M., Martin, R., Bakke, O. & Long, E. O. Antigen presentation mediated by recycling of surface HLA-DR molecules. *Nature* **375**, 603–6 (1995).

163. Pinet, V. M. & Long, E. O. Peptide loading onto recycling HLA-DR molecules occurs in early endosomes. *Eur. J. Immunol.* **28**, 799–804 (1998).
164. Kropshofer, H., Hämmerling, G. J. & Vogt, A. B. The impact of the non-classical MHC proteins HLA-DM and HLA-DO on loading of MHC class II molecules. *Immunol. Rev.* **172**, 267–78 (1999).
165. Sadegh-Nasseri, S. & Kim, A. Exogenous antigens bind MHC class II first, and are processed by cathepsins later. *Mol. Immunol.* **68**, (2015).
166. Hennecke, J. & Wiley, D. C. Structure of a complex of the human alpha/beta T cell receptor (TCR) HA1.7, influenza hemagglutinin peptide, and major histocompatibility complex class II molecule, HLA-DR4 (DRA*0101 and DRB1*0401): insight into TCR cross-restriction and alloreactivity. *J. Exp. Med.* **195**, 571–581 (2002).
167. Petersen, J. *et al.* Diverse T Cell Receptor Gene Usage in HLA-DQ8-Associated Celiac Disease Converges into a Consensus Binding Solution. *Structure* **24**, 1643–1657 (2016).
168. Cole, D. K. *et al.* Modification of the carboxy-terminal flanking region of a universal influenza epitope alters CD4⁺ T-cell repertoire selection. *Nat. Commun.* **3**, 665 (2012).
169. Lippolis, J. D. *et al.* Analysis of MHC class II antigen processing by quantitation of peptides that constitute nested sets. *J. Immunol.* **169**, 5089–97 (2002).
170. Chicz, R. M. *et al.* Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. *J. Exp. Med.* **178**, 27–47 (1993).
171. Richards, K. a. *et al.* Seasonal Influenza Can Poise Hosts for CD4 T-Cell Immunity to H7N9 Avian Influenza. *J. Infect. Dis.* 1–9 (2014). doi:10.1093/infdis/jiu662
172. Public Health England. Surveillance of influenza and other respiratory viruses in the United Kingdom: Winter 2014 to 2015. 29 (2016).
173. World Health Organization. *Influenza virus infections in humans (February 2014). Note* (2014).
174. De Graaf, M. & Fouchier, R. A. M. Role of receptor binding specificity in influenza

- A virus transmission and pathogenesis. *EMBO Journal* **33**, 823–841 (2014).
175. Nobusawa, E., Ishihara, H., Morishita, T., Sato, K. & Nakajima, K. Change in receptor-binding specificity of recent human influenza A viruses (H3N2): a single amino acid change in hemagglutinin altered its recognition of sialyloligosaccharides. *Virology* **278**, 587–596 (2000).
 176. Dawood, F. S. *et al.* Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. *Lancet. Infect. Dis.* **12**, 687–95 (2012).
 177. Li, K. S. *et al.* Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature* **430**, 209–213 (2004).
 178. Watanabe, T., Watanabe, S., Maher, E. a., Neumann, G. & Kawaoka, Y. Pandemic potential of avian influenza A (H7N9) viruses. *Trends Microbiol.* **22**, 623–631 (2014).
 179. Yuen, K. Y. *et al.* Clinical features and rapid viral diagnosis of human disease associated with avian influenza A H5N1 virus. *Lancet (London, England)* **351**, 467–71 (1998).
 180. Gao, R. *et al.* Human infection with a novel avian-origin influenza A (H7N9) virus. *N. Engl. J. Med.* **368**, 1888–97 (2013).
 181. Bui, C. *et al.* A Systematic Review of the Comparative Epidemiology of Avian and Human Influenza A H5N1 and H7N9 - Lessons and Unanswered Questions. *Transbound. Emerg. Dis.* **63**, 602–620 (2016).
 182. Rossman, J. S. & Lamb, R. a. Influenza virus assembly and budding. *Virology* **411**, 229–36 (2011).
 183. Ye, Q., Krug, R. M. & Tao, Y. J. The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature* **444**, 1078–82 (2006).
 184. Ng, A.K., K.-L. Zhang, H. et A. Structure of the influenza virus A H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. *FASEB J.* **22**, 3638–3647 (2008).
 185. Zheng, W. & Tao, Y. J. Structure and assembly of the influenza A virus

- ribonucleoprotein complex. *FEBS Lett.* **587**, 1206–14 (2013).
186. Calder, L. J., Wasilewski, S., Berriman, J. A. & Rosenthal, P. B. Structural organization of a filamentous influenza A virus. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 10685–90 (2010).
 187. Zhang, J. & Lamb, R. A. Characterization of the membrane association of the influenza virus matrix protein in living cells. *Virology* **225**, 255–66 (1996).
 188. Gómez-Puertas, P., Albo, C., Pérez-Pastrana, E., Vivo, A. & Portela, A. Influenza virus matrix protein is the major driving force in virus budding. *J. Virol.* **74**, 11538–47 (2000).
 189. Pflug, A., Guilligay, D., Reich, S. & Cusack, S. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* **516**, 355–360 (2014).
 190. Reich, S. *et al.* Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* **516**, 361–366 (2014).
 191. Resa-Infante, P., Jorba, N., Coloma, R. & Ortin, J. The influenza virus RNA synthesis machine: Advances in its structure and function. *RNA Biol.* **8**, 207–215 (2011).
 192. Fodor, E. The RNA polymerase of influenza a virus: mechanisms of viral transcription and replication. *Acta Virol.* **57**, 113–122 (2012).
 193. de Chasse, B. *et al.* The Interactomes of Influenza Virus NS1 and NS2 Proteins Identify New Host Factors and Provide Insights for ADAR1 Playing a Supportive Role in Virus Replication. *PLoS Pathog.* **9**, e1003440 (2013).
 194. O'Neill, R. E., Talon, J. & Palese, P. The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins. *EMBO J.* **17**, 288–296 (1998).
 195. Nayak, D. P., Balogun, R. a., Yamada, H., Zhou, Z. H. & Barman, S. Influenza virus morphogenesis and budding. *Virus Res.* **143**, 147–161 (2009).
 196. Chiu, C. & Openshaw, P. J. Antiviral B cell and T cell immunity in the lungs. *Nat. Immunol.* **16**, 18–26 (2014).
 197. Drake, J. W. Rates of spontaneous mutation among RNA viruses. *Proc. Natl.*

- Acad. Sci. U. S. A.* **90**, 4171–5 (1993).
198. Carrat, F. & Flahault, A. Influenza vaccine: The challenge of antigenic drift. *Vaccine* **25**, 6852–6862 (2007).
 199. Ampofo, W. K. *et al.* Strengthening the influenza vaccine virus selection and development process: Report of the 3rd WHO Informal Consultation for Improving Influenza Vaccine Virus Selection held at WHO headquarters, Geneva, Switzerland, 1-3 April 2014. *Vaccine* **33**, 4368–82 (2015).
 200. Heiny, A. T. *et al.* Evolutionarily conserved protein sequences of influenza A viruses, avian and human, as vaccine targets. *PLoS One* **2**, e1190 (2007).
 201. Ghedin, E. *et al.* Mixed infection and the genesis of influenza virus diversity. *J. Virol.* **83**, 8832–41 (2009).
 202. Garten, R. J. *et al.* Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science* **325**, 197–201 (2009).
 203. Smith, G. J. D. *et al.* Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* **459**, 1122–5 (2009).
 204. Animal farm: pig in the middle. *Nature* **459**, 889–889 (2009).
 205. Vallat, B. Flu: no sign so far that the human pandemic is spread by pigs. *Nature* **460**, 683–683 (2009).
 206. Society, A. P. On the Doctrine of Original Antigenic Sin Author (s): Thomas Francis , Jr . **104**, 572–578 (2014).
 207. Kim, J. H., Skountzou, I., Compans, R. & Jacob, J. Original antigenic sin responses to influenza viruses. *J. Immunol.* **183**, 3294–301 (2009).
 208. O'Donnell, C. D. *et al.* Humans and ferrets with prior H1N1 influenza virus infections do not exhibit evidence of original antigenic sin after infection or vaccination with the 2009 pandemic H1N1 influenza virus. *Clin. Vaccine Immunol.* **21**, 737–746 (2014).
 209. Liu, X. S. *et al.* Overcoming original antigenic sin to generate new CD8 T cell IFN-gamma responses in an antigen-experienced host. *J. Immunol.* **177**, 2873–2879 (2006).

210. Klenerman, P. & Zinkernagel, R. M. Original antigenic sin impairs cytotoxic T lymphocyte responses to viruses bearing variant epitopes. *Nature* **394**, 482–485 (1998).
211. Wooldridge, L. *et al.* A single autoimmune T cell receptor recognizes more than a million different peptides. *J. Biol. Chem.* **287**, 1168–77 (2012).
212. Adams, J. J. *et al.* Structural interplay between germline interactions and adaptive recognition determines the bandwidth of TCR-peptide-MHC cross-reactivity. *Nat. Immunol.* **advance on**, 1–10 (2015).
213. Wilkinson, T. M. *et al.* Preexisting influenza-specific CD4+ T cells correlate with disease protection against influenza challenge in humans. *Nat. Med.* **18**, 274–80 (2012).
214. Theaker, S. M. *et al.* T-cell libraries allow simple parallel generation of multiple peptide-specific human T-cell clones. *J. Immunol. Methods* **430**, 43–50 (2016).
215. Bulek, A. M. *et al.* TCR/pMHC Optimized Protein crystallization Screen. *J. Immunol. Methods* **382**, 203–10 (2012).
216. Tungatt, K. *et al.* Antibody Stabilization of Peptide-MHC Multimers Reveals Functional T Cells Bearing Extremely Low-Affinity TCRs. *J. Immunol.* **194**, 463–474 (2015).
217. Dolton, G. *et al.* More tricks with tetramers: a practical guide to staining T cells with peptide-MHC multimers. *Immunology* **146**, 11–22 (2015).
218. Garcia, K. C. *et al.* An alphabeta T cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. *Science* **274**, 209–19 (1996).
219. Hennecke, J., Carfi, A. & Wiley, D. C. Structure of a covalently stabilized complex of a human alphabeta T-cell receptor, influenza HA peptide and MHC class II molecule, HLA-DR1. *EMBO J.* **19**, 5611–5624 (2000).
220. Novak, E. J. *et al.* Tetramer-guided epitope mapping: rapid identification and characterization of immunodominant CD4+ T cell epitopes from complex antigens. *J. Immunol.* **166**, 6665–70 (2001).
221. Lehner, B. P. J. *et al.* Human HLA-A0201-restricted Cytotoxic T Lymphocyte

- Recognition of Influenza A Is Dominated by T Cells Bearing the V β 17 Gene Segment. *J. Exp. Med.* **181**, 79–91 (1995).
222. Powell, T. J. *et al.* Examination of Influenza Specific T Cell Responses after Influenza Virus Challenge in Individuals Vaccinated with MVA-NP+M1 Vaccine. *PLoS One* **8**, e62778 (2013).
223. Fabb, S. a *et al.* Generation of novel human MHC class II mutant B-cell lines by integrating YAC DNA into a cell line homozygously deleted for the MHC class II region. *Hum. Mol. Genet.* **6**, 1295–304 (1997).
224. DeMars, R., Chang, C. C., Shaw, S., Reitnauer, P. J. & Sondel, P. M. Homozygous deletions that simultaneously eliminate expressions of class I and class II antigens of EBV-transformed B-lymphoblastoid cells. I. Reduced proliferative responses of autologous and allogeneic T cells to mutant cells that have decreased expressi. *Hum. Immunol.* **11**, 77–97 (1984).
225. Erlich, H., Lee, J. S., Petersen, J. W., Bugawan, T. & DeMars, R. Molecular analysis of HLA class I and class II antigen loss mutants reveals a homozygous deletion of the DR, DQ, and part of the DP region: implications for class II gene order. *Hum. Immunol.* **16**, 205–19 (1986).
226. Salter, R. D., Howell, D. N. & Cresswell, P. Genes regulating HLA class I antigen expression in T-B lymphoblast hybrids. *Immunogenetics* **21**, 235–246 (1985).
227. Sette, A. *et al.* Invariant Chain Peptides in Most HLA-DR Molecules of an Antigen-Processing Mutant. **411**, (1992).
228. Riberdy, J. M. & Cresswell, P. The antigen-processing mutant T2 suggests a role for MHC-linked genes in class II antigen presentation. *J. Immunol.* **148**, 2586–90 (1992).
229. Ceman, S., Rudersdorf, R., Long, E. O. & Demars, R. MHC class II deletion mutant expresses normal levels of transgene encoded class II molecules that have abnormal conformation and impaired antigen presentation ability. *J. Immunol.* **149**, 754–61 (1992).
230. Rist, M., Smith, C., Bell, M. J., Burrows, S. R. & Khanna, R. Cross-recognition of

- HLA DR4 alloantigen by virus-specific CD8 α T cells : a new paradigm for self- / nonself-recognition. **114**, 2244–2253 (2009).
231. Servenius, B. *et al.* Molecular map of the human HLA-SB (HLA-DP) region and sequence of an SB alpha (DP alpha) pseudogene. *EMBO J.* **3**, 3209–14 (1984).
232. Amria, S. *et al.* HLA-DM negatively regulates HLA-DR4-restricted collagen pathogenic peptide presentation and T cell recognition. *Eur. J. Immunol.* **38**, 1961–1970 (2008).
233. Chen, L. *et al.* Immunodominant CD4⁺ T-cell Responses to Influenza A Virus in Healthy Individuals Focus on Matrix 1 and Nucleoprotein. *J. Virol.* **1**, 11760–73 (2014).
234. Hayward, A. C. *et al.* Natural T Cell Mediated Protection Against Seasonal and Pandemic Influenza: Results of the Flu Watch Cohort Study. *Am. J. Respir. Crit. Care Med.* **191**, 150406071958005 (2015).
235. Lovitch, S. B. Conformational isomers of a peptide – class II major histocompatibility complex. 293–313 (2005).
236. Lovitch, S. B., Pu, Z. & Unanue, E. R. Amino-terminal flanking residues determine the conformation of a peptide-class II MHC complex. *J. Immunol.* **176**, 2958–2968 (2006).
237. Campion, S. L. *et al.* Proteome-wide analysis of HIV-specific naive and memory CD4⁺ T cells in unexposed blood donors. *J. Exp. Med.* **211**, 1273–80 (2014).
238. Su, L. F. & Davis, M. M. Antiviral memory phenotype T cells in unexposed adults. *Immunol. Rev.* **255**, 95–109 (2013).
239. Neu, K. E., Henry Dunand, C. J. & Wilson, P. C. *Heads, stalks and everything else: how can antibodies eradicate influenza as a human disease? Current Opinion in Immunology* **42**, 48–55 (2016).
240. Air, G. M. Influenza neuraminidase. *Influenza Other Respi. Viruses* **6**, 245–256 (2012).
241. Krammer, F. & Palese, P. Influenza virus hemagglutinin stalk-based antibodies and vaccines. *Curr. Opin. Virol.* **3**, 521–530 (2013).

242. DiPiazza, A., Richards, K. A., Knowlden, Z. A. G., Nayak, J. L. & Sant, A. J. The Role of CD4 T Cell Memory in Generating Protective Immunity to Novel and Potentially Pandemic Strains of Influenza. *Front. Immunol.* **7**, 10 (2016).
243. Babon, J. A. B. *et al.* 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. *Hum. Immunol.* **70**, 711–21 (2009).
244. Leddon, S. A., Richards, K. A., Treanor, J. J. & Sant, A. J. Abundance and specificity of influenza reactive circulating memory follicular helper and non-follicular helper CD4 T cells in healthy adults. *Immunology* **146**, 157–162 (2015).
245. Arzt, Steffi S., Baudin, F., Barge, A., T. & Ruigrok. Combined results from solution studies on intact influenza virus M1 protein and from a new crystal form of its N-terminal domain show that M1 is an elongated monomer. *Virology* **279**, 439–446 (2001).
246. Freed, E. & Martin, M. *Fields Virology*. *Fields Virology* (Wolters Kluwer Health/Lippincott Williams & Wilkins, 2013). doi:9781451105636
247. Strutt, T. T. M. *et al.* Multipronged CD4+ T-cell effector and memory responses cooperate to provide potent immunity against respiratory virus. *Immunol. ...* 149–164 (2013).
248. Swain, S. L., McKinstry, K. K. & Strutt, T. M. Expanding roles for CD4⁺ T cells in immunity to viruses. *Nat. Rev. Immunol.* **12**, 136–48 (2012).
249. Stewart-Jones, G. B. E., McMichael, A. J., Bell, J. I., Stuart, D. I. & Jones, E. Y. A structural basis for immunodominant human T cell receptor recognition. *Nat. Immunol.* **4**, 657–663 (2003).
250. Valkenburg, S. A. *et al.* Molecular basis for universal HLA-A*0201-restricted CD8+ T-cell immunity against influenza viruses. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 4440–5 (2016).
251. Yang, J. *et al.* CD4+ T cells recognize unique and conserved 2009 H1N1 influenza hemagglutinin epitopes after natural infection and vaccination. *Int. Immunol.* **25**,

- 447–457 (2013).
252. Roti, M. *et al.* Healthy Human Subjects Have CD4+ T Cells Directed Against H5N1 Influenza Virus. (2008).
253. Rothbard, J. B. *et al.* Structural model of HLA-DR1 restricted T cell antigen recognition. *Cell* **52**, 515–23 (1988).
254. Angelo, D. B. S., Robinson, E., Janeway, C. A. & Denzin, L. K. Recognition of core and flanking amino acids of MHC class II-bound peptides by the T cell receptor. 2510–2520 (2002).
255. O'Brien, C., Flower, D. R. & Feighery, C. Peptide length significantly influences in vitro affinity for MHC class II molecules. *Immunome Res.* **4**, 1–7 (2008).
256. Unanue, E. R. *Antigen Presentation in the Autoimmune Diabetes of the NOD Mouse. Annual Review of Immunology* **32**, (2014).
257. Gallimore, A. *et al.* A protective cytotoxic T cell response to a subdominant epitope is influenced by the stability of the MHC class I/peptide complex and the overall spectrum of viral peptides generated within infected cells. *Eur. J. Immunol.* **28**, 3301–3311 (1998).
258. Gallimore, A. *et al.* Induction and exhaustion of lymphocytic choriomeningitis virus-specific cytotoxic T lymphocytes visualized using soluble tetrameric major histocompatibility complex class I-peptide complexes. *J. Exp. Med.* **187**, 1383–93 (1998).
259. Engels, B. *et al.* Relapse or eradication of cancer is predicted by peptide-major histocompatibility complex affinity. *Cancer Cell* **23**, 516–526 (2013).
260. van Ham, S. M. *et al.* Human histocompatibility leukocyte antigen (HLA)-DM edits peptides presented by HLA-DR according to their ligand binding motifs. *J Exp.Med.* **184**, 2019–2024 (1996).
261. Andreatta, M. *et al.* Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. *Immunogenetics* **67**, 641–650 (2015).
262. Lundegaard, C. *et al.* NetMHC-3.0: accurate web accessible predictions of

- human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic Acids Res.* **36**, W509-12 (2008).
263. Lundegaard, C., Lund, O., Buus, S. & Nielsen, M. Major histocompatibility complex class I binding predictions as a tool in epitope discovery. *Immunology* **130**, 309–18 (2010).
264. Karosiene, E., Lundegaard, C., Lund, O. & Nielsen, M. NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. *Immunogenetics* **64**, 177–86 (2012).
265. Hammer, J., Takacs, B. & Sinigaglia, F. Identification of a motif for HLA-DR1 binding peptides using M13 display libraries. *J. Exp. Med.* **176**, 1007–13 (1992).
266. Reid, R. a *et al.* CD8(+) T-cell recognition of a synthetic epitope formed by t-butyl modification. *Immunology* 495–505 (2014). doi:10.1111/imm.12398
267. Jones, E. Y., Fugger, L., Strominger, J. L. & Siebold, C. MHC class II proteins and disease: a structural perspective. *Nat. Rev. Immunol.* **6**, 271–282 (2006).
268. Yang, J. *et al.* Multiplex mapping of CD4 T cell epitopes using class II tetramers. *Clin. Immunol.* **120**, 21–32 (2006).
269. Yang, J. *et al.* Expression of HLA-DP0401 molecules for identification of DP0401 restricted antigen specific T cells. *J. Clin. Immunol.* **25**, 428–436 (2005).
270. Chow, I. T. *et al.* DRB1*12:01 presents a unique subset of epitopes by preferring aromatics in pocket 9. *Mol. Immunol.* **50**, 26–34 (2012).
271. Broughton, S. E. *et al.* Biased T Cell Receptor Usage Directed against Human Leukocyte Antigen DQ8-Restricted Gliadin Peptides Is Associated with Celiac Disease. *Immunity* **37**, 611–621 (2012).
272. Cameron, T. O., Cohen, G. B., Islam, S. A. & Stern, L. J. Examination of the highly diverse CD4+ T-cell repertoire directed against an influenza peptide: A step towards TCR proteomics. *Immunogenetics* **54**, 611–620 (2002).
273. Kuwana, M. Autoreactive CD4 T cells to b 2 -glycoprotein I in patients with antiphospholipid syndrome. *Autoimmun. Rev.* **2**, 192–198 (2003).
274. Madi, A. *et al.* T-cell receptor repertoires share a restricted set of public and

- abundant CDR3 sequences that are associated with self-related immunity. *Genome Res.* **24**, 1603–1612 (2014).
275. Venturi, V. *et al.* TCR beta-chain sharing in human CD8+ T cell responses to cytomegalovirus and EBV. *J. Immunol.* **181**, 7853–7862 (2008).
276. Quigley, M. F. *et al.* Convergent recombination shapes the clonotypic landscape of the naive T-cell repertoire. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 19414–9 (2010).
277. Venturi, V. *et al.* Sharing of T cell receptors in antigen-specific responses is driven by convergent recombination. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 18691–6 (2006).
278. Zhou, V. *et al.* The functional CD8 T cell memory recall repertoire responding to the influenza A M158-66 epitope is polyclonal and shows a complex clonotype distribution. *Hum. Immunol.* **74**, 809–817 (2013).
279. Moss, P. A. *et al.* Extensive conservation of alpha and beta chains of the human T-cell antigen receptor recognizing HLA-A2 and influenza A matrix peptide. *Proc. Natl. Acad. Sci.* **88**, 8987–90 (1991).
280. Petrova, G. V, Naumova, E. N. & Gorski, J. The polyclonal CD8 T cell response to influenza M158-66 generates a fully connected network of cross-reactive clonotypes to structurally related peptides: a paradigm for memory repertoire coverage of novel epitopes or escape mutants. *J. Immunol.* **186**, 6390–6397 (2011).
281. Grant, E. J. *et al.* Lack of heterologous cross-reactivity towards HLA-A*02:01 restricted viral epitopes is underpinned by distinct $\alpha\beta$ T cell receptor signatures. *J. Biol. Chem.* jbc.M116.753988 (2016). doi:10.1074/jbc.M116.753988
282. Menezes, J. S. *et al.* A public T cell clonotype within a heterogeneous autoreactive repertoire is dominant in driving EAE. *J. Clin. Invest.* **117**, 2176–2185 (2007).
283. Fazilleau, N. *et al.* Persistence of autoreactive myelin oligodendrocyte glycoprotein (MOG)-specific T cell repertoires in MOG-expressing mice. *Eur. J. Immunol.* **36**, 533–543 (2006).
284. Petersen, J. *et al.* T-cell receptor recognition of HLA-DQ2-gliadin complexes associated with celiac disease. *Nat. Struct. Mol. Biol.* **21**, 480–8 (2014).

285. Holland, C. J., Cole, D. K. & Godkin, A. Re-Directing CD4(+) T Cell Responses with the Flanking Residues of MHC Class II-Bound Peptides: The Core is Not Enough. *Front. Immunol.* **4**, 172 (2013).
286. Benati, D. *et al.* Public T cell receptors confer high-avidity CD4 responses to HIV controllers. *J. Clin. Invest.* **126**, 2093–2108 (2016).
287. Yoshida, K. *et al.* Restricted T-cell receptor β -chain usage by T cells autoreactive to β 2-glycoprotein I in patients with antiphospholipid syndrome. *Blood* **99**, (2002).
288. Gallimore, a, Dumrese, T., Hengartner, H., Zinkernagel, R. M. & Rammensee, H. G. Protective immunity does not correlate with the hierarchy of virus-specific cytotoxic T cell responses to naturally processed peptides. *J. Exp. Med.* **187**, 1647–1657 (1998).
289. Rossjohn, J. *et al.* T cell antigen receptor recognition of antigen-presenting molecules. *Annu. Rev. Immunol.* **33**, 169–200 (2015).
290. Dushek, O. Elementary Steps in T Cell Receptor Triggering. *Frontiers in Immunology* **2**, (2012).
291. Birnbaum, M. E. E. *et al.* Deconstructing the Peptide-MHC Specificity of T Cell Recognition. *Cell* **157**, 1073–1087 (2014).
292. Cohn, M. Whither T-suppressors: if they didn't exist would we have to invent them? *Cell. Immunol.* **227**, 81–92 (2004).
293. Garcia, K. C., Adams, J. J., Feng, D. & Ely, L. K. The molecular basis of TCR germline bias for MHC is surprisingly simple. *Nat. Immunol.* **10**, 143–7 (2009).
294. Song, I. *et al.* Broad TCR repertoire and diverse structural solutions for recognition of an immunodominant CD8+ T cell epitope. *Nat. Struct. & Mol. Biol.* (2017). doi:10.1038/nsmb.3383
295. Broughton, S. E. *et al.* Biased T cell receptor usage directed against human leukocyte antigen DQ8-restricted gliadin peptides is associated with celiac disease. *Immunity* **37**, 611–21 (2012).
296. Danke, N. a & Kwok, W. W. HLA class II-restricted CD4+ T cell responses directed against influenza viral antigens postinfluenza vaccination. *J. Immunol.* **171**, 3163–

- 9 (2003).
297. Cole, D. K. *et al.* Human TCR-binding affinity is governed by MHC class restriction. *J. Immunol.* **178**, 5727–34 (2007).
298. Willcox, B. E. *et al.* TCR binding to peptide-MHC stabilizes a flexible recognition interface. *Immunity* **10**, 357–65 (1999).
299. Munoz, F. M. The impact of influenza in children. *Semin. Pediatr. Infect. Dis.* **13**, 72–8 (2002).
300. Elderfield, R. A. *et al.* Accumulation of human-adapting mutations during circulation of A(H1N1)pdm09 influenza virus in humans in the United Kingdom. *J. Virol.* **88**, 13269–83 (2014).
301. McMichael, A. J., Gotch, F. M., Noble, G. R. & Beare, P. A. S. Cytotoxic T-Cell Immunity to Influenza. *N. Engl. J. Med.* **309**, 13–17 (1983).
302. Nayak, J. L. *et al.* CD4+ T-cell expansion predicts neutralizing antibody responses to monovalent, inactivated 2009 pandemic influenza A(H1N1) virus subtype H1N1 vaccine. *J. Infect. Dis.* **207**, 297–305 (2013).
303. Nayak, J. & Richards, K. The effect of influenza A (H5N1) pre-pandemic priming on CD4 T-cell responses. *J. Infect. ...* 1–29 (2014).
304. Soema, P. C., Kompier, R., Amorij, J.-P. & Kersten, G. F. A. Current and next generation influenza vaccines: Formulation and production strategies. *Eur. J. Pharm. Biopharm.* **94**, 251–263 (2015).
305. Baz, M. *et al.* Nonreplicating Influenza A Virus Vaccines Confer Broad Protection against Lethal Challenge. *MBio* **6**, e01487-15 (2015).
306. Morgan, S. B. *et al.* Aerosol Delivery of a Candidate Universal Influenza Vaccine Reduces Viral Load in Pigs Challenged with Pandemic H1N1 Virus. *J. Immunol.* **196**, 5014–23 (2016).
307. Francis, J. N. *et al.* A novel peptide-based pan-influenza A vaccine: A double blind, randomised clinical trial of immunogenicity and safety. *Vaccine* **33**, 396–402 (2015).
308. Lillie, P. J. *et al.* Preliminary assessment of the efficacy of a T-cell-based influenza

- vaccine, MVA-NP+M1, in humans. *Clin. Infect. Dis.* **55**, 19–25 (2012).
309. Antrobus, R. D. *et al.* Clinical Assessment of a Novel Recombinant Simian Adenovirus ChAdOx1 as a Vectored Vaccine Expressing Conserved Influenza A Antigens. *Mol. Ther.* **22**, 668–674 (2014).
310. Gurwith, M. *et al.* Safety and immunogenicity of an oral, replicating adenovirus serotype 4 vector vaccine for H5N1 influenza: a randomised, double-blind, placebo-controlled, phase 1 study. *Lancet Infect. Dis.* **13**, 238–250 (2013).