Plasticity at the T cell receptor-peptide-major histocompatibility complex class I interface

A thesis submitted to Cardiff University for the candidature of Doctor of Philosophy

January 2021

Jade Rhianna Hopkins
Acknowledgements

The research described in this thesis would not have been possible without the contribution and support of my supervisory team, colleagues, family, and friends. Any attempt at writing would be insufficient to express the extent of my gratitude but I attempt:

Foremost, I would like to thank my supervisor Professor Andrew Sewell for allowing me this opportunity, for the continuous support and guidance over the years, and for always finding time for me – especially in these last zoom-filled few months! Thank you to my co-supervisors Dr David Cole and Dr Chris Pudney for establishing the GW4 collaboration, and your respective protein biochemistry and REES expertise. I would also like to express immeasurable gratitude to Dr Garry Dolton for your knowledge, patience, and guidance. Your dedication is inspiring and I cannot thank you enough.

I would like to acknowledge my colleagues and collaborators. Thank you to all of the T cell modulation group past and present, but a special mention to Anna, Thalia, and Mat for teaching me ‘how to lab’, and to Tom for computational help with the directed TCR mutants. Thank you to Dr Barbara Szomolay for developing the PS-CPL tool that was vital to this thesis. I would also like to thank all of the Pudney group for being so welcoming during my time at Bath, particularly Dr Dragana Catici for teaching me how to use the crazy sailor machine. I am grateful to the CBS team, Dr Catherine Naseriyan and Dr Ann Kift-Morgan for your training and continuing support. Special thanks to Professor Colin Dayan and Professor Susan Wong for your help with accessing diabetic blood samples.

Thank you to the Medical Research Council and the Wellcome Trust for funding my PhD, and to all at the GW4 BioMed DTP team for your support. I would like to acknowledge the support of Professor Ann Ager, Professor Phil Tayor, and Dr Jason Twohig for taking the time to read my annual progress reports and always providing helpful advice.

Finally, thank you to Mum, Dad, Grandma, Ryan, Jess, Hannah, and Annie for reminding me that the world outside the lab exists. My last, biggest, and most inexpressible thank you to Aaron for literally everything, but in particular the endless cups of tea and reassurance, and to my cats Liam and Kevin for being completely indifferent to the whole thing and offering no help whatsoever.

Diolch o galon i chi gyd.
This thesis is dedicated to Gloria Jean Richards.

“I cannot fix on the hour, or the spot, or the look, or the words, which laid the foundation. It is too long ago. I was in the middle before I knew that I had begun.”

– Jane Austen, Pride and Prejudice.
Publications


Summary

**Background** – The interaction between the T cell receptor (TCR) on the surface of αβ T cells and the peptide-major histocompatibility complex (pMHC) on the surface of target cells helps αβ T cells to defend the host from virtually any foreign pathogen. To achieve such extensive coverage necessitates substantial T cell crossreactivity. While essential for providing complete immune coverage, T cell crossreactivity can also have negative consequences as it is believed to contribute to a large number of autoimmune diseases and can have fatal consequences when T cells are used therapeutically. The therapeutic deployment of the TCR-pMHC interaction therefore requires thorough understanding of the biochemical characteristics that underpin the interaction.

**Results** – I began my studies by defining the crossreactive profile of an autoreactive T cell clone. Using combinatorial peptide library (CPL) screening, I showed that the HLA-A*0201-restricted, insulin-reactive InsB4 TCR, derived from a patient with type 1 diabetes (T1D) could be strongly activated by peptides arising from three common human pathogens. I further showed that EBV-insulin crossreactivity could be found in some T1D patients but not in healthy controls. The potential for virus-specific T cells to recognise self-antigens in autoimmune patients highlights the importance of understanding the underlying driving features of TCR plasticity. To this end, I set about investigating how two biochemical characteristics can influence the TCR-pMHC interaction, TCR sequence and pMHC flexibility. To explore how TCR sequence impacts TCR function, I profiled the crossreactive nature of an HLA-A*0201-restricted, HIV-1-specific TCR and compared its peptide recognition repertoire to mutant versions of the same TCR with CDR3 amino acid substitutions. These studies inadvertently resulted in the discovery that this TCR could also respond a self-antigen expressed on many HLA-A*0201+ cancer cell lines. Recognition of cancer cells, but not the cognate HIV-derived peptide, could be removed by a single CDR3 amino acid substitution. Finally, I studied how the dynamic flexibility of pMHC could be influenced by the peptide cargo by using red-edge excitation shift to examine a collection of well-studied analogues of a preproinsulin-derived peptide in the context of HLA A*0201.

**Conclusions** – My research highlights both the negative (viral/autoimmunity) and the positive (viral/cancer) aspects of T cell crossreactivity and demonstrates that plasticity in both the TCR and cognate pMHC ligands are likely to play a role in the range of TCR-pMHC interactions that are capable of triggering a T cell response.
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_2$M</td>
<td>Beta 2 microglobin</td>
</tr>
<tr>
<td>ABC</td>
<td>Antigen-binding cassette</td>
</tr>
<tr>
<td>ACT</td>
<td>Adoptive cell transfer</td>
</tr>
<tr>
<td>ADR</td>
<td>Alloimmune defence receptor</td>
</tr>
<tr>
<td>AIDS</td>
<td>Acquired immunodeficiency syndrome</td>
</tr>
<tr>
<td>AMIGO</td>
<td>Amphoterin-induced protein 2 precursor</td>
</tr>
<tr>
<td>APC</td>
<td>Antigen presenting cell</td>
</tr>
<tr>
<td>APL</td>
<td>Altered peptide ligand</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>bNAb</td>
<td>Broadly neutralising antibody</td>
</tr>
<tr>
<td>C</td>
<td>Celsius</td>
</tr>
<tr>
<td>CAR</td>
<td>Chimeric antigen receptor</td>
</tr>
<tr>
<td>Cas9</td>
<td>CRISPR-associated protein-9</td>
</tr>
<tr>
<td>CCR5</td>
<td>C-C chemokine receptor type 5</td>
</tr>
<tr>
<td>CD</td>
<td>Cluster of differentiation</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary deoxyribonucleic acid</td>
</tr>
<tr>
<td>CDR</td>
<td>Complementarity-determining region</td>
</tr>
<tr>
<td>CFSE</td>
<td>Carboxyfluorescein succinimidyl ester</td>
</tr>
<tr>
<td>CMV</td>
<td>Cytomegalovirus</td>
</tr>
<tr>
<td>COVID</td>
<td>Coronavirus disease</td>
</tr>
<tr>
<td>CPL</td>
<td>Combinatorial peptide library screen</td>
</tr>
<tr>
<td>CRISPR</td>
<td>Clustered regularly interspaced short palindromic repeats</td>
</tr>
<tr>
<td>Cr</td>
<td>Chromium</td>
</tr>
<tr>
<td>CSM</td>
<td>Centre of spectral mass</td>
</tr>
<tr>
<td>CST</td>
<td>Clonal selection theory</td>
</tr>
<tr>
<td>CTLA-4</td>
<td>Cytotoxic T-lymphocyte-associated protein 4</td>
</tr>
<tr>
<td>CXCR4</td>
<td>C-X-C chemokine receptor type 4</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco’s Modified Eagle Medium</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl-sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EBV</td>
<td>Epstein-Barr virus</td>
</tr>
<tr>
<td>EC</td>
<td>Half maximal effective concentration</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-linked immunosorbent assays</td>
</tr>
<tr>
<td>ELISPOT</td>
<td>Enzyme-linked immunospot</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic reticulum</td>
</tr>
<tr>
<td>ERAAAP</td>
<td>ER aminopeptidase associated with antigen processing</td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescence-activated cell sorting</td>
</tr>
<tr>
<td>FBS</td>
<td>Foetal bovine serum</td>
</tr>
<tr>
<td>FITC</td>
<td>Fluorescein isothiocyanate</td>
</tr>
<tr>
<td>FPLC</td>
<td>Fast Protein Liquid Chromatography</td>
</tr>
<tr>
<td>GAD</td>
<td>Glutamic acid decarboxylase</td>
</tr>
<tr>
<td>GBM</td>
<td>Glomerular basement membranes</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
</tbody>
</table>
GPU: Graphics processing unit
H: Hour
HeBS: HEPES-buffered saline
HCMV: Human cytomegalovirus
HDX: Hydrogen/deuterium exchange
HER-2: Human epidermal growth factor receptor 2
HIV: Human immunodeficiency virus
HLA: Human leukocyte antigen
HSV: Herpes simplex virus
HTLV: Human T-lymphotropic virus
ICAM: Intercellular adhesion molecule
IFN: Interferon
Ig: Immunoglobulin
IL: Interleukin
IMP-2: Insulin-like growth factor 2 mRNA binding protein-2
IPTG: Isopropyl β-D-1-thiogalactopyranoside
IRES: Internal ribosomal entry sites
ITAM: Immunoreceptor tyrosine-based activation motif
IU: International units
IV: Influenza virus
K_d: Dissociation constant
LB: Lysogeny broth
LCL: Lymphoblastic cell lines
M1: Matrix protein 1
MACS: Magnetic activated cell-sorting
MAGE: Melanoma antigen
MAIT: Mucosal-associated invariant T cell
MART-1: Melan-A/Melanoma antigen recognised by T cells
MBP: Myelin basic protein
MD: Molecular dynamics
MFI: Mean fluorescence intensity
MHC: Major histocompatibility complex
MIP: Macrophage inflammatory protein
Min: Minutes
mL: Millilitre
MM: Malignant melanoma
MR1: Major histocompatibility class I-related
mRNA: Messenger ribonucleic acid
NEIB: National eye institute bank
NFAT: Nuclear factor of activated T cells
NMR: Nuclear magnetic resonance
NOD: Non-obese diabetic
NR: No response
Ns: Non-significant
NY-ESO: New York esophageal squamous cell carcinoma-1
OD: Optical density
PBMC: Peripheral blood mononuclear cells
PBS: Phosphate buffered saline
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PD-1/2</td>
<td>Programmed cell death protein-1/2</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein data bank</td>
</tr>
<tr>
<td>PD-L</td>
<td>Programmed death ligand</td>
</tr>
<tr>
<td>PE</td>
<td>Phycoerythrin</td>
</tr>
<tr>
<td>PerCP</td>
<td>Peridinin chlorophyll protein</td>
</tr>
<tr>
<td>PFA</td>
<td>Paraformaldehyde</td>
</tr>
<tr>
<td>PHA</td>
<td>Phytohaemagglutinin</td>
</tr>
<tr>
<td>PLC</td>
<td>Peptide loading complex</td>
</tr>
<tr>
<td>pMHC</td>
<td>Peptide-major histocompatibility complex</td>
</tr>
<tr>
<td>PPI</td>
<td>Preproinsulin</td>
</tr>
<tr>
<td>RAG</td>
<td>Recombination activating gene</td>
</tr>
<tr>
<td>RBC</td>
<td>Red blood cell</td>
</tr>
<tr>
<td>rCD2</td>
<td>Rat CD2</td>
</tr>
<tr>
<td>REES</td>
<td>Red edge excitation shift</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RPM</td>
<td>Revolutions per minute</td>
</tr>
<tr>
<td>RPMI</td>
<td>Roswell Park Memorial Institute</td>
</tr>
<tr>
<td>RSS</td>
<td>Recombination signal sequence</td>
</tr>
<tr>
<td>SABR</td>
<td>Signalling and antigen-presenting bifunctional receptor</td>
</tr>
<tr>
<td>SARS-CoV</td>
<td>Severe acute respiratory syndrome-associated coronavirus</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulphate–polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SEM</td>
<td>Standard error of the mean</td>
</tr>
<tr>
<td>SPR</td>
<td>Surface plasmon resonance</td>
</tr>
<tr>
<td>TAA</td>
<td>Tumour-associated antigen</td>
</tr>
<tr>
<td>TACE</td>
<td>TNF converting enzyme</td>
</tr>
<tr>
<td>TAE</td>
<td>TRIS-acetate-EDTA</td>
</tr>
<tr>
<td>TAP</td>
<td>Transporter associated with antigen processing</td>
</tr>
<tr>
<td>TAPI-0</td>
<td>TNF processing inhibitor-0</td>
</tr>
<tr>
<td>TCR</td>
<td>T cell receptor</td>
</tr>
<tr>
<td>TdT</td>
<td>Terminal deoxynucleotidyl transferase</td>
</tr>
<tr>
<td>TE</td>
<td>TRIS-EDTA</td>
</tr>
<tr>
<td>TGF</td>
<td>Transforming growth factor</td>
</tr>
<tr>
<td>TIL</td>
<td>Tumour-infiltrating lymphocytes</td>
</tr>
<tr>
<td>TNF</td>
<td>Tumour necrosis factor</td>
</tr>
<tr>
<td>Tregs</td>
<td>T regulatory cells</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>V</td>
<td>Volts</td>
</tr>
<tr>
<td>VDJ</td>
<td>Variable, diversity, joining region</td>
</tr>
<tr>
<td>VZV</td>
<td>Varicella-zoster virus</td>
</tr>
<tr>
<td>WBS</td>
<td>Welsh blood service</td>
</tr>
<tr>
<td>WPRE</td>
<td>Woodchuck hepatitis post-transcriptional regulatory element</td>
</tr>
<tr>
<td>WT</td>
<td>Wildtype</td>
</tr>
</tbody>
</table>

All units of measurement in this thesis follow the International System of Units, unless otherwise specified. Amino acids are referred to by their standard single- or three-letter code.
Table of contents

1 – Introduction .................................................................................................................. 1
  1.1 – The immune system ............................................................................................... 1
    1.1.1 – Overview ........................................................................................................ 1
  1.2 – αβ T cells ............................................................................................................. 2
    1.2.1 – Overview of T cell types .............................................................................. 2
    1.2.2 – αβ CD8+ T cell function ........................................................................... 3
    1.2.3 – TCR gene rearrangement .......................................................................... 7
    1.2.4 – Maturation and thymic selection .............................................................. 9
  1.3 – The major histocompatibility complex (MHC) .................................................. 10
    1.3.1 – Overview ....................................................................................................... 10
    1.3.2 – The HLA loci .............................................................................................. 11
    1.3.3 – pMHC structure .......................................................................................... 12
    1.3.4 – Peptide presentation by MHC .................................................................... 15
  1.4 – The TCR-pMHC interaction ................................................................................. 20
    1.4.1 – The co-receptors .................................................................................... 24
    1.4.2 – MHC-restriction .......................................................................................... 26
    1.4.3 – Specificity and crossreactivity within the TCR-pMHC interaction .......... 30
    1.4.4 – How is crossreactivity achieved? ............................................................... 31
  1.5 – Application of the TCR-pMHC interaction ......................................................... 32
    1.5.1 – The TCR-pMHC interaction in immunity ............................................... 32
    1.5.2 – CD8+ T cells in autoimmunity ................................................................... 33
  1.6 – Aims and objectives ............................................................................................ 41

2 – Materials and methods .............................................................................................. 43
  2.1 – Protein production .............................................................................................. 43
    2.1.1 – Buffers and media used in protein production ..................................... 43
    2.1.2 – Vectors and protein sequences ................................................................. 45
    2.1.3 – Making chemically competent cells ..................................................... 46
    2.1.4 – Transformation of competent cells ...................................................... 47
    2.1.5 – Amplifying DNA via mini preparation .................................................. 47
    2.1.6 – Expression of constructs and purification of inclusion bodies .............. 48
    2.1.7 – Refolding soluble protein from inclusion bodies ................................... 51
    2.1.8 – Purification of soluble protein by fast protein liquid chromatography (FPLC) .. 51
  2.2 – Biochemical assays ............................................................................................. 53
2.2.1 – Dynamic light scattering to exclude aggregated samples .................................. 53
2.2.2 – Red edge excitation measurements of HLA-A2 complexes ............................. 53
2.2.3 – Buffer exchange using a PD-10 desalting column ........................................... 53
2.2.4 – Pressure/temperature dependent fluorimetry .................................................. 54
2.2.5 – Surface plasmon resonance (SPR) .................................................................... 54
2.3 – Molecular cloning .................................................................................................. 54
2.3.1 – Buffers and media used in molecular cloning and lentiviral production .......... 56
2.3.2 – 868 construct and design of mutants ................................................................. 56
2.3.3 – Cloning constructs into the pELNS lentiviral vector ......................................... 57
2.3.4 – Sequencing constructs for validation ............................................................... 61
2.3.5 – Amplifying DNA via maxi preparation ............................................................. 62
2.4 – Lentiviral production and cell transduction .......................................................... 63
2.4.1 – Production of lentiviruses in HEK 293T cells ................................................... 63
2.4.2 – Transduction of immortal cell lines ................................................................. 64
2.4.3 – Concentration of lentiviral supernatant ........................................................... 64
2.4.4 – Transducing primary CD8 T cells ..................................................................... 64
2.4.5 – Purification based on surface marker expression ............................................ 65
2.5 – Cell culture ........................................................................................................... 66
2.5.1 – Buffers and media used in cell culture work throughout this thesis ............... 66
2.5.2 – Culturing of adherent cells lines ..................................................................... 67
2.5.3 – Culturing of suspension cell lines ..................................................................... 67
2.5.4 – Isolation of CD8⁺ primary T cells from whole blood ......................................... 68
2.5.5 – T cell purification based on multimer-binding specificity ................................. 70
2.5.6 – Expansion of CD8⁺ primary T cells and subsequent culturing ....................... 70
2.5.7 – Cell counting .................................................................................................... 71
2.5.8 – Freezing cells for storage in liquid nitrogen .................................................... 71
2.6 – Cell-based functional assays .................................................................................. 73
2.6.1 – Antibodies used for flow cytometry throughout this thesis ............................... 73
2.6.2 – Flow cytometry: staining for surface markers .................................................. 74
2.6.3 – Flow cytometry: pMHC multimer staining ....................................................... 75
2.6.4 – Enzyme-linked immunosorbert assays (ELISA) of T cell function .................. 78
2.6.5 – Combinatorial peptide library (CPL) screens ................................................. 80
2.6.6 – TNF processing inhibitor-0 (TAPI-0) assays .................................................... 80
2.6.7 – Chromium-51 (^{51}Cr)-release cytotoxicity assay ............................................ 81
3 – Insulin-specific T cells isolated from patients with type 1 diabetes strongly recognise common viral epitopes ........................................................................ 82
3.1 – Introduction .................................................................................................................. 82
3.1.1 – Type 1 diabetes ......................................................................................................... 82
   3.1.1.1 – T1D is a CD8⁺ T cell mediated disease .............................................................. 83
3.1.2 – T cells are, by necessity, highly crossreactive ....................................................... 85
   3.1.2.1 – Benefits of T cell crossreactivity ...................................................................... 86
   3.1.2.2 – Drawbacks of T cell crossreactivity .................................................................. 87
3.1.3 – Measuring T cell crossreactivity ............................................................................ 89
   3.1.3.1 – CPL screens for measurement of T cell crossreactivity ................................. 91
   3.1.3.2 – CPL screens can identify novel peptide epitopes ........................................... 92
3.1.4 – Aims ......................................................................................................................... 94
3.2 – Results ......................................................................................................................... 95
   3.2.1 – Work leading up to the project ............................................................................ 95
   3.2.1.1 – The HLA-A2-restricted HLVEALYLV epitope is presented at the surface of pancreatic β cells ........................................................................................................... 95
   3.2.1.2 – CPL scanning identifies the peptide repertoire of the InsB4 T cell clone ......... 97
   3.2.1.3 – Fungal-derived peptide agonists of the InsB4 T cell clone .............................. 99
   3.2.1.4 – Bacterial-derived peptide agonists of the InsB4 T cell clone ......................... 101
   3.2.2 – Fungal and bacterial peptides act as functional ligands for the InsB4 T cell clone ......................................................................................................................... 103
   3.2.3 – Peptide sequences derived from human viruses act as potent agonists of the InsB4 T cell ............................................................................................................... 105
   3.2.3.1 – The InsB4 TCR binds to viral epitopes strongly ............................................. 108
   3.2.3.2 – Viral epitopes are correctly processed and presented ...................................... 110
3.2.4 – InsB4 recognises EBV-infected HLA-A2⁺ B cells (LCL lines) ............................... 112
3.2.5 – Insulin-specific T cells in an HLA-A2⁺ T1D patient recognise EBV ....................... 113
3.3 – Discussion ..................................................................................................................... 117
   3.3.1 – The InsB4 T cell kills human pancreatic β cells .................................................. 117
   3.3.2 – Molecular mimicry may trigger autoimmune disease ...................................... 117
   3.3.3 – InsB4 can be activated by pathogen-derived peptides ...................................... 118
   3.3.4 – Viral epitopes strongly activate InsB4 T cells .................................................... 119
   3.3.5 – T1D/insulin crossreactivity in T1D patients ....................................................... 121
   3.3.6 – Summary .............................................................................................................. 122
4 – Exploring how mutation of key peptide contact residues in CDR3 impact TCR peptide specificity ................................................................................................................ 123
   4.1 – Introduction .............................................................................................................. 123
   4.1.1 – The 868 TCR ....................................................................................................... 123
   4.1.2 – Probing TCR ‘rules of engagement’ .................................................................... 126
4.1.3 – Aims .................................................................................................................. 127
4.2 – Results ................................................................................................................ 129
4.2.1 – Designing the mutant 868 TCRs ................................................................. 129
  4.2.1.1 – TCR 3, BT96F ......................................................................................... 134
  4.2.1.2 – TCR 6, BV97T ......................................................................................... 134
  4.2.1.3 – TCR 22, BD95S ....................................................................................... 134
  4.2.1.4 – TCR 24, BD95I T96R V97S ................................................................. 134
4.2.2 – Validation of 868 mutants ................................................................. 137
4.2.3 – Minimal differences between the recognition footprints of 868, TCR 3, and TCR 6 ................................................................. 139
4.2.4 – TCR 22 and TCR 24 show differing specificity than the parent 868 TCR ... 142
4.2.5 – Identifying the HLA-A2 self-epitope that the 868 TCR responds to .... 144
4.2.6 – 868 TCR recognises epitopes that are not seen by TCR 22 ................. 147
4.2.7 – Attempts to validate the lengsin epitope ............................................... 149
4.2.8 – Attempts to isolate lengsin-reactive T cells ....................................... 154
4.2.9 – Preliminary characterisation of two further mutants of the 868 TCR ................................................................................................. 157
4.3 – Discussion ......................................................................................................... 163
  4.3.1 – TCR transduced T cells ........................................................................... 163
  4.3.2 – TCRs 3 and 6 showed little change in peptide preference .............. 164
  4.3.3 – TCRs 22 and 24 have altered peptide preferences ......................... 166
  4.3.4 – The 868 TCR may crossreact with a tumour-associated antigen .... 168
  4.3.5 – T cells expressing TCRs 9 and 11 might exhibit new specificities .... 171
  4.3.6 – Future optimisation to reduce polyclonal background noise ......... 171
  4.3.7 – Summary ................................................................................................. 172
5 – Peptide cargo modulates pMHC flexibility .................................................. 175
  5.1 – Introduction ................................................................................................. 175
    5.1.1 – The flexibility of the pMHC is generally overlooked ..................... 175
    5.1.2 – The pMHC as a dynamic molecule .................................................. 175
    5.1.3 – The REES phenomenon .................................................................. 177
    5.1.4 – Aims ..................................................................................................... 179
  5.2 – Results ......................................................................................................... 181
    5.2.1 – Producing stable pMHC complexes ............................................... 181
    5.2.2 – HLA-A2-ALWGPDPAAA complexes show a significant REES signal .... 183
    5.2.3 – ΔCSM varies with different peptides suggesting the molecular flexibility of each complex is different ................................................. 185
    5.2.4 – Peptide cargo modulates energy barriers between conformational states ... 189
5.2.5 – Peptide-MHC contacts might drive global flexibility of the complex .......... 191
5.3 – Discussion........................................................................................................ 193
5.3.1 – Evidence for a flexible pMHC .................................................................... 193
5.3.2 – REES as a tool for exploring pMHC flexibility ............................................ 194
5.3.3 – Summary ....................................................................................................... 195

6 – General Discussion ............................................................................................ 198
6.1 – Understanding the TCR-pMHC interaction .................................................... 198
6.2 – Investigations into the TCR-pMHC interaction ............................................. 198
   6.2.1 – Display systems for studying the TCR-pMHC interface ......................... 199
   6.2.2 – Rational design based on knowledge of the TCR-pMHC interaction ........ 201
   6.2.3 – Further methods for peptide discovery .................................................... 202
      6.2.3.1 – T-Scan ................................................................................................. 202
      6.2.3.2 – Trogocytosis ..................................................................................... 202
      6.2.3.3 – SABR ............................................................................................... 203
   6.2.4 – Peptide-MHC plasticity: The other side of the coin ................................. 203
6.3 – T cell therapy ..................................................................................................... 205
   6.3.1 – Adoptive cell transfer (ACT) .................................................................... 205
   6.3.2 – TCR gene transfer .................................................................................... 207
6.4 – Concluding remarks ......................................................................................... 208

7 – Appendix and supplementary information ......................................................... 210
7.1 – Human immunodeficiency virus ..................................................................... 210
7.2 – Optimised expression of transduced TCRs at the T cell surface .................... 217
7.3 – Optimising TCR gene transfer ....................................................................... 219
7.4 – Supplementary figures ................................................................................... 222

References .............................................................................................................. 232
List of Figures

Figure 1.1: The structure of the αβ TCR ................................................................. 3
Figure 1.2: V(D)J recombination ................................................................. 7
Figure 1.3: The structure of MHC-I .................................................................. 13
Figure 1.4: Peptide loading of MHC .......................................................... 18
Figure 1.5: The TCR-pMHC interaction ...................................................... 22
Figure 1.6: The co-receptors ........................................................................ 25
Figure 1.7: The advantages and disadvantages of crossreactivity ............... 35
Figure 1.8: Molecular mimicry in autoimmunity ......................................... 39
Figure 2.1: Schematic of the pGMT7 plasmid used for protein production .... 46
Figure 2.2: Protein production via *E. coli* expression systems ................. 50
Figure 2.3: Purification of soluble protein .................................................... 52
Figure 2.4: Schematic of surface plasmon resonance ................................... 55
Figure 2.5: TCR construct design for lentiviral work .................................. 57
Figure 2.6: pELNS vector for lentiviral work ............................................. 58
Figure 2.7: Representative colony PCR ....................................................... 61
Figure 2.8: Typical gating strategy for analysis of T cells ......................... 74
Figure 2.9: Schematic explanation of tetramer staining ............................. 76
Figure 2.10: Standard curve for a MIP-1β ELISA ....................................... 78
Figure 3.1: Pathogen-triggered autoimmunity ............................................ 88
Figure 3.2: Typical 9-mer CPL format .......................................................... 91
Figure 3.3: An overview of the TCR ligand discovery process .................... 92
Figure 3.4: The InsB4 T cell clone recognises an insulin-derived epitope .... 95
Figure 3.5: InsB4 recognises and kills pancreatic β cells ........................... 96
Figure 3.6: Characterising the peptide preferences of InsB4 ...................... 97
Figure 3.7: InsB4 recognises fungal peptides with greater sensitivity than the index peptide .......................................................... 99
Figure 3.8: InsB4 recognises bacterial peptides with greater sensitivity than the index peptide .......................................................... 101
Figure 3.9: Pure peptide titrations for the top pathogenic peptides .......... 102
Figure 3.10: InsB4 TCR binding of pathogen-derived agonist peptides is confirmed by SPR and pMHC multimer staining................................................................. 104

Figure 3.11: InsB4 recognises viral peptides with greater sensitivity than the index peptide.................................................................................................................. 106

Figure 3.12: Viral-derived peptides act as strong agonists for the insulin-specific InsB4 T cell clone........................................................................................................... 107

Figure 3.13: InsB4 TCR binding of pathogenic agonists is confirmed by SPR and multimer staining................................................................................................. 109

Figure 3.14: Putative viral epitopes are efficiently processed and presented to InsB4 T cells in the context of HLA-A2........................................................................ 111

Figure 3.15: InsB4 responds to EBV-transformed lymphoblastic cell lines (LCLs) in an HLA-A2 dependent manner........................................................................ 112

Figure 3.16: Insulin-specific cells in an HLA-A2’ T1D donor can recognise three viral epitopes.................................................................................................................. 114

Figure 3.17: EBV/Insulin B chain crossreactivity is not present in healthy donors........... 116

Figure 4.1: The interaction between 868 and HLA-A2-SLYNTVAL........................................ 130

Figure 4.2: The classification of amino acids based on chemical characteristics............. 132

Figure 4.3: The 868 mutants used in this chapter................................................................ 136

Figure 4.4: Validating expression of the lentiviral constructs........................................... 138

Figure 4.5: TCRs 3 and 6 show little variation in their ability to recognise the cognate HLA-A2-SLYNTVAL epitope........................................................................ 140

Figure 4.6: TCRs 3 and 6 show little variation under CPL screening.................................... 141

Figure 4.7: 868, TCRs 22 and 24 show differences in peptide preference........................ 143

Figure 4.8: CPL screening of TCR 22 and 24 in comparison to 868................................. 145

Figure 4.9: 868 and TCR 22 are predicted to have varying peptide preference footprints .................................................................................................................. 146

Figure 4.10: Recognition of two putative self-derived epitopes by 868 TCR-expressing cells appears to have been ablated by the βD95S mutation in TCR 22............... 148

Figure 4.11: Attempts to validation the lengsin-derived epitope SQILTPPQL...................... 151

Figure 4.12: 868-transduced CD8+ T cells respond to MOLT-3 cells expressing HLA-A2...... 152

Figure 4.13: 868-transduced CD8+ T cells respond to a panel of potential APCs in an HLA-A2-dependent manner.............................................................................. 153

Figure 4.14: Lengsin-reactive T cells were not found in TIL samples................................. 155

Figure 4.15: Potential lengsin-reactive T cells found in healthy PBMC samples............... 156

Figure 4.16: TCR 9 and TCR 11......................................................................................... 158
Figure 4.17: TCRs 9 and 11 exhibit different peptide preferences than the 868 parental TCR

Figure 4.18: Recognition of putative epitopes unique to mutant TCRs

Figure 4.19: Novel specificities in mutant TCRs

Figure 5.1: Production of pure pMHC complexes

Figure 5.2: HLA-A2 complexes show a significant REES signal

Figure 5.3: ΔCSM varies for pMHC complexes carrying different peptide cargo

Figure 5.4: Fluorescence intensity spectra for each HLA-A2 complex

Figure 5.5: ΔG varies for pMHC complexes carrying different peptide cargo

Figure 5.6: Peptide-MHC contacts might drive global flexibility of the complex

Supplementary Figure 7.1: TCR 24-transduced CD8+ T cells do not respond to the M1 influenza-derived GILGFVFTL epitope

Supplementary Figure 7.2: 868-transduced CD8+ T cells respond to untransduced HLA-A2+ MOLT-3 cells, repeat assay

Supplementary Figure 7.3: 868-transduced CD8+ T cells respond to a range of HLA-A2+ immortalised cell lines

Supplementary Figure 7.4: Pressure/temperature-dependence of HLA-A2-ALWGPDPAAA complex

Supplementary Figure 7.5: Pressure/temperature-dependence of HLA-A2-YQFGPDPFPTA complex

Supplementary Figure 7.6: Pressure/temperature-dependence of HLA-A2-RQFGPDWIVA complex

Supplementary Figure 7.7: Pressure/temperature-dependence of HLA-A2-RQWGPDPAAAV complex

Supplementary Figure 7.8: Pressure/temperature-dependence of HLA-A2-MWGPDPPLYV complex

Supplementary Figure 7.9: Pressure/temperature-dependence of HLA-A2-RQFGPDFTI complex
List of Tables

Table 2.1: A summary of the buffers and media used during protein production throughout this thesis.......................................................... 44
Table 2.2: The buffers and media used for molecular cloning and lentiviral production...... 56
Table 2.3: A summary of the primers used throughout this thesis............................................ 62
Table 2.4: Buffers and media used for cell culture throughout this thesis.............................. 67
Table 2.5: A summary of cell lines used in this thesis.............................................................. 68
Table 2.6: The antibodies used for flow cytometry throughout this thesis............................ 72
Table 3.1: The pathogenic peptides used throughout Chapter 3........................................... 119
Table 4.1: Contact table for the 868-A2-SLYNTVATL interaction........................................ 131
Table 4.2: The TCRs used in Chapter 4................................................................................. 133
Table 4.3: CPL data was used to screen the 9-mer peptide universe for the top 20 predicted peptide agonists of each TCR................................................................. 160
Table 5.1: The peptides used in Chapter 5............................................................................. 181
Table 5.2: The change in CSM over the 292-310 nm excitation range for each protein................................................................................. 187
Supplementary Table 7.1 The SASA of the 6 pMHC complexes is similar......................... 225
1 – Introduction

1.1 – The immune system
1.1.1 – Overview
The human immune system is capable of defending against virtually any foreign antigen by means of a highly specific immune response, despite often never having encountered the offending foreign antigen before (Davis and Bjorkman, 1988). Whilst maintaining this extensive protective coverage against harmful pathogens, the immune system must remain tolerant to self and to non-harmful foreign entities such as ingested food (aberrancies manifest as allergies) and commensal bacteria (Bluestone, 2011).

This sophisticated balance between protection and tolerance is achieved via a series of defensive barriers. Anatomical barriers such as the skin and the oral mucosa provide a physical barrier as the first line of defence against infection (Walker, 2004; Bangert, Brunner and Stingl, 2011). If these barriers are breached the complement system – a group of soluble proteins present in blood plasma – can initiate an immune defence (Jordon, 1982). Complement can be triggered directly by recognition of molecules on the surface of a pathogen (Gialeli, Gungor and Blom, 2018), or indirectly by binding to opsonising antibodies (Jordon, 1982). The triggered complement cascade can result in either the direct lysis of pathogens, via formation of the membrane-attack complex (MAC) (Gialeli, Gungor and Blom, 2018), or can facilitate pathogen removal by recruiting phagocytes that express complement receptors (Jordon, 1982).

In addition to the complement system, a pathogen that manages to breach the anatomical barriers must face numerous other immune defences. These are broadly divided into the innate and the adaptive arms of the immune system (Parkin and Cohen, 2001). The innate immune system, which can be triggered by the complement cascade (Jordon, 1982), consists of immune cells such as macrophages and dendritic cells (Parkin and Cohen, 2001). The response of the innate immune system is extremely rapid following a pathogenic threat (Parkin and Cohen, 2001). The innate immune system has been extensively reviewed elsewhere and will not be discussed further here (Medzhitov and Janeway, 2000; Beutler, 2004; Akira, Uematsu and Takeuchi, 2006; Bangert, Brunner and Stingl, 2011; Riera Romo, Pérez-Martínez and Castillo Ferrer, 2016).
The focus of this thesis is αβ T cells which are members of the adaptive immune system. The adaptive immune system, found in jawed vertebrates, evolved after the innate immune system (Pancer and Cooper, 2006). It is populated by lymphocytes known as B and T cells (Parkin and Cohen, 2001). Whilst the humoral immunity provided by B cells is important, the majority of B cells require growth and differentiation factors secreted by T cells (Schimpl and Wecker, 1971, 1972; Hamaoka and Ono, 1986). B cells are reviewed in the following articles (Lebien and Tedder, 2008; Mauri and Bosma, 2012; Hoffman, Lakks and Chalasani, 2016), this introduction will henceforth centre on the discussion of αβ T cells.

1.2 – αβ T cells
1.2.1 – Overview of T cell types
T cells express a T cell receptor (TCR). The TCR is a heterodimer; made up of two of four distinct TCR polypeptides; α, β, γ, or δ. These polypeptides were first described in the 1980s and primarily form two different heterodimers αβ and γδ expressed in αβ T cells and γδ T cells respectively (Kappler et al., 1983; Hedrick et al., 1984; Bank et al., 1986; Brenner et al., 1986; Loh et al., 1987; Davis and Bjorkman, 1988). γδ T cells are extensively reviewed (Xiong and Raulet, 2007; Fahl, Coffey and Wiest, 2014; Yazdanifar et al., 2020). ‘Conventional’ αβ T cells predominate in the human peripheral blood (Davis and Bjorkman, 1988; Zou et al., 2017); their specificity for, and interaction with their target cells is the focus of this thesis.

There are thought to be ~10^{12} αβ T cells in the human body (Arstila et al., 1999). Conventionally, αβ T cells are characterised based on their surface expression of one of two glycoproteins called cluster of differentiation (CD)4 and CD8 (Parnes, 1989). As well as being a means of identification and characterisation, these surface glycoproteins are important to T cell function (Miceli, Von Hoegen and Parnes, 1991; Gavichandran and Burakoff, 1994). The CD4 and CD8 glycoproteins, or ‘co-receptors’, help T cells to receive signals (Doyle and Strominger, 1987; Norment et al., 1988; Rosenstein et al., 1989) and can determine how the T cell reacts, whether with direct cytotoxic functions (CD8^+ T cells), or with assistive or regulatory functions (CD4^+ T cells) (Parnes, 1989; Gavichandran and Burakoff, 1994).

CD4^+ T cells can be further subdivided into cells types; those that assist effector arms of the immune system or those that regulate immunity (Tse and Dutton, 1977; Golubovskaya and Wu, 2016), the latter being important in the prevention of autoimmunity, see Section 1.5.2 (Strelkauskas et al., 1978). In contrast, CD8^+ T cells, often known as cytotoxic T cells, play a more direct role in the removal of pathogens (Nakayama et al., 1979; Zhang and Bevan,
Unorthodoxly, expression of CD4 and CD8 is not always mutually exclusive and CD4+CD8+ T cells have been suggested to play a role in immunity to HIV and cancer (Gallagher, Fazekas de and Miller, 1989; Jiménez et al., 2002; Desfrançois et al., 2010; Frahm et al., 2012).

1.2.2 – αβ CD8+ T cell function
The structure of the αβ TCR is key to understanding αβ TCR function. The first TCR crystal structures, published in 1996 (Garboczi et al., 1996; Garcia et al., 1996), and an improved method for refolding soluble forms of the TCR in 2003 (Boulter et al., 2003) led to greater understanding of αβ TCR structure. The α and β polypeptide chains that comprise the αβ TCR are themselves comprised of several distinct regions (Kappler et al., 1983; Garboczi et al., 1996; Garcia et al., 1996). Each chain has a large extracellular component, a small hydrophobic domain towards the C-terminus of the chain which forms the single membrane-spanning region, and a short cytoplasmic tail (Bentley and Mariuzza, 1996; Garboczi et al., 1996; Garcia et al., 1996). The α and β polypeptide chains are covalently linked by disulphide bonds between their extracellular domains, adjacent to the plasma membrane, see Figure 1.1A (Garboczi et al., 1996; Garcia et al., 1996).

Figure 1.1: The structure of the αβ TCR. A) a schematic representation of the αβ TCR. Showing α chain (dark green) and β chain (light green) linked by an interchain disulphide bond. Each chain has a transmembrane domain and a short cytoplasmic tail. B) 3D crystal structure (868 TCR, PDB: 5NMD) showing the variable and constant domains of the α chain and β chain. Hypervariable CDR loops within the variable domain shown in red. C) The TCR non-covalently associates with a CD3 hexamer to facilitate signalling.
Each TCR chain consists of a variable domain (V) and a constant domain (C), see Figure 1.1B (Garboczi et al., 1996; Garcia et al., 1996, 1998; Garcia, Teyton and Wilson, 1999). The constant (C) domains, present on both the α and β polypeptide chains, are conserved between different TCRs (Garboczi et al., 1996; Ding et al., 1998; Garcia et al., 1998; Garcia, Teyton and Wilson, 1999) (although in the case of the β chain one of two possible, almost identical, constant domains is selected, see Section 1.2.3 (Rowen, Koop and Hood, 1996; Attaf et al., 2015)). A key feature of the variable (V) domains are the hypervariable loops, known as complementarity-determining regions (CDR), Figure 1.1B in red. There are three CDRs on each chain, CDR1-3 (Davis and Bjorkman, 1988; Garboczi et al., 1996; Garcia et al., 1996; Garcia, Teyton and Wilson, 1999). The six CDR loops are responsible for forming the antigen-binding site of the TCR (Garboczi et al., 1996; Ding et al., 1998; Garcia et al., 1998). The αβ TCR is generally restricted in that it usually only recognises peptide antigen when they are presented on major histocompatibility complex (MHC) molecules which are expressed on the surface of other cells, this is known as MHC-restriction, see Section 1.3 (Rosenthal and Shevach, 1973; Zinkernagel and Doherty, 1974; Davis and Bjorkman, 1988). Successful interaction between an αβ TCR and a peptide-MHC (pMHC) can result in T cell activation, and lead to the effective killing of the target cell or the release of soluble immune messengers such as cytokines (Davis and Bjorkman, 1988; Zhang and Bevan, 2011). TCR activation is coupled to intracellular signalling by means of CD3. CD3 is a nonpolymorphic surface-expressed hexamer made up of γ, δ, ε and ζ subunits (Figure 1.1C). The CD3γε-CD3δε-CD3ζζ hexamer non-covalently associates with the TCR constant domain forming an essential part of the signalling complex (Frank et al., 1990; Straus and Weiss, 1993; Ghendler et al., 1998; Dong et al., 2019). The cytoplasmic tails of CD3 subunits contain immunoreceptor tyrosine-based activation motifs (ITAMs) (Reth, 1989; Weiss, 1993; Dong et al., 2019). Upon ligation of a pMHC complex, conformational change is thought to expose CD3 ITAMs to targeting by cytosolic Src family protein tyrosine kinases (such as Lck) (Rudd et al., 1988; Barber et al., 1989; June et al., 1990; Straus and Weiss, 1993). An alternative model of T cell activation, the kinetic segregation model, proposes that in the absence of a target cell, phosphorylation of ITAMs is minimised by the presence of bulky, cell surface-expressed phosphatases such as CD45 and CD148. The close cell-cell contact brought about by TCR-pMHC interaction excludes the large phosphatases from the vicinity of the contact zone, allowing the TCR ITAMs to be phosphorylated (Davis and van der Merwe, 1996, 2006; Van Der Merwe et al., 2000; Choudhuri et al., 2005, 2009). The conformational changes of CD3 and the kinetic segregation of phosphatases may not be mutually exclusive and both
facilitate phosphorylation of CD3 ITAMs which recruits the Syk kinase family member ZAP-70 which propagates downstream signalling (Chan et al., 1992; Bu, Shaw and Chan, 1995). Thus, in this manner the TCR, which has no enzymatic function of its own, is able to initiate a cascade of signalling events that result in activation of the T cell.

Phosphorylation of CD3 ITAMs is assisted by the co-receptors, CD4 and CD8 (Barber et al., 1989; Veillette et al., 1989) because, intracellularly, they are able to bind to Lck (Rudd et al., 1988; Barber et al., 1989; Van Laethem et al., 2007). Extracellularly, CD8 and CD4 assist TCR binding to two different types of MHC molecule; MHC class I and MHC class II respectively (Doyle and Strominger, 1987; Norment et al., 1988) (although despite this generalisation, TCRs have been known to be capable of binding both MHC-I and MHC-II ligands (Yin et al., 2011)). MHCs are discussed in detail in Section 1.3. When a TCR binds a pMHC the CD8/CD4 co-receptor is recruited along with its accompanying pool of sequestered Lck (Veillette et al., 1988; Abraham et al., 1991; Van Laethem et al., 2007). In this way, the CD4/CD8-MHC interaction recruits Lck for the phosphorylation of CD3ζ ITAMS (Purbhoo et al., 2001). Thus, the CD8/CD4 co-receptor both facilitates T cell activation, by providing a concentrated pool of signalling molecules, and simultaneously biases TCR-signalling towards MHC-dependence (Turner et al., 1990; Van Laethem et al., 2007), this raises an interesting debate as to whether the TCR is inherently MHC-restricted, or whether the bias is imposed by the co-receptor (see Section 1.4.2). Interestingly, Lck is thought to associate more frequently with the co-receptors in memory T cells than naïve T cells. Bachmann et al., found that more Lck co-immunoprecipitated with CD8 isolated from memory or effector T cells than CD8 isolated from naïve T cells. The resulting distributional difference between experienced and inexperienced cells may contribute to the more rapid response of the former (Bachmann et al., 1999).

In addition to CD8, CD4, and CD3 there are a number of surface molecules that can also influence the success of the TCR-pMHC interaction. For example, CD80 and CD86 molecules on the surface of antigen presenting cells (APCs), such as dendritic cells, can interact with either CD28 or cytotoxic T-lymphocyte-associated protein 4 (CTLA-4; also known as CD152) on the surface of T cells and have a co-stimulatory or co-inhibitory effect respectively (Linsley, Brady, Grosmaire, et al., 1991; Linsley, Brady, Urnes, et al., 1991). An additional co-stimulatory interaction is the binding of CD2 on the T cell surface to CD58 on the APC surface (Selvaraj et al., 1987; Pan et al., 2019). In contrast, binding of PD-1 on the T cell surface to PD-L1 or PD-L2 on the surface of other cells can provide inhibitory signals (Ishida et al., 1992; Freeman et al., 2000;Latchman et al., 2001; Brown et al., 2003; Rodig et al., 2003). The role
of co-stimulation in regulating T cell activation and preventing autoimmunity is discussed further in Section 1.5.2. Clearly, the success of an interaction between the TCR and its cognate pMHC can be exquisitely fine-tuned through the regulation of accompanying surface molecules.

Following activation, a CD8+ T cell will respond by undergoing clonal expansion – it will proliferate in order to expand the pool of CD8+ T cells that can recognise that specific foreign antigen (Butz and Bevan, 1998; Badovinac, Haring and Harty, 2007). In a positive feedback manor, activated T cells produce a range of proinflammatory cytokines and chemokines that propagate the immune response. For example, interleukin (IL)-2 which promotes cell-cycle progression (Mier and Gallo, 1980; Smith, 1988; Williams, Tyznik and Bevan, 2006) and differentiation (Bachmann et al., 2007; Pipkin et al., 2010), and macrophage inflammatory protein (MIP)-1β (also known as CCL4) a chemoattractant that encourages T cell migration to the site of response (Taub et al., 1993; Bystry et al., 2001; Kenway-Lynch et al., 2014). Activated T cells can also produce interferon (IFN)-γ (Wheelock, 1965; Boehm et al., 1997) and tumour necrosis factor (TNF) (Kolb and Granger, 1968) which have many immune functions including stimulating the upregulation of MHC expression in surrounding cells (Boehm et al., 1997), the suppression of viral replication (Mestan et al., 1986; Bouley et al., 1995; Finke et al., 1995; van den Broek et al., 1995; Seo and Webster, 2002), and enhancing the proliferation of CD8+ T cells (Yokota, Geppert and Lipsky, 1988).

Activated CD8+ T cells also have a more direct role in pathogen removal. They can induce apoptosis in target cells via two main pathways: the secretory perforin-granzyme pathway, and the Fas pathway (Podack, Hengartner and Lichtenheld, 1991; Henkart, 1994; Heusel et al., 1994; Kägi et al., 1994; Ebnet et al., 1995). The former involves Ca2+-dependant degranulation, engagement of the TCR causes the contents of cytosolic granules to be released in an localised manner directly onto the target cell (Weiss et al., 1984; Burkhardt et al., 1993; Berke, 1995). Stored within these granules are perforin – a lytic protein that polymerises to form pores in the surface membrane of the target cell – and granzymes – serine proteases that enter the target cells through the newly created pores and damage cellular machinery to induce the intrinsic apoptotic pathway (Heusel et al., 1994; Kägi et al., 1994; Berke, 1995). The latter involves the upregulation of FasL on the surface of the activated CD8+ T cell (Shresta et al., 1998). The interaction of FasL with Fas (expressed on the surface of the target cell) induces the extrinsic apoptotic pathway in the target cell (Kägi et al., 1995; Jensen et al., 1998).
In summary, the CD8+ T cell is capable of a vast array of antipathogenic activities, and central to it all is the engagement of its TCR.

1.2.3 – TCR gene rearrangement
The TCR repertoire is thought to consist of ~10⁷ unique receptors (Arstila et al., 1999; Turner et al., 2006). A unique gene for every unique TCR would require more genomic material than would fit in the cell nucleus. Instead, this sizeable diversity arises from the random combination of a finite number of gene segments - a process known as somatic gene rearrangement (Hozumi and Tonegawa, 1976). This enables a limited number of genes to give rise to a vast range of different TCRs (Griesser et al., 1988; Alt et al., 1992; Hodges et al., 2003; Attaf et al., 2015). The TCR gene segments lie far apart in the genomic configuration of deoxyribonucleic acid (DNA), but are brought together to form a continuous transcript by somatic recombination (Alt and Baltimore, 1982; Tonegawa, 1983; Hodges et al., 2003).

**Figure 1.2: V(D)J recombination.** **A)** shows the recombination events required to produce a TCR α chain. Recombination occurs between a V segment and a J segment, the intervening DNA is excised. The recombined DNA is transcribed into pre-mRNA. Subsequent RNA splicing places the C segment next to the VJ. **B)** shows the recombination events required to produce a TCR β chain. An additional recombination event, D-J, occurs first. Followed by V-DJ recombination. One of two C segments is juxtapositioned during RNA splicing.
The variable domain of the TCR α-chain (encoded at the tra locus), consists of two recombined gene segments known as variable (V) and junctional (J) gene segments. V-J recombination brings together a randomly selected V segment and a randomly selected J segment (Figure 1.2A) (Market and Papavasiliou, 2003). The DNA is broken and reannealed to create the continuous exon that will encode the V domain of the TCR. The stretch of DNA is transcribed into pre-messenger ribonucleic acid (mRNA), and subsequent RNA splicing removes the genetic material between the VJ sequence and the single constant domain segment present at the tra locus, prior to translation into polypeptide (Alt et al., 1992; Market and Papavasiliou, 2003). There are 70 V segments and 61 J segments at the tra locus, resulting in a large number of potential combinations and giving rise to TCR α-chain diversity (Market and Papavasiliou, 2003).

The variable domain of the TCR β-chain, in contrast, is encoded by three gene segments. At the trb locus there is an additional recombination step in which one of two diversity (D) segments are selected (Griesser et al., 1988; Alt et al., 1992; Arstila et al., 1999; Market and Papavasiliou, 2003), see Figure 1.2B. For the TCR β-chain there are 52 V segments, 13 J segments, and two possible constant segments (Market and Papavasiliou, 2003). Given the vast number of potential combinations, it is not difficult to envision how such diversity arises.

Somatic recombination is strictly regulated and carried out by a group of enzymes collectively called the recombinase. To ensure recombination occurs in the correct places each gene segment is flanked with recombination signal sequences (RSSs) (Alt and Baltimore, 1982; Alt et al., 1992; Bassing, Swat and Alt, 2002; Hodges et al., 2003). The RSSs are recognised by two key members of the recombinase, recombination activating gene (RAG)-1 and RAG-2 (Schatz, Oettinger and Baltimore, 1989; Oettinger et al., 1990; Alt et al., 1992). This recognition was visualised by a cryo-electron microscopic structure in 2015 (Ru et al., 2015).

Recombination at the V-J or V-D-J junctions is imperfect, and nucleotides often get deleted during the process (Sakano et al., 1979; Weigert et al., 1980; Alt and Baltimore, 1982; Alt et al., 1992). V-J or V-D-J junctions are also subject to the addition of palindromic (P)- or non-template (N)-nucleotides by another member of the recombinase – terminal deoxynucleotidyl transferase (TdT) (Alt and Baltimore, 1982; Alt et al., 1992; Hodges et al., 2003; Peralta-Zaragoza, Recillas-Targa and Madrid-Marina, 2004). The resulting heightened diversity at the junctional regions (Tonegawa, 1983; Alt et al., 1992; Gilfillan et al., 1993) gives rise to the ‘hypervariable’ CDR3 loops of both TCR chains (Jorgensen et al., 1992; Hodges et al., 2003). In contrast, the ‘variable’ CDR1 and CDR2 loops of both chains are
germline encoded and arise from variable gene segments (Jorgensen et al., 1992; Hodges et al., 2003).

The rearrangement of the genome is tightly controlled as dysregulation can result in transformation of lymphocytes (Mijušković et al., 2015). V(D)J recombination has been extensively reviewed (Alt and Baltimore, 1982; Griesser et al., 1988; Bassing, Swat and Alt, 2002; Market and Papavasiliou, 2003; Attaf et al., 2015).

1.2.4 – Maturation and thymic selection

Thymocytes are positively selected in the thymic cortex if they are able to recognise a self-pMHC, ligation with the pMHC saves the precursor cell from cell death (Von Boehmer, Haas and Jerne, 1978; Boehmer, 1988). Only those that can recognise self-antigen in the context of self-MHCs survive because those that are unable to recognise a self-peptide are likely to be equally inefficient at recognising foreign antigen, these thymocytes are said to “die by neglect” (Von Boehmer, Teh and Kisielow, 1989; Attaf et al., 2015).

In the thymic medulla, a second phase of thymic selection – known as negative selection – destroys any potentially harmful T cells that recognise pMHC too strongly (Kappler, Roehm and Marrack, 1987; Von Boehmer, Teh and Kisielow, 1989). It is thought to be the strength, or dwell-time, of the TCR-pMHC interaction which ultimately decides cell fate (Kalergis et al., 2001; Gascoigne, 2014; Stepanek et al., 2014). A weak interaction is essential to prevent cell death and promote positive selection however too strong an interaction results in negative selection and apoptosis (Alam et al., 1996; Williams et al., 1999; Starr, Jameson and Hogquist, 2003).

The microenvironment in which the interaction occurs (Koble and Kyewski, 2009; Klein et al., 2014) is also important in determining which thymocytes survive to populate the periphery (Ashton-Rickardt et al., 1994; Sebzda et al., 1994). Epithelial cells in the thymic cortex possess a special proteasome – known as the thymoproteasome – which produces peptides unique
to the thymus for positive selection (Takada, Kondo and Takahama, 2017). With regards to negative selection, the developing T cell repertoire will only be tolerised to antigen it is exposed to. The peripheral pMHCs on display may vary in different microenvironments due to differences in protein expression, post-translational modification, or level of surface presentation (Lanzavecchia, 1995; Salemi et al., 1995; Simitsek et al., 1995; Wood and Elliott, 1998). This may explain the lack of tolerance towards ‘cryptic’ self-antigens (Moudgil and Sercarz, 1994). Exposure of cryptic self-antigens, for example during an infection or inflammation in the local microenvironment (Lanzavecchia, 1995; Salemi et al., 1995), can drive the development of autoimmunity (Lanzavecchia, 1995; Fairchild, Pope and Wraith, 1996), discussed in detail in Section 1.5.2.

In addition to eradicating harmful self-reactive T cells, thymic selection determines another key stage of thymocyte maturation. Prior to positive selection thymocytes express both CD4 and CD8, whether their final phenotype will be CD8⁺ or CD4⁺ depends on whether they encounter a peptide presented on MHC-I or MHC-II respectively (Sha et al., 1988; Scott et al., 1989; Von Boehmer, Teh and Kisielow, 1989).

1.3 – The major histocompatibility complex (MHC)

1.3.1 – Overview

The MHCs were first discovered in the mouse in 1936 (Gorer, 1936; Klein, 1986). They are a family of cell surface-expressed glycoproteins (Parham et al., 1977; Bjorkman, M A Saper, et al., 1987) that play a vital role in the immune response, via their activation of αβ T cells (Heber-Katz, Hansburg and Schwartz, 1983). Human MHCs are also known as the human leukocyte antigens (HLA). MHCs present peptide fragments, generated by the cleavage of proteins (Babbitt et al., 1985; Guillet et al., 1986; Townsend, Rothbard, et al., 1986), thus allowing the αβ T cells to scrutinise the proteome at the cell surface (Davis and Bjorkman, 1988; Germain, 1994; Garcia, Teyton and Wilson, 1999).

The classical MHC molecules are divided into two broad families; MHC class I and MHC class II (Hood, Steinmetz and Malissen, 1983; Davis and Bjorkman, 1988). The former is expressed on almost all nucleated cells and presents peptides from endogenous sources to CD8⁺ T cells (Bjorkman, M A Saper, et al., 1987; David-Watine, Israël and Kourilsky, 1990; Ting and Baldwin, 1993; Van Den Elsen, 2011). In contrast, MHC class II expression is largely restricted to APCs and thymic epithelial cells (Glimcher and Kara, 1992; Van Den Elsen et al., 2003). MHC class II molecules present peptides from exogenous sources to CD4⁺ T cells, thus T cell
immunity can sample both the intracellular and the extracellular proteome (Swain, 1981; Spits et al., 1982; Swain et al., 1984; Van Den Elsen, 2011). Whilst other cell types do not constitutively express MHC class II, expression can be induced by the cytokines produced during an immune response such as IFN-γ (Piskurich et al., 1998; Holling et al., 2002).

1.3.2 – The HLA loci

The MHCs are encoded by one of the most diverse set of alleles in the human genome (Campbell and Trowsdale, 1993; Beck et al., 1999; Kumánovics, Takada and Lindahl, 2003; Horton et al., 2004). The two aforementioned classical MHC classes are further divided into three subfamilies apiece. HLA-A, HLA-B, and HLA-C in the case of class I, and HLA-DR, HLA-DQ, and HLA-DP in the case of class II. Each subfamily itself is highly polymorphic (Campbell and Trowsdale, 1993; Beck et al., 1999; Kumánovics, Takada and Lindahl, 2003; Mungall et al., 2003; Horton et al., 2004; Trowsdale and Knight, 2013). This enormous diversity is surmised to be an evolutionary safeguard to ensure that detection of any potential foreign pathogen is covered across the human population (Horton et al., 2004). The vast majority of this diversity occurs in the peptide-binding groove of the MHC molecule (Bjorkman, M. A. Saper, et al., 1987; Kumánovics, Takada and Lindahl, 2003; Trowsdale and Knight, 2013). It is the sequence of the peptide-binding groove that determines which peptide fragments can be loaded into the groove and presented to the T cells (Barber et al., 1995; Gao et al., 2001).

Despite the huge number of available HLA alleles each individual only expresses a small subset of MHC molecules (Stern and Wiley, 1994). As such, the complement of MHC molecules expressed by an individual determines the peptide fragments that that individual is able to present to their T cell repertoire. Even a single amino acid difference between two MHC molecules can alter the epitopes that can be presented (Stern et al., 1994; Barber et al., 1995). Thus, possession of certain HLA alleles can confer susceptibility or resistance to viral disease, based on the ability, or lack thereof, to present key pathogenic epitopes (Fabio et al., 1992; Gao et al., 2001; Geldmacher et al., 2009; Kløverpris et al., 2012; Kløverpris, Cole, et al., 2015). Expression of certain HLA alleles has also been linked to susceptibility to autoimmune disease (Todd, Bell and McDevitt, 1987; Mignot et al., 1997; Price et al., 1999; Chabas et al., 2003; Larsen and Alper, 2004; Trowsdale, 2005; Horton et al., 2008), see Section 1.5.2.
1.3.3 – pMHC structure

The first MHC structure (HLA-A*0201, HLA-A2 hereafter) was solved in 1987 (Bjorkman, M A Saper, et al., 1987). This was achieved by purifying the protein from a human Epstein-Barr virus (EBV)-transformed lymphoblastoid cell line, JY. As such, the groove contained a mixture of peptides (Bjorkman, M A Saper, et al., 1987). By 1992 MHC molecules could be artificially refolded with a single peptide to produce clear crystal structures (Fremont et al., 1992; Garboczi, Hung and Wiley, 1992; Wilson and Fremont, 1993).

MHC class I molecules are heterodimers consisting of a heavy α-chain associated with a smaller β2-microglobin (β2M), Figure 1.3A. The α-chain is a transmembrane glycoprotein folded into three domains, denoted α1, α2, and α3, Figure 1.3B (Bjorkman, M A Saper, et al., 1987).

Domains α1 and α2 form the peptide binding groove - two parallel α-helices linking a seven-stranded β-sheet ‘floor’ (Bjorkman, M A Saper, et al., 1987). The α2 helix possesses a distinctive kink at residue 162 (Bjorkman, M A Saper, et al., 1987). The binding groove is closed at either end and can generally accommodate peptides of 8-13 amino acids in length, Figure 1.3C. Peptides longer than eight amino acids are accommodated in a bulged fashion Figure 1.3D (Chicz et al., 1992; Henderson et al., 1992; Matsumura et al., 1992; Chen et al., 1994).

Within the peptide-binding groove are six pockets (known as pockets A-F), Figure 1.3E-F (Bjorkman, M A Saper, et al., 1987; Bjorkman, M. A. Saper, et al., 1987; Bjorkman and Parham, 1990). These pockets are responsible for most of the contacts with the peptide, and can accommodate its side chains (Bjorkman, M. A. Saper, et al., 1987). These pockets also house most of the allelic variety of the MHC-I molecules and are chemically distinct, giving them varying binding preferences for different amino acids. The pockets are largely responsible for determining which peptides can be presented to TCRs (Bjorkman, M. A. Saper, et al., 1987; Van Bleek and Nathenson, 1991). Altering the peptide sequence even slightly can affect how well it is able to be presented by a particular MHC molecule (Van Bleek and Nathenson, 1991).
Figure 1.3: The structure of MHC-I. A) shows a schematic representation of the MHC-I heavy chain (blue) associated with β2M (pink) and presenting a peptide (yellow). B) The heavy chain is further divided into three globular domains donated α1, α2, and α3. C) α1 and α2 form the binding groove in which the peptides sit. D) Peptides sit within this groove in a bulged fashion, anchored at their termini with their central residues protruding. E+F) Deep ‘pockets’ within the peptide binding groove floor form interactions with peptide side chains helping to anchor it in place. Images generated using PDB: 2BCK. Overlaid peptides in (D): ALWGPDPAAD (PDB 3UTQ) magenta, AQWGPDPAAD (PDB 5COD) yellow, YLGPFDPPTI (PDB 5CG1) pink, ELAGILTV (PDB 1JF1) grey, SLLMWITQC (PDB 2BNR) blue, and AAGIGILTV (PDB 3QE1) green.
The third globular domain, known as the α3 domain, is an immunoglobulin-like domain containing the membrane-spanning portion of the molecule (Bjorkman, M A Saper, et al., 1987). Together α1, α2, and α3 are known as the MHC-I heavy chain. The heavy chain non-covalently associates with β2M to form a heterodimer (Bjorkman, M A Saper, et al., 1987; Tysoe-Calnon, Grundy and Perkins, 1991). The α3 domain also has an important role in the association with the CD8 co-receptor. A negatively charged loop formed by residues 222-228 within the α3 domain was suggested as the binding site for CD8 (Salter et al., 1990). This was later confirmed with atomic resolution crystal structures (Gao et al., 1997; Wang, Natarajan and Margulies, 2009).

As mentioned briefly above, class I MHC molecules generally present short peptides (8-13mers) (Cerundolo et al., 1991; Falk et al., 1991; Jardetzky et al., 1991; Madden et al., 1991; Hunt et al., 1992; Chen et al., 1994; Rist et al., 2013; Trolle et al., 2016). This length-restriction is imposed by the binding groove of MHC-I which adopts a closed conformation, with the termini of the peptide typically forming clusters of hydrogen bonds with the B and F pockets of the MHC (Bjorkman, M A Saper, et al., 1987; Saper, Bjorkman and Wiley, 1991; Garboczi, Hung and Wiley, 1992; Falk et al., 1994; Fidelis Maier, 1994; Stern and Wiley, 1994; Garboczi et al., 1996). The principle anchor residues for the peptide-MHC-I interaction are generally located at the termini of the peptides (often peptide positions 2 and the C-terminal residue) (Bjorkman, M A Saper, et al., 1987; Garboczi, Hung and Wiley, 1992; Stern and Wiley, 1994; Garboczi et al., 1996; Ding et al., 1999; Hausmann et al., 1999; Rudolph, Stanfield and Wilson, 2006). In addition to the clusters of hydrogen bonds that occur at primary anchor residues, other contacts can be made elsewhere between the MHC molecule and the peptide. These contacts can increase the overall binding affinity between peptide and MHC and subtly adjust the peptide position (Fremont et al., 1992; Silver et al., 1992; Jardetzky et al., 1994; Stern and Wiley, 1994; Cole et al., 2016; Cole, Fuller, et al., 2017).

The centre of the peptide tends to arch away from the MHC molecule and thus interacts with it less (Fremont et al., 1992; Guo et al., 1992; Silver et al., 1992; Madden, Garboczi and Wiley, 1993). This central peptide region is further away from the MHC-I molecules and is therefore more exposed for recognition by the TCR (Madden et al., 1992; Silver et al., 1992; Madden, Garboczi and Wiley, 1993). The peptide side chains that are not nestled within the MHC pockets can protrude upwards and interact with the TCR (Garboczi et al., 1996; Garcia et al., 1996; Ding et al., 1999). Even the smallest sequence difference within the peptide can lead to a striking alteration in the resulting conformation (Madden, Garboczi and Wiley, 1993;
Borbulevych et al., 2009). Longer peptides can bind MHC-I by protruding further out of the groove, known as super-bulged peptides (Guo et al., 1992; Speir et al., 2001; Tynan, Borg, et al., 2005; Tynan, Burrows, et al., 2005; Tynan et al., 2007), or by extending out of the ‘closed’ groove in a manner akin to MHC-II peptide presentation (Motozono, Pearson, et al., 2015; Pymm et al., 2017).

MHC class II molecules are also heterodimers, consisting of two transmembrane glycoproteins – α and β. Both chains fold into two domains – α1 and α2, or β1 and β2 respectively – with the groove-like binding site formed by the α1 and β1 domains (Brown et al., 1993). In contrast to MHC class I molecules, the binding sites of class II molecules are ‘open-ended’ meaning that the termini of the peptide can protrude at either side and longer peptides (13-25 amino acids) can be accommodated (Chicz et al., 1992; Brown et al., 1993; Jardetzky et al., 1994; Stern and Wiley, 1994; Stern et al., 1994). The termini of MHC-II-bound peptides are able to play a role in the TCR interaction whereas in MHC-I presentation they tend to be buried (Sant’Angelo et al., 2002; Zavala-Ruiz et al., 2004; MacLachlan et al., 2019). Unlike class I-restricted proteins that tend to have anchor residues at the peptide termini, class II restricted peptide anchor residues can be distributed throughout the peptide (Stern and Wiley, 1994; Stern et al., 1994; Sant’Angelo et al., 2002; MacLachlan et al., 2019)

1.3.4 – Peptide presentation by MHC

The peptides presented by MHC-I molecules are largely derived from cellular proteins that have been degraded by proteases or – following ubiquitination – by the proteasome (Townsend, Gotch and Davey, 1985; Townsend et al., 1986), Figure 1.4A. In addition to degraded cellular proteins, another, more immediate source of presentable material is defective ribosomal products (DRiPs). DRiPs are N-terminal polypeptide fragments that are generated when the ribosomal translation process fails; they cannot be used by the cell and are rapidly delivered to the proteasome (Yewdell, Antón and Bennink, 1996; Berglund et al., 2007; Yewdell, 2011). Presentation of DRiPs offers the advantage of very early detection of viral infection, soon after the start of viral protein translation (Yewdell, 2011).

Proteins, both self and viral in the case of an infected cell, can be broken down by the host proteasome. The proteasome became theoretically implicated in MHC-I antigen processing through a sequential summation of evidence, reviewed in 1992 by Goldberg and Rock (Goldberg and Rock, 1992). It was previously known that MHC class II presentation required the uptake of extracellular antigens and their degradation in acidic lysosomal compartments.
(Allen, Babbitt and Unanue, 1987; Puri and Factorovich, 1988). Yet MHC-I presentation did not seem to require acidophilic lysosomal proteases (Morrison et al., 1986). In contrast, degradation of antigen for presentation by MHC-I was thought to occur in the cytosol because antigens that were injected into the cytosol were effectively processed (Townsend, Gotch and Davey, 1985; Townsend, Bastin, et al., 1986; Moore, Carbone and Bevan, 1988; Yewdell, Bennink and Hosaka, 1988). An identified ‘MHC-linked low molecular mass protein’ was found to be similar in size and sequence to the 20S subunit of the proteasome (Brown, Driscoll and Monaco, 1991; Glynne et al., 1991; Kelly et al., 1991; Martinez and Monaco, 1991). It was also encoded by a gene located within the MHC locus (Brown, Driscoll and Monaco, 1991; Ortiz-Navarrete et al., 1991), this pointed towards a role for the proteasome in MHC-I peptide presentation. In 1993 the role of the proteasome was demonstrated experimentally. Michalek et al. used a temperature sensitive ubiquitinase to demonstrate that MHC-I presentation utilises ubiquitinase-mediated proteasome degradation (Michalek et al., 1993). This observation was supported by subsequent similar observations (Rock et al., 1994; Hughes et al., 1996). It is worth noting that the proteasome is not essential for the presentation of all MHC-I epitopes (Bai and Forman, 1997).

In addition to the standard ‘housekeeping’ proteasome, there is the immunoproteasome (expression of which can be upregulated by the IFNγ produced in an antiviral response) (Brown, Driscoll and Monaco, 1991; Glynne et al., 1991; Martinez and Monaco, 1991; Aki et al., 1994; Griffin et al., 1998; Kingsbury, Griffin and Colbert, 2000; Huber et al., 2012). The immunoproteasome cleaves proteins in a more suitable manner for MHC-I ligation than the standard proteasome as it favours cleavage after hydrophobic amino acids, reflecting the preference of MHC-I for C terminal binding (Romero et al., 1991; Falk et al., 1994; Gaczynska et al., 1994). Whilst the immunoproteasome cleaves peptides at more MHC-I-compatible positions, both proteasomes are important to the generation of MHC-I ligands and cells lacking an immunoproteasome can still process peptides for MHC-I-presentation (Yewdell et al., 1994; Cerundolo et al., 1995).

Peptides cleaved by the respective proteasomes may not be at optimal length for MHC-I presentation (8-13mers), consequently they can be cleaved further by cytosolic peptidases Figure 1.4A (Craiu et al., 1997; Mo et al., 1999; Reits et al., 2004; Hulpke and Tampé, 2013). The relevance of this further cleavage to MHC-I presentation of antigen is particularly interesting and has generated some debate as to its usefulness. Some argue that the overall effect of the peptidases might be negative to the adaptive immune response as they may destroy more MHC-I ligands than they create (Chapiro et al., 2006; York et al., 2006; Marcilla,
On the other hand, this peptidase-dependent method of antigen generation may actually be able to compensate for an absence of proteasomal activity (Glas et al., 1998). It is interesting to speculate that these apparently degenerate mechanisms for protein breakdown may function as evolutionary safeguards against pathogens.

Peptides are loaded onto MHC-I complexes in the endoplasmic reticulum (ER) by a collection of proteins known as the peptide loading complex (PLC), Figure 1.4B (Hulpke and Tampé, 2013). Following degradation peptide fragments are translocated into the ER by the transporter associated with antigen processing (TAP). TAP is a member of the antigen-binding cassette (ABC) transporter superfamily that resides in the ER membrane (Spies et al., 1990; Kleijmeer et al., 1992). TAP is a heterodimer composed of a TAP1 subunit and a TAP2 subunit (Spies et al., 1992). It carries two nucleotide-binding domains that bind adenosine triphosphate (ATP) (Spies et al., 1992), and two transmembrane domains that help anchor it in the ER membrane (Spies et al., 1990). The gene encoding TAP (peptide supply factor gene) was mapped to the MHC loci in 1990 (Deverson et al., 1990; Monaco, Cho and Attaya, 1990; Spies et al., 1990; Trowsdale et al., 2008). In 1991, experimental evidence showed that TAP expression is required for normal levels of surface MHC expression (Spies and DeMars, 1991). Further supporting evidence for the necessity of TAP was generated in 1992 using mutant cell lines deficient in the transporter; deficiencies resulted in unstable MHC-I and significantly reduced presentation of intracellular peptides (Kelly et al., 1992). In 1994, immunoprecipitation experiments demonstrated that the TAP complex associates with the MHC heavy chain (Ortmann, Androlewicz and Cresswell, 1994; Suh et al., 1994). In 1996, HLA-A2 was conclusively shown to interact with TAP. Disruption of this interaction by point mutation in HLA-A2 resulted in unstable HLA-A2-β2M dimers at the cell surface (Lewis et al., 1996). In the presence of mutated TAP, MHC molecules were loaded with suboptimal peptides; without TAP binding to tether them to the ER they rapidly progress to the cell surface before peptide optimisation (mediated by molecular chaperones and discussed below) can occur (Lewis and Elliott, 1998).
Figure 1.4: Peptide loading of MHC. A) Cytosolic proteins or DRiPs are broken down by proteasomes into peptides which are further cleaved by cytosolic peptidases. Cleavage products are translocated into the ER by the ABC transporter family member, TAP. B) Within the ER free MHC heavy chain become associated with β2M with the aid of protein chaperones. This complex is brought into close contact with translocated peptides. This juxtaposition is brought about via a PLC scaffold made up of TAP, tapasin, ERp57, and calreticulin. Prior to MHC loading peptides can be trimmed to optimal length by ERAP1/2.

TAP performs the characteristic cycling of ATP binding and hydrolysis which induces a conformational change in TAP that allows peptide fragments to be transported across the ER membrane (Kelly et al., 1992; Androlewicz, Anderson and Cresswell, 1993; Shepherd et al., 1993; Knittler et al., 1999; Boyle et al., 2015). TAP has a peptide-binding region (Nijenhuis and Hämmerling, 1996), which binds peptides in an ATP-independent manner (Van Endert et al., 1994). Binding and hydrolysis of ATP then facilitates the translocation of the bound peptide (Neefjes, Momburg and Hämmerling, 1993; Van Endert et al., 1994). Similar to MHC-I binding, TAP can bind promiscuously to peptide fragments due to the requirement of relatively few anchor residues. The remaining residues in the peptide are then open to diversity (Van Endert et al., 1995; Uebel et al., 1997; Herget et al., 2011), although interestingly TAP is stereospecific – meaning that it is unable to bind peptides composed of D-amino acids (Grommé et al., 1997; Uebel et al., 1997). TAP can transport peptides of a range of different lengths because it is thought that the ends of the peptide are bound to
TAP, whilst the centre protrudes in an extended kink conformation – not interacting with TAP (Uebel et al., 1995; Van Endert et al., 1995; Koopmann et al., 1996; Herget et al., 2011). Once in the ER, potential antigenic peptides can be cleaved further (Snyder, et al., 1994; Elliott, et al., 1995) by an ER aminopeptidase associated with antigen processing, ERAP1/2 (known as ERAAP in mice). This enzyme serves to trim peptides for insertion into the MHC-I binding pocket (Saric et al., 2002; Serwold et al., 2002; York et al., 2002; Saveanu et al., 2005; Hulpke and Tampé, 2013). Polymorphisms in this enzyme can predispose to autoimmune diseases through alterations in the peptide repertoire that is presented at the cell surface to T cells (Kemming et al., 2019).

Also in the ER lumen are free MHC-I heavy chains, the folding of which is assisted and stabilised by an ER-resident chaperone protein called calnexin and the ER-resident thiol oxidoreductase ERp57 enzyme which catalyses the formation of disulphide bonds Figure 1.4B (Degen and Williams, 1991; Hochstenbach et al., 1992; Hughes and Cresswell, 1998). These chaperones stabilise intermediate forms of the quaternary structure, and promote the association MHC-I heavy chain with β2M (Degen and Williams, 1991; Ortmann, Androlewicz and Cresswell, 1994). Following association with β2M, calnexin is exchanged for calreticulin, another ER-resident chaperone protein that associates with MHC-I/β2M, it also associates with ERp57 and serves to facilitate recruitment of this intermediate complex to the aforementioned TAP via an additional member of the PLC, tapasin (Ortmann, Androlewicz and Cresswell, 1994; Sadasivan et al., 1996).

Tapasin forms a heterodimer with ERp57 (Oliver et al., 1999; Zhang, Baig and Williams, 2006; Santos et al., 2007), the crystal structure of which was reported in 2009 (Dong et al., 2009). Tapasin facilitates the physical proximity of the MHC-I/β2M complex and TAP (Ortmann, Androlewicz and Cresswell, 1994; Sadasivan et al., 1996; Ortmann et al., 1997) by simultaneous binding of TAP and calreticulin (Ortmann, Androlewicz and Cresswell, 1994; Sadasivan et al., 1996; Ortmann et al., 1997; Koch et al., 2004). In this way, the Tapasin-ERp57 complex tethers the TAP – which is transporting peptides – to the empty MHC-I molecule which will receive the peptides. Tapasin has binding sites on the α2 and α3 domains of the MHC class I heavy chains (Peace-Brewer et al., 1996; Kulig et al., 1998; Suh et al., 1999; Varela-Rohena et al., 2008) and on the N-terminal domains of the TAP subunits (Koch et al., 2004). The presence of tapasin increases the quantity of peptides that can be presented by most MHC molecules (Barber et al., 2001). However, in addition to its tethering role, tapasin has a role in optimisation – ensuring that high-affinity peptides are preferentially presented,
thus mediating both quantity and quality of peptide presentation (Barber et al., 2001; Williams et al., 2002; Howarth et al., 2004; Chen and Bouvier, 2007; Wearsch and Cresswell, 2007; Praveen et al., 2010). Tapasin is thought to bind to the MHC-I binding groove, enforcing an open conformation and causing lower-affinity peptides to dissociate. Higher affinity peptides can overcome this energy barrier and their binding induces maturing conformational changes in the MHC that induces its release from the PLC (Elliott et al., 1991, 1992; Rigney et al., 1998; Chen and Bouvier, 2007). Kinetically stable pMHC complexes can then be released from the PLC to make their way to the cell surface for scrutiny by TCRs (Chen and Bouvier, 2007; Wearsch and Cresswell, 2007; Hulpke and Tampé, 2013). Tapasin is assisted in its role by an additional, related, MHC-I-peptide editor, TAPBR (Boyle et al., 2013; Hermann et al., 2013, 2015).

1.4 – The TCR-pMHC interaction
At the heart of the adaptive immune response is the interaction between the αβ TCR and the pMHC. This interaction was alluded to by a series of studies in the 1970s (discussed below). One such study demonstrated that an interaction between cytotoxic T cells and somatic cells infected with lymphocytic choriomeningitis only occurred if the two cell types share expression of at least one H-2 allele (a mouse MHC variant) (Zinkernagel and Doherty, 1974).

The TCR itself remained elusive for another decade (Hedrick et al., 1984; Yanagi et al., 1984). By the 1980s, monoclonal antibodies raised against T cell lines had identified disulphide linked heterodimers with immunoglobulin-like constant and variable domains (Acuto et al., 1983; McIntyre and Allison, 1983). These antibodies were shown to be able to stimulate and inhibit T cell functions, suggesting they targeted the unknown surface receptor (Allison, McIntyre and Bloch, 1982; Infante et al., 1982; Haskins et al., 1983; Kaye et al., 1983; Meuer et al., 1983; Samelson, Germain and Schwartz, 1983). Hendrick et al. used 32P-labelled cDNA probes to isolate genes that were expressed in T but not B cells, encoded a membrane-bound polypeptide that would be amendable to somatic gene recombination (similar to what had been described for B cell receptors (Hozumi and Tonegawa, 1976; Tonegawa, 1983)); and would have variable and constant regions. Using this method they successfully isolated the TCR (Hedrick et al., 1984).

The TCR-pMHC interaction serves to bring the T cell into proximity with its target cell, and facilitates T cell activation (Davis and Bjorkman, 1988; Garcia, Teyton and Wilson, 1999). In the case of the CD8+ T cell, TCR interactions with pMHC-I can mediate killer cell functions
such as target cell lysis (see Section 1.2.2) (Davis and Bjorkman, 1988; Garcia, Teyton and Wilson, 1999). The interaction is also central to the process of maturation and selection that thymocytes undergo whilst developing in the thymus (see Section 1.2.3).

Multiple structural studies of the TCR-pMHC interaction have collectively revealed a relatively conserved binding mode, Figure 1.5A (Garboczi et al., 1996; Garcia et al., 1996, 1998; Ding et al., 1998; Teng et al., 1998). The variable domain of the TCR α-chain docks over the α2 helix of the MHC and the N-terminal region of the presented peptide. The variable domain of the TCR β-chain, in contrast, docks over the α1 helix of the MHC and the C-terminal region of the peptide (Hong et al., 1992; Sun et al., 1995; Sant’Angelo et al., 1996; Garcia, Teyton and Wilson, 1999; Rudolph, Stanfield and Wilson, 2006).

The ‘contact zone’ on the TCR is formed by its variable domains, the six hypervariable CDR loops form the antigen-binding site Figure 1.5A inset (see Section 1.2.2) (Garboczi et al., 1996; Garcia et al., 1996; Ding et al., 1998). The germline-encoded CDR1 and CDR2 loops mainly contact the MHC surface (Garboczi et al., 1996; Garcia et al., 1996; Ding et al., 1998), although CDR1 and CDR2 loops have been shown to contact both the MHC and the peptide (Ding et al., 1998; Cole et al., 2009; Cole, Van Den Berg, et al., 2017). CDR3 loops principally form contacts with the presented peptide (Garboczi et al., 1996; Garcia et al., 1996; Ding et al., 1998). As the peptide tends to sit in the MHC binding groove anchored at either end with the central region of the peptide bulging out of the groove (Garcia et al., 1996; Ding et al., 1998; Hausmann et al., 1999; Speir et al., 2001), most interactions occur between the TCR and central peptide residues with upward facing side chains (Degano et al., 2000; Chen et al., 2005; Tynan, Borg, et al., 2005; Cole et al., 2016; Cole, Fuller, et al., 2017; Madura et al., 2019), although many exceptions to this rule exist (Hahn et al., 2005; Yili Li et al., 2005; Ekeruche-Makinde et al., 2012; Gras et al., 2016; Harris et al., 2016).
Figure 1.5: The TCR-pMHC interaction. A) shows the consensus binding mode between TCR and pMHC. The antigen binding site is formed by the six CDR loops, three CDRα and three CDRβ (inset, red). B) The αβ TCR adopts an approximately diagonal orientation when interacting with pMHC. Image generated using PDB SMEN. Crossing angles calculated by Dr Bruce MacLachlan.

When the αβ TCR binds the pMHC class I complex it tends to adopt an approximately diagonal orientation relative to the peptide, Figure 1.5B, although there is much variability with little difference in the ability of the complex to promote T cell activation (Garboczi et al., 1996; Ding et al., 1998; Teng et al., 1998; Rudolph and Wilson, 2002; Rudolph, Stanfield and Wilson, 2006).

The consensus manner of binding thus described led to the proposal of a two-step binding model in 2002 (Wu et al., 2002). This model suggests that the TCR initially forms transient contacts with the surface of the MHC, holding the TCR in close proximity and thus allowing the ‘scanning’ of the presented peptide (Wu et al., 2002). This model has been subject to much controversy. For example, in the case of super-bulged peptides where centrally bulging
peptides in the peptide-binding groove can act as an obstacle, impeding TCR access to the MHC (Tynan, Burrows, et al., 2005). The consensus binding mode has been challenged further by Cole et al. who showed that affinity-enhanced TCRs whose enhanced affinity was mediated by an increase in TCR-MHC contacts still exhibited exquisite peptide-specificity (Cole et al., 2014). These results were inconsistent with the two-step binding model (Wu et al., 2002) leading the authors to propose an alternative model, known as the “synchronised docking” model, in which there is no temporal difference between the binding of TCR-MHC and that of TCR-peptide (Cole et al., 2014).

The interaction between the TCR and pMHC can be highly plastic. To facilitate engagement of a pMHC ligand, a TCR can undergo substantial conformational changes – principally within its CDR loops (Garcia et al., 1998; Reiser et al., 2003; Colf et al., 2007; Ayres et al., 2016). It may be supposed that a requirement for conformational change might result in a less potent interaction due to the entropic cost of the movement. However, Cole et al. showed in 2017 that the naturally-occurring high-affinity HIV-specific 868 TCR undergoes extensive CDR rearrangement in order to engage its ligand (Cole, Fuller, et al., 2017).

Flexibility can also occur on the side of the MHC (Hawse et al., 2013; Van Hateren et al., 2017; Hopkins et al., 2020). The human T-lymphotropic virus (HTLV)-1 Tax peptide (LLFGYPVYV), and the Saccharomyces cerevisiae peptide Tel1p (MLWGYLQYV) are both presented by HLA-A2 to the human αβ TCR, A6 (Borbulevych et al., 2009). When both peptides were crystallised in complex with HLA-A2 they showed superimposable structures. However, when these pMHC-I pairs were bound to the A6 TCR they showed a substantially different interface of contacts, facilitated by a conformational change of the HLA-A2 α-helices when in complex with Tel1p but not Tax (Borbulevych et al., 2009). Many other examples of MHC flexibility have been described (Hülsmeyer et al., 2002; Pöhlmann et al., 2004; Tynan, Borg, et al., 2005; Borbulevych et al., 2011; Kass, Buckle and Borg, 2014).

Peptide conformational changes can also occur within this interaction, such as the flattening of super-bulged peptides (Tynan et al., 2007), and the ‘flexing’ of the peptide to occupy different binding groove pockets (Madura et al., 2019)

The natural affinity of functional TCR-pMHC-I interactions tends to range between \( K_D \approx 0.1–100 \) μM (Weber et al., 1992; Corr et al., 1994; Sykulev et al., 1994; Alam et al., 1996; Garcia et al., 1997, 2001; Kersh et al., 1998; Boniface et al., 1999; Willcox et al., 1999; Ding et al., 1999; van der Merwe and Davis, 2003; Cole et al., 2007; Bridgeman et al., 2011). This is low relative to antibody-ligand affinities (Goldman et al., 1997; Conte, Chothia and Janin, 1999).
The range of off-rates observed for the TCR-pMHC interaction is narrow (~0.01-0.73 s⁻¹), with a mean of 0.24 s⁻¹ (Bridgeman et al., 2011). It has been suggested that the off-rate of the interaction must fall within a narrow ‘window’ so that the interaction is long enough to allow the induction of intracellular signal transduction but brief enough to allow each pMHC to be ‘scanned’ by, and trigger, multiple TCRs in a short time (Valitutti et al., 1995; Cole et al., 2007).

1.4.1 – The co-receptors
Another important consideration for the TCR-pMHC interaction is the role of the co-receptors. As mentioned briefly above, CD4 and CD8 both phenotypically characterise the two major subsets of αβ T cells and dictate which class of MHC they interact with. CD4 specifically recognises MHC class II molecules, whilst CD8 specifically recognises MHC class I molecules (Wettstein et al., 1978; Swain and Panfili, 1979; Swain, 1980; Engleman et al., 1981).

An interesting series of observations prompted the suggestion of an interaction between MHC-I and CD8 long before it was confirmed crystallographically. Firstly, it was observed that mature T cells tended to express either CD8 or CD4 (Cantor and Boyse, 1977; Reinherz et al., 1979, 1980). Amounting evidence in the 1970s and 80s demonstrated that MHC-I-expressing cells associated with CD8⁺ T cells and MHC-II with CD4 (Swain, 1983). The interaction between co-receptor and respective MHC was then demonstrated by cell-adhesion experiments (Doyle and Strominger, 1987; Norment et al., 1988). Antibodies against CD4 or CD8 could block T cell functions against MHC-II⁺ or MHC-I⁺ cells respectively (Engleman et al., 1981; MacDonald, Thiernesse and Cerottini, 1981; Swain, 1981; Krensky et al., 1982; Meuer, Schlossman and Reinherz, 1982; Wilde et al., 1983; Swain et al., 1984; Rosenstein et al., 1989). It was further observed that gene transfer of CD8 could redirect a MHC-II-reactive T cell to MHC-I⁺ target cells (Gabert et al., 1987).

In 1988, Connolly et al. showed that a point mutation in the α3 domain of the MHC molecule (E227K) could abrogate binding to CD8 and subsequent killing by cytotoxic lymphocytes (Connolly et al., 1988). A negatively charged loop formed by residues 222-228 within the α3 domain was suggested as the binding site for CD8 (Salter et al., 1990). This theoretical binding location was further supported by the crystallisation of CD8, demonstrating that the proposed binding manner would be feasible (Leahy, Axel and Hendrickson, 1992). Finally, this successive series of events culminated in the 1997 crystallisation of a HLA-A2-CD8
complex that visually demonstrated MHC-I binding to the co-receptor (Gao et al., 1997). CD4 has also been crystallised in complex with MHC-II (Yin, Wang and Mariuzza, 2012).

Figure 1.6: The co-receptors. A) Schematic representation of the CD4 and CD8 co-receptors. CD4 exists as a single polypeptide chain folded into four immunoglobulin (Ig)-like domains. CD8 exists as a heterodimer. Both co-receptors have transmembrane domains, and cytoplasmic tails that associate with the cytosolic Src family protein tyrosine kinase, Lck. B) This association allows the juxtapositioning of Lck and CD3 ITAMs when either co-receptor binds their respective MHC.

On conventional cytotoxic T cells, CD8 exists as a heterodimer made up of CD8α and CD8β chains, Figure 1.6A (Fung-Leung et al., 1991; Moebius et al., 1991; Casey Crooks and Littman, 1994; Nakayama et al., 1994). As discussed above, CD8 binds primarily to the α3 domain of MHC-I (Gao et al., 1997; Wang, Natarajan and Margulies, 2009). CD4 binds the α2 and β2 regions of MHC-II near to the membrane of the antigen presenting cell (Wang et al., 2001, 2011). Thus, in both cases, the peptide-MHC complex can be simultaneously bound by both the TCR and its associated co-receptor, Figure 1.6B.

CD8 interacts with pMHC-I with relatively low affinity ($K_D \sim 150\ \mu M$) (Wyer et al., 1999; Huang, Edwards, et al., 2007; Li, Yin and Mariuzza, 2013), with respect to most TCR-pMHC interactions (Wooldridge, Clement, et al., 2010; Wooldridge, Laugel, et al., 2010). Evidence presented by Wooldridge et al. in 2010 suggested that this low affinity is actually an evolutionary advantage, and in fact essential for the maintenance of TCR specificity
Artificially increasing the strength of the CD8-MHC interaction resulted in non-specific binding of the TCR to pMHC and T cell activation, irrespective of the peptide presented. Thus, the low affinity of the CD8-MHC interaction is essential to maintain specificity of CD8+ T cells (Wooldridge, Clement, et al., 2010).

The co-receptors are thought to enhance the kinetics of the TCR-pMHC engagement (Veillette et al., 1989; Holler and Kranz, 2003; Wooldridge et al., 2005; Wooldridge, Clement, et al., 2010; Wooldridge, Laugel, et al., 2010). This is achieved in a number of ways. Despite its low affinity for MHC-I, CD8 can contribute to an enhanced overall avidity of the interaction (Holler and Kranz, 2003; Li, Yin and Mariuzza, 2013). Laugel et al. showed that TCRs with high affinities for their cognate pMHC could be labelled with pMHC tetramers despite mutations to remove CD8 from the interaction (Laugel et al., 2007). However, in incidences where the TCR-pMHC affinity was low (>200µM), the presence of the co-receptor was essential to T cell activation (Laugel et al., 2007).

By recruiting downstream signalling molecules to the localised site of the interaction (see Section 1.2.2), CD8 can lower the threshold needed for T cell activation by facilitating the propagation of the signalling cascade (Artyomov et al., 2010; Jiang et al., 2011). Blocking the binding of CD8 to the α3 domain of MHC-I can reduce some, but not all, T cell functions (Xu et al., 2001). It has been postulated that the evolutionary benefit conferred by co-receptor-mediated recruitment of downstream signalling molecules might have influenced the binding conformation of the TCR-pMHC-I interaction. As discussed in Section 1.2.2, the cytoplasmic tail of the co-receptor sequesters Lck (Veillette et al., 1989; Van Laethem et al., 2013). Lck phosphorylates the ITAMs on the cytoplasmic tail of the TCR-associated CD3 (Barber et al., 1989; Straus and Weiss, 1993; Purbhoo et al., 2001; Artyomov et al., 2010).

Thus, interactions that align CD8 in proximity to CD3 (see Figure 1.6B) may be evolutionarily favourable and contribute to the conserved docking mode observed in so many TCR-pMHC interactions (Garboczi et al., 1996; Ding et al., 1998; Rudolph and Wilson, 2002; Rudolph, Stanfield and Wilson, 2006; Van Laethem et al., 2007; Li, Yin and Mariuzza, 2013).

1.4.2 – MHC-restriction
The propensity of αβ TCRs to bind antigen in the context of MHC molecules is known as MHC-restriction (Zinkernagel and Doherty, 1974; Kappler and Marrack, 1976). MHC-restriction was first demonstrated by a series of studies in the 1970s. In 1973, Rosenthal and Shevach
highlighted the requirement of antigen presentation on histocompatible macrophages for T cell activation (Rosenthal and Shevach, 1973). Ern and Feldmann subsequently demonstrated that histocompatible, but not allogenic, macrophages could activate CD4+ T cells (Erb and Feldmann, 1975b, 1975a, 1975c). Thymic grafting between mice of different strains was used to show that T cells develop MHC-restriction in the thymus during their maturation process (Fink and Bevan, 1978). The K and D regions were isolated as the specific regions within the H-2 complex that needed to match between target cell and T cell in order to facilitate T cell activation (Doherty, Blanden and Zinkernagel, 1976). A number of studies were published showing that the T cell response to virus occurred only with MHC-matched target cells (Doherty, Zinkernagel and Ramshaw, 1974; Shearer, 1974; Wainberg et al., 1974; Zinkernagel and Doherty, 1974; Koszinowski and Ertl, 1975; Lewandowski, Gerhard and Palmer, 1976). The preference for self-MHC was further demonstrated in mice, where heterozygous AB stem cells were used to repopulate a non-lethally irradiated strain A mouse. T cell activity was mainly detected against cells expressing the MHC of strain A cells, little or no reactivity was detected against strain B cells (Bevan, 1977; Zinkernagel, G. Callahan, et al., 1978; Zinkernagel, G. N. Callahan, et al., 1978).

The Dual Recognition Theory was proposed by Katz, in order to explain MHC-restriction. It suggested that a T cell might express two receptors – one to recognise the self-MHC and one to recognise the foreign antigen (Katz and Armerding, 1976; Katz, Dixon and Kunkel, 1977; Fink and Bevan, 1978). This hypothesis was largely superseded by the Altered-Self Theory, proposed by Doherty et al. which suggested that T cells had only one receptor that could recognise the antigen-MHC as a complex (Doherty and Zinkernagel, 1976; Doherty, Blanden and Zinkernagel, 1976). The altered-self theory to explain MHC-restriction was ultimately confirmed with the publication of TCR-pMHC complex structures demonstrating how the αβ TCR can simultaneously engage self-MHC and foreign peptide (Garboczi et al., 1996; Garcia et al., 1996). How MHC restriction is imposed is generally debated by two key theories, the Germ-line Encoded Theory and the Selection Theory (La Gruta et al., 2018).

The germ-line encoded theory of MHC restriction proposes that αβ TCRs have evolved a preference for MHC recognition (Jerne, 1971; La Gruta et al., 2018). This theory is supported by the conserved binding mode observed for the interaction that was revealed by the first crystal structures in the 1990s (Garboczi et al., 1996; Ding et al., 1998). The germ-line encoded theory poses that MHC-restriction is driven by pairwise interactions between evolutionary conserved residues in both of the interacting molecules (Huseby et al., 2005). These so-called ‘interaction codons’ are a product of side-by-side evolution of the MHC and
TCR, and structural evidence in their favour was published in 2007 by Feng et al. in a study involving four Vβ8.2 TCRs (Feng et al., 2007). Feng et al. observed a number of structurally superimposable interactions amongst these complexes, and proposed that these were germline-encoded interaction codons (Feng et al., 2007). A similar observation was made of 12 TCR-pMHC complexes in 2016 (Adams et al., 2016). Interaction codons were also proposed to explain why certain TCR chains tend to interact with certain MHC alleles, explaining why TCRs tend to be restricted to a single MHC allele (Garcia et al., 2009).

Subsequent examples of interaction codons bolstered support for the germ-line encoded theory of MHC restriction (Dai et al., 2008; Marrack et al., 2008; Scott-Browne et al., 2009; Yin, Li and Mariuzza, 2012). In 2005, Tynan et al. observed that three MHC positions (65, 69, and 155) invariably made contact with the TCR in all TCR-pMHC-I structures that had been described at the time. They postulated that this might represent the minimum binding footprint required for the interaction and termed the three residues the “restriction triad” (Tynan, Burrows, et al., 2005). Mutating the conserved contact regions was shown to significantly diminish the size of the naïve T cell pool (Scott-Browne et al., 2009, 2011; Yin et al., 2011). In further evidence for the germ-line encoded theory, other classes of TCR interact with other classes of MHC molecules. For example, mucosal-associated invariant T cells (MAITs) which recognise metabolites in the context of MHC class I-related (MR1), and CD1-restricted T cells which respond to lipid antigen (McMichael et al., 1979; Treiner et al., 2003; Gold et al., 2010; Rossjohn et al., 2012; Godfrey et al., 2015; Van Rhijn et al., 2015). The fact that other classes of TCR also interact with ligands in the context of MHC-like models is often cited evidence in support of the germ-line theory.

The contrasting argument, the selection theory, proposes that MHC-restriction is not an intrinsic feature of TCRs but rather a feature than is imposed upon them during the selection process in the thymus, see Section 1.2.3 (La Gruta et al., 2018). Support for this theory comes from studies showing that the interaction codons are not essential, and structural compensations can be made to accommodate mutations in these residues (Van Laethem, Tikhonova and Singer, 2012). In fact, Rossjohn’s group queried their own 2005 study (Tynan, Burrows, et al., 2005) which first proposed the restriction triad by demonstrating that these residues could be mutated with minimal destructive effect on the interaction (Burrows et al., 2010). Further support for the selection theory comes from evidence that only around 30% of the preselection TCR repertoire is MHC-restricted (Merkenschlager et al., 1997; Zerrahn, Held and Raulet, 1997), suggesting that the selection process itself is responsible for these relatively few TCRs dominating the mature repertoire. Additionally several incidences of
unconventional αβ TCRs have been described in the literature (Beringer et al., 2015; Gras et al., 2016), including αβ TCRs that are not actually MHC-restricted (Barnd et al., 1989; Jerome, Domenech and Finn, 1993; Magarian-Blander et al., 1998; Hanada et al., 2011).

The selection theory proposes that MHC-restriction occurs because the co-receptors, CD8 and CD4, sequester a key component of TCR signalling – Lck (Van Laethem, Tikhonova and Singer, 2012; Van Laethem et al., 2013; Rangarajan and Mariuzza, 2014; La Gruta et al., 2018). As briefly discussed in Section 1.2.2, Lck is a protein tyrosine kinase that initiates the signal transduction pathway in both immature thymocytes and mature T cells (Rudd et al., 1988). Successful TCR signalling requires Lck, which tends to associate with the cytoplasmic tails of the co-receptors (Turner et al., 1990; Miceli, Von Hoegen and Parnes, 1991; Van Laethem, Tikhonova and Singer, 2012; Van Laethem et al., 2013; Rangarajan and Mariuzza, 2014; La Gruta et al., 2018). Lck initiates T cell signal transduction by phosphorylating regions within the cytosolic tail of the TCR/CD3 complex, resulting in the recruitment of downstream signalling molecules such as ZAP-70 (Meuer et al., 1983; Abraham et al., 1991; Chan et al., 1992; Van Laethem et al., 2013).

When a TCR recognises a ligand in the context of an MHC molecule, the co-receptor also binds the MHC (Rosenstein et al., 1989), as such extracellular MHC-co-receptor engagement brings intracellular co-receptor bound Lck into the vicinity of the ITAMs on the cytoplasmic tail of CD3 (see Section 1.2.2) (Artyomov et al., 2010). This intracellular juxtaposition gives MHC co-receptor-dependent signalling an advantage over co-receptor-independent signalling, in which “free” Lck must be passively captured by an activated TCR before signal transduction can begin (Lee, Loh and Lacy, 1992; Van Laethem, Tikhonova and Singer, 2012; Van Laethem et al., 2013; Rangarajan and Mariuzza, 2014; La Gruta et al., 2018). Deletion of the co-receptors, allowed thymocytes to differentiate into mature αβ T cells carrying TCRs that could recognise non-MHC ligands (Van Laethem et al., 2007, 2013; Van Laethem, Tikhonova and Singer, 2012). Furthermore, van Laethem et al. showed that if the two cysteine residues in the N-terminal region of the Lck protein (that are used to associate with CD8 and CD4) are mutated a MHC-independent αβ TCR repertoire is generated (Van Laethem et al., 2013). The selection theory proposes that V(D)J recombination produces a broad range of αβ TCRs but that only those that are MHC-restricted survive thymic selection (Van Laethem et al., 2007).

The germ-line encoded and the selection theories of MHC-restriction both offer up compelling evidence and it is likely to be a combination of the two that contribute to the
MHC-restriction so commonly seen amongst αβ TCRs. Regardless of how it is imposed, the characteristic MHC-restriction is undoubtedly a driving factor behind the specificity of TCRs in the peripheral population of T cells.

1.4.3 – Specificity and crossreactivity within the TCR-pMHC interaction
In order to maintain host health, the immune system must comprise of enough TCRs to specifically recognise any potential pathogen that could be encountered. Incomplete immune coverage would quickly be exploited by opportunistic pathogens. Ideas concerning how this comprehensive immune coverage is achieved have evolved over time.

In 1955 Neils Kaj Jerne described the natural selection theory of antibody formation (Jerne, 1955). Jerne postulated that instead of manufacturing antibodies (and by extrapolation, TCRs) in response to a newly-encountered antigen (Talmage, 1957), a host already possessed the full cohort of immunoreceptors that would be required to combat any foreign antigen it might encounter (Jerne, 1955). The exposure of the immune system to an antigen would result in the selection of the immunoreceptor best fit to fight it, from the existing pool of antibodies (Jerne, 1955). In 1959, Frank Macfarlane Burnet proposed the clonal selection theory of acquired immunity (Burnet, 1959). Burnet suggested that that when an antigen was encountered the immune cell specific for that antigen would undergo clonal expansion (Burnet, 1959). Whilst cementing the concept of specificity in the acquired immune response, this theory propagated the assumption that an immune receptor would be highly specific, responding to a single antigen (Burnet, 1959, 1976).

The clonal selection theory satisfied the question of specificity and was generally accepted by the field. However, in 1998 Don Mason and others demonstrated that the number of theoretical pMHC targets that could be encountered far exceeds the number of unique TCRs that exist within the host’s immune system (Mason, 1998; Arstila et al., 1999; Sewell, 2012). There are $10^{12}$ T cells in humans, carrying $\sim 10^7$ unique TCRs. Between them these TCRs must recognise $>10^{15}$ possible foreign antigen in order to provide comprehensive immune coverage and avoid potentially fatal immunological gaps (Sewell, 2012; Wooldridge, 2013).

Mason argued that whilst somatic gene recombination could theoretically produce sufficient variability (Arstila et al., 1999; Turner et al., 2006; Bhati et al., 2014), the number of individual T cells (each expressing a single highly specific TCR) required to meet the clonal selection requirements would be more than the host body could physically accommodate (Mason, 1998). Drawing on other evidence (Bhardwaj et al., 1993; Reay, Kantor and Davis, 1994;
Wucherpfennig and Strominger, 1995; Kersh and Allen, 1996), Mason called for the abandonment of the belief that every single immunogenic peptide must have its own unique TCR. He instead proposed that TCRs could recognise multiple peptides, (Mason, 1998), and that crossreactivity must be an essential feature of T cell immunity.

1.4.4 – How is crossreactivity achieved?
Many examples of crossreactivity in T cells have since been described in the literature. In 1995, a myelin basic protein (MBP)-specific T cell clone isolated from a multiple sclerosis patient was shown to recognise a number of other peptides (Wucherpfennig and Strominger, 1995). These included 7 viral- and 1 bacterial-derived peptide, indicating that molecular mimicry between T cell epitopes and self-antigen might underlie some incidences of autoimmunity, see Section 1.5.2 (Wucherpfennig and Strominger, 1995). The following year Kersh and Allen showed that the 3.L2 T cell clone could respond to 12 altered peptide ligands (APLs) (Kersh and Allen, 1996).

One way that TCRs might achieve crossreactivity is by focused binding to peptide ‘hotspots’. This describes a situation in which the TCR only recognises a few residues of the peptide. The same TCR could theoretically recognise all other peptides which contained the essential motif, allowing for variation at other residues of the peptide outside of the hotspot. For example, the type I diabetes-associated 1E6 T cell clone can recognise over 1 million peptides (Wooldridge et al., 2012) and does so largely by focusing on a central three-residue peptide motif, GPD (Cole et al., 2016).

TCRs might also exhibit crossreactivity by being flexible in their binding mode. For example, the murine 2C TCR adopts different conformations when in complex with self (L^d-QL9) and non-self (K^b-dEV8) targets (Colf et al., 2007). The two peptides included in this study were entirely distinct, differing at every residue. When interacting with QL9 the 2C TCR focused mainly on the C-terminal region. Whereas the interaction with dEV8 spanned the length of the peptide (Colf et al., 2007). In another example of a TCR seeing two peptides with completely distinct sequences, the BM3.3 TCR recognised INFDFNTI and RGYVYQGL peptides in the context of H-2K^b via CDR3a loop flexibility (Reiser et al., 2003). Crossreactivity in the context of antipathogenicity and autoimmunity is discussed in further detail in Sections 1.5 and 4.1.
1.5 – Application of the TCR-pMHC interaction
The TCR-pMHC interaction has a fundamental role in the immune response, it can be highly effective against pathogens and neoplasms. The TCR or pMHC can also be manipulated and employed as a therapeutic. The generic T cell response, including mechanisms of cytotoxicity, is discussed in Section 1.2.2.

1.5.1 – The TCR-pMHC interaction in immunity
Viruses are obligate intracellular pathogens and the CD8+ T cell response to viruses is well characterised. In fact, the murine lymphocytic choriomeningitis virus has provided a model system from which much of what is known about T cell responses has been derived (Marker and Volkert, 1973; Moskophidis et al., 1987; Ashton-Rickardt et al., 1994; van den Broek et al., 1995; Butz and Bevan, 1998; Gallimore et al., 1998; Williams, Tznik and Bevan, 2006). The TCR-pMHC interaction in the context of a number of example viral responses is discussed below.

Influenza virus (IVs) are common human pathogens (Grant et al., 2016). Influenza viruses undergo rapid antigenic drift allowing them to continually evade the immune system (Taubenberger and Kash, 2010). Despite this, CD8+ T cells are demonstrably important in controlling IV infections (McMichael et al., 1983; Sridhar et al., 2013; Wang et al., 2015). A large number of influenza-specific CD8+ T cell epitopes have been described (Wu et al., 2011; Grant et al., 2013, 2014).

Given the CD8+ T cell involvement in the response to IVs, efforts have been made to produce polypeptide vaccines that trigger a CD8+ T cell response (Moriyama et al., 2017; Skibinski et al., 2018; Zheng et al., 2019). In 2018 Miles et al. proposed that non-natural amino acids, that were resistance to gastric degradation, could be delivered by oral administration (Miles et al., 2018). A T cell clone specific for a matrix protein 1 (M1) epitope, the HLA-A2–restricted GILGFVFTL58–66 (GIL), was subjected to combinatorial peptide library (CPL) screening (see Section 3.1.2) with a D-amino acid library (Miles et al., 2018). The resulting D-amino acid peptide (gppqwnnpp; lower case letters used to indicate the D isomer in single letter amino acid code) bore little resemblance to the original GILGFVFTL peptide sequence, yet could stimulate GILGFVFTL-specific T cells clones and importantly was resistant to serum degradation for over 1 hour (Miles et al., 2018). Oral administration of gppqwnnpp was shown to protect mice against a lethal challenge with the influenza PR8 strain. This work
exemplifies how greater understanding of specific TCR-pMHC interactions in clinically relevant contexts has the potential to influence vaccine design.

In another example, strains of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) have caused two major outbreaks since the turn of the century, an epidemic in 2003 (Peiris et al., 2003) and a global pandemic in 2019 (Dong and Bouey, 2020). Following the 2003 epidemic, SARS-CoV-specific CD8+ T cell responses were detected in patients who had recovered from the infection (Huang, Cao, et al., 2007; Yang et al., 2007). A number of HLA-A2-restricted spike protein epitopes were identified (Wang et al., 2004; Tsao et al., 2006; Lv et al., 2009). These include HLA-A2-ILPDLKPT(5787–795) and HLA-A2-KLPDDFMGCV(5411–420) (Zhao et al., 2010). These HLA-A2-restricted spike protein epitopes trigger specific CD8+ T cells responses that can protect the host from infection (Zhao et al., 2010). Since the beginning of the 2019 pandemic, wide ranging CD8+ T cell responses to SARS-CoV have been detected in recovered patients (Kared et al., 2021). Furthermore, vaccines developed against SARS-CoV have been shown to successfully trigger T cell responses (Ramasamy et al., 2020).

Specific TCR-pMHC interactions have been described in many other viral infections including Epstein-Barr virus (EBV) (Steven et al., 1997; Hislop et al., 2002, 2005; Pudney et al., 2005; Forrest et al., 2018), varicella zoster virus (VZV; chicken pox) (Frey et al., 2003; Chiu et al., 2014), and herpes simplex virus-1 (HSV-1) (Ott et al., 2017; Treat et al., 2017; Russell et al., 2018). Understanding the TCR-pMHC interactions involved in various human pathologies can be useful in directing the quest for therapeutic interventions. A viral pathogen of particular interest to this thesis is the human immunodeficiency virus (HIV) which is discussed in detail in Appendix Section 7.1.

1.5.2 – CD8+ T cells in autoimmunity

Viruses will evolve quickly to try to evade the immune system (viral escape mutations are discussed in the context of HIV in Section 4.1.1). However, not all viral escape mutations successfully avoid the immune system. As discussed in Section 1.4.3, this is because many TCRs are highly promiscuous. T cell crossreactivity (Mason, 1998; Sewell, 2012), can be of significant benefit as it can prevent immunological gaps, and allow the recognition of escape mutants or as-yet unencountered pathogens (Lee et al., 2004; Varela-Rohena et al., 2008; Wen et al., 2017). However, this comprehensive immune coverage comes at a price.
Autoimmunity arises when crossreactive T cells unintentionally recognise self-antigen, see Figure 1.7.

Autoimmune diseases develop in approximately 5% of people (Goodnow, 2007). They are commonly classified based on the nature of their target antigen (Khan and Ghazanfar, 2018). They can be systemic – such as rheumatoid arthritis (RA) and multiple sclerosis (MS) – or organ-specific – such as type 1 diabetes (T1D), discussed in detail in Section 3.1.1, or Goodpasture’s syndrome (Khan and Ghazanfar, 2018).

Given the extensive variation generated by V(D)J gene recombination (see Section 1.2.3) (Tonegawa, 1983; Alt et al., 1992; Lewis, 1994), the production of self-reactive TCRs is almost inevitable (Schwartz, 2012; Khan and Ghazanfar, 2018). To account for this, a number of regulatory mechanisms have evolved, including negative selection in the thymus, peripheral suppression through regulatory T cell subsets, and the requirement for co-stimulation for full T cell activation (Schwartz, 2012). Collectively these mechanisms achieve ‘self-tolerance,’ a notion first introduced in 1945 to describe how the immune system is able to ignore self (Owen, 1945).

In the 1950s, studies on transplant rejections in mice supported the idea that the immune system could tolerate self, whilst maintaining aggression against foreign antigens (Billingham, Brent and Medawar, 1954). In the context of Burnet’s clonal selection theory, proposed in the same decade, the logical assumption was that self-reactive clones must be eliminated by an undiscovered mechanism that functions to prevent autoimmunity (Burnet, 1959; Schwartz, 2012). Proof-of-concept for the clonal elimination theory occurred in 1987 when Kappler et al. showed that a specific TCR present in the immature thymocyte population would not emerge in the periphery if a high affinity ligand for that TCR was expressed in the thymus (Kappler, Roehm and Marrack, 1987). In 1994, staining for DNA strand breaks was used to demonstrate that negative selection occurs through apoptosis, and to localise the event to the thymic medulla (Surh and Sprent, 1994). The, now well-recognised, process of negative selection is discussed in Section 1.2.3.
Figure 1.7: The advantages and disadvantages of crossreactivity. T cell crossreactivity allows a limited number of specific TCRs to provide comprehensive immune coverage, protecting the host from any potential immune challenge that could be encountered — including viral escape mutants and neoplastic changes in host cells. The downside to crossreactivity is the risk of autoimmunity.

Despite this fairly robust system for removing self-reactive T cells, autoimmunity does occur. Autoreactive T cells can be found in the peripheral circulation of healthy individuals (Kitze et al., 1988; Naquet et al., 1988; Mason and Fowell, 1992; Fowell and Mason, 1993; King and Sarvetnick, 1997). It is also evident that if a self-antigen is not expressed in the thymus, a T cell cannot be tolerised to it, this includes both self-antigens that do not occur in the thymus and splice variants that may differ between tissues (Akkaraju et al., 1997; Pugliese et al., 1997; Klein et al., 2000; Kamradt and Avrion Mitchison, 2001; Schwartz, 2012). Some antigens exist in immune privileged sites or are protected from the immune system by the blood-brain barrier, these can cause an immune reaction should they become aberrantly expressed elsewhere (Pachter, De Vries and Fabry, 2003; Benhar, London and Schwartz, 2012; Li, Wang and Han, 2012; Khan and Ghaanfar, 2018). These observations necessitate the existence of additional regulatory mechanisms to suppress autoreactive T cells that
escape thymic selection. Thymic selection is often referred to as central tolerance, to distinguish from autoreactive suppression that occurs in the periphery – known as peripheral tolerance (Schwartz, 2012).

Peripheral tolerance is executed by a subset of CD4-expressing T cells known as T regulatory cells (Tregs) (Schwartz, 2012). Cells with suppressive function were shown to populate the periphery in the 1990s (Powrie and Mason, 1990; Modigliani et al., 1995; Sakaguchi et al., 1995). Such cells intentionally recognise self-antigen and negatively influence any response in the vicinity of the ligand. This so-called ‘bystander suppression’ involves the secretion of inhibitory cytokines such as IL-4, IL-10, and TGF-β (Schwartz, 2012).

If, despite peripheral suppression, a T cell does recognise an antigen in the periphery an additional layer of protection exists. A naïve T cell generally requires additional co-stimulation to achieve full activation, in the absence of these additional stimuli expressed by professional APCs presenting foreign antigen it will become senescent or undergo apoptosis (known as clonal deletion) (Lafferty and Woolnough, 1977; Ragazzo, 2001; Steinman, Hawiger and Nussenzweig, 2003; Schwartz, 2012). This ‘two signal’ model allowing discrimination between self and non-self has long been appreciated (Bretscher and Cohn, 1970; Lafferty, Misko and Cooley, 1974). In 1987, Jenkins and Schwartz demonstrated that chemically-fixed APCs presenting the appropriate antigen failed to activate T cells. These chemical-fixed presenting cells were metabolically inactive and thus unable to produce proinflammatory cytokines or upregulate their expression of co-stimulatory molecules (Jenkins and Schwartz, 1987). From these, and other, experiments it became clear that engagement of a suitable pMHC target is insufficient to activate naïve T cells without additional costimulatory signals (Bach et al., 1977; Lafferty and Woolnough, 1977; Roska and Lipsky, 1985; June et al., 1987; Mueller, Jenkins and Schwartz, 1989). The two signal model is discussed further, in the context of T1D, in Section 3.1.2.

One well-described co-stimulatory interaction is that of CD28 on T cell surfaces with CD80/CD86 on the surface of APCs (Gmünder and Lesslauer, 1984; Lesslauer et al., 1986; Aruffo and Seed, 1987; Linsley, Clark and Ledbetter, 1990; Howard, Rochelle and Seldin, 1991; Naluai et al., 2000). CD28 is constitutively expressed on CD8+ and CD4+ T cells (Gross, St John and Allison, 1990; Lee et al., 1990), but expression of CD80/CD86 on APCs such as dendritic cells depends upon their encountering foreign antigen (Freeman, Borriello, et al., 1993; Freeman, Gribben, et al., 1993; Inaba et al., 1994). CD28 engagement contributes to the intracellular T cell signalling pathways and enhances the overall avidity of the interaction
between the two cells (Riha and Rudd, 2010). Its importance can be demonstrated by the immunocompromisation seen in CD28-deficient mice (Shahinian et al., 1993). Additionally, enhancement of this regulatory pathway can amplify the response against traditionally weaker antigen such as tumour antigen (Chen et al., 1992; Baskar et al., 1993; Townsend and Allison, 1993; Allison, Hurwitz and Leach, 1995).

In a further regulatory effort, CTLA-4 is upregulated in response to activation (Alegre, Frauwirth and Thompson, 2001). CTLA-4 is a homolog of CD28 and thus also a ligand for the CD80/CD86 proteins, but provides a negative feedback loop (Khan and Ghazanfar, 2018). Blocking CTLA-4 by means of antibodies has an encouraging effect on the anti-tumour immune response (Walunas et al., 1994; Leach, Krummel and Allison, 1996). Whereas triggering CTLA-4 signalling inhibits the T cell response (Walunas et al., 1994; Krummel and Allison, 1995). It competes with CD28 for CD80/CD86 thus balancing the positive and negative signals (Alegre, Frauwirth and Thompson, 2001).

Another noteworthy regulatory signalling mechanism is programmed cell death protein (PD)-1. PD-1 is expressed on activated T cells (Ishida et al., 1992; Agata et al., 1996; Vibhakar et al., 1997). It was first described in 1992 (Ishida et al., 1992). Deficiencies in PD-1 led to mice displaying increased rates of multiple autoimmune disorders and a loss of peripheral tolerance (Nishimura et al., 1999, 2001; Ansari et al., 2003). Its ligands, PD-L1 and PD-L2 are expressed on APCs and activated lymphocytes (Carreno and Collins, 2002; Eppihimer et al., 2002; Mazanet and Hughes, 2002; Sheppard et al., 2004). Interaction of PD-1 with its ligands evokes negative feedback on T cell activation, including reducing proliferation and cytokine secretion (Freeman et al., 2000; Latchman et al., 2001; Carter et al., 2002; Brown et al., 2003). PD-1 activation results in a reduction in the phosphorylation of early T cell signalling molecule ZAP-70 – thus mediating its effects by dampening down signalling from engaged TCRs (Sheppard et al., 2004).

Despite these regulatory mechanisms to prevent it, autoimmunity does occur. This indicates that these regulatory systems must be surmountable, resulting in the aberrant activation of lymphocytes (Khan and Ghazanfar, 2018). This can occur in a number of ways, including molecular mimicry and the emergence of cryptic self.

T cells with the potential for autoreactivity can ignore a self-target that is insufficiently potent to induce activation (Akkaraju et al., 1997). The naïve T cell can be activated by a stronger – often pathogen derived – antigen that is a molecular mimic of the self-antigen and with
which it can cross-react, **Figure 1.8** (Raman *et al.*, 2016; Cole, Van Den Berg, *et al.*, 2017; Khan and Ghazanfar, 2018).

Many examples of molecular mimicry between pathogenic peptides and self-peptides have been described in the literature. An example is the ASVKVLLGRKDSERG peptide epitope, derived from herpes simplex virus (HSV)-1. It was shown to be recognised by autoreactive T cells that could target a corneal antigen in mice. Mutated strains of the virus, lacking this epitope, were not recognised (Zhao *et al.*, 1998). In another example, group A streptococcal proteins were shown to be capable of triggering T cell responses against the cardiac protein myosin (Dale and Beachey, 1986).

Molecular mimicry has been cited as a trigger for multiple sclerosis (MS). Levin *et al.* observed that Epstein-Barr virus (EBV) infection increases the susceptibility to MS in young adults (Levin *et al.*, 2003). A case of molecular mimicry between an EBV peptide and a myelin-derived self-peptide has been demonstrated and shown to be able to trigger TCR-dependent demyelination (Lang *et al.*, 2002). In this manner, crossreactivity following exposure to particular viral infections could feasibly contribute to the development of MS (Olson *et al.*, 2001; Harkiolaki *et al.*, 2009; Münz *et al.*, 2009).
Figure 1.8: Molecular mimicry in autoimmunity. A schematic representation showing how exposure to a true pathogenic antigen can cause a T cell to become activated and crossreact with self-antigens to which it had previously been ignorant. For illustrative purposes this figure envisions a viral antigen acting as a molecular mimic for an antigen expressed on pancreatic β cells.

Other examples of molecular mimicry occur in cases of rheumatoid arthritis (RA). RA has been associated with alphavirus infections (Laine, Luukkainen and Toivanen, 2004; Rulli et al., 2007; Toivanen, 2008). This has prompted bioinformatical investigation into potential cases for molecular mimicry (Venigalla, Premakumar and Janakiraman, 2020). Venigalla et al. analysed the proteomes of RA-associated alphaviruses to identify homology to human proteins – particularly those specifically associated with joint cartilage and RA (Venigalla, Premakumar and Janakiraman, 2020). From this analysis they identified epitopes with
predicted ability to bind RA-associated class II MHC molecules. Potential epitopes include the viral epitope \textit{YSGGRFTIPTGA\textsubscript{GKPGDDG}RPIFDN} as a homolog to the GKPGDDG sequence found in human collagen type II (Venigalla, Premakumar and Janakiraman, 2020). T cell clones and antibodies directed against this protein had already been identified in RA patients (Londei \textit{et al.}, 1989; Mullazehi \textit{et al.}, 2012). The viral epitope \textit{SKDVYANTQLVQRPAAGTVHVPYS} is also a potential homolog to the QSPAAGTVQGRVP sequence from human cartilage intermediate layer protein 1 (Venigalla, Premakumar and Janakiraman, 2020). Antibodies against this protein have been previously described in the context of RA (Tsuruha \textit{et al.}, 2001).

Molecular mimicry has also been described in the context of Goodpasture’s syndrome, an autoimmune disorder in which the principle target is type IV collagen, found in the glomerular basement membranes (GBM) of the kidneys and in the alveoli of the lungs (Hudson \textit{et al.}, 1989; Merkel \textit{et al.}, 1996). A role for T cells in Goodpasture’s syndrome was indicated by T cell infiltration in diseased kidneys, which correlated with disease severity (Bolton \textit{et al.}, 1987; Hu \textit{et al.}, 2016, 2017). The symptoms can be alleviated through administration of anti-T cell therapies such as inhibition of the CD28-CD80/CD86 interaction (Reynolds \textit{et al.}, 2000, 2004), or via anti-CD4/CD8 antibodies (Reynolds \textit{et al.}, 2002). T cell clones specific for the GBM have been detected (Merkel \textit{et al.}, 1996; Derry \textit{et al.}, 2008), and a number of T cell epitopes have been described (Merkel \textit{et al.}, 1996; Wu \textit{et al.}, 2003). In 2006, Arends \textit{et al.} identified microbial peptides that resembled known T cell epitopes and could induce glomerular inflammation in rats suggesting a role for molecular mimicry in yet another autoimmune disease (Arends \textit{et al.}, 2006).

During an immune response the cytokine-rich environment encourages endogenous cells to increase pMHC expression at the cell surface, exposing antigen that would not normally be presented (Lehmann \textit{et al.}, 1992; Schwartz, 2012). This emergence of ‘cryptic self’ has the potential to propagate autoimmunity as T cells would not have been tolerised to unencountered antigen in the thymus (Lehmann \textit{et al.}, 1992; Schwartz, 2012). These T cells are specific – not to pathogenic molecular mimics but to self-antigens that are not normally exposed to immune scrutiny (Moudgil and Sercarz, 1994; Itariu and Stulnig, 2014). T cells with specificities to non-dominant self-epitopes can escape negative selection (Schild \textit{et al.}, 1990; Cibotti \textit{et al.}, 1992; Slifka \textit{et al.}, 2003), these T cells then reside in the periphery harmlessly until the, fairly unlikely but potential, emergence of their cryptic self-antigen (Gammon and Sercarz, 1989; Lipham \textit{et al.}, 1991).
A final noteworthy mechanism for escaping the extensive self-tolerance regulations is through the expression of dual TCRs. This describes a cell that has multiple Vα or Vβ chains rearranged and expressed. Differing pairings can provide the cell with differing specificities (Heath and Miller, 1993; Elliott and Altmann, 1995; Padovan et al., 1995). T cells expressing two fully rearranged TCRβ chains were first isolated from mice in 1988 (Matis, Ezquerra and Coligan, 1988; Triebel et al., 1988). Dual α-chain expression was demonstrated shortly afterwards (Furutani et al., 1989; Kuida et al., 1991). In 1993, the existence of dual TCRs in humans was demonstrated (Padovan et al., 1993). It is thought that ~10% of αβ TCRs co-express two α-chains (Heath et al., 1995; Niederberger et al., 2003; Schuld and Binstadt, 2019), whilst ~1% of αβ TCRs co-express two β-chains (Balomenos et al., 1995; Davodeau et al., 1995; Sieh and Chen, 2001; Brady, Steinel and Bassing, 2010).

In the context of autoimmunity, an interesting example of dual TCR expression was described in 2010 by Qingyong et al. They described a case of MS in mice that could be induced by infection with vaccinia virus (Ji, Perchellet and Goverman, 2010). This causality could not be explained by locating crossreactive molecular mimics between viral peptides and peptides within the MBP sequence (Ji, Perchellet and Goverman, 2010). Instead a co-expressed, alternative Vβ chain was identified within virus-specific T cells, allowing for two unique TCR pairs. One pair was specific for MBP and the other for vaccinia virus. Activation of the clone by exposure to vaccinia virus initiated cytotoxicity against MBP via its co-expressed TCR (Ji, Perchellet and Goverman, 2010).

Mechanisms for the breakdown of self-tolerance emphasise the importance of investigating the specific TCR-pMHC interactions involved in autoimmune settings. An autoimmune disorder of particular interest to this thesis is type 1 diabetes mellitus (T1D). It is discussed in detail in Section 3.1.

1.6 – Aims and objectives
Functionally, the TCR-pMHC interaction encompasses an exceptional degree of plasticity and crossreactivity that enables comprehensive immune coverage to protect the host from any potential, often as yet unencountered, immune threat (see Section 1.5.1). Yet this crossreactive nature of the TCR-pMHC interaction can result in aberrant recognition of self in autoimmune diseases (see Section 1.5.2).
Ultimately, further investigation into the plasticity of the TCR-pMHC interaction will enhance understanding of the nature of T cell antigen recognition. Here, three main aspects of the TCR-pMHC interaction are investigated:

1) Can understanding the inherent crossreactivity of a particular TCR-pMHC interaction aid the identification of specific pathogenic triggers for an autoimmune disease?
2) How would amino acid substitutions in the CDR3 loops of a TCR influence the repertoire of peptides it can interact with?
3) Finally, on the other side of the interaction, how is the molecular flexibility of the MHC influenced by the peptide cargo it carries?

To address these questions, I aimed to:

- Investigate whether the T1D-associated InsB4 T cell clone could crossreact with peptides from pathogenic proteomes to address the hypothesis that pathogenic molecular mimics might act as triggers of T1D.
- Use the extensively described interaction between the 868 TCR and its cognate HIV-1 Gag-derived peptide SLYNTVATL in the context of HLA-A2 to investigate how single point mutations in the amino acid sequence of a CDR3 loops might alter the peptide repertoire of the TCR.
- Probe the flexible nature of the pMHC using a T1D-relevant epitope (HLA-A2-ALWGPDPAAA) and a series of APLs designed around this epitope to investigate whether the red edge excitation shift (REES) phenomenon might be able to provide information regarding the peptide-dependent flexibility of the pMHC complex and add to the tool-kit of experimental techniques that can be used to assess pMHC flexibility.
2 – Materials and methods

2.1 – Protein production

2.1.1 – Buffers and media used in protein production

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lysogeny broth (LB) agar plates</td>
<td>10 g/L tryptone (VWR Chemicals), 5 g/L yeast extract (VWR Chemicals), 10 g/L NaCl (Fisher Scientific), 15 g/L bacteriological agar (Oxoid). Supplemented following autoclaving with 50 μg/mL carbenicillin (Fisher Bioreagents).</td>
</tr>
<tr>
<td>TYP media</td>
<td>16 g/L tryptone (VWR Chemicals), 16 g/L yeast extract (VWR Chemicals), 5 g/L NaCl (Fisher Scientific), 3.3 g/L potassium phosphate dibasic (Sigma Aldrich). Supplemented following autoclaving with 50 μg/mL carbenicillin (Fisher Bioreagents).</td>
</tr>
<tr>
<td>Lysis buffer</td>
<td>10 mM TRIS pH 8.1, 10 mM MgCl₂ (Sigma), 150 mM NaCl (Fisher Scientific), 10% glycerol (Fisher Chemical).</td>
</tr>
<tr>
<td>Triton wash</td>
<td>0.5 % Triton X (Sigma), 50 mM TRIS pH 8.1, 100 mM NaCl (Fisher Scientific), 2 mM ethylenediaminetetraacetic acid EDTA (Sigma Aldrich).</td>
</tr>
<tr>
<td>Resuspension buffer</td>
<td>50 mM TRIS pH 8.1, 100 mM NaCl (Fisher Scientific), 2 mM EDTA (Sigma Aldrich).</td>
</tr>
<tr>
<td>Guanidine buffer</td>
<td>6 M guanidine (Fisher Bioreagents), 50 mM TRIS pH 8.1, 100 mM NaCl (Fisher Scientific), 2 mM EDTA (Sigma Aldrich).</td>
</tr>
<tr>
<td>pMHC refold buffer</td>
<td>50 mM TRIS pH 8.1, 2 mM EDTA (Sigma Aldrich), 400 mM L-arginine (Sigma Aldrich), pH adjusted to 8 with HCl (Sigma Aldrich). 0.74 g/L cysteamine (Sigma) and 0.83 g/L cystamine (Aldrich) added prior to use.</td>
</tr>
<tr>
<td>TCR refold buffer</td>
<td>50 mM TRIS pH 8.1, 2 mM EDTA (Sigma Aldrich), 2.5 M urea (Sigma), pH adjusted to 8 with HCl (Sigma Aldrich). 0.74 g/L cysteamine (Sigma) and 0.83 g/L cystamine (Aldrich) added prior to use.</td>
</tr>
<tr>
<td>2M TRIS pH 8.1</td>
<td>242.24 g/L TRIS-base (Sigma), pH adjusted to 8.1 with HCl (Sigma Aldrich), filtered 0.45 μM.</td>
</tr>
<tr>
<td>0.5M EDTA pH 8</td>
<td>146.12 g/L EDTA (Sigma Aldrich), pH adjusted to 8 with NaOH (Fisher Chemical).</td>
</tr>
<tr>
<td>Buffer/Media</td>
<td>Composition/Details</td>
</tr>
<tr>
<td>-------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Dialysis buffer</td>
<td>20 L dH₂O 10 mM TRIS, pH 8.1.</td>
</tr>
<tr>
<td>Ion exchange buffer ‘A’</td>
<td>10 mM TRIS, filtered 0.45 mM, pH 8.1.</td>
</tr>
<tr>
<td>Ion exchange buffer ‘B’</td>
<td>10 mM TRIS, 1 M NaCl (Fisher Scientific), filtered 0.45 mM, pH 8.1.</td>
</tr>
<tr>
<td>Biomix A</td>
<td>0.5 M bicine buffer, pH 8.3</td>
</tr>
<tr>
<td>Biomix B</td>
<td>100 mM ATP, 100 mM magnesium acetate, 500 μM d-Biotin.</td>
</tr>
<tr>
<td>Phosphate buffered saline (PBS)</td>
<td>Oxoid, tablets composed of: 8 g/L sodium chloride, 0.2 g/L potassium chloride, 1.15 g/L di-sodium hydrogen phosphate, 0.2 g/L potassium dihydrogen phosphate.</td>
</tr>
<tr>
<td>HEPES buffer</td>
<td>150 mM NaCl (Fisher Scientific), 50 mM HEPES (Fisher Bioreagents).</td>
</tr>
<tr>
<td>Crystal buffer</td>
<td>10 mM TRIS pH 8.1, 10 mM NaCl (Fisher Scientific).</td>
</tr>
<tr>
<td>Sodium dodecyl sulphate (SDS) reducing sample buffer</td>
<td>Lithium dodecyl sulphate at pH 8.4 (Novex, Thermo Fisher), 20 mM dithiothreitol (DTT) (Fisher Chemical).</td>
</tr>
<tr>
<td>SDS non-reducing sample buffer</td>
<td>Lithium dodecyl sulphate at pH 8.4 (Novex, Thermo Fisher).</td>
</tr>
<tr>
<td>Psi Broth</td>
<td>5 g/L yeast extract (VWR Chemicals), 20 g/L tryptone (VWR Chemicals), 5 g/L magnesium sulphate (Fisher Chemical). pH to 7.6 with potassium hydroxide (AnaIR) before autoclaving.</td>
</tr>
<tr>
<td>Transformation buffer I</td>
<td>30 mM potassium acetate (0.588 g) (Sigma), 100 mM rubidium chloride (2.42 g) (Sigma), 10 mM calcium chloride (0.294 g) (Sigma Aldrich), 50 mM magnesium chloride (2.0 g) (Fisher Chemical), 15 % glycerol (w/v) (30 mL) (Fisher Chemical). pH to 5.8 with acetic acid (Fisher Scientific) before filtering 0.45 μm and storing at 4°C. Measurements given for making 200 mL.</td>
</tr>
<tr>
<td>Transformation buffer II</td>
<td>10 mM MOPS (0.21 g) (Sigma), 75 mM calcium chloride (1.1 g) (Sigma Aldrich), 15 % glycerol (w/v) (15 mL) (Fisher Chemical). pH to 6.5 with sodium hydroxide (Fisher Chemical) before filtering 0.45 μm and storing at 4°C. Measurements given for making 100 mL.</td>
</tr>
</tbody>
</table>

**Table 2.1**: A summary of the buffers and media used during protein production throughout this thesis.
2.1.2 – Vectors and protein sequences

The carbenicillin-resistant pGMT7 vector was used for expression of proteins of interest in bacterial cells, Figure 2.1. pGMT7 is an extensively used bacterial expression vector that utilises the T7 RNA polymerase promotor (Rosenberg et al., 1987; William Studier et al., 1990; Banham and Smith, 1993; Reid et al., 1996). The constructs shown below were cloned into pGMT7 using BamH1 and EcoR1 restriction sites. In addition to conferring resistance to carbenicillin, pGMT7 allows inducible expression of the construct via isopropyl β-D-1-thiogalactopyranoside (IPTG) and the lac operon (Hughes et al., 1991). This expression system was used to produce soluble HLA-A2 and human β2M. In some cases, construct sequences were modified slightly from wildtype sequences to optimise in vitro experiments (Boulter et al., 2003). For tetramerisation studies an N-terminal 15 amino acid biotinylation sequence was added to HLA-A2 (GLNDIFEAQKIEWHE) (AviTag™, Avidity).

HLA-A2wt:

MGSHSMYFFTSVSRPGRGEPRFIAVGYVDDTQFVRFDSDAASQRMetEPRAPWIEQEGPEY WDGETRKVKAHSQTHRVDLGTRLRGYYNQSEAGSHTVQRMYGCDVGSDFRFLRGYHQYAYDGK DYIALKEDLRSTADMAQQTTKHKEAAHVAEQLRAYLEGTCWETLRRYENGKELQRTD APKTHMTHHAVSDHEATLRCWALSEYFAAEITLTVQRDEDQ TDTELVEFTPAGDQFQKWA AVVVPSSQEQRHCHVQHEGLPKPLTRWEH

HLA-A2tag:

MGSHSMYFFTSVSRPGRGEPRFIAVGYVDDTQFVRFDSDAASQRMetEPRAPWIEQEGPEY WDGETRKVKAHSQTHRVDLGTRLRGYYNQSEAGSHTVQRMYGCDVGSDFRFLRGYHQYAYDGK DYIALKEDLRSTADMAQQTTKHKEAAHVAEQLRAYLEGTCWETLRRYENGKELQRTD APKTHMTHHAVSDHEATLRCWALSEYFAAEITLTVQRDEDQ TDTELVEFTPAGDQFQKWA AVVVPSSQEQRHCHVQHEGLPKPLTRWEH GLNDIFEAQKIEWHE

β2M:

MIQRTPKIQVYSRHPENGKSNFLNCYVSFGPHPSDIEVDDLKNGERIEKVEHSLSFSKDW FYLYYTFTEPTKDEYACRVNHVTLSSQPCKKVWDREDM
Figure 2.1: Schematic of the pGMT7 plasmid used for protein production. The desired protein construct is cloned into the pGMT7 using the BamH1 and EcoR1 restriction sites. Successful transformation of bacterial cells is detected via the conferred ampicillin resistance. The presence of the lac operon allowed inducible protein production.

2.1.3 – Making chemically competent cells
A vial of the previous (or original) batch of competent cells (either BL21 DE3, TOP10, or XL-10 Gold) was used to inoculate an agar plate supplemented with the relevant antibiotic (chloramphenicol, streptomycin, or chloramphenicol respectively), plates were grown at 37°C overnight. The following day a single colony was used to inoculate 50 mL Psi broth supplemented with the relevant antibiotic, this was grown overnight in an orbital shaking
incubator (37°C, 100 revolutions per minute; RPM) (Thermo Fisher). 1 mL of the resulting culture was used to inoculate a fresh 100 mL of Psi broth supplemented with the relevant antibiotic. This was then grown in an orbital shaking incubator (37°C, 100 RPM) (Thermo Fisher) until it reached an optical density at 600 nm (OD_{600nm}) of 0.45. Cells were pelleted by spinning in a pre-cooled centrifuge 6371 x g (6000 RPM) for 10 minutes (Beckman Coulter Avanti J-E centrifuge). Cells were then washed in 40 mL of Transformation Buffer I and incubated on ice for 15 minutes. Cells were pelleted by spinning in a pre-cooled centrifuge 6371 x g (6000 RPM) for 10 minutes (Beckman Coulter Avanti J-E centrifuge). Cells were then resuspended in 4 mL Transformation Buffer II, aliquoted into sterile microcentrifuge tubes, and snap frozen. Performed by Anna Fuller, Aaron Wall, or Théo Morin.

2.1.4 – Transformation of competent cells
For transformation, 20 μL aliquots of bacterial cells were first thawed on ice for 5 minutes. 1 μL (containing 50-100 ng of DNA) of vector DNA was added to the 20 μL bacterial cells followed by a further 5-minute incubation on ice. To encourage uptake of vector DNA, bacterial cells were then heat-shocked by a 2-minute incubation at 42°C. Following heat-shock, cells were allowed to recover on ice for 5 minutes before the full volume was plated onto LB agar plates supplemented with 50 μg/mL carbenicillin. Plates were incubated at 37°C overnight to allow growth of colonies.

BL21 DE3 chemically competent *Escherichia coli* cells were used to produce MHC chains in the form of inclusion bodies for artificial refolding. TOP10 chemically competent *E. coli* cells were used to amplify DNA. XL-10 Gold ultracompetent cells were used to amplify the larger pELNS plasmid for viral work, discussed in Section 2.3.

2.1.5 – Amplifying DNA via mini preparation
Competent, TOP10, cells were transformed with the relevant DNA as described in Section 2.1.4. Single colonies were selected to inoculate 5 mL of TYP media and grown overnight in an orbital shaking incubator (37°C, 220 RPM) (Thermo Fisher). The following day, bacterial cells were harvested by centrifuging the cultures at 16,000 x g for 5 minutes (Eppendorf Centrifuge 5424). DNA extraction was then carried out using a PureLink Quick Plasmid Miniprep kit (Thermo Fisher) according to the manufacturer’s instructions. The resulting pellets were resuspended in 250 μL of Resuspension Buffer (R3) supplemented with 20 mg/ml RNase. Cells were then lysed in 250 μL Lysis Buffer (L7), incubating at room
temperature for 5 minutes to allow for lysis. 350 μL of Precipitation Buffer (N4) was used to precipitate protein and genomic DNA from the resulting lysate, leaving circular plasmid DNA in solution. Lysates were then centrifuged at 16,000 x g for 10 minutes (Eppendorf Centrifuge 5424) to pellet the precipitant. Supernatants were loaded on to HiPure spin columns (Thermo Fisher) assembled with wash tubes to collect flow through. Samples were centrifuged at 12,000 x g for 1 minute (Eppendorf Centrifuge 5424) to bind DNA to the column. Columns were then washed to remove impurities, first with 500 μL of Wash Buffer (W10) supplemented with 60 %v/v ethanol, then with 700 μL of Wash Buffer (W9) supplemented with 70 %v/v ethanol. Columns were centrifuged at 12,000 x g for 1 minute (Eppendorf Centrifuge 5424) once more to remove residual Wash Buffer and ethanol. To recover DNA, columns were incubated in 30 μL of nuclease-free water for 5 minutes at room temperature before being centrifuged at 12,000g for 2 minutes (Eppendorf Centrifuge 5424). Concentrations were measured using a spectrophotometer (Geneflow). For DNA, optical density was measured at 260 nm and concentration was calculated by the instrument using Beer’s Law: A=ε*l*C. The extinction coefficient of the protein was then used to determine final concentration. Samples were sequenced using Eurofins Geomics’ Mix2Seq service, for protein preparation all constructs were within the pGMT7 vector so could be sequenced with the T7 forward primer (TAATACGACTCACTATAGGG). For longer constructs, the SP6 primer can also be used for sequencing (ATTTAGGTGACACTATAG).

2.1.6 – Expression of constructs and purification of inclusion bodies
Following overnight growth on agar at 37°C (as described above), single colonies were selected (Figure 2.2A) and grown in 10 mL TYP media supplemented with 50 μg/mL carbenicillin (37°C, 220 RPM) until an OD$_{600nm}$ of 0.5 was reached (Figure 2.2A). The following day these starter cultures were expanded in 1 L of TYP media supplemented with 50 μg/mL carbenicillin until an OD$_{600nm}$ of 0.5 was reached (Figure 2.2A). Based on previous experiments by my laboratory this OD is considered to be indicative of the peak of the bacterial exponential growth phase, and so is the optimal point at which to induce expression of the protein of interest before the cells begin to enter the death phase. A 500 μL ‘pre-induction’ sample of culture was taken to later be run on an SDS gel to verify protein expression (Figure 2.2B). Protein expression was induced using 500 μL of 1 M IPTG (to give a working concentration 0.5 mM). 3 hours post-induction a 500 mL sample of culture was taken to later be run on an SDS gel to verify protein expression (Figure 2.2B), bacterial cells were then harvested by centrifuging in an Eppendorf 5810R centrifuge for 20 minutes at
2755 x g (3700 RPM). The resulting cell pellets were resuspended in 40 mL of lysis buffer for overnight freeze-thaw to encourage cell lysis. To further aid cell lysis, thawed cells were sonicated using a Bandelin Sonopuls sonicator for 20 minutes at 50 % power, using a 2 x 10 % cycle. Dirtier lysates were subject to a second round of freeze-thaw and sonication. The cell lysate was treated with 200 μL of 20 mg/mL DNAse enzyme (working concentration 0.1 mg/mL) and incubated for 1 hour at 37°C in an orbital shaking incubator (220 RPM) (Thermo Fisher). For purification, the inclusion bodies were washed three times, twice in 200 mL of a triton wash buffer, and finally in 200 mL of resuspension buffer. Between each wash, the supernatant was discarded, and the pellet was resuspended in the next buffer by mechanical homogenisation (VWR, VDI 25). A final 500 μL sample of inclusion body dissolved in resuspension buffer was taken to be run on an SDS gel to verify protein purification (Figure 2.2B). Inclusion bodies were pelleted by centrifuging at 17,696 X g (10,000 RPM) for 20 minutes (Beckman Coulter Avanti J-E centrifuge). Finally, the pelleted inclusion bodies were dissolved in 10 mL guanidine buffer. To remove any insoluble debris inclusion bodies were centrifuged in an Eppendorf 5810R centrifuge for 30 minutes at 2755 x g (3700 RPM). Inclusion body concentration was measured using a spectrophotometer (Geneflow) at 280 nm, as described in Section 2.1.5. The extinction coefficient was then used to determine final concentration. On average 50 mg of purified inclusion bodies was yielded per 1 L of TYP culture.

To verify protein expression and purification, an SDS gel was used to compare pre- and post-IPTG induction samples to a ‘post-wash’ sample of inclusion bodies in resuspension buffer, this gel was used to verify purity of the inclusion bodies (Figure 2.2B). 4 μL of SDS reducing sample buffer or SDS non-reducing sample buffer was added to 20 μL of sample and incubated at 95°C for 10 minutes. To remove bubbles, samples were centrifuged at 16,000 x g for 30 seconds. 20 μL of each sample was loaded onto a gel (BOLT 4-12 % Bis-Tris Pius, Invitrogen), with the pre-stained protein ladder (Elite pre-stained protein ladder, 6.5-270 kDa, Generon). Gels were run at 165 V for 35 minutes in 1X BOLT MES SDS-PAGE Running Buffer (Invitrogen). To visualise bands gels were stained with 25 mL Quick Coomassie Stain (Generon) before being destained with water.
Figure 2.2: Protein production via *E. coli* expression systems. **A)** Bacterial cells that have received the pGMT7 vector containing the protein of interest are able to grow on carbenicillin agar plates. Single colonies are selected for expansion in liquid culture before protein production is induced using IPTG. **B)** Protein production and subsequent washing is verified using SDS-PAGE. Rows A, B, and C represent pre-induction, post-induction, and resuspension buffer samples respectively, in reducing conditions. Rows D, E, and F contain the same under non-reducing conditions. A band at ~43 kDa indicates production of HLA-A2 heavy chain.
2.1.7 – Refolding soluble protein from inclusion bodies
pMHC refold buffer were made up according to recipes in Table 2.1, filtered 1.2 μM (Fisher Scientific) and 0.45 μM (Sartorius) and pre-cooled to 4°C. For pMHC refolds, 30 mg HLA-A2 α chain inclusion bodies, 30 mg human β2M inclusion bodies, and 4 mg of a relevant synthetic peptide (Peptide Synthetics), per 1 L refold buffer, was incubated with 10 mM DTT (Fisher Chemical) at 37°C for 30 minutes. The inclusion body mixture was then diluted into the cold refold buffer and stirred at 4°C for 6-8 hours or overnight. Refold mixtures were then dialysed twice in 20 L of 10 mM TRIS pH 8.1 to ensure the conductivity of the refold was <1 mS/cm. Dialysed refold mixtures were then re-filtered (as above) prior to purification (Section 2.1.8).

2.1.8 – Purification of soluble protein by fast protein liquid chromatography (FPLC)
All FPLC procedures were carried out using an ÄKTA Pure (GE Healthcare) and Unicorn 6.3 software (GE Healthcare). A Poros 50 HQ anion exchange column (Life Technologies) was washed with 100 mL of ion exchange buffer ‘B’ (see Table 2.1) before being equilibrated using 100 mL of ion exchange buffer ‘A’. Unless otherwise stated, a flow rate of 20 mL/minute and a pressure alarm of 5 MPa was used throughout. A protein sample was then loaded onto the column. The sample was eluted, using increasing concentrations of ion exchange buffer ‘B’ (Figure 2.3A), into 5 mL FPLC tubes (Greiner Bio-One) held in a fraction collector (F9-R, GE Healthcare). Fractions were then analysed by SDS-PAGE (as described in Section 2.1.6) for presence of pure protein of interest (Figure 2.3B). Fractions verified to contain protein of interest were concentrated to ~1 mL using a Amicon Ultra 4 mL 10 kDa centrifugal filter (Merck) by centrifuging at 2755 x g (3700 RPM, Eppendorf 5810R centrifuge) (for approximately 20 minutes). If pMHC samples required biotinylation, samples were instead concentrated to 700 mL using the same method. Following this, 100 μL each of Biomix A, Biomix B, and 500 μM d-Biotin (Avidity) were added to the sample along with 2 μL of BirA biotin-protein ligase at 3 mg/mL (Avidity). The solution was left at room temperature overnight to allow for biotinylation of the HLA heavy chain.

A SuperdexTM 200 Increase 10/300 GL size-exclusion column (VWR) was used to purify sample by size (removing aggregates and excess biotin) and for buffer exchange, the column was pre-equilibrated in the appropriate buffer (PBS, unless otherwise stated). The aforementioned ~1 mL sample was transferred into a 1 mL syringe (BD Plastipak), avoiding air bubbles, and manually loaded onto the column. As above, sample was eluted into 5 mL
FPLC tubes (Figure 2.3C). Fractions were analysed by SDS-PAGE (Section 2.1.6) for presence of pure protein of interest (Figure 2.3D), verified fractions were concentrated to ~1 mL using a Amicon Ultra 4 mL 10 kDa centrifugal filter (Merck) by centrifuging at 2755 x g (3700 RPM, Eppendorf 5810R centrifuge) (for approximately 20 minutes). Sample concentrations were measured using a spectrophotometer (Geneflow) at 280 nm, as described above. Samples were used in ongoing experiments, or frozen at -20°C for storage.

Figure 2.3: Purification of soluble protein. Protein purification occurs in two stages. A) shows a typical ion exchange chromatography trace. B) Fractions corresponding to peaks are run on SDS-PAGE under both reducing and non-reducing conditions to verify presence of desired protein. C) Fractions are concentrated and subject to size exclusion chromatography to remove any residual impurities. D) Fractions corresponding to peaks are once again verified by SDS-PAGE.
2.2 – Biochemical assays

2.2.1 – Dynamic light scattering to exclude aggregated samples
Dynamic light scattering (Zetasizer) was used to confirm that freeze-thawed proteins had not aggregated or unfolded. 50 μL sample was incubated at 4°C before measurements were taken. Peaks between 1-10 nm represented refolded proteins. Samples displaying peaks of larger values were discarded as aggregated protein.

2.2.2 – Red edge excitation measurements of HLA-A2 complexes
Fluorescence experiments were performed as previously described (Catici et al., 2016). REES measurements were performed using a Perkin Elmer LS50B Luminescence Spectrometer (Perkin Elmer, Waltham, MA, USA) which was connected to a circulating water bath in order to maintain a temperature of 15°C (±1°C). Samples were equilibrated at this temperature for 5 minutes. Tryptophan emission was measured between 325 nm and 500 nm with excitation and emissions slit widths of 3.5 nm. The excitation wavelength was initially set at 292 nm and subsequently increased at intervals of 1 nm for a total of 19 scans (in triplicate), thus the highest excitation wavelength used was 310 nm. The PBS buffer spectra was subtracted from the sample spectra at each experimental condition. The centre of spectral mass was calculated as follows:

\[ CSM = \frac{\sum (f_i \times \lambda_{Em})}{\sum (f_i)} \]

where \( f_i \) is the measured fluorescence intensity and \( \lambda_{Em} \) is the emission wavelength.

2.2.3 – Buffer exchange using a PD-10 desalting column
Samples prepared in Cardiff were in a PBS buffer (produced as described in Section 2.1). For pressure/temperature dependent fluorimetry measurements carried out in Bath samples were exchanged into a HEPES buffer (50 mM HEPES, 150 mM NaCl, pH 7.4) using a PD-10 desalting column containing Sephadex™ G-25 medium (GE Healthcare), according to the manufacturer’s instructions. The column storage solution was discarded and the column equilibrated in the desired HEPES buffer – to achieve this, a column volume of HEPES buffer was allowed to flow through the column by gravity pull four times. The sample was then added to the column and allowed to enter the packages bed. The volume was topped up to 2.5 mL with HEPES buffer, the flow through was discarded. The sample was then eluted by adding an additional 3.5 mL HEPES buffer. Eluate was collected in a series of test tubes which
were analysed by a spectrophotometer (measuring at 280 nm) for the presence of protein. Relevant eluate tubes were combined and the concentration measured, as described above.

2.2.4 – Pressure/temperature dependent fluorimetry
Pressure/temperature measurements were performed using an ISS high-pressure cell (ISS, Champaign, IL, USA) fitted with a custom fibre optic wires connecting to the fluorimeter and the water bath. pMHC complexes were excited at 295 nm, tryptophan emission was measured between 325-500 nm. Emission and excitation slits were set to 15 nm to minimise the signal to noise ratio (due to optimal set up of the pressure cell). Initial measurements were made at 10°C and increased in 5°C increments, up to 30°C. The pressure dependence at each temperature was measured at 50, 400, 800, 1200, and 1600 bar. Measurements were taken in triplicate. Following each full pressure/temperature range, repeat scans were taken at lower pressure/temperature conditions to ensure extreme the pressure/temperature conditions had not denatured the protein. For all measurements the appropriate buffer controls (HEPES) were subtracted prior to data processing.

2.2.5 – Surface plasmon resonance (SPR)
SPR experiments were carried out by Aaron Wall using a BIAcore T200™ and CM5 sensor chips (GE Healthcare) as described previously (Cole et al., 2008; MacLachlan et al., 2017; Whalley et al., 2020). A typical SPR immunoassay is shown schematically in Figure 2.4. Soluble pMHC molecules were bound to a CM5 Sensor Chip (GE Healthcare). A negative control was bound to flow cell 1. Flow cells 2-4 were occupied by either test samples or a positive control (if available). Approximately 500 response units (RUs) of each pMHC was immobilised on the chip. Equilibrium binding analysis experiments were then performed at 25°C by injecting TCR at varying concentrations over the chip. A flow rate of 30 µL/min was used. The $K_D$ values of each TCR-pMHC interaction were calculated by assuming 1:1 binding and Michaelis-Menten kinetics. Data was analysed using GraphPad Prism and fitted to a global fit algorithm:

$$AB=B\times AB_{MAX}/(K_D+B)$$
Figure 2.4: Schematic of surface plasmon resonance. Biotinylated pMHC monomers are immobilised on a CM5 sensor chip (gold) due to a high affinity interaction between biotin, streptavidin, and a dextran matrix that coats the surface of the chip. The chip is coupled to a prism that refracts light from a light source, this is then measured by an optical detector. Known concentrations of soluble TCR are injected over the chip at a controlled flow rate. Successful interaction between TCR and pMHC alters the refractive properties of the sensor chip, this change is measured by the optical detector. Variations in light intensity from baseline is used to monitor interaction between the proteins. This measurement is carried out at a range of TCR concentrations to allow plotting of a response curve.
2.3 – Molecular cloning

2.3.1 – Buffers and media used in molecular cloning and lentiviral production

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>LB media</td>
<td>10 g/L tryptone (VWR Chemicals), 5 g/L yeast extract (VWR Chemicals), 10 g/L NaCl (Fisher Scientific). Supplemented following autoclaving with 50 μg/mL carbenicillin (Fisher Bioreagents).</td>
</tr>
<tr>
<td>TAE buffer</td>
<td>242 g TRIZMA-BASE (Sigma), 57.1 mL acetic acid (Fisher Scientific), 100 mL 0.5 M EDTA pH 8, made up to 1 L with ddH₂O.</td>
</tr>
<tr>
<td>0.1 X TRIS-EDTA (TE) buffer</td>
<td>1 mM TRIS, 0.1 mM EDTA in ddH₂O, pH 8.0. Sterilised with 0.22 μM filter (Merck).</td>
</tr>
<tr>
<td>CaCl₂</td>
<td>2.5 M CaCl₂ = 9.18 g CaCl₂·2H₂O (Sigma), in 25 mL H₂O. Sterilised with 0.22 μM filter (Merck).</td>
</tr>
<tr>
<td>2xHEPES-buffered saline</td>
<td>0.28 M NaCl (Fisher Scientific), 0.05 M HEPES (Fisher Scientific), 1.5 mM anhydrous Na₂HPO₄ (Fisher Bioreagents) in ddH₂O. pH to 7.0 using NaOH (Fisher Chemical). Sterilised with 0.22 μM filter (Merck).</td>
</tr>
<tr>
<td>Buffered water</td>
<td>125 μL, 1 M HEPES (Fisher Scientific), in 50 mL ddH₂O. Sterilised with 0.22 μM filter (Merck).</td>
</tr>
<tr>
<td>MACS buffer</td>
<td>5 g/L Bovine serum albumin (Sigma), 2 mM EDTA. Sterilised with 0.22 μM filter (Merck).</td>
</tr>
</tbody>
</table>

Table 2.2: The buffers and media used for molecular cloning and lentiviral production.

2.3.2 – 868 construct and design of mutants

Constructs containing 868 α- and β-chains was ordered from GeneArt, Thermo Fisher. The TCR chains were separated by a 2A self-cleaving sequence (Kim et al., 2011; Liu et al., 2017), and flanked by XbaI and XhoI restriction sites at the start of the α-chain and end of the β-chain respectively, see Figure 2.5. This was in keeping with their intended viral vector, pELNS, which contains a marker gene, Rat CD2 (rCD2), flanked by restriction sites XhoI and SalI, and is separated from the TCR chains by a second 2A sequence. The 2A sequences ensure equal expression of both TCR chains and the marker gene (Kim et al., 2011; Liu et al., 2017). In addition, the 2A sequence between the TCR chains contains a SmaI restriction site in order to allow chains to be taken in and out if required.
868 mutants were designed based on the complex structure and table of contacts published by Cole et al. in 2017 (Cole, Fuller, et al., 2017) PDB 5NME. Amino acid substitutions are discussed in detail in Section 4.2.1. In short, key contact residues were substituted to amino acids of different chemical properties. For example, in the case of TCR22 (BD5S), the negatively charged aspartic acid was exchanged for the polar, uncharged serine. TCR constructs were codon optimised for human expression and ordered from GeneArt, Thermo Fisher.

Figure 2.5: TCR construct design for lentiviral work. TCR α and β chains were flanked with XbaI and Xhol restriction sites to allow cloning into pELNS transfer vector. Chains were separated by 2A self-cleaving peptides to aid equal expression. A Small restriction site sits between the TCR chains to allow editing of the TCR pair should the need arise. Once in the pELNS vector the TCR construct is immediately followed by marker gene rCD2. The two are separated by a 2A cleavage site to aid equal expression.

2.3.3 – Cloning constructs into the pELNS lentiviral vector
As discussed above, the 3rd generation lentiviral vector, pELNS contains the rCD2 marker gene, see Figure 2.6. CD2 is a T cell surface antigen involved in cell adhesion, it interacts with CD48 and CD58 (LFA-3) (Sandrin et al., 1993; van der Merwe et al., 1994; Davis and van der Merwe, 1996; Dustin et al., 1997). In addition to the marker gene, pELNS contains an elongation factor (EF)-1α promoter than controls expression of the inserted gene, a Kozak sequence, a woodchuck hepatitis post-transcriptional regulatory element (WPRE), and an ampicillin resistance gene. The EF-1α promoter allows for constitutive expression in a number of different eukaryotic cell types, and has been shown to induce more efficient transduction and improved long-term transgene expression than the cytomegalovirus CMV promoter (X. Wang et al., 2017). It is thought that human EF-1α is more resistant to silencing
The Kozak sequence is a eukaryotic protein translation initiation site upstream of most protein encoding genes (Kozak, 1984, 1987), its presence in DNA vectors can significantly upregulate expression of the transgene (Ólafsdóttir et al., 2008). The WPRE was isolated in 1998 through shared homology with the hepatitis B virus post-transcriptional regulatory element (Huang and Liang, 1993; Donello, Loeb and Hope, 1998). Both regulatory elements serve to increase expression of associated genes at the post-transcriptional level, by facilitating trafficking of RNA transcripts from the nucleus to the cytoplasm (Huang and Liang, 1993; Donello, Loeb and Hope, 1998). Inclusion of the WPRE sequence in viral vectors has been shown to enhance expression of the transgene (Schwenter, Déglon and Aebischer, 2003; Klein et al., 2006). The ampicillin resistance is principally employed during the amplification of the DNA via maxiprep (Section 2.3.5).

Figure 2.6: pELNS vector for lentiviral work. The TCR construct is cloned into the pELNS vector using XbaI and XhoI restriction sites. It is preceded by an EF1α promoter and a Kozak sequence to boost expression. It is immediately followed by marker gene, rCD2. rCD2 is flanked by XhoI and SalI restriction sites to facilitate exchange of the marker gene should the need arise. The pELNS vector also contains a WPRE to further enhance expression of the transgene and an ampicillin resistance gene to allow easy recognition of successful cell uptake during amplification of the vector in bacterial cells.
The wildtype 868 construct and all mutants were ordered from the GeneArt service by Thermo Fisher. Constructs arrived in either pMK or pMA transport vectors, which confer resistance to kanamycin or ampicillin respectively. All constructs were immediately amplified via mini prep, using XL-10 Gold cells and agar plates supplemented with the relevant antibiotic (Section 2.1.5). The DNA was then digested in order to remove the construct from the transport vector; 1-5 μg of DNA was mixed with 1 μL XbaI (Thermo Scientific), 1 μL XhoI (Thermo Scientific), and 2 μL 10X Fast Digest buffer (Thermo Scientific). The solution was made up to 20 μL with nuclease-free water (Ambion). 5-10 μg of the recipient vector, pELNS, was digested in the same manner. The digestion mixtures were centrifuged briefly and incubated at 37°C overnight. The following morning the digestion mixtures were loaded onto a 1% agarose gel (Sigma), along with the Hyperladder 1 kb (Bioline), and run at 90 V for 45 minutes. This gel was produced by dissolving 0.7 g agarose (Sigma) in 70 mL TAE buffer, when the gel was sufficiently cooled 2.5 μL Midori Green DNA dye (Nippon Genetics Europe) was added and it was poured into the mould. A 1 kb DNA Hyperlader (Bioline) was also included. The gel was visualised using an UV illuminator (Nippon Genetics Europe).

The relevant bands were cut out of the gel using a scalpel, and the gel fragments were digested using a Promega Wizard SV Gel and PCR Clean-up System according to the manufacturer’s instructions. Gel slices were weighed and 10 μL of Membrane Binding Solution was added per 10 mg. This was incubated at 65°C, vortexing every few minutes until the gel had completely dissolved. This solution was then loaded onto a SV Minicolumn/Collection Tube set up and incubated at room temperature for 1 minute to allow DNA to bind to the column. The tube was centrifuged 16,000 x g for 1 minute to remove impurities (Eppendorf Centrifuge 5424). The column was washed twice with Membrane Wash Solution (supplemented with ethanol), and the empty column was centrifuged once more to remove any residual ethanol. The DNA was eluted in 20 μL nuclease-free water (Ambion), and the concentration measured using a spectrophotometer (Geneflow) as described in Section 2.1.5. The molar concentration was calculated using the Promega biomath tool.

Pure digested construct DNA was then ligated into the digested pELNS vector. The digested insert and vector were mixed at a ratio of 150 fmol:30 fmol, with the insert in molecular excess. 2 μL 10x T4 ligase buffer (Thermo Scientific), and 1μL T4 ligase (Thermo Scientific) was added to the ligation mixture and the total volume made up to 20 μL with nuclease-free water (Ambion). The ligation solution was mixed gently and incubated at room temperature
for 2 hours. A negative control in which the insert DNA was replaced with an equal volume of nuclease-free water (Ambion) was also set up.

5 μL ligation mixture was then used to transform XL-10 Gold cells, undigested empty pELNS acted as a positive control and the water-containing ligation mixture as a negative control. XL-10 Gold cells were allowed to defrost on ice for 5 minutes before the DNA was added. They were then incubated for a further 30 minutes on ice before a 30 second heat shock at 42°C to encourage uptake of DNA. In order to amplify cell numbers 500 μL of pre-warmed SOC media (Clontech) was added and tubes were placed in an orbital shaking incubator for 1 hour (37°C, 220 RPM) (Thermo Fisher). This culture was used to inoculate agar plates supplemented with 50 μg/mL carbenicillin, they were grown overnight at 37°C.

The following morning, individual colonies were selected for colony polymerase chain reaction (PCR) to confirm the presence of the desired insert. For lentiviral work, constructs were cloned into the pELNS vector and the pELNS F1 and pELNS R1 primers (Eurofins Genomics) (see Table 2.3) were used for colony PCR. For each colony a PCR mixture was prepared, containing: 12.5 μL DreamTaq MasterMix (Thermo Scientific), 1 μL pELNS F1, 1 μL pELNS R1, 10.5 μL nuclease-free water (Ambion). Sterile P2 tips were used to inoculate each mix. The tubes were briefly centrifuged to remove air bubbles and then placed in a thermal cycler for the following conditions:

- 94°C 10 minutes
- 94°C 20 seconds
- 57°C 20 seconds
- 72°C 1 minute
- 72°C 7 minutes

(\{ x 20 cycles \})

Each PCR mixture was then run on a 1% agarose gel, prepared as described above. The gel was visualised using an UV illuminator (Nippon Genetics Europe). Lanes with bands of the appropriate size (~1836 base pairs) indicate positive ligation reaction, see Figure 2.7.

Plasmids were extracted from positive colonies using a PureLink Quick Plasmid Miniprep kit (Thermo Fisher), as described above (Section 2.1.5). Concentrations were measured using a spectrophotometer (Geneflow). Samples were sequenced using Eurofins Geomics’ TubeSeq service with the relevant primers (see Table 2.3).
Figure 2.7: Representative colony PCR. The transgene is 1836 base pairs in size. A colony PCR uses primers to amplify a region within the transfer vector and provides a very clear demonstration of which colonies have successfully taken up the vector. In this example gel, the construct has successfully been incorporated into all colonies apart from colonies G, K, and M. Well A contains a Hyperladder 1 kb (Bioline).

2.3.4 – Sequencing constructs for validation
Isolated plasmids were sequenced to confirm presence of the inserted construct using Eurofins Genomics. Following mini prep, 15 μL of DNA at 100 ng/μL was sent along with 2 μL pELNS F2 primer (Table 2.3) at 10 pmol/μL using the Mix2Seq service (Eurofins genomics). Sequences were checked using the NCBI Nucleotide Blast tool. Following maxi prep (Section 2.3.5), 15 μL of DNA at 100 ng/μL was sent to Eurofins genomics along with 2 μL each of pELNS F2, Alpha Constant, Beta Fw, and Rat Reverse primers at 10 pmol/μL (Table 2.3) using the TubeSeq service. Overlapping sequence results were checked manually using SnapGene Viewer to confirm the absence of any mutation aside from the intended change.
<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>pELNS F1</td>
<td>GAGTTTGGATCTTGGTTCATTC</td>
<td>Sequencing constructs in pELNS, and PCR.</td>
</tr>
<tr>
<td>pELNS F2</td>
<td>CTTCATTTTCAGGGTTCGTCGTG</td>
<td>Sequencing constructs in pELNS</td>
</tr>
<tr>
<td>pELNS R1</td>
<td>GCATCAAGGAGCGGTATACAC</td>
<td>Sequencing constructs in pELNS, and PCR</td>
</tr>
<tr>
<td>pELNS R2</td>
<td>CCAGGAGGCGTTGATTGTGCGAC</td>
<td>Sequencing constructs in pELNS</td>
</tr>
<tr>
<td>CMV Fw</td>
<td>CGCAATGGGCGAGGCGGTGCGTG</td>
<td>Sequencing pMD2.G and pMDLG</td>
</tr>
<tr>
<td>M13 Fw</td>
<td>GTAAAACGACGACGAGGACAGG</td>
<td>Sequencing pRSV Rev</td>
</tr>
<tr>
<td>pSF Fw</td>
<td>TACCATCCACTCGACACACC</td>
<td>Sequencing constructs in pSF</td>
</tr>
<tr>
<td>pSF Rev</td>
<td>AAGCGCGGTATCCAGCATAGC</td>
<td>Sequencing constructs in pSF</td>
</tr>
<tr>
<td>Early Alpha C</td>
<td>GACTCCAAGTCCAGCGACAA</td>
<td>Sequencing 868TCR</td>
</tr>
<tr>
<td>Alpha Constant</td>
<td>AACCTGCTGATGACACTGAG</td>
<td>Sequencing 868TCR</td>
</tr>
<tr>
<td>Beta Fw</td>
<td>GTTGGGCTCTGCTGCTGCTCT</td>
<td>Sequencing 868TCR</td>
</tr>
<tr>
<td>Rat Rev</td>
<td>AACTTGACCCCGCATATGCGAT</td>
<td>Sequencing 868TCR</td>
</tr>
</tbody>
</table>

Table 2.3: A summary of the primers used throughout this thesis. Primers sourced from Eurofins Genomics.

2.3.5 – Amplifying DNA via maxi preparation

Competent, XL-10 Gold, cells were transformed with the relevant DNA as described in Section 2.1.4. Single colonies were selected from the resulting LB agar plates to inoculate 250 mL of LB media, these were grown overnight in an orbital shaking incubator (37°C, 220 RPM) (Thermo Fisher). The following morning, bacterial cells in these cultures were harvested by centrifuging at 2755 x g (3700 RPM, Eppendorf 5810R centrifuge) for 30 minutes, DNA extraction was then carried out using a PureLink HiPure Plasmid Filter Maxiprep kit (Thermo Fisher) according to the manufacturer’s instructions. 10 mL of Resuspension Buffer (R3) supplemented with RNAse A was used to dissolve the pellet, before an equal volume of warmed Lysis Buffer (L7) was used to lyse cells. Samples were incubated at room temperature for 5 minutes to allow for lysis. 10 mL of Precipitation Buffer (N3) was added to the lysates to precipitate protein and genomic DNA, leaving circular plasmid DNA in solution. Samples were loaded onto pre-equilibrated PureLink HiPure Maxi Columns (Thermo Fisher) and allowed to filter by gravity flow. The column was washed using 50 mL
of Wash Buffer (W8) before 15 mL of Elution Buffer (E4) was used to recover the purified plasmid DNA. 10.5 mL of propan-2-ol (Fisher Chemical) was added, and the eluate was centrifuged at 16,000 x g for 30 minutes (Beckman Coulter Avanti J-E) at 4°C to precipitate the DNA. Finally, the DNA was washed using 5 mL 70% ethanol (Fisher Chemical), before being re-pelleted by centrifuging at 16,000 x g for 5 minutes (Beckman Coulter Avanti J-E). Air-dried pellets were resuspended in 500 μL of nuclease-free water (Ambion). As above, concentrations were measured using a spectrophotometer (Geneflow). Samples were sequenced using Eurofins Geonomics’ TubeSeq service with the relevant primers, Table 2.3.

2.4 – Lentiviral production and cell transduction
2.4.1 – Production of lentiviruses in HEK 293T cells
3rd generation lentiviruses were produced using HEK 293T cells plated in T175 flasks (Grenier Bio-one) at approximately 60-80% confluency. HEK 293T D10 media (Table 2.4) was refreshed 6 hours before transduction. The transfer plasmid used was pELNS (kindly provided by Dr. James Riley, University of Pennsylvania) carrying 868 constructs and marker gene rCD2, separated by self-cleaving 2A peptide sequences (Section 2.3.2). 15 μg of transfer plasmid was co-transduced with 7 μg of envelope plasmid pMD2.G (Addgene plasmid #12259) and 18 μg of both packaging plasmids pMDLg/pRRE and pRSV-Rev (Addgene plasmids #12251 and #12253 respectively). Plasmids were co-transduced into HEK 293T cells using calcium phosphate precipitation. To achieve this DNA was first diluted in a solution of 1.1 mL 0.1 X TE Buffer and 580 μl buffered water (Table 2.2). 188 μL of 2.5 M CaCl₂ was added to the diluted DNA before 1.9 mL 2 X HEPES Buffered Saline (Table 2.2) was added to the solution in a dropwise manner. During this addition the solution was aerated via a vortex (Vortex Genie 2, Scientific Industries) to prevent sudden changes to pH and buffer salt concentrations. This ensured formation of a fine CaCl₂-DNA precipitate and minimise DNA clumping which would hamper cell uptake. Precipitation occurred during a 20-minute incubation at room temperature, before the solution was transferred to HEK 293T cells via Pasteur pipette (Starlab). HEK 293T D10 media was refreshed 16 hours following transduction. Media containing lentiviral product was harvested at 48 hours and 72 hours post-transduction, and filtered through 0.45 μM filters (Sigma) to remove cells and debris. Lentiviral product was stored at 4°C for up to 1 week, during which time transduction efficiency was tested in the Jurkat cell line (Section 2.4.2). Following this confirmation, lentivirus supernatants were concentrated and either used immediately for transduction of primary T cells or frozen at -80°C for long term storage (Section 2.4.3).
2.4.2 – Transduction of immortal cell lines
Antigen presenting cells (such as Jurkat or MOLT-3 cells) were plated in 24 well plates (Grenier) at a density of 100,000 cells/well. 1 mL of lentiviral supernatant (prior to concentration) was added to each test well. Cells were then spinfected by centrifuging at 500 x g for 2 hours (Heraeus Megafuge 1.0R). Lentiviral supernatant was removed the following morning and replaced with R10 media.

Lentiviral efficiency was evaluated by transducing the T cell leukaemia Jurkat cell line (clone E6.1). Jurkat cells were cultured for 1 week following transduction before lentiviral uptake was assessed by surface marker staining and flow cytometry (Section 2.6.2).

2.4.3 – Concentration of lentiviral supernatant
Lentivirus supernatants were concentrated by centrifugation at 140,000 x g (28,000 RPM) for 2 hours (Beckman Coulter Optima™ ultracentrifuge, SW28 rotor). The lentiviral pellet was resuspended in 300 µL 200 international units (IU) T cell media and either used immediately for transduction of CD8+ T cells, or frozen at -80°C for storage.

2.4.4 – Transducing primary CD8 T cells
In preparation for lentiviral transduction, CD8+ T cells were isolated from peripheral blood mononuclear cells (PBMCs) as described in Section 2.5.4. CD8+ T cells were plated in 48 well plates (Grenier) at a density of 500,000 cells/well in 200 IU media. Cells were activated overnight to increase susceptibility to viral uptake (Bilal, Vacaflores and Houtman, 2015). This was achieved by mixing T cells with CD3/CD28 activator beads (Dynabeads, Life Technologies) at a ratio of 3:1 (beads:cells), thus a density of 1.5 million beads/well. The total volume of each well was 1 mL 200 IU media. T cells were incubated at 37°C 5% CO2 overnight (BB 15 CO2 incubator, Thermo Scientific).

The following morning 700 µL of media was removed from each well and 100 µL of concentrated lentivirus added. 0.25 µL of polybrean (Santa Cruz Biotechnology) at a concentration of 10 mg/mL was added to each well to give a working concentration of 5 µg/mL. The total volume in each well was made up to 500 µL with 200 IU media, and cells were incubated overnight at 37°C (BB 15 CO2 incubator, Thermo Scientific).

The following morning 500 µL of T cell transduction media was added to each well. Subsequently T cells were fed three times a week (Monday, Wednesday, and Friday) by
replacing half of the media. T cells were fed with T cell transduction media for at least two weeks following transduction, before moving on to 200 IU media.

Following the initial one-week period CD8⁺ T cells were assessed for lentiviral uptake by rCD2 surface marker staining and flow cytometry (Section 2.6.2). If staining indicated that uptake of the transgenes had been successful, then transduced CD8⁺ T cells were separated from untransduced CD8⁺ T cells by means of rCD2 surface marker purification (Section 2.4.5) and expanded (Section 2.5.5).

2.4.5 – Purification based on surface marker expression

Transduced CD8⁺ T cells were isolated from their untransduced peers by means of surface expression of the marker gene, rCD2, using the miniMACS separation protocol (Miltenyi Biotec) according to the manufacturer’s instructions. T cells were harvested, washed with 15 mL of cold MACS buffer, and pelleted by centrifuging at 400 x g for 5 minutes (Heraeus Megafuge 1.0R). Pelleted cells were resuspended in 50 μL of cold MACS buffer and 2 μL of anti-rCD2-PE (clone OX-34, Biolegend). Cells were incubated on ice for 20 minutes. MACS buffer, the centrifuge, and cells were kept cold at all times to prevent beads from being internalised by cells.

Excess antibody was removed by washing with 15 mL of cold MACS buffer. Following centrifuging at 400 x g for 5 minutes (Heraeus Megafuge 1.0R) cells were resuspended in 80 μL of cold MACS buffer and 20 μL of anti-PE microbeads (Miltenyi Biotec) per 10⁷ cells and incubated at 4°C for 15 minutes.

Excess beads were removed by washing with 15 mL of cold MACS buffer. Following centrifuging at 400 x g for 5 minutes (Heraeus Megafuge 1.0R) cells were resuspended in 500 μL of cold MACS buffer. This suspension was passed through a 30 μM filter (Celltrics, Wolf Labs) to remove aggregated cells before being loaded onto pre-equilibrated miniMACS separation columns (Miltenyi Biotec). The column was washed with 500 μL of cold MACS buffer three times to remove rCD2⁻ cell populations. The column was then removed from the magnetic holder and the rCD2⁺ population eluted in 500 μL of 200 IU media using the plunger supplied with the column. rCD2⁺ T cells were counted and plated at the appropriate density in 24 well plates (Grenier).

This protocol can be used to isolate subpopulations expressing other surface markers by replacing the rCD2-PE antibody with a PE-conjugated antibody targeting the marker of
interest. In contrast, marker-negative populations can be collected by preserving the flow through.

2.5 – Cell culture

2.5.1 – Buffers and media used in cell culture work throughout this thesis

<table>
<thead>
<tr>
<th>Buffer/media</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>D10</td>
<td>Dulbecco’s Modified Eagle Medium (DMEM) – high glucose (Sigma Life Science), 10% heat inactivated foetal bovine serum (FBS) (Gibco Life Technologies), 100 U/ml penicillin (Sigma-Aldrich) 100 µg/mL streptomycin (Sigma-Aldrich), 2 mM L-glutamine (Sigma-Aldrich).</td>
</tr>
<tr>
<td>D-PBS-EDTA</td>
<td>Dulbecco’s PBS (D-PBS) (Sigma-Aldrich) supplemented with 2 mM EDTA.</td>
</tr>
<tr>
<td>R10</td>
<td>RPMI 1640 (Sigma-Aldrich), 10% FBS, 100 U/ml penicillin (Sigma-Aldrich) 100 µg/mL streptomycin (Sigma-Aldrich), 2 mM L-glutamine (Sigma-Aldrich).</td>
</tr>
<tr>
<td>R0</td>
<td>RPMI 1640 (Sigma-Aldrich), 100 U/ml penicillin (Sigma-Aldrich) 100 µg/mL streptomycin (Sigma-Aldrich), 2 mM L-glutamine (Sigma-Aldrich).</td>
</tr>
<tr>
<td>R5</td>
<td>RPMI 1640 (Sigma-Aldrich), 5% FBS, 100 U/ml penicillin (Sigma-Aldrich) 100 µg/mL streptomycin (Sigma-Aldrich), 2 mM L-glutamine (Sigma-Aldrich).</td>
</tr>
<tr>
<td>200IU T cell culture media</td>
<td>R10 supplemented with 25 ng/ml interleukin (IL)-15 (Miltenyi Biotech), 200 IU/mL IL-2 (Proleukin), 1 x non-essential amino acids solution (Sigma-Aldrich), 1 mM sodium pyruvate (Sigma-Aldrich), 10 mM HEPES buffer (Sigma-Aldrich).</td>
</tr>
<tr>
<td>20IU T cell expansion media</td>
<td>R10 supplemented with 25 ng/ml interleukin (IL)-15 (Miltenyi Biotech), 20 IU/mL IL-2 (Proleukin), 1 x non-essential amino acids solution (Sigma-Aldrich), 1 mM sodium pyruvate (Sigma-Aldrich), 10 mM HEPES buffer (Sigma-Aldrich).</td>
</tr>
<tr>
<td>Transduction media</td>
<td>200 IU T cell culture media supplemented with 20% FBS (final concentration).</td>
</tr>
<tr>
<td>Red blood cell (RBC) lysis buffer</td>
<td>155 mM Ammonium chloride (NH₄Cl) (Acros Organics), 10 mM potassium bicarbonate (KHCO₃) (Sigma-Aldrich), 0.1 mM EDTA at pH 8.0, solution made up to 500 mL with ddH₂O, pH to 7.2-7.4 with HCl (Sigma-Aldrich).</td>
</tr>
</tbody>
</table>
2.5.2 – Culturing of adherent cells lines
Adherent cell lines (such as HEK 293T cells and mm909.24 cells) were sourced from the American Type Cell Collection and cultured as recommended in their respective media (Table 2.4) in incubators maintaining 37°C and 5% CO₂ humidity (BB 15 CO₂ incubator, Thermo Scientific). Whilst being maintained in culture, cell was routinely refreshed Mondays and Fridays. In general, adherant cells were split once a confluency of 80% was exceeded using a non-enzymatic detachment method. Old cell culture media was removed and the cells were washed in 2 mM EDTA in Dulbecco’s PBS (Sigma-Aldrich) to remove any remaining media and detached cells. Fresh EDTA-D-PBS was added to the flask in sufficient volume to coat the surface. Cells were incubated at 37°C until all cells had detached, approximately 10 minutes. Following this, the EDTA in the D-PBS was neutralised by adding an equal volume of cell culture media to the flask. Detached cells were then pelleted by centrifuging at 400 x g for 5 minutes (Heraeus Megafuge 1.0R) in a falcon tube. The pellet was resuspended in fresh cell culture media, and seeded back at a density suitable for the immediate experimental requirements. If not needed in the near future, adherent cells were generally split 1:20.

2.5.3 – Culturing of suspension cell lines
The suspension cell lines used throughout this thesis are summarised in Table 2.5. Suspension cells were grown in R10 media (Table 2.4) in incubators maintaining 37°C and 5% CO₂ humidity (BB 15 CO₂ incubator, Thermo Scientific). Cells were sourced from the American

<table>
<thead>
<tr>
<th>Buffer Type</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freezing buffer</td>
<td>90 % FBS (Gibco Life Technologies), 10 % dimethyl-sulfoxide (DMSO) (Sigma-Aldrich).</td>
</tr>
<tr>
<td>MACS buffer</td>
<td>0.5 % Bovine serum albumin (Sigma-Aldrich), 2 mM EDTA, in D-PBS (Sigma-Aldrich).</td>
</tr>
<tr>
<td>Fixing buffer</td>
<td>4 % paraformaldehyde (Fisher Chemicals) in PBS.</td>
</tr>
<tr>
<td>Fluorescence-activated cell sorting (FACS) buffer</td>
<td>2 % FBS in PBS.</td>
</tr>
<tr>
<td>Wash buffer</td>
<td>0.05 % Tween-20 (Merck) in PBS.</td>
</tr>
<tr>
<td>Reagent diluent</td>
<td>1 % Bovine serum albumin (Sigma-Aldrich) in PBS.</td>
</tr>
</tbody>
</table>

Table 2.4: Buffers and media used for cell culture throughout this thesis.
Type Cell Collection and cultured as recommended. Whilst being maintained in culture, suspension cells were split Mondays and Fridays routinely or when media colour indicated a change in pH due to a build-up of waste products. Cells were split by resuspending fully and then removing half of the media to be replaced by fresh media. If cell lines were being expanded for experimental purposes fresh media was added incrementally each day, and cells were split over multiple flasks to maintain appropriate culture volumes.

<table>
<thead>
<tr>
<th>Cell line name</th>
<th>Media</th>
<th>Use</th>
<th>Cell type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jurkat</td>
<td>R10</td>
<td>Testing lentiviral products.</td>
<td>Immortalised T cells, suspension.</td>
</tr>
<tr>
<td>T2</td>
<td>R10</td>
<td>Antigen presenting target cells.</td>
<td>Lymphoblastic cell line, suspension.</td>
</tr>
<tr>
<td>MOLT-3</td>
<td>R10</td>
<td>Transduction of lentiviral constructs.</td>
<td>Immortalised T cells, suspension.</td>
</tr>
<tr>
<td>HEK 293T</td>
<td>D10</td>
<td>Production of lentiviral particles.</td>
<td>Human embryonic kidney cells, adherent.</td>
</tr>
<tr>
<td>C1R</td>
<td>R10</td>
<td>Transduction of lentiviral constructs.</td>
<td>B lymphoblastic cell line, suspension.</td>
</tr>
<tr>
<td>MOLT-3</td>
<td>R10</td>
<td>Transduction of lentiviral constructs.</td>
<td>T lymphoblastic cell line, suspension.</td>
</tr>
<tr>
<td>K562</td>
<td>R10</td>
<td>Transduction of lentiviral constructs.</td>
<td>Myelogenous leukemia line, suspension.</td>
</tr>
<tr>
<td>MM909.24</td>
<td>R10</td>
<td>Transduction of lentiviral constructs.</td>
<td>Melanoma cell line, adherent.</td>
</tr>
</tbody>
</table>

Table 2.5: A summary of cell lines used in this thesis.

2.5.4 – Isolation of CD8+ primary T cells from whole blood
EDTA-treated buffy coats from three healthy donors were acquired regularly from the Welsh Blood Service (WBS) in accordance with the Human Tissue Act and local ethical approval. PBMCs were isolated using a Histopaque™ (Stemcell Technologies) density gradient approach. Blood was delivered by the WBS in the afternoon, diluted 1:2 with R10 and rolled overnight at room temperature in 50 mL falcons. The following morning 13 mL Histopaque™
was added to the bottom compartment of 50 mL SepMate Tubes (Stemcell Technologies). 25 mL of diluted blood was added slowly to the SepMate Tube to minimise mixing, and tubes were topped up to 50 mL with R0 to further dilute blood (approximately 10-12 mL). SepMate tubes were centrifuged at 1200 x g for 10 minutes (Heraeus Megafuge 1.0R). The top layer of primarily lymphocytes and some RBCs was transferred to a 50 mL falcon and diluted further with R0. Cells were pelleted by centrifuging at 800 x g for 10 minutes (Heraeus Megafuge 1.0R). All pellets from the same donor were combined and dissolved in the same 25 mL of RBC lysis buffer, see Table 2.4. Tubes were then incubated at 37°C for 10 minutes in a water bath. Following this, the tubes were topped up to 50 mL with R0 and centrifuged 400 x g for 5 minutes (Heraeus Megafuge 1.0R). This step was repeated with a fresh 25 mL of RBC lysis buffer if, following centrifugation, the pelleted cells were still visibly red. RBC-clear pellets were resuspended in 25 mL of R10 for counting.

If primary CD8+ T cells were not required, isolated PBMCs were non-fatally irradiated to be used for T cell expansions (Section 2.5.6). If primary CD8+ T cells were required for lentiviral transduction, they were isolated at this point using the miniMACS separation protocol (Miltenyi Biotec) according to the manufacturer’s instructions. For this protocol MACS buffer, the centrifuge, and cells where kept cold at all times to prevent CD8 microbeads (Miltenyi Biotec) from being internalised by cells.

Lymphocytes were washed with 15 mL of cold MACS buffer, and pelleted by centrifuging at 400 x g for 5 minutes (Heraeus Megafuge 1.0R). Pelleted cells were resuspended in 80 μL of cold MACS buffer and 20 μL of CD8 microbeads (Miltenyi Biotec) per 10^7 cells and incubated at 4°C for 15 minutes, the number of expected CD8+ cells was estimated based on the frequencies of CD8+ cells recovered by previous attempts.

Excess beads were removed by washing with 15 mL of cold MACS buffer. Following centrifuging at 400 x g for 5 minutes (Heraeus Megafuge 1.0R) cells were resuspended in 500 μL of cold MACS buffer. This suspension was passed through a 30 μM filter (Celltrics, Wolf Labs) to remove aggregated cells before being loaded onto pre-equilibrated miniMACS separation columns (Miltenyi Biotech). The column was washed with 500 μL of cold MACS buffer three times to remove CD8- cell populations. The column was then removed from the magnetic holder and the CD8+ population eluted in 500 μL of 200 IU media using the plunger supplied with the column. CD8+ cells were counted and plated at the appropriate density in 24 well plates (Greiner). At this point cells were ready for the lentiviral transduction procedure described in Section 2.4.4.
2.5.5 – T cell purification based on multimer-binding specificity

CD8+ T cells that are able to bind to certain pMHC molecules can be isolated in a similar manner to that described in Section 2.4.5 using streptavidin-conjugated tetramers made as described in Section 2.6.3. They are isolated using the miniMACS separation protocol (Miltenyi Biotec) according to the manufacturer’s instructions. T cells were harvested, washed with 15 mL of cold MACS buffer, and pelleted by centrifuging at 400 x g for 5 minutes (Heraeus Megafuge 1.0R). Pelleted cells were resuspended in 50 μL of cold MACS buffer and 50 μL of 100 nM Protein Kinase Inhibitor (Dasatinib, Axon Medchem), cells were incubated at 37°C for 10 minutes. Subsequently 5 μg of tetramer was added and cells were incubated on ice for 30 minutes. MACS buffer, the centrifuge, and cells where kept cold at all times to prevent beads from being internalised by cells.

Excess tetramer was removed by washing with 15 mL of cold MACS buffer. Following centrifuging at 400 x g for 5 minutes (Heraeus Megafuge 1.0R) cells were resuspended in 80 μL of cold MACS buffer and 20 μL of anti-PE microbeads (Miltenyi Biotec) per 10⁷ cells and incubated at 4°C for 15 minutes.

Excess beads were removed by washing with 15 mL of cold MACS buffer. Following centrifuging at 400 x g for 5 minutes (Heraeus Megafuge 1.0R) cells were resuspended in 500 μL of cold MACS buffer. This suspension was passed through a 30 μM filter (Celltrics, Wolf Labs) to remove aggregated cells before being loaded onto pre-equilibrated miniMACS separation columns (Miltenyi Biotech). The column was washed with 500 μL of cold MACS buffer three times to remove tetramer− cell populations. The column was then removed from the magnetic holder and the tetramer+ population eluted in 500 μL of 200 IU media using the plunger supplied with the column. Tetramer+ cells were counted and plated at the appropriate density in 24 well plates (Grenier).

Alternatively, towards the end of my studies the laboratory acquired a SONY MA900 Multi-Application cell sorter which can be used to sort tetramer+ populations. T cells were stained with 5 μg of PE-tetramer, as described above, prior to analysis. Tetramer+ populations were cultured in 200 IU T cell media supplemented with Amphotericin B/Fungizone (Merck) following sorting. All FACS-based cell-sorting performed by Dr Cristina Ruis Rafael.

2.5.6 – Expansion of CD8+ primary T cells and subsequent culturing

CD8+ T cells were routinely expanded in per T25 flasks (Greiner) at 37°C, 5% CO₂. Up to 1 x 10⁶ CD8+ T cells were co-cultured with 15 x 10⁶ cells from an allogenic, non-lethally irradiated
mix of PBMCs from three different WBS donors (known as a feeder mix). Feeder mix cells were isolated as described above (Section 2.5.4) and irradiated at 3000-3100 cGy. Each flask was made up to a total volume of 15 mL with 20 IU media and 1 µg/mL phytohemagglutinin (PHA, PAN Biotech), see Table 2.4.

Expansions were incubated at 37°C, 5 % CO₂, tilted at an angle to allow cells to congregate and benefit from cell-cell proximity, for five days. On the 5th day, half of the media (7.5 mL) was exchanged for a fresh 7.5 mL of 20 IU. Seven days into the expansion cells were harvested, counted, and plated at the appropriate density (3-4 million per well) in 24 well plates (Greiner), in 2 mL of 200 IU media per well see Table 2.4.

T cells were subsequently cultured by changing half the media three times a week; Monday, Wednesday, and Friday. Following day 14 of the expansion cells were considered fit for downstream functionality assays. If necessary, cells were frozen within four weeks of expansion (Section 2.5.7).

2.5.7 – Cell counting
For counting, cells were harvested, washed, and resuspended in R10 – the volume of which was dependent on the estimated number of cells. If cells were too densely packed to obtain accurate counts they were diluted further. 10 µL of resuspended cells was then mixed with 10 µL of 0.4 % trypan blue solution (Sigma Aldrich) (dilution factor, 2). Cells were added to a haemocytometer and counted based on the assumption that live cells would not take up trypan blue and so would appear bright, whereas dead cells would appear blue under the microscope. The following calculation was used:

\[
\text{Average cells per grid} \times \text{dilution factor} \times 10^4 = \text{cells/mL}
\]

2.5.8 – Freezing cells for storage in liquid nitrogen
For freezing, cells were counted (Section 2.5.6) and centrifuged at 400 x g for 5 minutes to remove cell culture media (Heraeus MegaFuge 1.0R). Cells were then resuspended in freezing buffer (Table 2.4) and placed, in 1 mL aliquots, in internal thread cryovials (Nunc). Cells were initially frozen at -80°C in isopropanol-filled controlled-rate freezing pots (Mr Frosty, Nalgene). Isopropanol was refreshed every five uses, as per the manufacturer’s recommendation. After at least 24 hours at -80°C frozen cells were transferred to liquid nitrogen for long-term storage.
For thawing, cells were removed from liquid nitrogen and rapidly defrosted by incubation in a 37°C water bath. As soon as they had thawed, cells were transferred into 10 mL of warmed R10 to dilute freezing buffer. Cells were then centrifuged at 400 x g for 5 minutes to remove freezing buffer (Heraeus Megafuge 1.0R). Cells were resuspended in their required media, counted, and plated or placed in flasks according to their preference.
2.6 – Cell-based functional assays

2.6.1 – Antibodies used for flow cytometry throughout this thesis

<table>
<thead>
<tr>
<th>Antibody specificity</th>
<th>Conjugated fluorochrome</th>
<th>Clone</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>rCD2</td>
<td>FITC, PE</td>
<td>OX-34</td>
<td>Biolegend</td>
</tr>
<tr>
<td>CD8</td>
<td>FITC, APC, PE</td>
<td>REA734</td>
<td>Miltenyi Biotec</td>
</tr>
<tr>
<td>CD8</td>
<td>PE, PE-Vio770, APC-Vio770</td>
<td>BW135/80</td>
<td>Miltenyi Biotec</td>
</tr>
<tr>
<td>CD8</td>
<td>APC-H7, Per-CP</td>
<td>SK1</td>
<td>BD Biosciences, Biolegend</td>
</tr>
<tr>
<td>CD4</td>
<td>PE, APC-Vio770</td>
<td>VIT4</td>
<td>Miltenyi Biotec</td>
</tr>
<tr>
<td>CD4</td>
<td>FITC, APC</td>
<td>REA623</td>
<td>Miltenyi Biotec</td>
</tr>
<tr>
<td>CD4</td>
<td>PerCP, PE</td>
<td>M-T466</td>
<td>Miltenyi Biotec</td>
</tr>
<tr>
<td>CD3</td>
<td>PerCP, PE</td>
<td>BW264/56</td>
<td>Miltenyi Biotec</td>
</tr>
<tr>
<td>CD19</td>
<td>Pacific Blue</td>
<td>HIB19</td>
<td>Biolegend</td>
</tr>
<tr>
<td>HLA-A2</td>
<td>APC, FITC, PE</td>
<td>BB7.2</td>
<td>Miltenyi Biotec</td>
</tr>
<tr>
<td>TNF</td>
<td>PE-Vio770, PE-Cy7, PerCP/Cyanine5.5, APC</td>
<td>cA2, mAb11</td>
<td>Miltenyi Biotec, BD Biosciences</td>
</tr>
<tr>
<td>CD107a</td>
<td>PE, FITC</td>
<td>H4A3</td>
<td>BD Biosciences</td>
</tr>
<tr>
<td>Anti-PE</td>
<td>Unconjugated</td>
<td>PE001</td>
<td>Biolegend</td>
</tr>
<tr>
<td>TCR Vbeta5a</td>
<td>Unconjugated</td>
<td>1C1</td>
<td>Thermo Fisher</td>
</tr>
</tbody>
</table>

Anti-Mouse Ig  | PE          | Polyclonal | BD Biosciences |

**Table 2.6**: The antibodies used for flow cytometry throughout this thesis.

All antibodies were raised in mice against human proteins except for rCD2 (mouse anti-rat), and anti-mouse Ig (Goat anti-mouse). The latter was used in conjugation with the anti-human TCR Vbeta5a. PE = phycoerythrin; PerCP = peridinin chlorophyll protein; FITC = fluorescein isothiocyanate; APC = allophycocyanin.
2.6.2 – Flow cytometry: staining for surface markers

Approximately 50,000 cells were used per staining condition. Cells were harvested, transferred to FACS tubes (Elkay Labs), and washed twice with 3 mL of autoclave-sterilised PBS (made as described in Table 2.1) to remove any remaining components of cell media. Tubes were centrifuged at 800 x g for 2 minutes (Heraeus Megafuge 1.0R) to pellet cells before PBS was removed and cells were stained with an amine-reactive dye, LIVE/DEAD™ Fixable Violet Dead Stain (Thermo Fisher), referred to henceforth as vivid. Tubes were vortexed briefly to mix and incubated at room temperature in the dark for 5 minutes. Primary fluorochrome-conjugated antibodies were then added to stain the required surface antigen (Table 2.6). Each tube was vortexed briefly to mix and incubated with the surface antibody for 20 minutes on ice in the dark. Excess antibody and vivid was removed with 2x 3 mL PBS washes. If not immediately analysed, cells were fixed by adding 50 µL of paraformaldehyde (PFA) to each tube. Tubes were incubated for 20 minutes on ice in the dark before excess PFA was removed with 2x 3 mL PBS washes.

Compensation was achieved using anti-mouse Ig Compensation Beads (BD Biosciences). Beads were vortexed, diluted in PBS, and aliquoted 25 µL per tube. Using a single fluorochrome per tube, 2 µL of relevant antibody was added to the beads. Staining was analysed using a FACS Canto II machine (BD Biosciences) and data subsequently analysed using FlowJo (Tree Star Inc, Ashland OR).

A typical gating strategy is shown in Figure 2.8. In general, forward and side scatter was used to isolate lymphocytes. Doublets and debris were removed by gating on single cells. Dead cells were then removed based on staining with vivid, before further analysis of surface marker expression. The placing of gates was aided by biologically relevant negative controls (also known as fluorescence minus one; FMO) (for example untransduced CD8⁺ T cells when assessing viral uptake).
Figure 2.8: Typical gating strategy for analysis of T cells. First lymphocytes are isolated using their forward and side scatter properties. Then the area to height ratio is used to gate on single cells, excluding doublets and debris. Finally, vivid uptake is used as a marker to exclude dead cells, and CD3 is used to indicate T cells.

2.6.3 – Flow cytometry: pMHC multimer staining
Biotinylated pMHC monomers were refolded as described in Section 2.1. Tetramerisation was achieved by incubating pMHC monomers with PE-conjugated streptavidin at a ratio of 1:4 was used with streptavidin in molar excess (Invitrogen, Thermo Fisher) on ice. Streptavidin was added to pMHC monomer incrementally, five additions each separated by 20 minute incubations to encourage fulfilment of all streptavidin valences. Dextramers were assembled by incubation of pMHC monomers with dextran-streptavidin-PE, at a molar ratio...
of 3:1, for 30 minutes in the dark at room temperature. Tetramers were made to a final concentration of 0.1 µg/µL using PBS.

Tetramer staining was performed as previously described (Wooldridge et al., 2009; Dolton et al., 2015; Tungatt et al., 2015). Approximately 50,000 cells were used per staining condition. Cells were harvested, transferred to FACS tubes (Elkay Labs), and washed twice with 3 mL of FACS buffer (Table 2.4) to remove any remaining components of cell media. Tubes were centrifuged at 800 x g for 2 minutes (Heraeus Megafuge 1.0R) to pellet cells before FACS buffer was removed. 50 µL of 100 nM Protein kinase inhibitor (Dasatinib, Axon Medchem) was added to each tube and tubes were incubated at 37°C for at least 10 minutes. Protein kinase inhibitors enhance tetramer staining by reducing TCR downregulation and internalisation of bound pMHC tetramers (Lissina et al., 2009).

0.5 µg of tetramer (with respect to the pMHC component) was then added to each tube, tubes were vortexed briefly and incubated on ice for 30 minutes. Cells were washed twice in PBS to remove excess tetramer before a fluorochrome-unconjugated anti-PE antibody was added. This addition serves to stabilise the TCR-pMHC interaction at the cell surface, thus enhancing staining (Tungatt et al., 2015). Cells were incubated with the unconjugated antibody on ice for 20 minutes before washing in PBS to remove excess.

Finally, cells were stained with vivid and surface marker antibodies as described in Section 2.6.2. If not immediately analysed, cells were fixed by adding 50 µL of PFA to each tube. Tubes were incubated for 20 minutes on ice in the dark before excess PFA was removed with 2x 3 mL PBS washes. Staining was analysed using a FACS Canto II machine (BD Biosciences), see Figure 2.9, and data subsequently presented using FlowJo (Tree Star Inc, Ashland OR). Negative controls included staining with an irrelevant tetramer, and the use of no tetramer (only surface marker antibodies), these controls were used to aid the gating strategy.
Figure 2.9: Schematic explanation of tetramer staining. The multivalent nature of streptavidin is used to associate four pMHC molecules with a fluorochrome, in this thesis the fluorochrome used is invariably PE but other conjugated streptavidins are available. TCR-pMHC interaction then tags the fluorochrome to the surface of the T cell. This schematic represents the basis of tetramer staining and does not include ‘tricks’ to improve staining, these are described by Tungatt et al. 2015. For simplicity, pMHC tetramers are shown making a single binding event.
2.6.4 – Enzyme-linked immunosorbent assays (ELISA) of T cell function
Enzyme-linked immunosorbent assays (ELISA) were used to determine T cell responsivity to
the various stimuli such as peptides or cell lines transduced with potential target proteins via
lentivirus. This assay uses DuoSet human ELISA kits (R + D Systems) to measure production
of MIP-1β or TNF in response to stimuli as a proxy for responsiveness.

Assays were carried out as per the manufacturer’s instructions. T cells were harvested,
washed in R0, and rested in R5 overnight before the assay. The following morning, T cells
were plated in a 96 well plate (Greiner) at a density of 30,000 cells/well. They were co-
incubated with 60,000 antigen presenting cells (generally HLA-A2+ T2 cells) and 10 μL of
peptide dilution or 60,000 target cells. Plates were kept at 37°C and 5 % CO₂ humidity (BB 15
CO₂ incubator, Thermo Scientific) overnight.

Meanwhile, half-area flat-bottomed 96 well plates (Greiner) were coated in 50 μL of anti-
human MIP-1β or TNF antibodies (R + D Systems) at a concentration of 1.0 μg/mL diluted in
PBS. These plates were sealed and incubated overnight at room temperature.

The following morning, these assay plates were washed three times with 190 μL wash buffer
(Table 2.4) using a plate washer (Wellwash Versa, Thermo Fisher). To prevent non-specific
binding 150 μL reagent diluent (Table 2.4) was used to block assay plates, incubating at room
temperature for at least 1 hour.

The co-incubation plates containing the T cells and targets were centrifuged 400 x g for 5
minutes (Heraeus Megafuge 1.0R) to pellet cells to prevent interference with the assay. 50
μL of supernatant was then taken from each well into a fresh 96 well plate (Greiner) and
diluted by adding 70 μL R5 (Table 2.4). Following the aforementioned 1-hour blocking period,
assay plates were washed three times with 190 μL wash buffer using a plate washer
(Wellwash Versa, Thermo Fisher). 50 μL of diluted supernatant was added to assay plates
and allowed to bind to the immobilised antibody by incubation at room temperature for 1
hour and 15 minutes. Recombinant human standards for the assayed cytokine (either MIP-
1β or TNF) were also added to the assay plate. A serial dilution of standards (starting at 2000
pg/mL) was created by diluting in reagent diluent. This serial dilution would facilitate
generation of a standard curve from which the concentration of MIP-1β or TNF in
experimental wells could be calculated. An example standard curve is shown in Figure 2.10.
Remaining supernatant was frozen at -20°C for storage and repeats if necessary.

Unbound supernatant was removed by washing three times with 190 μL wash buffer using a
plate washer (Wellwash Versa, Thermo Fisher). 50 μL of the appropriate biotinylated
detection antibody (R + D Systems) diluted to 50 ng/mL was then added to the plate. The plate was incubated at room temperature for 1 hour and 15 minutes. Excess antibody removed by washing three times with 190 µL wash buffer using a plate washer (Wellwash Versa, Thermo Fisher).

Next, the plate was incubated with 50 µL of streptavidin-conjugated horseradish peroxidase (R + D Systems) per well for 20 minutes in the dark. Following this incubation, the plate was washed three times with 180 µL wash buffer using a plate washer (Wellwash Versa, Thermo Fisher). Colour reagents A + B (R + D Systems) which were then mixed at a ratio of 1:1, 50 µL of this mixture was immediately added to the plate in the dark to minimise bleaching. Typically, plates took 5-10 minutes to develop. The reaction was halted by the addition of 25 µL 1 M sulphuric acid (Stop Solution, R + D Systems). The OD₄₅₀nm of each well was then immediately recorded using a Bio-Rad® iMark microplate reader and a correction set to 570nm.

![Figure 2.10: Standard curve for a MIP-1β ELISA.](image)

**Figure 2.10: Standard curve for a MIP-1β ELISA.** The OD of known concentrations of MIP-1β is measured and used to generate a standard curve. The linear regression of the resulting data (y=m*x) is then used to indicate MIP-1β concentrations from sample wells.
2.6.5 – Combinatorial peptide library (CPL) screens

CPL screens were performed using an ELISA format (Section 2.6.4) 30,000 rested T cells were co-incubated with 60,000 T2 antigen presenting cells and 10 µL of a peptide mix at a final concentration of 100 µg/mL from a 9-mer combinatorial peptide library (CPL) screen manufactured by Pepscan Presto Ltd. Peptide and T2 cells had previously been pulsed by incubation at 37°C and 5% CO₂ humidity (BB 15 CO₂ incubator, Thermo Scientific) for 2 hours. MIP-1β production in response to each peptide mix was then measured using ELISA as a proxy for T cell responsivity as described in Section 2.6.4.

A 9-mer CPL consists of individual peptide mixes in which, for each peptide position, a single L-amino acid is fixed. All other positions vary from any of the remaining 19 natural L-amino acids (cysteine is excluded from the non-fixed pool due to its propensity to form disulphide bonds causing aggregates within the peptide mix) (Wooldridge et al., 2012; Ekeruche-Makinde et al., 2013; Szomolay et al., 2016). In a 9-mer CPL there are 180 peptide mixes (9 positions x 20 possible proteogenic L-amino acids) (Ekeruche-Makinde et al., 2013). Within each mix there are 1.7 x 10^10 peptide (19^9). The total number of peptides in the full assay is 4.8 x 10^{11} ((9+19) x 19^9) (Ekeruche-Makinde et al., 2013).

2.6.6 – TNF processing inhibitor-0 (TAPI-0) assays

A TAPI-0 assay assesses T cell functionality based on their ability to produce TNF in response to antigen (Haney et al., 2011). The method was first described in 2011 and is thought to be particularly useful for low sensitivity TCRs, or those in which background noise is high – such as polyclonal T cell populations with transduced TCRs (Haney et al., 2011). An additional benefit in the context of this thesis, is that analysis is achieved by flow cytometry allowing gating on the rCD2⁺ population. The assay is based on the knowledge that TNF is originally produced in a membrane-bound form (Björnberg et al., 1994; Haney et al., 2011). It is subsequently cleaved by TNF converting enzyme (TACE) and released as a soluble protein. Inhibiting TACE traps TNF at the cell surface where it can be stained (Björnberg et al., 1994; Haney et al., 2011). Presence of CD107a at the cell surface can also be used as a marker of T cell activation (Betts et al., 2003; Rubio et al., 2003; Haney et al., 2011).

An excess of the required number of CD8⁺ T cells were rested overnight in RS media. The following day target cells were plated at a density of 60,000 cells/well in a 96 U-well plate (Greiner) in a total volume of 50 µL. Rested CD8⁺ T cells were then harvested and added to the plate at a density of 30,000 cells/well in a total volume of either 40 µL (or 39 µL if 1 µL of
peptide was to be included in the experiment). Negative controls lacking target cells or peptide (depending on the nature of the assay) were also included, along with a positive control in which T cells were activated by phytohaemagglutinin (PHA) (PAN Biotech).

A TAPI-0 mix consisting of 0.5 µL of TAPI-0 (Chem Cruz) (giving a working concentration of 30 µM), 1.5 µL anti-TNF-PE-Vio770 (clone cA2), 1.5 µL anti-CD107a-FITC (clone H4A3), and 6.5 µL R5 per well was made. 10 µL of this mixture was then added to each well.

Cells were incubated in the dark at 37°C and 5% CO₂ humidity (BB 15 CO₂ incubator, Thermo Scientific) for a minimum of 4 hours to allow T cells to respond to antigen. Following this cells were washed in PBS, and stained with vivid and surface marker antibodies as described in Section 2.6.2. Staining was analysed using a FACS Canto II machine (BD Biosciences) and data subsequently presented using FlowJo (Tree Star Inc, Ashland OR).

2.6.7 – Chromium-51 (⁵¹Cr)-release cytotoxicity assay
Assay performed by Dr Garry Dolton. Cytotoxicity of CD8⁺ T cells can be measured by ⁵¹Cr-release assay. In this assay target cells are harvested, washed, and labelled with 30 µCi of ⁵¹Cr (sodium chromate, Perkin Elmer) per 1 x 10⁶ cells. Labelling occurs over 1 hour at 37°C. Cells are subsequently washed with D-PBS to remove excess ⁵¹Cr and resuspended in R10. These cells are then incubated at 37°C to allow for ⁵¹Cr leaching from cells. Cells are then washed, resuspended in R10, and plated at 2000 cells/well (usually in duplicate) in 96 U-well plates. Target cells were also incubated in 5% Triton X-100 to determine maximum ⁵¹Cr release, and in R10 to determine spontaneous ⁵¹Cr release. Effector CD8⁺ T cells, suspended in R10, were added to wells to achieve the desired effector:target ratio in a total volume of 150 µL. Co-incubation over 4 hours occurred at 37°C 5% CO₂. Radioactivity of harvested supernatants was measured using a 1450-Microbeta™ counter (Perkin Elmer). Indirect T cell cytotoxicity against target cells (% specific lysis) was calculated thus:

\[
\% \text{Specific lysis} = \frac{\text{Experimental } ⁵¹\text{Cr release} - \text{Spontaneous } ⁵¹\text{Cr release}}{\text{Maximum } ⁵¹\text{Cr release} - \text{Spontaneous } ⁵¹\text{Cr release}} \times 100
\]
3 – Insulin-specific T cells isolated from patients with type 1 diabetes strongly recognise common viral epitopes

3.1 – Introduction
I originally applied to undertake a PhD in the Sewell laboratory with Dr David Cole as I was very interested in the presentation of peptide antigens to αβ TCRs by MHC molecules. When I started my project, the laboratory had recently partially characterised an insulin B chain-specific CD8⁺ T cell clone that was isolated from the blood of a patient with type 1 diabetes (T1D).

3.1.1 – Type 1 diabetes
T1D, also known as insulin-dependent diabetes mellitus, commonly – but not exclusively – manifests in childhood (Katsarou et al., 2017). The disease is a chronic metabolic disorder caused by defects in insulin secretion (Karamanou, 2016). The absence of insulin causes overproduction of glucose in the liver via glycogenolysis and gluconeogenesis, and decreased uptake of glucose by peripheral tissue, manifesting symptomatically as hyperglycaemia, polyuria, excessive thirst, and weight loss (Gepts, 1965; Eisenbarth, 1986; Atkinson, Eisenbarth and Michels, 2014). Metabolic ketosis caused by excessive fat breakdown occurs due to the need for alternative fuel source (Eiselein, Schwartz and Rutledge, 2004). Untreated, symptoms lead to coma and death (Eiselein, Schwartz and Rutledge, 2004). T1D is clinically managed by lifelong insulin administration to regulate blood glucose levels (Bluestone, Herold and Eisenbarth, 2010). Despite this effective treatment, complications arise over the lifetime of the patient as a result of vascular changes caused by long term exposure to elevated blood glucose levels (Eiselein, Schwartz and Rutledge, 2004). These include an increased risk of both cardiac and vasculature morbidities, renal disease, visual impairment and neuropathy. Life expectancy of T1D patients is reduced by 25% compared to healthy contemporaries (Eiselein, Schwartz and Rutledge, 2004), and the disease is expected to become the 7th leading cause of death by 2030 (Mathers and Loncar, 2006; Karamanou, 2016).

Descriptions of patients suffering from excessive thirst and polyuria date back to ancient times when an imbalance of bodily humours was believed to be the cause of illness. The sugar content of diabetic urine was first noted in the 5th century BC (Nduati et al., 2011). In 1889, Minkowski and von Mering removed the pancreases of four dogs and observed that
they subsequently developed polyuria, with urine containing a high sugar content (Mering and Minkowski, 1890). Reimplanting pancreatic tissue subcutaneously prevented hyperglycemia. This landmark study demonstrated that the pancreas was responsible for glucose homeostasis, and at fault in T1D (Mering and Minkowski, 1890). This role was subsequently refined to the islets of Langerhans within the pancreas (Obstet, 1920).

Building on the observations of Minkowski and von Mering, Banting, Best, and colleagues isolated insulin from dog pancreases in 1922 (Banting et al., 1922). They found that, when suitably extracted, ‘pancreatic extract’ had the ability to reduce blood glucose and the glucose present in urine when administered to diabetic dogs. Banting et al. further demonstrated that the pancreatic extract of another species – fetal calves – would achieve the same effect (paving the way for the future treatment of human diabetics with pig insulin), and that extracts of other organs such as liver and spleen did not (Banting et al., 1922). They named their pancreatic extract insulin and used it to treat a 14 year old human boy, who had previously been adhering to the best available treatment of the day – ‘eat as little as possible’ (Karamanou et al., 2014). Daily injections produced immediate improvement that regressed if injections ceased (Banting et al., 1922). This was a life-changing advancement for diabetics. However, lifelong insulin-dependency, and the resulting economic cost (Björk, 2001; Ettaro et al., 2004), means that the quest for a cure to T1D could not stop at this ‘miracle’ treatment.

3.1.1.1 – T1D is a CD8+ T cell mediated disease

Further investigations into the underlying pathology of T1D led to an appreciation for the strong associations of the disease with particular MHC alleles. In 1987, Todd et al. described how a polymorphism at position 57 of the HLA-DQβ chain could influence susceptibility to T1D (Todd, Bell and McDevitt, 1987). DQβ alleles with alanine, serine, or valine at this position were associated with T1D, whereas alleles with aspartic acid at this position were not (Todd, Bell and McDevitt, 1987). A genome wide association search involving over 2000 diabetic patients identified certain HLA-I genes associated with T1D susceptibility (Nejentsev et al., 2007). HLA-B*39, HLA-B*18, and HLA-A*24 were associated with T1D susceptibility whilst HLA-B*27, HLA-A*01, HLA-A*11, and HLA-A*31 were found to be protective (Nejentsev et al., 2007). The susceptibility alleles confer a relative risk comparable to, or greater than, the more classically associated T1D genes, such as INS and PTPN22, with some alleles being associated with younger disease onset (Nejentsev et al., 2007). HLA-A2, one of the most abundant Caucasian alleles, is also notably associated with T1D (Noble et al., 2002; Howson et al., 2009; Noble and Valdes, 2011). In 2002, Marron et al. demonstrated marked
acceleration of disease development in non-obese diabetic (NOD) mice following transgenic introduction of the HLA-A*02:01 gene (Marron et al., 2002). Subsequently, in 2005, DiLorenzo et al. showed that abrogation of MHC class I expression in animal models could confer disease protection (DiLorenzo and Serreze, 2005). These studies indicate that MHC class I-mediated events contribute to the aetiology of T1D.

MHC-association is a hallmark of many autoimmune disorders, (Huey et al., 1993; Karell et al., 2003; Yeo et al., 2007; Chemin et al., 2016) so observations of MHC-association in T1D provided strong evidence of an autoimmune component to the disease. Further proof came with the demonstration of inflammation and leukocyte infiltration within pancreatic islets. Post-mortem studies of patients close to disease onset showed CD8+ T cells present in the islet mononuclear cell infiltrate (Gepts, 1965; Bottazzo et al., 1985). CD8+ T cells were quantified as the most abundant leukocyte in the infiltrate, with macrophages, CD4+ T helper cells, and Tregs also present (Willcox et al., 2009). Interestingly, only a small fraction of islets exhibit leukocyte infiltration at any one time (Willcox et al., 2009) and islets appear to be destroyed progressively over time corresponding to an advancing decline in insulin production (Foulis, Farquharson and Meager, 1987). Once islets become insulin-deficient the immune infiltrate dramatically declines suggesting that viable β cells are required to maintain the presence of immune cells (Willcox et al., 2009).

There is abundant evidence in support of a role for CD8+ T cells in T1D. Adoptive transfer of CD8+ T cells in animal models was shown to induce disease state (DiLorenzo and Serreze, 2005). The CD8+ T cells present in diabetic islets of mice were shown to be islet-specific, autoreactive T cells (Trudeau et al., 2003; Wong et al., 2007; Énée et al., 2008). This was followed by similar demonstration in human diabetic islets (Coppierets et al., 2012). Many diabetic autoantigens have been described, and autoreactive T cell clones capable of β cell destruction isolated (Liu et al., 1999; Li et al., 2010; Delong et al., 2011; Nikoopour et al., 2011; Unger et al., 2012; Chujo et al., 2013). For example, in 1995 Panina-Bordignon et al. demonstrated specific T cell cytotoxicity against a glutamic acid decarboxylase 65 (GAD65)-derived peptide (GAD114-123). They noted that the peptide was presented by HLA-A2 and cytotoxicity was only detected in T cells derived from HLA-A2+ T1D patients (Panina-Bordignon et al., 1995). In another example, CD8+ T cells specific for zinc transporter 8 (ZnT8) were isolated from diabetic children (Énée et al., 2012). ZnT8 is expressed in β cells and involved in the regulation of insulin storage and secretion (Kawasaki, 2012). The preproinsulin (PPI)-derived epitope LWMRLLPLL (residues 3-11) is presented on the T1D-associated HLA-A24 molecule and is recognised by CD8+ T cell clones found in diabetic
patients (Kronenberg et al., 2012). An autoantigen that holds importance to this chapter is the HLV EALYLV peptide. This peptide is derived from the insulin B chain (residues 10-18) and is also presented on HLA-A2. HLV EALYLV was first described as a T1D-relevant epitope in 2005 (Pinkse et al., 2005). It is recognised by a T cell clone called InsB4 that was isolated from a T1D patient in 2016 (Theaker et al., 2016). Another T1D epitope used as a model system in this thesis is the HLA-A2-restricted PPI-derived peptide, ALWGPDPAAA (ALW) (15-24). PPI15-24 is a key diabetic epitope (Velthuis et al., 2010) encoded within the signal peptide of PPI. ALWGPDPAAA was isolated by elution of peptides from INS-expressing HLA-A2+ cells and is recognised by a patient-derived CD8+ T cell clone, 1E6 (Skowera et al., 2008). Tellingly, human pancreatic β cells only become targets for 1E6 when maintained in glucose-containing media where they manufacture insulin from stored PPI (Skowera et al., 2008). Inflammatory environments promote MHC class I overexpression in the islets (Foulis, Farquharson and Meager, 1987) which could contribute to disease propagation through epitope spreading and the emergence of cryptic antigens.

The constellation of evidence described above strongly implicates CD8+ T cells in the cause of T1D. One model of how autoimmune diseases are triggered (see Section 1.5.2) is that pathogen-specific T cells inadvertently recognise self-antigens because they are crossreactive.

3.1.2 – T cells are, by necessity, highly crossreactive

CD8+ T cells, including those that implicated in various autoimmune diseases, are crossreactive (as described in Section 1.4). The notion of T cell crossreactivity has supplanted the previously accepted model of how lymphocytes contribute to adaptive immunity known as the Clonal Selection Theory (CST), which was proposed prior to 1960 and later refined (Burnet, 1959, 1976). The CST propagated the idea that each lymphocyte is highly specific for a single antigenic ligand. This paradigm faced mounting evidence to the contrary in the 1990s as understanding of the nature of the ligands recognised by antigen receptors, especially the αβ TCR, increased (Bhardwaj et al., 1993; Wucherpfennig and Strominger, 1995; Mason, 1998; Misko et al., 1999). Counterevidence questioning the CST included structural proof of a single TCR recognising multiple peptide ligands (Hagerty and Allen, 1995; Reiser et al., 2003). In fact, a single TCR can even tolerate polymorphisms in the MHC component of its target (Hennecke and Wiley, 2002; Luz et al., 2002). Further evidence in favour of crossreactive T cells stems from the fact that a single pMHC molecule can select for
a relatively wide ranging T cell repertoire (Ignatowicz, Kappler and Marrack, 1996), and a T cell that has been activated by one peptide can then respond to others (Ignatowicz et al., 1997; Galloway et al., 2019).

Don Mason’s seminal opinion piece, published in 1998, documented the theoretical evidence for why crossreactivity is an essential characteristic of the TCR (Mason, 1998). Mason suggested that in order to achieve full immune coverage a single TCR would likely recognise between $10^6$ and $10^7$ different pMHC targets (Mason, 1998). The CST was further counteracted by calculations by Arstila et al. in 1999, describing how the $10^{12}$ T cells present in the human body likely carry one of just 25 million ($2.5 \times 10^7$) distinct TCRs (Arstila et al., 1999). The 20 proteogenic L-amino acids can be combined as 9-mers in $5.1 \times 10^{11}$ different ways with at least 1% of these combinations amenable to be presented by a given MHC molecule at the cell surface (Mason, 1998; Wooldridge, 2013). Thus, even the most conservative calculations, those that assume that there is only one MHC per individual and that it only presents 9-mer peptides, require that each TCR see hundreds of different peptides to provide full recognition coverage. Therefore, even lacking the abundant experimental evidence of T cell crossreactivity, it would be a logical necessity to provide comprehensive immune coverage to the host (Mason, 1998; Sewell, 2012).

3.1.2.1 – Benefits of T cell crossreactivity

T cell crossreactivity can facilitate recognition of viral escape mutants, post-translationally modified peptides, or neoepitopes allowing the immune system to remain in control of infections or malignancy despite a high mutational load (Elliott, Bodmer and Townsend, 1996; Glithero et al., 1999). For example, the HIV-1-specific 868 TCR can recognise a number of common SLYNTVATL escape mutants (Cole, Fuller, et al., 2017). A tumour-responsive T cell clone, MEL5, recognises epitopes from the Melan-A/MART-1 protein in melanoma. MEL5 is able to bind two overlapping epitopes expressed on melanoma cells in the context of HLA-A2 (Borbulevych et al., 2007; Madura et al., 2019). Interestingly, Wen et al. described an example of cross-species crossreactivity (also known as heterologous immunity) by showing that Dengue-responsive CD8$^+$ T cells could confer protection against Zika virus (Wen et al., 2017). In another example, Chiu et al. identified an immunodominant varicella-zoster virus (VZV) epitope and subsequently found homologs in two other α-herpesviruses (herpes simplex virus (HSV)-1, and HSV-2) and a γ-herpesvirus, Epstein-Barr Virus (EBV) (Chiu et al., 2014). They found that PBMCs from VZV-seropositive individuals could respond to the 3
homolog epitopes regardless of the individuals prior exposure to each virus, suggesting cross-species crossreactivity (Chiu et al., 2014).

T cell crossreactivity is believed to assist the immune system spatially and temporally. An infected cell can only be scanned by so many T cells in a given time scale, and it is important that foreign peptides are detected rapidly to help control the spread of the infection. Each T cell being capable of recognising many (even more than a million (Wooldridghe et al., 2012)) peptides can speed up detection time as any foreign peptide will likely be recognised by many different T cells (Sewell, 2012).

T cell crossreactivity can also be exploited therapeutically as peptide mimics can be used as vaccines to prime endogenous T cell populations. This can include use of more potent ‘super agonists’ as vaccines. Another PhD student in my research group recently used CPL screening to generate a super agonist peptide for a melanoma-reactive T cell clone isolated from the tumour-infiltrating lymphocytes (TIL) infusion product used to induce complete remission in a Stage IV malignant melanoma patient (Galloway et al., 2019). The ST8.24 T cell clone used in this study was expanded in patient blood following cancer remission (Galloway et al., 2019). T cell priming from HLA-A2+ PBMC with a super agonist peptide for the ST8.24 T cell clone generated more robust killing of melanoma cells than those primed with the natural antigen (Galloway et al., 2019). In another example, the crossreactive nature of T cells has allowed for an artificial D-amino acid peptide to be used as a vaccine to prime T cells to respond to an influenza epitope (Miles et al., 2018). D-amino acid peptides have the advantage of being much longer lived in serum and resistant to degradation by gastric enzymes thereby making oral delivery possible (Miles et al., 2018). Thus, well-understood crossreactivity can be harnessed to provide effective and intelligently designed therapeutics.

3.1.2.2 – Drawbacks of T cell crossreactivity
Unfortunately, T cell crossreactivity can be the root cause of a number of negative clinical and biological outcomes. T cell crossreactivity can cause potentially fatal side effects when T cells or TCRs are used as a therapeutic. For example, an affinity-enhanced TCR targeting the multi-cancer associated MAGE-A3 antigen presented on HLA-A*01 led to patient deaths due to severe myocardial damage (Linette et al., 2013). Extensive investigation led to the identification of crossreactivity involving an unrelated cardiac protein, titin (Linette et al., 2013). The titin epitope responsible was soon identified and shown to also be presented by HLA-A*01 (Cameron et al., 2013). Crystal structures showed that the two pMHCs were
molecular mimics (Raman et al., 2016). Other examples of unintentional therapeutic crossreactivity have been reported (Johnson et al., 2009; Parkhurst et al., 2011), demonstrating the need for a more complete understanding of TCR peptide preferences to enable prediction of such crossreactivities. Experimental procedures such as those described in this chapter can be used to gain better understanding of TCR peptide preferences.

Aberrant T cell crossreactivity is also believed to initiate autoimmunity (see Section 1.5.2). Naive T cells require strong TCR-pMHC interactions to activate them, so are not normally activated by self-ligands which tend to be weak (van der Merwe and Davis, 2003; Bridgeman et al., 2011). Furthermore, naive T cells generally require a ‘second signal’ as an in-built safety mechanism (Jenkins and Schwartz, 1987). The role of CD28 and the co-inhibitory receptors in preventing autoimmunity is discussed in detail in Section 1.5.2. Naive T cells are, therefore, activated by professional APCs that present pathogenic ligands in the context of other co-stimulatory signals, this allows a pathogenic epitope to break tolerance (Shahinian et al., 1993; Riha and Rudd, 2010). Effector-memory T cells, in contrast, can be re-activated by less potent ligands than those required to active naive T cells and can be activated in the absence of a ‘second signal’ (Sewell, 2012). Thus, pathogenic T cell epitopes, to which the immune system should, and will, respond to can trigger autoimmunity through the activation of T cells with promiscuous TCRs, see Figure 3.1 (Ruff et al., 2019). These pathogenic epitopes may be molecular mimics of self-epitopes that would otherwise have gone unnoticed by the T cells with the capacity to crossreact to them (Ruff et al., 2019). 1E6, a CD8+ T cell clone implicated in T1D, recognises over a million peptides (Wooldridge et al., 2012). Some of these peptides could be found in proteome databases for human pathogens (Wooldridge et al., 2012; Cole et al., 2016). One peptide epitope in particular, RQFGPDWIVA from the C. asparagiforme proteome, was recognised with 4-fold stronger affinity than the index T1D-associated epitope (Wooldridge et al., 2012; Cole et al., 2016). This raises the possibility that the autoreactive T cell in question, 1E6, could have initially been activated by a much more potent ligand of pathogenic origin and, as a result of its promiscuity, have subsequently activated in response to the insulin epitope.
3.1.3 – Measuring T cell crossreactivity

The ability to measure and define T cell crossreactivity is important both to the discovery of new TCR-pMHC interactions, and in the safety considerations of their therapeutic application (Cameron et al., 2013). Understanding the ‘rules’ that govern the plasticity of a TCR-pMHC interaction, including those that dictate the promiscuity of a TCR, may help to identify autoimmune triggers (Yi Li et al., 2005; Bulek et al., 2012; Wooldridge et al., 2012; Cole et al., 2016), facilitate design of more potent TCRs as therapeutics (Molloy, Sewell and Jakobsen, 2005; Harper et al., 2018; Sanderson et al., 2020), and help prevent off-target toxicity in T cell therapeutics (Cameron et al., 2013; Linette et al., 2013). As a result, a number of methodologies for assessing T cell crossreactivity have been developed.

T cell crossreactivity can be demonstrated via the crystallisation of complex interactions, showing a single TCR in complex with MHC molecules presenting distinct peptides (Ding et al., 1999; Reiser et al., 2003; Yin et al., 2011). Positive interactions can also be shown using SPR (Harkiolaki et al., 2009; Cole, Van Den Berg, et al., 2017). T cell assays can be used to measure response to distinct peptide ligands in many different ways including cytokine production (Whalley et al., 2020), specific lysis (Miles et al., 2018), CD69-upregulation as
measured by flow cytometry (Ishizuka et al., 2009), and \( ^3 \)H incorporation or carboxyfluorescein succinimidyl ester (CFSE) loss as proxies for T cell proliferation (Kersh et al., 1998; Galloway et al., 2019). These methods are useful, on the small-scale, to confirm suspected crossreactivities. They do not, however discover or demonstrate the extent of T cell crossreactivity.

In an attempt to discover crossreactivities, Ishizuka et al. assessed nearly 30,000 TCR-pMHC interactions using a large library of synthetic peptides (Ishizuka et al., 2009). They evaluated the specificities of 15 T cell clones by co-incubating pools of the peptides, the T cell clones, and an APC line. They then assessed reactivity using CD69 upregulation, \( ^3 \)H incorporation, and IFN\( \gamma \) enzyme-linked immunospot (ELISpot). Despite this labour-intensive effort, they were only able to identify a single example of crossreactivity (Ishizuka et al., 2009). 30,000 peptides represents just a fraction of the available MHC-binding peptide universe (Sewell, 2012; Wooldridge, 2013) so this result is unsurprising, however it does aptly demonstrate the need for very high throughput methodology for demonstrating and assessing T cell crossreactivity.

T cell specificity for a range of pMHC ligands can be demonstrated by staining with fluorescently labelled multimers and subsequent flow cytometry analysis (Wooldridge et al., 2009; Dolton et al., 2015). However, this tends to be limited by the available panel of fluorophores (Lee et al., 2020). To overcome this limitation, pMHC multimers can be tagged with DNA barcodes (Bentzen et al., 2016). The barcodes act as individual identifiers that allow the peptide agonists to be determined following staining and cell sorting. This method gives a more extensive insight into the crossreactive repertoire of a particular T cell clone. Bentzen et al. were able to screen over 1000 different peptides using the technique (Bentzen et al., 2016). Measuring the full extent of T cell crossreactivity, however, requires greater breadth that even DNA barcoded pMHC can offer. Furthermore, technologies for assessing T cell crossreactivity involving any sort of pMHC multimers are expensive and are therefore still limited to very small fraction of the possible peptide universe. True assessment of T cell crossreactivity requires that a TCR is screened against all possible agonists – something made possible by combinatorial peptide library (CPL) screens (Hemmer et al., 1997; Hiemstra et al., 1997; Wooldridge et al., 2012).
3.1.3.1 – CPL screens for measurement of T cell crossreactivity

CPL screen technologies were pioneered in the early 1990s by Dr Clemencia Pinilla and colleagues (Houghten et al., 1991; Pinilla et al., 1992). Originally developed to probe the ligand preferences of monoclonal antibodies, CPL screens consist of mixtures of free peptides in solution. Within each mixture a single amino acid is fixed at a single position, with all other positions allowed to vary amongst the 19 proteogenic L-amino acids, Figure 3.2 (cysteine is generally excluded from the non-fixed amino acid pools due to its propensity to cause cross-linking via the formation of disulphide bonds) (Houghten et al., 1991; Pinilla et al., 1992; Hemmer et al., 1998; Ninovasquez, 2004). A significant advantage to the use of CPL screens for ligand identification is the lack of a need for any prior knowledge of the likely ligand candidates (Houghten et al., 1991). Preceding methods involved time-consuming synthesis and screening of individual peptides (Merrifield, 1963; Geysen, Meloen and Barteling, 1984), which were generally designed around the known index peptide (Houghten, 1985) and thus did not account for the ability of TCRs to see peptides of little or no similarity to the known ligand (Reiser et al., 2003). For such methods, screening of 1000 peptides was a substantial achievement (Fodor et al., 1991). In contrast, Houghten et al. had screened T cell recognition of a library of 34 million 6-mer peptides as early as 1991 (Houghten et al., 1991).

Since their advent, CPL screens have been used to study the ligand preferences of many TCRs (Hemmer et al., 1998; Wooldridge et al., 2012; Szomolay et al., 2016). In 2012 Wooldridge et al. used CPL screening to demonstrate that the aforementioned 1E6 T cell clone (Skowera et al., 2008; Bulek et al., 2012), can recognise more than one million peptides with a potency equal to or better than the ‘index’ PPI-derived sequence (Wooldridge et al., 2012). CPL screens tend to be length-matched to the index peptide, if known, as the crossreactive behaviour of CD8+ T cells with regards to peptide sequence does not generally extend to peptides of different length. Ekeruche-Makinde et al. showed that, in fact, MHC class I-restricted TCRs tend to have strong peptide-length preferences, and rarely respond to peptides of non-preferred length (Ekeruche-Makinde et al., 2013). Ekeruche-Makinde et al. subjected 4 well-characterised TCRs to 6 CPL screens each (8-mers to 13-mers). They found that each TCR showed the highest reactivity to the CPL that matched the length of its index peptide. For example, ILA1 showed highest reactivity in the 9-mer CPL and little-to-no reactivity in the other screens. Whilst MEL5 and 1E6 showed highest reactivity in the 10-mer screen (Ekeruche-Makinde et al., 2013). MHC class I molecules do not tend to show the same tight restriction regarding the length of peptides they can present (a single MHC class I molecule can generally present peptides of 8-13 amino acids in length) (Burrows et al., 2008;
Bell et al., 2009) and so it is likely that this length-preference is imposed by the TCR component of the interaction.

<table>
<thead>
<tr>
<th>Amino Acid Residues</th>
<th>0 Fixed</th>
<th>X Degenerate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peptide sublibrary 1-20</td>
<td>O</td>
<td>X X X X X X X</td>
</tr>
<tr>
<td>Peptide sublibrary 21-40</td>
<td>X</td>
<td>O X X X X X X</td>
</tr>
<tr>
<td>Peptide sublibrary 41-60</td>
<td>X X</td>
<td>O X X X X X X</td>
</tr>
<tr>
<td>Peptide sublibrary 61-80</td>
<td>X X X</td>
<td>O X X X X X X</td>
</tr>
<tr>
<td>Peptide sublibrary 81-100</td>
<td>X X X X</td>
<td>O X X X X X</td>
</tr>
<tr>
<td>Peptide sublibrary 101-120</td>
<td>X X X X</td>
<td>O X X X X</td>
</tr>
<tr>
<td>Peptide sublibrary 121-140</td>
<td>X X X X</td>
<td>O X X</td>
</tr>
<tr>
<td>Peptide sublibrary 141-160</td>
<td>X X X X X</td>
<td>O X</td>
</tr>
<tr>
<td>Peptide sublibrary 161-180</td>
<td>X X X X X X X</td>
<td>O</td>
</tr>
</tbody>
</table>

**Figure 3.2: Typical 9-mer CPL format.** The diagram represents the 180 of peptide mixtures that comprise a 9-mer CPL scan. Each peptide mixture contains one of the 20 proteogenic amino acids fixed at a one position (represented by O), with all other positions allowed to vary amongst the 19 proteogenic L-amino acid excluding cysteine (represented as X). Cysteine is generally excluded from X due to its propensity to cause cross-linking via the formation of disulphide bonds.

3.1.3.2 – CPL screens can identify novel peptide epitopes
In 2016, Szomolay et al. developed a publicly accessible webtool to facilitate ‘CPL-driven database searching’ (Szomolay et al., 2016). This method uses raw CPL screen data to search proteome databases and identify likely ligands of the T cell clone in question (Szomolay et al., 2016). Unlike previous CPL-based ligand identification, this webtool does not just identify likely ligands but ranks them in order of recognition likelihood. An overview of the TCR ligand discovery process is given in Figure 3.3. The first stage of the process involves isolating specific CD8⁺ T cells from PBMC and cloning by dilution to obtain a single T cell (Theaker et al., 2016). If the T cell cannot be cloned it can be manufactured by TCR replacement (Legut et al., 2018). Once a monoclonal population has been obtained it is subject to a sizing scan to determine the peptide length preference (Ekeruche-Makinde et al., 2013). A length-matched CPL screen is then conducted, the output of which is used to search proteomic databases (Szomolay et al., 2016). Finally predicted peptide ligands require validation and
epitope confirmation, as demonstrated in this chapter. CPL-driven database searching can assess the entire peptide universe and thus has the potential to identify ligands that are stronger than the known index peptide, such as a pathogenic trigger for an autoreactive T cell (Wooldridge et al., 2012; Szomolay et al., 2016; Whalley et al., 2020).

Figure 3.3: An overview of the TCR ligand discovery process. **Stage 1:** T cells are first isolated from PBMC based on their specificity. These are then diluted to less than 1 cell per well and plated to obtain a T cell clone. **Stage 2:** The clone is subject to sizing scan to uncover its peptide length preferences. Then an appropriately length-matched CPL is carried out. CPL data is used to search proteomic databases for likely ligands. **Stage 3:** These peptide ligands then require validation by expressing the parent protein in APC to confirm epitope processing and presentation.
3.1.4 – Aims
I hypothesised that a previously-characterised T1D-relevant T cell clone might be crossreactive and would potentially recognise peptides from pathogenic proteomes. To investigate this, I used CPL-driven database searching to explore the crossreactive profile of this T cell clone and then used the resulting profile to search for potential pathogenic peptides that act as strong agonists. This approach has potential to identify the ‘initial trigger’ that stimulates T cell activation, breaks self-tolerance, and promotes autoimmune development. Specifically, I aimed to:

- Investigate the crossreactive nature of the InsB4 T cell clone through CPL-driven database screening of pathogenic databases.
- Validate any notable peptides as legitimate HLA-A2-presented epitopes.
- Explore the potential for crossreactivity between the cognate antigen and any identified pathogenic epitopes in T1D patient blood.
3.2 – Results

3.2.1 – Work leading up to the project

InsB4 is a T1D-associated T cell clone that was isolated from a diabetic patient in 2016 by Dr Garry Dolton (Theaker et al., 2016). InsB4 is specific for a peptide from the insulin B chain (residues 10-18), HLVEALYLV. Expression of the insulin B chain is restricted to pancreatic islet β cells, making it a suitably specific target for autoreactive CD8+ T cells in the context of T1D (Harding and Ron, 2002; Pinkse et al., 2005). HLVEALYLV is a naturally processed HLA-A2 restricted human T cell epitope that was first described in 2005 (Pinkse et al., 2005), it is therefore a highly relevant model system in which to investigate T cell crossreactivity. Following characterisation of the epitope and demonstration of strong HLA-A2 binding, Pinkse et al. used tetramer staining to demonstrate the presence of HLA-A2-HLVEALYLV-specific CD8+ T cells in both healthy donors and diabetics. They observed that in healthy donors the immune system ‘remains ignorant,’ (Pinkse et al., 2005) potentially until exposure to a pathogenic trigger.

This last piece of evidence, that HLA-A2-HLVEALYLV-reactive CD8+ T cell clones found in healthy humans remain ignorant of their insulin-derived ligand is intriguing. Why is it that T cell clones could be activated and contributing to autoimmunity in one individual but remain naive in another? To investigate this, we turned to the ample evidence for pathogenic triggers of autoimmunity (discussed in detail in Section 1.5.2).

3.2.1.1 – The HLA-A2-restricted HLVEALYLV epitope is presented at the surface of pancreatic β cells.

My laboratory confirmed that the InsB4 T cell clone recognised the HLA-A2-restricted HLVEALYLV peptide (B chain residues 10-18). Figure 3.4A shows the dose-dependent response of the clone to this peptide, as measured by MIP-1β ELISA. The InsB4 T cell also stains well with HLA-A2-HLVEALYLV PE-conjugated pMHC dextramers using an optimised staining procedure (Wooldridge et al., 2009; Dolton et al., 2015, 2018). When cells were gated on live CD3+CD8+ single lymphocytes, mean fluorescence intensity (MFI) values of 73.2, 108, and 7384 were given for staining with no dextramer, staining with HLA-A2-GILGFVFTL derived from the influenza virus matrix protein (irrelevant dextramer), and staining with HLA-A2-HLVEALYLV dextramer respectively, Figure 3.4B.

HLVEALYLV was then shown to be a bona fide epitope that can be appropriately processed and expressed at the surface of cells expressing PPI. K562 cells were transduced with HLA-A2 and PPI (thereby acting as surrogate β cells). These surrogate β cells were recognised and
killed by InsB4 but were not killed by an irrelevant CMV-specific T cell clone, Figure 3.5A. As a positive control, both clones were shown to respond when $10^{-5}$ M of their cognate peptide (pp65 residues 495–503 for the irrelevant clone) was added, Figure 3.5A. This specific killing by InsB4 was dependent on the expression of PPI as K562 cells expressing only HLA-A2, or those expressing HLA-A2 and another transduced protein (GAD65), were not killed, Figure 3.5B. These data suggest that the HLV EALYLV is correctly processed from the PPI protein and expressed at the cell surface for scrutiny by T cells.

My laboratory subsequently demonstrated specific killing of real human pancreatic β cells by InsB4 T cells, Figure 3.5C. T1D-specific clones InsB6 (sister clone) and 1E6 (Bulek et al., 2012) were used as positive controls. CMV-specific clone, CMV.1, was used as a negative control.

Figure 3.4: The InsB4 T cell clone recognises an insulin-derived epitope. A) MIP-1β response of the HLA-A2-restricted clone InsB4 to increasing concentrations of the insulin B chain HLV EALYLV peptide. 30,000 InsB4 T cells were cultured in the presence of 60,000 T2 cells as APC and varying concentrations ($10^{-9}$ M to $10^{-5}$ M) of HLV EALYLV overnight before the supernatant was harvested for measurement of MIP-1β presence by ELISA analysis. Performed in duplicate, error bars depict standard error of the mean (SEM). Assay performed by Garry Dolton. B) 50,000 InsB4 T cells were stained with 0.5 µg HLV EALYLV dextramer (relative to the pMHC component) with a PE-conjugated backbone. Cells previously gated on live CD3+CD8+ single lymphocytes. MFI values shown for no dextramer control (grey shaded), irrelevant dextramer (red), and HLV EALYLV dextramer (blue) are the mean of 10,000 events. Irrelevant pMHC used in this experiment is HLA-A2-GILGFVFTL, an M1-derived immunodominant influenza peptide.
Figure 3.5: InsB4 recognises and kills pancreatic β cells. A) K562 cells transduced with HLA-A2 and preproinsulin are used as a proxy for pancreatic β cells. InsB4 clone kills surrogate β cells, as measured by chromium release cytotoxicity assay at a T cell to target cell ratio of 3:1 over 5 h. CMV.1 clone used as an irrelevant control. 10^{-5}M of index peptide (pp65 residues 495–503 for the CMV.1 clone) was added exogenously for each clone to provide a positive control. B) InsB4 clone only specifically kills K562 cells transduced with both HLA-A2 and PPI. K562 cells transduced with HLA-A2 alone (no protein) or HLA-A2 with an irrelevant protein (GAD65) were not recognised. Chromium release cytotoxicity assay at a T cell to target cell ratio of 3:1 over 5 h. C) InsB4 kills real human pancreatic β cells, as measured by chromium release cytotoxicity assay at a T cell to target cell ratio of 3:1 over 5 h. T1D-specific clones InsB6 (sister clone) and 1E6 used as positive control. CMV-specific clone CMV.1 used as a negative control. All assays performed in duplicate, error bars depict SEM. Assays performed by Garry Dolton.

3.2.1.2 – CPL scanning identifies the peptide repertoire of the InsB4 T cell clone
The index peptide for InsB4, HLVEALYLV, is a 9-mer. The preference of InsB4 for 9-mer peptides was confirmed by sizing scan. This method uses random mixtures of peptides 8, 9, 10, 11, 12, and 13 amino acids in length and measures MIP-1β response to each pool by ELISA. InsB4 showed an extremely strong preference for 9-mer peptides, Figure 3.6A.

Based on these data, a 9-mer CPL screen was performed on the InsB4 clone to investigate amino acid preferences at each peptide position. Response to each of the 180 peptide mixtures in the CPL, as measured by MIP-1β production, is shown in Figure 3.6B. At many peptide positions, particularly positions 1, 5, 6, and 7, the index amino acids (indicated by the red bar) were not the preferred amino acid – suggesting that a stronger peptide agonist might be readily discoverable. These data are summarised as a logo plot in Figure 3.6C. It is noteworthy that the glutamic acid residue found at position 4 in the index peptide is so clearly preferred at this position. This suggests that the 4E residue of the peptide might be a key TCR contact residue for the interaction.
Figure 3.6: Characterising the peptide preferences of InsB4. A) Sizing scan of the InsB4 clone, based on MIP-1β production. 30,000 InsB4 T cells were cultured in the presence of 60,000 T2 cells as APC and peptides of varying lengths overnight before the supernatant was harvested for measurement of MIP-1β presence by ELISA analysis. Performed in duplicate, error bars depict SEM. Assay performed by Garry Dolton. B) CPL screen of the InsB4 T cell clone. 30,000 InsB4 T cells were co-cultured with 60,000 T2 cells as APC and 100 µg/mL peptide mix from a 9-mer CPL screen. Supernatants were harvested and ELISA analysis performed to measure MIP-1β presence as a proxy for T cell response. Red bars indicate the sequence of the index HLVEALYL peptide. Performed in duplicate, error bars depict SEM. Assay performed by Garry Dolton. C) Logo plot representing amino acid preferences of InsB4 at each peptide position.
3.2.1.3 – Fungal-derived peptide agonists of the InsB4 T cell clone
CPL data, along with a publicly assessable webtool (Szomolay et al., 2016; Whalley et al., 2020), were then used to screen pathogenic proteome databases. This screening method, developed by Szomolay et al. allowed ranked predictions of peptide ligands for the InsB4 TCR.

The top 20 scoring fungal peptides ranked by a likelihood score, are shown in Figure 3.7. I assessed the recognition of these peptides in crude form (~40% purity) by InsB4 T cells alongside that of the insulin-derived HLVEALYLV peptide by MIP-1β ELISA, Figure 3.7. Half maximal effective concentration (EC₅₀) values, as calculated by GraphPad Prism software, confirmed that many of the predicted fungal peptides were more potent ligands than HLVEALYLV.
Figure 3.7: InsB4 recognises fungal peptides with greater sensitivity than the index peptide. The top 20 ranked peptides from the fungal proteomic database predicted to be InsB4 ligands generated as described by Szomolay et al. (Szomolay et al., 2016). Greater ‘score’ indicates prediction for a more potent ligand. Peptides were tested for their ability to trigger MIP-1β production. 30,000 InsB4 T cells were cultured in the presence of 60,000 T2 cells as APC and varying concentrations (10^{-12} M to 10^{-5} M) of each peptide overnight before the supernatants were harvested for measurement of MIP-1β presence by ELISA analysis. The EC_{50} for each peptide is shown in the accompanying table. NR = no response. Measurements performed in duplicate, error bars depict SEM.
3.2.1.4 – **Bacterial-derived peptide agonists of the InsB4 T cell clone**

CPL data was also used to screen the pathogenic bacterial database collated by Szomolay *et al.* (Szomolay *et al.*, 2016). The top 20 scoring bacterial peptides ranked by a likelihood score, are shown in **Figure 3.8**. I assessed the ability of these peptides in crude form (~40% purity) to activate InsB4 T cells, alongside that of the insulin-derived HLVEALYL peptide, by MIP-1β ELISA, **Figure 3.8**. EC$_{50}$, as calculated by GraphPad Prism software, suggested that many of the predicted bacterial peptides were more potent ligands than HLVEALYL.

At the time of my joining the lab, the above data were being assembled into a paper that has recently been published (Whalley *et al.*, 2020). The paper describes a novel graphics processing unit (GPU)-based algorithm that can be used to rapidly screen databases, such as the peptide databases assembled by my colleague Dr Barbara Szomolay (Szomolay *et al.*, 2016). This algorithm was used to identify agonist ligands for the InsB4 T cells that might be processed and presented from fungal and bacterial pathogens (Whalley *et al.*, 2020). As a new member of the group, my initial training in T cell assays involved performing the ELISAs that were used to verify the predicted peptides, as shown above, as such I was listed as a co-author on the Whalley *et al.* study which described agonist peptide ligands derived from bacterial and fungal proteomes (Whalley *et al.*, 2020).
Figure 3.8: InsB4 recognises bacterial peptides with greater sensitivity than the index peptide. The top 20 ranked peptides from the bacterial proteomic database predicted to be InsB4 ligands generated as described by Szomolay et al. (Szomolay et al., 2016). Greater ‘score’ indicates prediction for a more potent ligand. Peptides were tested for their ability to trigger MIP-1β production. 30,000 InsB4 T cells were cultured in the presence of 60,000 T2 cells as APC and varying concentrations (10^{-11} M to 10^{-5} M) of each peptide overnight before the supernatant was harvested for measurement of MIP-1β presence by ELISA analysis. The EC_{50} for each peptide is shown in the accompanying table. ∼ = ambiguous EC_{50} based on GraphPad Prism software analysis. Measurements performed in duplicate, error bars depict SEM.

<table>
<thead>
<tr>
<th>Peptide Score</th>
<th>EC_{50} [M]</th>
<th>Peptide Sequence</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>-24.60</td>
<td>1.1 x 10^{-5}</td>
<td>ELVEALAYLV</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>-12.66</td>
<td>2.9 x 10^{-6}</td>
<td>LLLERYLLYV</td>
<td>Klebsiella pneumonia</td>
</tr>
<tr>
<td>-12.70</td>
<td>3.2 x 10^{-5}</td>
<td>IMENILFF</td>
<td>Enterobacter cloacae</td>
</tr>
<tr>
<td>-12.83</td>
<td>5.1 x 10^{-6}</td>
<td>LLLPERFLL</td>
<td>Bacillus cereus</td>
</tr>
<tr>
<td>-12.87</td>
<td>3.3 x 10^{-6}</td>
<td>IMGERALLLYV</td>
<td>Helicobacter pylori</td>
</tr>
<tr>
<td>-12.88</td>
<td>1.5 x 10^{-7}</td>
<td>IMERILLYYL</td>
<td>Haemophilus influenzae</td>
</tr>
<tr>
<td>-12.90</td>
<td>8.8 x 10^{-7}</td>
<td>IMGERALLLF</td>
<td>Streptococcus mitis</td>
</tr>
<tr>
<td>-12.92</td>
<td>2.2 x 10^{-7}</td>
<td>IMGERALLF</td>
<td>Helicobacter pylori</td>
</tr>
<tr>
<td>-12.93</td>
<td>2.8 x 10^{-7}</td>
<td>IMERILLF</td>
<td>Clostridium novyi</td>
</tr>
<tr>
<td>-12.94</td>
<td>3.2 x 10^{-7}</td>
<td>IMEROLLF</td>
<td>Pseudomonas aeruginosa</td>
</tr>
<tr>
<td>-12.97</td>
<td>4.6 x 10^{-5}</td>
<td>IMEROLLFY</td>
<td>Streptococcus oralis</td>
</tr>
<tr>
<td>-10.05</td>
<td>7.8 x 10^{-7}</td>
<td>ELVRYLLLYV</td>
<td>Streptococcus salivarius</td>
</tr>
<tr>
<td>-10.07</td>
<td>1.8 x 10^{-8}</td>
<td>LLLERYLLYV</td>
<td>Clostridium butyricum</td>
</tr>
<tr>
<td>-10.08</td>
<td>2.2 x 10^{-7}</td>
<td>LLRERYLLLYV</td>
<td>Helicobacter pylori</td>
</tr>
<tr>
<td>-10.13</td>
<td>1.9 x 10^{-8}</td>
<td>MILLERLL</td>
<td>Pseudomonas aeruginosa</td>
</tr>
<tr>
<td>-10.14</td>
<td>1.3 x 10^{-6}</td>
<td>IMGERALLA</td>
<td>Clostridium butyricum</td>
</tr>
<tr>
<td>-10.16</td>
<td>3.5 x 10^{-8}</td>
<td>LLLERYLLYY</td>
<td>Haemophilus influenzae</td>
</tr>
<tr>
<td>-10.17</td>
<td>&lt;0.08</td>
<td>LLLEROLLFA</td>
<td>Pseudomonas aeruginosa</td>
</tr>
<tr>
<td>-10.24</td>
<td>1.1 x 10^{-5}</td>
<td>MLLVNLLYLV</td>
<td>Pseudomonas aeruginosa</td>
</tr>
<tr>
<td>-10.26</td>
<td>6.0 x 10^{-7}</td>
<td>LLRERYLLYLV</td>
<td>Pseudomonas aeruginosa</td>
</tr>
<tr>
<td>-10.28</td>
<td>4.2 x 10^{-6}</td>
<td>LLRERYLLYLV</td>
<td>Salmonella enterica</td>
</tr>
</tbody>
</table>
3.2.2 – Fungal and bacterial peptides act as functional ligands for the InsB4 T cell clone

I next aimed to validate the top scoring fungal and bacterial peptides. The top fungal peptide (the *Candida albicans* peptide (MIVENWPLL)) and the top bacterial peptide (the *Helicobacter pylori* peptide (MLLENGLLA)) were ordered as pure preparations (>95% purity; Peptide Synthetics) and I re-assessed the MIP-1β response of InsB4 by ELISA (peptide concentrations ranging from $10^{-15}$ to $10^{-4}$ M), **Figure 3.9**. The EC$_{50}$ value for each titration was generated using GraphPad Prism software analysis. Each pathogenic peptide proved to be a more potent agonist for InsB4 that the T1D-associated index peptide, HLVEALYLV.

<table>
<thead>
<tr>
<th>Peptide Score</th>
<th>EC$_{50}$ (M)</th>
<th>Peptide Sequence</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>-24.60</td>
<td>$1.6 \times 10^{-4}$</td>
<td>HLVEALYLV</td>
<td><em>Homo sapiens</em></td>
</tr>
<tr>
<td>-14.31</td>
<td>$1.6 \times 10^{-3}$</td>
<td>MIVENWPLL</td>
<td><em>Candida albicans</em></td>
</tr>
<tr>
<td>-13.13</td>
<td>$2.7 \times 10^{-4}$</td>
<td>MLLENGLLA</td>
<td><em>Helicobacter pylori</em></td>
</tr>
</tbody>
</table>

**Figure 3.9**: Pure peptide titrations for the top pathogenic peptides. The top bacterial and fungal peptides based on crude peptide titration EC$_{50}$ values were ordered as pure preparations. Peptides were tested for their ability to trigger MIP-1β production. 30,000 InsB4 T cells were cultured in the presence of 60,000 T2 cells as APC and varying concentrations ($10^{-15}$ M to $10^{-4}$ M) of each peptide overnight before the supernatant was harvested for measurement of MIP-1β presence by ELISA analysis. The EC$_{50}$ for each peptide is shown in the accompanying table, based on GraphPad Prism software analysis. Measurements performed in duplicate, error bars depict SEM.
Binding of the InsB4 TCR to the two pathogenic agonists was then confirmed by SPR. Each of the pathogenic peptides had an affinity for InsB4 that was measurable by SPR, whereas the index HLV_EALYLV ligand bound too weakly to be measured. HLA-A2-MIVENVPLL (C. albicans) and HLA-A2-MLLENGLLA (H. pylori) had affinities (K_d) for InsB4 of 60.7 µM and 168.5 µM respectively (Figure 3.10A). These data confirm the interaction between InsB4 and each of the pathogenic epitopes in the context of HLA-A2.

I next demonstrated that the InsB4 T cell clone successfully bound by PE-conjugated HLA-A2 multimers made with the two pathogenic peptides. MFI values (displayed on the respective histograms) support the notion that InsB4 binds more strongly to the pathogenic peptides than to the index HLV_EALYLV peptide (light blue) which only stains as a dextramer, Figure 3.10B.

**Figure 3.10**: InsB4 TCR binding of pathogen-derived agonist peptides is confirmed by SPR and pMHC multimer staining. A) SPR responses of the InsB4 TCR against the top fungal and top bacterial peptides as soluble HLA-A2 complexes. Responses against an irrelevant peptide HLA-A2-NLSAIGIFST were used to remove background response. 10 serial dilutions of the InsB4 were used, with a maximum concentration of 407.0 µM. Index A2-HLV_EALYLV binding to InsB4 was too weak to measure. Assay performed by Aaron Wall. B) 50,000 InsB4 T cells were stained with 0.5 µg of tetramer or dextramer (relative to the pMHC component), with a PE-conjugated backbone. Cells previously gated on live CD3^+^CD8^+^ single lymphocytes. The human telomerase reverse transcriptase-derived ILAFKFLHEL was used as a negative control. Numbers on the histograms correspond to MFI values, mean of 10,000 events.
3.2.3 – Peptide sequences derived from human viruses act as potent agonists of the InsB4 T cell

After validating the bacterial and fungal crude peptide ligands that were described in the Whalley et al. paper (Whalley et al., 2020), the majority of my own work with the InsB4 T cell clone revolved around examining viral-derived agonist peptides. Some of these ligands are known viral epitopes (discussed in Section 3.3.4) which acted as extremely potent agonists for the InsB4 T cell, making them strong candidates for being the triggering event that might have initiated T1D.

I began by using the CPL data and the Szomolay webtool (Szomolay et al., 2016), to screen for potential ligands within the pathogenic viral proteome. The top scoring peptides from the viral database were synthesised (Peptide Synthetics) in crude form (~40% purity) and I assessed their ability to evoke a response from InsB4 T cells, Figure 3.11. InsB4 was incubated with titrations of each peptide (concentrations ranging from $10^{-12}$ to $10^{-5}$ M) and T2 cells overnight, the resulting MIP-1β production was measured by ELISA. All viral peptides gave measurable responses despite many having little sequence homology with the HLVEALYLV peptide. In fact, many proved to be far more potent ligands for InsB4 than the index HLVEALYLV peptide itself and were substantially more potent than the fungal and bacterial peptides that had previously been identified (Figure 3.7 and 3.8). These data, once again, show that InsB4 is capable of responding to more potent peptides than the T1D-associated epitope, and suggest that a pathogen-derived agonist may exist. Given that the top viral peptides were so much more potent than even the bacterial and fungal peptides, these data support the hypothesis that the initial trigger that activated InsB4 T cells in vivo could be virus-derived.

The three most potent viral epitopes for activating InsB4 T cells (ILLIEGFIV; varicella-zoster virus (referred to as chickenpox virus henceforth), ILLIEGIFV; HSV-1, and ILLIEGIFI; EBV) are highlighted by grey shading in Figure 3.11. These peptides were ordered as pure preparations (>95 % purity; Peptide Synthetics) and re-assessed for activation of the InsB4 T cell clone by MIP-1β ELISA over concentrations ranging from $10^{-15}$ to $10^{-4}$ M, Figure 3.12. The EC$_{50}$ value for each peptide was generated using GraphPad Prism software analysis. Each viral peptide proved to be a substantially more potent agonist for InsB4 than the T1D-associated index peptide, HLVEALYLV. Indeed, all three viral peptides were 100,000-fold more potent at activating the InsB4 T cell clone than the insulin-derived sequence known to be presented at the surface of HLA-A2* β cells.
106

Figure 3.11: InsB4 recognises viral peptides with greater sensitivity than the index peptide. The top 19 ranked peptides from the viral proteomic database predicted to be InsB4 ligands, generated as described by Szomolay et al. (Szomolay et al., 2016). Greater ‘score’ indicates prediction for a more potent ligand. Peptides were tested for their ability to trigger MIP-1β production. 30,000 InsB4 T cells were cultured in the presence of 60,000 T2 cells as APC and varying concentrations (10^{-12} M to 10^{-5} M) of each peptide overnight before the supernatant was harvested for measurement of MIP-1β presence by ELISA analysis. The EC_{50} for each peptide is shown in the accompanying table. ∼ = ambiguous EC_{50} based on GraphPad Prism software analysis. Measurements performed in duplicate, error bars depict SEM. Grey shading indicates the viral super agonists discussed henceforth.
Figure 3.12: Viral-derived peptides act as strong agonists for the insulin-specific InsB4 T cell clone. The top viral peptides based on crude peptide titration EC₅₀ values were ordered as pure preparations (>95 % purity). Peptides were tested for their ability to trigger MIP-1β production. 30,000 InsB4 T cells were cultured in the presence of 60,000 T2 cells as APC and varying concentrations (10⁻¹⁵ M to 10⁻⁴ M) of each peptide overnight before the supernatant was harvested for measurement of MIP-1β presence by ELISA analysis. The EC₅₀ for each peptide is shown in the accompanying table, based on GraphPad Prism software analysis. Measurements performed in duplicate, error bars depict SEM.

<table>
<thead>
<tr>
<th>Peptide Score</th>
<th>EC₅₀ (M)</th>
<th>Peptide Sequence</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>-24.60</td>
<td>1.6 x 10⁻⁶</td>
<td>HLVEALYTV</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>-13.56</td>
<td>2.0 x 10⁻⁶</td>
<td>ILIEGIFPV</td>
<td>Chickenpox</td>
</tr>
<tr>
<td>-13.80</td>
<td>9.7 x 10⁻⁶</td>
<td>ILIEGIPPA</td>
<td>HSV-1</td>
</tr>
<tr>
<td>-13.98</td>
<td>2.4 x 10⁻⁶</td>
<td>ILIEGIPFY</td>
<td>EBV</td>
</tr>
</tbody>
</table>
3.2.3.1 – The InsB4 TCR binds to viral epitopes strongly

Binding of the InsB4 TCR to the three viral agonists was confirmed by SPR. Each peptide had an affinity for InsB4 that was measurable by SPR, whereas the index HLV EALYLV ligand bound too weakly to be measured. HLA-A2-ILLIEGIFFI, (EBV), HLA-A2-ILLIEGIFFYV (chickenpox virus), and HLA-A2-ILLIEGIFFA (HSV-1) had affinities (K<sub>D</sub>) for InsB4 TCR of 30.8 µM, 30.8 µM, and 76.6 µM, respectively (Figure 3.13A). These data confirm the interaction between InsB4 and each of the viral epitopes in the context of HLA-A2.

The InsB4 T cell clone also successfully stained with PE-conjugated HLA-A2 multimers made with the three viral peptides. As seen with the bacterial and fungal peptides, MFI values (displayed on the respective histograms) suggest that InsB4 binds more strongly to the viral peptides than to the index HLV EALYLV peptide (light blue) which only stains as a dextramer, Figure 3.13B.
**Figure 3.13:** InsB4 TCR binding of pathogenic agonists is confirmed by SPR and multimer staining. 

A) SPR responses of the InsB4 TCR against the viral peptides as soluble HLA-A2 complexes. Responses against an irrelevant peptide HLA-A2-NLSAIGIFST were used to remove background response. 10 serial dilutions of the InsB4 were used, with a maximum concentration of 221 μM. Index A2-HLV-EALYLV binding to InsB4 was too weak to measure. Assay performed by Aaron Wall. 

B) 50,000 InsB4 T cells were stained with 0.5 μg of tetramer or dextramer (relative to the pMHC component), with a PE-conjugated backbone. Cells previously gated on live CD3⁺CD8⁺ single lymphocytes. The human telomerase reverse transcriptase-derived ILAKFLHEL was used as a negative control. Numbers on the histograms correspond to MFI values, mean of 10,000 events.
3.2.3.2 – Viral epitopes are correctly processed and presented
Following confirmation of the physical interaction between the viral-derived peptides and
the InsB4 TCR, I next validated each of the three viral peptides as legitimately processed and
presented epitopes. IL1EGIFFV (chickenpox virus), IL1EGIFFA (HSV-1), and IL1EGIFFI (EBV) are
derived from viral proteins ribonucleotide reductase (chickenpox virus), ribonucleotide
reductase (HSV-1) and BARF-1 (EBV) respectively, see Table 1. These proteins were cloned
into a lentiviral vector, pELNS, (Section 2.3.3) and transduced into MOLT-3 cells (Section
2.4.2), in order to assess whether the peptide epitopes contained within these proteins
would be successfully processed and presented by HLA-A2 at the cell surface. The proteins
were transduced into both the wildtype HLA-A2+ MOLT-3 line, and a MOLT-3 line that had
previously been stably transduced with HLA-A2. Transduced MOLT-3 cells were cultured for
1 week following transduction before InsB4 reactivity was measured via TAPI-0 assay
(Section 2.6.6).

InsB4 T cells produced both TNF and CD107a in response to HLA-A2+ targets expressing all
three viral proteins, suggesting that all the epitopes were successfully processed and
presented from the full protein sequence, Figure 3.14. The response was HLA-A2-dependent
as InsB4 cells did not respond to HLA-A2- MOLT-3 cells that had been transduced with any of
the viral proteins. Flow cytometry analysis showed that 93.3, 83.9, and 72.2 % of the InsB4 T
cells in the assay responded to HLA-A2+ MOLT-3 targets expressing the target antigen from
EBV, chickenpox virus, and HSV-1 respectively. These data confirm that InsB4 T cells
recognise HLA-A2+ MOLT-3 target cells that have been transduced with the ribonucleotide
reductase genes from chickenpox virus and HSV-1 or BARF-1 from EBV.
Figure 3.14: Putative viral epitopes are efficiently processed and presented to InsB4 T cells in the context of HLA-A2. The three viral epitopes are correctly processed and presented by MOLT-3 cells allowing response from InsB4, as measured by TAPI-0 assay. MOLT-3 cells were transduced with lentiviral constructs encoding the three parental proteins from which the viral epitopes are derived. 30,000 InsB4 cells were co-cultured with 60,000 MOLT-3 cells for 4 hours before cells were harvested and assessed for surface expression of TNF and CD107a. Cells previously gated on live CD3+CD8+ single lymphocytes. Numbers on dot plots correspond to the percentage of cells in the gated population, mean of 10,000 events.
3.2.4 – InsB4 recognises EBV-infected HLA-A2\(^+\) B cells (LCL lines)

EBV is routinely used to transform lymphoblastic cell lines (LCLs). I therefore wondered if InsB4 would recognise EBV-transformed LCLs. **Figure 3.15** shows that the InsB4 clone responds to EBV-transformed LCL lines in an HLA-A2-dependent manner. TNF\(^+\) CD107a\(^+\) response to the two HLA-A2\(^+\) LCL lines, MM909.27 and MM909.36, was 1.80 % and 0.81 % respectively. Whereas 62.7 % and 46.9 % of the T cells responded to the two HLA-A2\(^+\) LCL lines MM909.26 and SAR.26 respectively. This result confirms that the InsB4 T cell can respond strongly to EBV-infected cells and suggests that herpesviruses like EBV may have been involved in the initial activation of this T cell in the T1D patient from which it was derived.

**Figure 3.15**: InsB4 responds to EBV-transformed lymphoblastic cell lines (LCLs) in an HLA-A2-dependent manner. 30,000 InsB4 T cells were cultured in the presence of 60,000 HLA-A2\(^+\) or HLA-A2\(^-\) LCLs for 4 hours before cells were harvested and assessed for surface expression of TNF and CD107a. Cells previously gated on live CD3\(^+\)CD8\(^+\) single lymphocytes. Numbers on dot plots correspond to the percentage of cells in the gated population, mean of 10,000 events.
3.2.5 – Insulin-specific T cells in an HLA-A2+ T1D patient recognise EBV

Thus far, I have demonstrated that the insulin-specific InsB4 T cell clone, isolated from a patient with T1D could respond strongly to epitopes derived from common human viruses. I next aimed to determine whether this crossreactivity might occur in other patients with T1D. Fortunately, I had a PBMC sample taken from a T1D patient, NDB119, taken just prior to the March COVID-19 lockdown.

Extracted lymphocytes from patient NDB119 were subjected to magnetic separation with HLA-A2-HLVEALYLV tetramer, as described in Section 2.5.5. HLA-A2-HLVEALYLV tetramer+ cells were expanded (Section 2.5.6) prior to tetramer staining and analysis by FACS as shown in Figure 3.16B. 8% of the resulting cell population could bind the HLA-A2-HLVEALYLV tetramer and, notably, 1.7%, 1.3%, and 1.0% of the CD8+ T cells isolated with the HLA-A2-HLVEALYLV tetramer could be bound by HLA-A2-LLIEGIFFI (EBV), HLA-A2-ILIEGIFFY (chickenpox virus), and HLA-A2-ILIEGIFFA (HSV-1) tetramers respectively. These data suggest that some insulin-specific T cells from patient NDB119 might bind the three viral epitopes that acted as strong agonists for the InsB4 T cell clone that had been isolated from different patient with T1D. Confirmation of crossreactivity will require TCR sequencing of the tetramer+ cells to show that identical TCRs are present in each respective tetramer+ population. Unfortunately, I was unable to sequence the TCRs from these T cell populations, but they have been stored and I hope to examine them fully in the future.

In order to extract maximum data from the NDB119 PBMC, I also stained them with the HLA-A2-LLIEGIFFI (EBV) tetramer after first removing the insulin-specific T cell population. After expansion of these cells over 50% of the resultant T cell line stained with HLA-A2 EBV tetramer. Over 30% of the cells from the expanded NDB119 line also stained with HSV-1 and chickenpox virus tetramers suggesting that many, but not all, EBV-specific T cells can respond to peptide sequences from other herpesviruses. This EBV-specific T cell line that had been grown after prior removal of insulin-specific T cells did not stain well with the HLA-A2-HLVEALYLV (insulin) tetramer. Taken together, this experiment shows that a further T1D patient has EBV-specific T cells that may crossreact with an insulin epitope. Future experiments could aim to sequence the TCRs from the tetramer+ T cells to confirm this potential crossreactivity.
Figure 3.16: Insulin-specific cells in an HLA-A2+ T1D donor can recognise three viral epitopes. A) A schematic overview of the procedure used in this work. B) HLA-A2-HLVEALYLVs from the NDB119 donor were isolated using magnetic separation and expanded. Following expansion 50,000 T cells were stained with 0.5 μg of various tetramer (relative to the pMHC component), with a PE-conjugated backbone. Numbers on dot plots represent percentage of tetramer-positive cells in the 10,000 events recorded. C) HLA-A2-HLVEALYLVs from the NDB119 donor were subject to HLA-A2-LLIEGIFFI tetramer binding and subsequent magnetic separation. HLA-A2-LLIEGIFFI cells were then expanded. Following expansion 50,000 T cells were stained with 0.5 μg (relative to pMHC component) of various tetramers, with PE-conjugated backbones. Numbers on dot plots represent percentage of tetramer-positive cells in the 10,000 events recorded. HLA-A2-SLYNTVATL tetramers used as a negative control.
Following my return to work after the summer 2020 lockdown I was unable to access to any more freshly donated samples from patients with T1D. I was, however, fortunate to benefit from three frozen PBMC samples from healthy donors, and a frozen T1D PBMC sample (LSWAIST 075AS) kindly donated by Professor Susan Wong. I also reasoned that if the EBV/insulin crossreactivity could be responsible for triggering disease, T cells specific for the EBV-derived peptide LLIEGIFFI derived from healthy donors would not cross-recognise the insulin-derived epitope HLVELYLV.

FACS-based cell sorting, performed by Dr Cristina Ruis Rafael, was used to isolate CD8+ T cells specific for the HLA-A2-LLIEGIFFI (EBV) tetramer from these PBMC samples. Following expansion of these HLA-A2-LLIEGIFFI+ cells I showed that the resulting cell populations from all three healthy donors (BB14, BB16, and BB19) did not stain with HLA-A2-HLVELYLV tetramer Figure 3.17A. Thus, I was unable to detect any insulin crossreactivity within HLA-A2-LLIEGIFFI-specific T cell populations from three healthy donors.

I next examined a frozen T1D PBMC sample, kindly donated by Professor Susan Wong, LSWAIST 075AS. Only 5.1% of the EBV tetramer sorted T cells stained with the HLA-A2-LLIEGIFFI tetramer (compared to >75% with all the healthy donor lines) (Figure 3.17B). This T1D line did not stain with the HLA-A2-HLVELYLV tetramer, even when a larger number of events were collected.

Overall, I have examined the peptide recognition landscape of an insulin-specific T cell clone isolated from a patient with T1D (Theaker et al., 2016) and proved that this T cell can strongly respond to genuinely processed epitopes from three herpesviruses that commonly infect children and/or adolescents. This same crossreactivity was then observed in the fresh PBMC from a further donor (Figure 3.16). EBV/insulin crossreactivity was not observed in three healthy donors, nor was it observed in a further HLA-A2+ T1D patient (Figure 3.17). It is interesting that I found EBV/insulin crossreactivity in two T1D donors but not a third as these data highlight the heterogeneity of T1D. Several β cell-specific epitopes have potential to be involved in triggering disease and it is likely that the initial causal trigger differs between T1D patients.
Figure 3.17: EBV/Insulin B chain crossreactivity is not present in healthy donors. A) Frozen PBMC samples from healthy donors were subject to HLA-A2-LLLEGIFFI tetramer-based cell sorting and subsequent expansion. Following this 50,000 T cells were stained with 0.5 μg of various tetramer (relative to the pMHC component), with a PE-conjugated backbone. Numbers on dot plots represent percentage of tetramer-positive cells in the 10,000 events recorded. B) A frozen PBMC sample from a T1D patient was subject to HLA-A2-LLLEGIFFI tetramer-based cell sorting and subsequent expansion. Following this 50,000 T cells were stained with 0.5 μg of various tetramer (relative to the pMHC component), with a PE-conjugated backbone. Numbers on dot plots represent percentage of tetramer-positive cells in the 100,000 events recorded.
3.3 – Discussion

3.3.1 – The InsB4 T cell kills human pancreatic β cells

In this chapter, I aimed to investigate the crossreactive nature of a previously-identified T1D-associated T cell clone following the experimental procedure depicted in Figure 3.3 (Pinkse et al., 2005; Theaker et al., 2016). First, my laboratory confirmed that the InsB4 T cell clone isolated from a patient with T1D (Theaker et al., 2016) recognised the HLVEALYLV insulin-derived peptide in the context of HLA-A2. The clone produced MIP-1β in response to the peptide presented by HLA-A2+ APCs, cytokine release was dose-dependent (Figure 3.4A). The InsB4 T cell clone also stained well with HLA-A2-HLVEALYLV PE-conjugated dextramers (Figure 3.4B). HLA-A2+ K562 cells were transduced with a PPI expression cassette to create artificial – or surrogate – pancreatic β cells. PPI-transduced K562 cells were lysed by InsB4 but not an irrelevant CMV-specific control T cell clone (Figure 3.5A). InsB4 T cells lysed K562 target cells transduced with PPI; they did not lyse untransduced cells or those transduced to express with an irrelevant protein (Figure 3.5B). Encouraged by the killing of surrogate β cells, we next demonstrated the ability of InsB4 to kill real human β cells (Figure 3.5C). Thus, the InsB4 T cell clone responds to the bona fide insulin B chain-derived, HLA-A2-restricted epitope, HLVEALYLV. Importantly, this T1D-patient derived T cell clone is capable of killing human pancreatic β cells and could thus contribute to the pathological loss of β cells characteristic of T1D.

3.3.2 – Molecular mimicry may trigger autoimmune disease

The HLA-A2-HLVEALYLV epitope was first described in 2005 when Pinkse et al. demonstrated the presence of HLA-A2-HLVEALYLV-specific CD8+ T cells in both T1D patients and healthy donors (Pinkse et al., 2005). Clearly, therefore, in some situations CD8+ T cells that have the capacity to recognise self-antigens remain inert and do not kill pancreatic β cells. In 2019, Cullina et al. reported that CD8+ T cells specific for a T1D-associated HLA-A2-presented epitope (ZnT8186-194) circulated at similar frequencies in healthy controls and T1D patients, but were enriched in the pancreases of T1D patients (Culina et al., 2018). The ability of these autoreactive T cells to remain inactive in healthy donors suggests the presence of autoreactive T cells and the auto-epitope alone is insufficient to trigger the response to β cells.

In 1985, Fujinami and Oldstone demonstrated that myelin basic protein (MBP) shared six consecutive amino acids with hepatitis B virus polymerase (HBVP) and that rabbits given injections of 8-mer and 10-mer HBVP peptides encompassing this region developed histology.
reminiscent of experimental allergic encephalomyelitis (Fujinami and Oldstone, 1985). Subsequent studies showed that the activation and clonal expansion of autoreactive lymphocytes was a critical step in the pathogenesis of autoimmune disease. Activated T cells but not resting naïve T cells could transfer autoimmunity in experimental models (Zamvil and Steinman, 1990). Microbially-derived peptides were shown to have sufficient homology with self-peptides to be able to activate autoimmune T cells (Wucherpfennig and Strominger, 1995; Zhao et al., 1998; Lang et al., 2002); a mechanism now referred to as molecular mimicry. The concept of molecular mimicry suggests that harmless naïve T cells that possess a TCR capable of binding to a self-antigen can be activated by pathogen-derived peptides with similar sequence (as depicted in Figure 3.1).

3.3.3 – InsB4 can be activated by pathogen-derived peptides
We set out to identify a potential ‘triggering’ epitope for InsB4. After confirming the peptide length preference (Ekeruche-Makinde et al., 2013) of InsB4 (9-mers, see Figure 3.6A), CPL screening was performed to investigate the peptide recognition repertoire of the clone (Figure 3.6B). These data (summarised in Figure 3.6C) demonstrated that at several peptide positions the index amino acids, HLV, are not the preferred amino acids. Positions 1, 5, 6, and 7 in particular showed strong preference for non-index amino acids. This observation supports the possibility that alternative peptides could act as ligands for InsB4. Ligands which I then endeavored to find using CPL-driven database searching (Szomolay et al., 2016). I was thus able to identify fungal, bacterial, and viral peptides that proved to be more potent antigens than the index HLV peptide (Figures 3.7, 3.8, and 3.11 respectively). This observation supports the hypothesis that InsB4 is initially activated by more potent pathogenic ligands displayed on ‘professional’ APCs that can generate the ‘second signal’ required to activate naïve T cells as depicted in Figure 3.1. In the absence of such molecular mimicry, autoreactive T cells can remain naïve in healthy donors (Pinkse et al., 2005). The pathogenic peptides described throughout this chapter are shown in Table 3.1. Viral peptides derived from viruses that commonly infect children and adolescents acted as very strong agonists of the InsB4 T cell. These peptides were of particular interest given that one recent study linked EBV to T1D (Harley et al., 2018). Furthermore, the three virus-derived peptides I identified here by unbiased CPL screening have previously been associated with each other. Chiu et al. endeavoured to identify HLA-A2-restricted immunodominant epitopes for the CD8\(^+\) T cell response to chickenpox virus. In doing so they not only isolated the ILIEGIFV peptide (derived from chickenpox virus ribonucleotide reductase) but found
homologs in other herpesviruses (\texttt{ILIEGIFFA} from HSV-1 ribonucleotide reductase and \texttt{LLIEGIFFI} from EBV BARF-1) and demonstrated the presence of CD8$^+$ T cell specificity to all three epitopes in PBMC of healthy subjects who had previously developed chickenpox virus, regardless of their EBV- or HSV-1-infection status, (Chiu \textit{et al.}, 2014). These data, particularly that from patients naïve to EBV or HSV-1, is suggestive of crossreactivity to these three epitopes.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Origin</th>
<th>Parental Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLVEALYLV</td>
<td>Homo sapiens</td>
<td>Insulin B chain</td>
</tr>
<tr>
<td>ILIEGIFFY</td>
<td>Chickenpox</td>
<td>Ribonucleotide reductase</td>
</tr>
<tr>
<td>ILIEGIFFA</td>
<td>HSV-1</td>
<td>Ribonucleotide reductase</td>
</tr>
<tr>
<td>LLIEGIFFI</td>
<td>EBV</td>
<td>BARF-1</td>
</tr>
<tr>
<td>MIVENVPLL</td>
<td>\textit{Candida albicans}</td>
<td>Putative protein</td>
</tr>
<tr>
<td>MLLENGLIA</td>
<td>\textit{Helicobacter pylori}</td>
<td>Putative protein</td>
</tr>
</tbody>
</table>

Table 3.1: The pathogenic peptides used throughout Chapter 3. Underlining indicates amino acid is identical to the index insulin-derived peptide sequence.

3.3.4 – Viral epitopes strongly activate InsB4 T cells
Using data from 27,000 newly diagnosed T1D patients, and 107,000 healthy controls Lai \textit{et al.} demonstrated that T1D patients have a 1.17-fold increase risk of developing shingles. Shingles occurs when the dormant varicella-zoster virus (chickenpox virus) reactivates in later life (Lai, Lin and Liao, 2019). This propensity could arise because the virus is more likely to have the opportunity to reactivate in an environment of immune dysfunction, or because T1D patients have an increased likelihood of harboring the virus compared to healthy controls. A number of other studies describe similar associations between T1D and shingles (Heymann \textit{et al.}, 2008; Okamoto \textit{et al.}, 2009; Suaya \textit{et al.}, 2014; Muñoz-Quiles \textit{et al.}, 2017; Queenan \textit{et al.}, 2018).
These intriguing associations, along with the fact that the viral peptides were the most potent of the predicted pathogenic ligands, led me to mainly focus my attention primarily on the viral agonists identified by the CPL screen. The top three viral peptides chosen were considerably stronger agonists than any predicted peptides from the much larger fungal and bacterial proteomics databases. The InsB4 T cell was shown to stain well with HLA-A2 tetramers of all five pathogenic peptides studied (ILIEGIFV; chickenpox virus, ILIEGIFFA; HSV-1, ILIEGIFFI, EBV; MIVENVPLL, C. albicans; and MLLENGLLA, H. pylori, see Table 1). Each putative pathogen-derived epitope stained InsB4 T cells more efficiently than the index HLA-A2-HLVEALYLV insulin sequence as a dextramer (Figures 3.10B and 3.13B). Furthermore, as soluble molecules, each viral agonist was able to bind InsB4 with measurable affinity (Figures 3.10A and 3.13A). The binding affinity between InsB4 and HLA-A2-HLVEALYLV was consistently too weak to be measured by SPR, further supporting the hypothesis that a more potent ligand might initially trigger the T cell clone allowing it to subsequently crossreact with weaker self-ligands. In combination, the data discussed thus far demonstrate that the putative pathogenic agonists both bind to and induce a cytokine response from InsB4, they appear to act as much more powerful agonists than the index insulin-derived sequence used to kill human β cells, HLVEALYLV.

The molecular mimicry hypothesis only remains valid if the identified pathogenic peptides are genuine epitopes that are correctly processed and presented at the cell surface in the context of HLA-A2, and are thus available for scrutiny by T cells like InsB4. To validate these epitopes, I transduced MOLT-3 cells with the parental proteins for each of the three viral peptides, see Table 1, and demonstrate that InsB4 only responds to MOLT-3 cells transduced with both HLA-A2 and one of the relevant viral proteins (Figure 3.14). These results strongly suggest that the peptides identified by CPL-driven database screening are real epitopes, capable of activating InsB4 and potentially triggering downstream autoimmunity through the crossreactive recognition of HLA-A2-HLVEALYLV. Formal confirmation that the pathogenic epitopes are displayed at the surface of infected cells will require peptide elution and mass spectrometry.

EBV-transformed lymphoblastic cell lines (LCLs) are used routinely in the work of my group. Given that InsB4 produced a cytokine response to BARF-1-transduced HLA-A2+ MOLT-3 cells I wondered whether InsB4 would recognise HLA-A2+ LCLs. I demonstrated that InsB4 produces both TNF and CD107a in response to LCLs in an HLA-A2-dependant manor (Figure 3.15). Thus, InsB4 can respond to EBV-infected cells.
3.3.5 – T1D/insulin crossreactivity in T1D patients
I began this project because my laboratory had recently partially characterised an insulin-specific patient-derived T cell clone, InsB4. I thought that InsB4 would provide an opportunity for me to investigate the plasticity of a clinically-relevant TCR-pMHC interaction. I wanted to probe the crossreactive nature of InsB4 as I am particularly interested in the theory of pathogenic autoimmune triggers. I demonstrated that InsB4 could crossreact with three viral epitopes and that it responded to these viral epitopes with substantially greater potency than the insulin epitope, HLVEALYLV (Figures 3.12 and 3.9).

InsB4 was isolated from a T1D patient (Theaker et al., 2016), and my demonstration of viral/autoimmune crossreactivity led me to wonder if similar crossreactivity could be found in other T1D patients. I thus showed that a HLVEALYLV-specific T cell line isolated from diabetic patient, NDB119, could be stained with HLA-A2-LLEIGIFFI tetramers (Figure 3.16). Confirmation of crossreactivity will require sequencing of the TCRs in the tetramer positive population but these data (Figure 3.16) hint at viral/autoimmune crossreactivity similar to that observed for InsB4. I subsequently demonstrated the absence of EBV/insulin crossreactivity in three healthy donors (Figure 3.17A). It is most interesting to me that EBV/insulin crossreactivity was absence from a further, frozen T1D donor sample, LSWAIST 075AS (Figure 3.17B). The underlying causes of T1D, of which I believe pathogenic molecular mimics play a substantial role, are likely to be extremely heterogeneous with many pathogenic epitopes triggering many different autoimmune T cells. In this chapter I propose three viral triggers for T1D but in many patients, LSWAIST 075AS included, disease could be initiated by crossreactivity with pathogenic epitopes hitherto undiscovered. This fascinating potential for a heterogeneous range of triggers for autoimmune diseases like T1D highlights the importance of studying the plasticity of the TCR-pMHC interaction.

Greater understanding of the underlying mechanisms of T1D, particularly the TCR-pMHC interactions involved in disease initiation and propagation, may lead to improved treatment options that somewhat alleviate the life-long dependence on injected insulin experienced by T1D patients. My laboratory is currently working with Dmitriy Chudakov, founder of the TCR database VDJdb (Bagaev et al., 2020) to understand the nature of diabetogenic TCRs. The ultimate aim is to be able to predict disease before the onset of symptoms (which generally only occur once >80% of β cells mass is lost). Loss of β cells is thought to occur over a period of ~3 years prior to the onset of symptoms. Insulin management can somewhat protect any remaining β cells from T cell destruction so early diagnosis is crucial to prognosis in terms of
extending the well-described honeymoon period that follows diagnosis (Abdul-Rasoul, Habib and Al-Khouly, 2006; Fonolleda et al., 2017; Zhong et al., 2020).

Knowing the initial pathogen triggers of T1D could lead to the development of a preventative vaccine (Metzler and Wraith, 1993, 1999; Liu et al., 1998; Burkhart et al., 1999; Streeter et al., 2015). However, it is important to recognise, that there are likely to be many potential triggers (in this chapter alone I propose five pathogenic peptides as potential triggers), and many TCR-pMHC interactions involved. The interactions described in this chapter focus on an insulin B chain-derived peptide, yet there are many other well-described self-antigens derived from proteins such as GAD65 (Lohmann et al., 1994), and ZnT8 (Wenzlau et al., 2007). Furthermore, it is important to recognise that T1D is likely a collection of similar disease states with highly heterogeneous causes, with molecularly mimicry such as that described in this chapter making up only one.

3.3.6 – Summary
In this chapter, I have shown that the T1D-associated InsB4 clone strongly crossreacts with pathogenic ligands from the viral, bacterial, and fungal proteomes. I have demonstrated that the identified viral peptides are genuine epitopes that are processed correctly and presented at the cell surface in the context of HLA-A2. I have also shown that InsB4 is activated by HLA-A2* EBV-transformed cell lines.

I next searched for the same EBV/insulin crossreactivity in other donors. I generated EBV-specific T cell lines from three healthy donors and two further patients with T1D. Insulin-specific T cells could be clearly detected within the EBV-specific T cell line from one T1D patient, NDB119. This crossreactivity was not detected in EBV-specific T cell lines from the other T1D patient (LSWAIST 075AS), nor was it present in three healthy controls suggesting that these T cell lines did not contain EBV/insulin crossreactive TCRs. The absence of EBV/insulin crossreactivity in the LSWAIST 075AS T1D patient highlights the likely highly heterogeneous nature of T1D initiation. There are likely many ways to trigger T1D via the T1D risk allele HLA-A2.

Future work should aim at confirming that individual TCRs in the NDB119 line respond to both EBV and insulin epitopes by sequencing the responsible TCRs or generating T cell clones. Overall, my work suggests that some insulin-reactive T cells in patients with T1D may have initially been triggered and expanded to provide immunity to EBV, chickenpox virus, or HSV-1.
4 – Exploring how mutation of key peptide contact residues in CDR3 impact TCR peptide specificity

4.1 – Introduction

4.1.1 – The 868 TCR

In Chapter 3 I demonstrated that autoreactive T cells can be strongly activated by crossreactive pathogenic peptide epitopes. This highlighted the importance of understanding how such crossreactivities occur and highlighted the benefit that would be gained from being able to predict crossreactivity from known characteristics of TCRs, such as their sequence.

As shown in Chapter 3, CPL screens can be used to find preferred agonists for a given TCR. I reasoned that this ability might allow me to explore how changing key peptide contact residues within a TCR altered its preferred peptide agonists. I hypothesised that altering the amino acids in the CDR3 loops of a well-characterised naturally-occurring TCR would alter its peptide recognition repertoire. In order to make an initial assessment of how amino acid substitutions at key peptide contacts might impact peptide specificity I needed to select a TCR where there was an atomic resolution structure of the TCR in complex with a cognate antigen. The HLA-A2-restricted 868 TCR was selected as this TCR was known to have an unusually high natural affinity for its cognate antigen and other members of my laboratory had recently solved the structure of this TCR with the natural, HIV-derived epitope and several common escape variants (Cole, Fuller, et al., 2017).

The 868 TCR recognises an immunodominant HLA-A2-restricted epitope from HIV-1 (Sewell et al., 1997). It originates from an HLA-A2* HIV-1-infected male, patient 9300868 (Sewell et al., 1997; Wilson et al., 1998). T cells from this patient responded well to the immunodominant HLA-A2-restricted p17 Gag epitope (SLYNTVATL, residues 77-85 of HIV p17 Gag) (Sewell et al., 1997). Using pMHC tetramers and Vβ-specific antibodies, the cytotoxic lymphocytes cultured from this patient were shown to undergo clonal, tetramer-specific expansions in response to SLYNTVATL (Wilson et al., 1998). The TRBV 5-6 βCASSDTVSYEQY TCR β chain was shown to be present in the majority of single cells sorted from the patient PBMC using an HLA-A2-SLYNTVATL tetramer (Wilson et al., 1998). Repeat attempts to generate a T cell clone from this line were unsuccessful and the 868 TCR was instead isolated by phage display of TCR sequences derived from a polyclonal HLA-A2-SLYNTVATL-specific T cell line grown from patient 9300868 (Varela-Rohena et al., 2008). The TCR was shown to
possess a TRAV 12-2 α chain with the CDR3 sequence, CAVRTNSGYALNF (Varela-Rohena et al., 2008; Cole, Fuller, et al., 2017). All HLA-A2-SLYNTVATL-tetramer+ T cells in the 868 SLYNTVATL-specific T cell line stained with antibodies specific for TRAV 12-2 and TRBV 5-6 variable domains (Varela-Rohena et al., 2008). The 868 TCR binds to its cognate HLA-A2-SLYNTVATL antigen with an extremely high affinity for a naturally occurring TCR (K_D = 210 nM at 37°C) (Cole, Fuller, et al., 2017).

The HIV-1 Gag protein from which SLYNTVATL is derived is a large viral precursor protein that is cleaved by viral protease enzymes into proteins required for viral replication. These include p24, the capsid protein, and p17 — or the matrix protein (MA) from which the SLYNTVATL epitope arises (Freed et al., 1994; Cannon et al., 1997). The p17 protein is encoded towards the amino-terminal of the Gag protein (Freed et al., 1994). It is thought to play a number of key roles during viral replication including localising the nascent viral particles to just under the lipid bilayer of the host cell where assembly generally takes place, and aiding the incorporation of Env glycoproteins into the portion of the host bilayer destined to become the new viral bilayer (Freed et al., 1994; Cannon et al., 1997).

The immunodominant response to SLYNTVATL is uncommon during the acute stage of HIV-1 infection and instead dominates during the long chronic infection phase (Goulder, Altfeld, et al., 2001). Responses to HLA-A2-SLYNTVATL occurs in over 70% of chronically infected HLA-A2+ patients, suggesting the epitope is favourably processed and presented at the cell surface of infected cells (Goulder, Sewell, et al., 1997; Brander et al., 1998; Goulder, Altfeld, et al., 2001). Accordingly, patient 9300868 was first diagnosed as HIV-1+ in 1993 but the SLYNTVATL-responsive T cells were isolated from a PBMC sample taken in 1996 when the patient would likely be in the chronic stage of the infection (Sewell et al., 1997).

Over the course of HIV-1 infection, a number of naturally occurring escape mutations arise within the SLYNTVATL epitope in HLA-A2+ individuals (Sewell et al., 1997). The evolution of these viral escape mutants has been well described (Brander et al., 1998; Jamieson et al., 2003; Iversen et al., 2006; Martinez-Hackert et al., 2006; Christie et al., 2009). A 2006 study showed that 58% of HLA-A2+ patients in the chronic phase of infection had at least one example of an SLYNTVATL escape mutation (Iversen et al., 2006). Mutations frequently occur at positions 3, 6, and 8 of the 9-mer epitope, yet rarely occur at positions 2, 9, or 4 (Sewell et al., 1997; Iversen et al., 2006; Tenzer et al., 2009; Cole, Fuller, et al., 2017). The mutations occur accumulatively over the course of the chronic infection (Iversen et al., 2006), with the triple mutant, SLENTIAYL, having been dubbed the ‘ultimate escape mutant’ (Iversen et al., 2006).
In the absence of a T cell selective pressure (e.g. when the virus is transmitted to a HLA-A2 individual) SLYNTVATL-mutated HIV-1 rapidly reverts to the wildtype sequence, suggesting the mutants result in a reduction in viral fitness (Iversen et al., 2006; Tenzer et al., 2009; Cole, Fuller, et al., 2017).

Viral mutations occur in order to escape recognition by host T cells (Sewell et al., 1997). In a 2006 study, 8% of patients did not mount responses to any mutant variant of SLYNTVATL tested at all (Iversen et al., 2006). As well as escaping recognition, mutations can have antagonistic effects (Sewell et al., 1997), in which they bind the TCR yet do not fully trigger downstream signalling – this occupancy but failure to activate the TCR prevents the TCR from encountering other variants that could potentially induce full activation (Purbhoo et al., 1998).

A common viral escape strategy is mutation of the residues that anchor the peptide to MHC so that the peptide is no longer presented at the surface of the infected cell for T cell inspection (Morikawa et al., 1998; Kelleher et al., 2001; Altfeld et al., 2005; Varela-Rohena et al., 2008). The primary anchor residues for HLA-A2 are the p2 and C-terminal residues of the peptide. MHC anchor mutations rarely arise in SLYNTVATL (Sette et al., 1994; Lee et al., 2004; Iversen et al., 2006) as the amino acids at peptide positions 2 and 9 are crucial to the correct formation of the p17 protein, meaning that mutation would have a detrimental impact on viral fitness (Freed et al., 1994; Cannon et al., 1997; Morikawa et al., 1998; Ono, Orenstein and Freed, 2000; Martinez-Hackert et al., 2006). It was therefore hypothesised that the common viral escape strategy at positions 3, 6, and 8 must have an impact on TCR binding (e.g. these positions must form key contacts with host TCRs) (Iversen et al., 2006).

Cole et al. found that, in fact, these residues were not the key contact residues. Structural analysis of the 868 TCR in complex with HLA-A2-SLYNTVATL and its escape variants demonstrated that most TCR contacts instead occur with the 4N peptide residue (Cole, Fuller, et al., 2017). This residue was noted in 1997 to rarely be mutated in SLYNTVATL escape variants (Sewell et al., 1997). Cole et al. demonstrated that immune escape is achieved by destabilising the pMHC structures on the surface of the cell over a long-term period (24 hours), a mechanism that had been missed in previous standard assays over a shorter duration (<7 hours) (Cole, Fuller, et al., 2017). This unexpected mechanism of immune escape highlights the danger of making assumptions about how TCRs bind in the absence of an atomic resolution structure and highlights the importance of perusing greater understanding of the molecular rules that govern the TCR-pMHC interaction.
The 868 TCR is restricted to HLA-A2, the most common HLA allele in the western population (Nunes et al., 2014; Gourraud et al., 2015) and ~75% of infected, HLA-A2+ individuals mount a response against HLA-A2-SLYNTVATL (Goulder, Sewell, et al., 1997; Brander et al., 1998; Goulder, Altfeld, et al., 2001; Iversen et al., 2006; Varela-Rohena et al., 2008). The interaction has been extensively studied and the crystal structure of the 868-A2-SLYNTVATL interaction, as well as complex structures of interactions with escape mutants have been solved (Cole, Fuller, et al., 2017). The TCR has already been subject to some degree of mutational studies. Varela-Rohena et al. introduced amino acid substitutions into the CDR3 loops of the 868 TCR and used a phage display system to identify variants with both enhanced affinity for the wildtype epitope and broader recognition of escape variants (Varela-Rohena et al., 2008). 868-A2-SLYNTVATL therefore makes an excellent model system with which to expand upon this work and investigate the influence of CDR3 sequence on peptide preference.

4.1.2 – Probing TCR ‘rules of engagement’

The sequences of TCR CDR loops are largely responsible for peptide recognition. Collectively, the six CDR loops form the antigen binding site; with the CDR3s of each chain being responsible for most – but not all – of the peptide contacts (Garboczi et al., 1996; Garcia et al., 1996; Ding et al., 1998; Hausmann et al., 1999; Cole et al., 2009). Subtle differences in the CDR loops of a TCR can alter its antigen specificity (Yang et al., 2011; Chen et al., 2017). For example, in 2013, Motozono et al. used a collection of patient-derived T clones specific for the immunodominant HIV-1 Nef epitope, VY8 (VPLRPMTY) presented by HLA-B*35:01 to show that, despite their shared recognition of this epitope, their individual peptide recognition ‘footprints’ varied considerably (Motozono et al., 2013). The ability to recognise HIV-1 escape mutants of this epitope varied from clone to clone (Motozono et al., 2013). In fact, in a similar experimental set up to that used in this chapter, the same group later used CPL screens to profile the crossreactive nature of T cells specific for the VY8 epitope (Motozono, Bridgeman, et al., 2015). Motozono et al. showed that single amino acid differences in the CDR3s could alter the crossreactive footprints of the T cell clones (Motozono, Bridgeman, et al., 2015).

In 2014, Smith et al. mutated CDR loop residues to successfully alter the specificity of the A6 TCR (Smith et al., 2014). The A6 TCR was chosen as, like the 868 TCR, it is structurally and biochemically well characterised (Garboczi et al., 1996; Ding et al., 1999; Borbulevych et al., 2009). The A6 TCR recognises the HTLV-1-derived Tax peptide (LLFGYPVYV) (Smith et al.,
Smith et al. aimed to re-direct the A6 TCR to recognise the unrelated cancer-associated MART-1 peptide (ELAGIGILTV) (Kawakami et al., 1994). These The LLFYPVVY and ELAGIGILTV peptides differ at every residue save the HLA-A2 anchor residues (P2 and the C-terminus of the peptide). In an example of using structural data to guide rational mutagenesis, they used molecular modelling of the published A6 structure to predict which CDR loop residues would most likely accommodate the MART-1 peptide. With just five amino acid substitutions they redirected the A6 TCR to see the desired peptide, completely ‘switching’ the specificity of the TCR (Smith et al., 2014). The complex structure of the mutant A6 (known as RD1) with the MART-1 epitope was subsequently solved in 2016 (Harris et al., 2016). It showed that the small amino acid substitutions that were introduced resulted in a significant alteration in binding mode, affecting interactions throughout the interface (Harris et al., 2016). Despite the altered binding mode, the re-directed TCR maintained the ability to trigger T cell activation in an antigen-dependent manner (Smith et al., 2014; Harris et al., 2016).

The crossreactive profile of the RD1 TCR was recently compared to a MART-1-specific TCR clone that had been isolated directly from a polyclonal population of melanoma-reactive T cells and subject to standard, random mutation-based, affinity maturation, known as T1 (Smith et al., 2013; Sharma et al., 2019). Both TCRs saw MART-1 comparably, but RD1 was less crossreactive than the randomly affinity-enhanced TCR which recognised a number of self-peptides. This study suggests that rationally designed TCRs can be functionally comparable but carry less potential for off-target crossreactivity than TCRs engineered by random mutagenesis methods (Sharma et al., 2019).

4.1.3 – Aims
TCR CDR loop sequences are paramount to function and peptide specificity. Mutation of just a few residues can enhance, alter, or – as demonstrated by Zhou et al. – completely abrogate function (Zhou et al., 2017). In this chapter, I investigated the impact of CDR3 amino acid substitutions on the function of the 868 TCR. In a manner akin to that described by Babad et al. in 2015, I used lentiviral vectors to ‘reprogram’ primary human CD8+ T cells with 868 TCR or a mutant version of this TCR (Babad et al., 2015). Babad et al. demonstrated that viral vectors containing TCR structures separated by 2A peptides, and produced using HEK 293T cells can efficiently confer antigen-specificity to primary polyclonal CD8+ T cells as measured by antigen-dependent cytokine release and cytotoxic activity (Babad et al., 2015).
Using transduced CD8+ T cells, I subjected the 868 TCR, and mutants thereof, to CPL-driven database searching as detailed in Section 3.1 first described by Szomolay et al. (Szomolay et al., 2016) and later refined by Whalley et al. (Whalley et al., 2020). Raw data from CPL screens were used to search proteomic databases and generate a ranked list of peptides predicted to be suitable ligands for a given TCR. This approach can also be used to rank potential TCR agonists from the entire peptide universe so as to predict the optimal epitopes for a given TCR. In order to begin exploring whether useful information regarding the relationship between sequence and function could be gained by this approach, I compared the CPL data generated with mutated 868 TCRs with that of the native parent TCR.

The specific aims of this project were:

- Use a lentiviral transduction vector, pELNS, to stably transduce the 868 TCR into primary CD8+ T cells.
- Investigate the crossreactive nature of the 868 TCR using these T cells and peptide length-matched CPL screens.
- Design lentiviral constructs containing mutated versions of the 868 TCR and use primary CD8+ T cells transduced with these constructs to assess the impact of CDR3 amino acid substitutions on peptide recognition.
4.2 – Results

4.2.1 – Designing the mutant 868 TCRs

Mutant constructs of the 868 TCR were designed based on structural data of the 868-A2-SLYNTVAL interaction (Figure 4.1A) (Cole, Fuller, et al., 2017). Figures 4.1B and 4.1C demonstrate how the SLYNTVAL peptide is presented in the groove of HLA-A2 for inspection by the 868 TCR. This structure, along with the accompanying table of contacts (Cole, Fuller, et al., 2017), reproduced in Table 4.1, was used to identify key contact residues responsible for the SLYNTVAL-specificity. As my aim was to alter the specificity of the TCR, key contact residues were substituted. It was assumed that substitutions to amino acids that possessed similar chemical properties (for example, glycine to alanine) might be well tolerated and therefore produce little change in peptide specificity so amino acid substitutions that radically changed the chemical properties or size of the residue were selected, for example replacing a polar amino acid with a hydrophobic amino acid, Figure 4.2. In total, 24 mutant versions of the 868 TCR were designed and synthesised (Table 4.2). Each mutant required cloning into the pELNS vector before analysis could commence, the progress of each mutant is indicated by grey shading in Table 4.2. Screening each TCR took a considerable amount of time, so I was only able to examine a few mutant TCRs during my PhD studies as my laboratory studies were interrupted by the March 2019 COVID-19 lockdown.

The first four mutant 868 TCRs that are discussed in this chapter, along with 868 itself, are shown in Figure 4.3A, each has a substitution or substitutions in the CDR3β loop. The TCRs, containing βT96F, βV97T, βD95S, and βD95I β96R V97S substitutions, are denoted TCR 3, TCR 6, TCR 22, and TCR 24 respectively, Table 4.2. The amino acid substitutions in each TCR are rationalised below.
Figure 4.1: The interaction between 868 and HLA-A2- Slyntvatl. A) The 868-A2-Slyntvatl interaction as solved by Cole et al. in 2017, PDB: 5NMD. B) shows a close up of the interface, demonstrating how 868 CDR loops interact with the Slyntvatl peptide. C) the Slyntvatl peptide is presented in a standard manner for MHC class I peptides, anchored at either end with a protruding centre.
<table>
<thead>
<tr>
<th>CDR loop</th>
<th>TCR residue</th>
<th>Peptide residue</th>
<th>MHC residue</th>
<th>vDW ≤4 Å</th>
<th>Electrostatic bonds ≤3.4 Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDR3α</td>
<td>Arg92</td>
<td>Asn4</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Arg92</td>
<td>Thr5</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thr93</td>
<td>Asn4</td>
<td></td>
<td>1</td>
<td>1 HB</td>
</tr>
<tr>
<td></td>
<td>Asn94</td>
<td>Lys66</td>
<td></td>
<td>3</td>
<td>1 HB</td>
</tr>
<tr>
<td></td>
<td>Asn94</td>
<td>Leu2</td>
<td></td>
<td>1</td>
<td>1 HB</td>
</tr>
<tr>
<td></td>
<td>Asn94</td>
<td>Tyr3</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asn94</td>
<td>Asn4</td>
<td></td>
<td>4</td>
<td>2 HB</td>
</tr>
<tr>
<td></td>
<td>Ser95</td>
<td>Gly62</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ser95</td>
<td>Arg65</td>
<td></td>
<td>2</td>
<td>1 HB</td>
</tr>
<tr>
<td></td>
<td>Gly96</td>
<td>Arg65</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gly96</td>
<td>Asn4</td>
<td></td>
<td>2</td>
<td>1 HB</td>
</tr>
<tr>
<td></td>
<td>Tyr97</td>
<td>Lys66</td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tyr97</td>
<td>Lys69</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tyr97</td>
<td>Asn4</td>
<td></td>
<td>7</td>
<td>1 HB</td>
</tr>
<tr>
<td>CDR3β</td>
<td>Thr96</td>
<td>Ala7</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thr96</td>
<td>Thr8</td>
<td></td>
<td>2</td>
<td>1 HB</td>
</tr>
<tr>
<td></td>
<td>Val97</td>
<td>Lys146</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Val97</td>
<td>Trp147</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Val97</td>
<td>Ala150</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Val97</td>
<td>Val152</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Val97</td>
<td>Ala7</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Val97</td>
<td>Thr8</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ser98</td>
<td>Ala150</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tyr99</td>
<td>Thr5</td>
<td></td>
<td>15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tyr99</td>
<td>Val6</td>
<td></td>
<td>6</td>
<td>2 HB</td>
</tr>
</tbody>
</table>

Table 4.1: Contact table for the 868-A2-SLYNTVATL interaction. Reproduced from (Cole, Fuller, et al., 2017).
Figure 4.2: The classification of amino acids based on chemical characteristics. In this work amino acid substitutions were generally made to an amino acid of differing chemical property in the hope of altering the peptide preference of the 868 TCR.
Table 4.2: The TCRs used in Chapter 4. Each TCR was given an identification number to prevent mix ups. The CDR3 loop mutation is indicated by underlining. Grey shading indicates the progress with each TCR.

<table>
<thead>
<tr>
<th>TCR mutation</th>
<th>ID</th>
<th>CDR3 sequences</th>
<th>Cloned into pELNS</th>
<th>Expressed in CD8+ T cells</th>
<th>CPL</th>
<th>Specificity testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>808 wt</td>
<td>(TCR 2)</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>βν96F</td>
<td>TCR 3</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>βν97R</td>
<td>TCR 4</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>βν97D</td>
<td>TCR 5</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>βν97T</td>
<td>TCR 6</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>βν102W</td>
<td>TCR 7</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>αν94W</td>
<td>TCR 8</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>αν92E</td>
<td>TCR 9</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>Partially complete</td>
</tr>
<tr>
<td>αν92L</td>
<td>TCR 10</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>αν93V</td>
<td>TCR 11</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>Partially complete</td>
</tr>
<tr>
<td>αν95V</td>
<td>TCR 12</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>αν96S</td>
<td>TCR 13</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>αν96D</td>
<td>TCR 14</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>αν96K</td>
<td>TCR 15</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>αν97I</td>
<td>TCR 16</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>αν91C</td>
<td>TCR 17</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>αν91W</td>
<td>TCR 18</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>αα98R</td>
<td>TCR 19</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>αα98E</td>
<td>TCR 20</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>βδ95I</td>
<td>TCR 21</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>βδ95S</td>
<td>TCR 22</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>βδ95I T96R</td>
<td>TCR 23</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>βδ95I T96R V97S</td>
<td>TCR 24</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>βδ95S T96R</td>
<td>TCR 25</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>βδ95S T96R V97S</td>
<td>TCR 26</td>
<td>α: CAVTRNSGYALNF β: CASSDTVSYEQY</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>
4.2.1.1 – TCR 3, βT96F
TCR 3 contains a mutation in the 96\textsuperscript{th} residue of the 868 β chain, βT96. In the wildtype TCR the threonine at this position makes two peptide contacts with SLYNTVATL residues, alanine (P7) and threonine (P8). The contact between βT96 and the threonine at P8 of the peptide is a hydrogen bond (defined according to the Cole \textit{et al.} table of contacts as having a bond distance of ≤3.4 Å). βT96, a relatively small polar amino acid, was mutated to phenylalanine, a bulky hydrophobic amino acid with a large aromatic ring. The modelled βT96F mutation is shown in \textbf{Figure 4.3B}.

4.2.1.2 – TCR 6, βV97T
TCR 6 contains a mutation in the residue adjacent to that mutated in TCR 3. Residue 97 in the 868 β chain is a valine, βV97. It makes two peptide contacts with the same two SLYNTVATL residues as βT96, alanine (P7) and threonine (P8). βV97 makes 4 additional contacts with the HLA-A2 molecule, \textbf{Table 4.1}. βV97, a hydrophobic amino acid, was mutated to threonine, a polar (hydrophilic) amino acid. The modelled βV97T mutation is shown in \textbf{Figure 4.3C}.

4.2.1.3 – TCR 22, βD95S
TCR 22 contains a mutation at the 95\textsuperscript{th} residue in the 868 β chain. According to the contact table published by Cole \textit{et al.} this residue does not form any contacts with the HLA-A2-SLYNTVATL target at all (Cole, Fuller, \textit{et al.}, 2017). This residue was selected for mutation to evaluate the impact of potential knock on effects that a non-contact residue might have on the recognition footprint of the TCR. In the wildtype sequence the 95\textsuperscript{th} residue is a negatively charged aspartic acid residue βD95, for this study it was mutated to a polar, uncharged, serine βD95S (\textbf{Figure 4.3D}).

4.2.1.4 – TCR 24, βD95I T96R V97S
Finally, TCR 24 contains three amino acid substitutions at a stretch of three consecutive 868 β chain residues, 95, 96, and 97. The wildtype sequence DTV was mutated to IRS (βD95I T96R V97S), \textbf{Figure 4.3E}. The βD95I mutation involves the replacement of a negatively charged amino acid for a hydrophobic amino acid. The βT96R mutation involves the replacement of a polar, uncharged amino acid with a positively charged amino acid, and the βV97S mutation involves the replacement of a hydrophobic amino acid with a polar amino acid. The model
of this triple mutation suggests a substantial structural rearrangement, Figure 4.3E. The amino acid substitutions chosen for TCR 24 were based on the published structure of a HLA-A2-restricted, influenza-specific TCR, JM22 (PDB:5HHO) (Valkenburg et al., 2016). The JM22 TCR recognises an epitope from the M1 influenza protein (GILGFVFTL, residues 58-66). The TCR consists of TRAV27 and TRBV19 chains (Valkenburg et al., 2016), in contrast to the TRAV12-2 and TRBV5-6 chains of the 868 TCR (Varela-Rohena et al., 2008). Despite this difference the CDR3β sequences differ by just three residues (CASSDTVSYEQYF; 868 and CASSIRSSYEQYF; JM22) (PDB:5HHO) (Valkenburg et al., 2016; Cole, Fuller, et al., 2017). Thus, the three amino acid substitutions in TCR 24 construct are sufficient to make the CDR3β sequence of 868 identical to that of JM22, whilst other factors – such as variable genes and the CDR3α sequence remain native to the 868 TCR. The similarity of the JM22 TCR was noted by Dr Thomas Whalley. We wondered whether these three amino acid substitutions alone would be sufficient to redirect the 868 TCR to recognising the M1 peptide. It is noteworthy that the SLYNTVATL and M1 peptides contain a VxTL motif at the C-terminus, whilst differing at all other residues (Valkenburg et al., 2016; Cole, Fuller, et al., 2017). According to the 868-A2-SLYNTVATL structure published by Cole et al., three of the six key contact points between 868 CDR3β and the SLYNTVATL peptide are within the VxTL motif (Cole, Fuller, et al., 2017). The βD95I T96R V97S triple mutation was, however, insufficient to redirect TCR 24 to the GILGFVFTL peptide (see Supplementary Figure 7.1), indicating that the other CDR loops play an important role in recognition of the influenza epitope by the JM22 TCR.
**Figure 4.3:** The 868 mutants used in this chapter. **A)** shows the CDR3 sequences of each of the first 5 TCR constructs used in this chapter, including the wildtype 868 sequence. All amino acid substitutions were in the CDR3 β sequence and are underlined. **B)** Model prediction of the βT96F substitution in TCR 3. **C)** Model prediction of the βV97T substitution in TCR 6. **D)** Model prediction of the βD95S substitution in TCR 22. **E)** Model prediction of the βD95I T96R V97S substitution in TCR 24.

<table>
<thead>
<tr>
<th>TCR</th>
<th>Mutation</th>
<th>CDR3 Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>868</td>
<td>None</td>
<td>αCAVRNSGYALNF βCASSDTVSYEQY</td>
</tr>
<tr>
<td>TCR 3</td>
<td>βT96F</td>
<td>αCAVRNSGYALNF βCASSDTVSYEQY</td>
</tr>
<tr>
<td>TCR 6</td>
<td>βV97T</td>
<td>αCAVRNSGYALNF βCASSDTTSYEQY</td>
</tr>
<tr>
<td>TCR 22</td>
<td>βD95S</td>
<td>αCAVRNSGYALNF βCASSDTVSYEQY</td>
</tr>
<tr>
<td>TCR 24</td>
<td>βD95I T96R V97S</td>
<td>αCAVRNSGYALNF βCASSDTTVSYEQY</td>
</tr>
</tbody>
</table>
4.2.2 – Validation of 868 mutants
As described in Section 2.3.2, the mutated TCR constructs were separated by 2A self-cleaving peptides to facilitate equal levels of expression (Figure 2.5). A rCD2 marker gene was also included to enable easy monitoring of cellular transduction by antibody staining. To ensure that my mutant TCRs were adequately expressed at the cell surface I first validated expression via lentiviral transduction in the T cell leukaemia line, Jurkat (clone E6.1). These readily available, immortalised CD4+ T cells were chosen for initial testing due to their ease of growth in culture compared to primary CD8+ T cells. This initial round of validation also served to prevent wastage of donated blood-derived primary CD8+ T cells on potentially unviable TCR constructs.

As described in Section 2.4.2, viral vectors were used to deliver the TCR and marker genes via spinfection. Figure 4.4A shows the surface expression rCD2 and TRBV 5-6 in Jurkat cells transduced with the 868 TCR and the four mutant TCRs described above. All 5 TCRs were successfully expressed and trafficked to the cell surface, with varying degrees of transduction efficiency ranging from 18% to 78.1% transduction (TCR 3 and 868 respectively), Figure 4.4A. Polybrean is frequently used to increase uptake of viral vectors. Polybrean was not used during initial testing in Jurkat cells as attempts to transduce with polybrean led to large amounts of Jurkat cell death and exceedingly slow recovery of cell numbers for staining. It was reasoned that, at this stage, validation of the expression of each construct required a yes/no answer rather than high transduction efficacy.

An additional validation step involved staining 868-transduced Jurkat cells with HLA-A2- SLYNTVATL tetramers to determine whether the transduced TCR was functional. 27.6% of rCD2+ 868-transduced Jurkat cells stained with the tetramer (Figure 4.4B), demonstrating specificity of the TCR – even in the absence of the CD8 co-receptor. The same 868-transduced Jurkat cells were then maintained in culture for over 1 month to evaluate the longevity of the transduction. The TCR construct would need to remain expressed in primary T cells for at least 1 month to allow time for purification, expansion, and functional testing. Surface expression was reassessed at the 1-month mark. The construct was still expressed at the surface, but a proportion of the Jurkat cells had lost expression of the transduced TCR (a fall from 78.1% to 26.3% of rCD2+TRBV Jurkat cells, Figure 4.4C, highlighting the need for regular enrichment of construct-positive cells, particularly before any functional testing (see Section 2.4.5).
Following validation of my mutant constructs in Jurkat cells, the lentiviral constructs were transduced into primary CD8\(^+\) T cells (see Section 2.4.4). Each construct was successfully expressed and trafficked to the cell surface with transduction efficiency ranging from 38.9% to 62.5%, Figure 4.4D. Following purification based on the rCD2 surface marker (see Section 2.4.5) surface expression increased to range between 73.8% to 85.4%, Figure 4.4D.

Figure 4.4: Validating expression of the lentiviral constructs. A) Jurkat cells were cultured in 10 mL R10 suspension flasks for 1 week following spinfection. Cells previously gated on live CD3\(^+\) single lymphocytes. Surface expression of lentiviral constructs was monitored based on simultaneous expression of marker gene, rCD2, and TRBV 5-6 chain. B) 50,000 868-transduced Jurkat cells were stained with 0.5 µg HLA-A2-SLYNTVATL-tetramer (relative to the pMHC component), with a PE-conjugated backbone. Cells previously gated on live CD3\(^+\) single lymphocytes. C) 868-transduced Jurkat cells were cultured in 10 mL R10 suspension flasks for 1 month following spinfection. Cells previously gated on live CD3\(^+\) single lymphocytes. Surface expression of lentiviral constructs was monitored based on simultaneous expression of marker gene, rCD2, and TRBV 5-6 chain. D) CD8\(^+\) T cells isolated from donated PBMCs and cultured in transduction media for 1 week prior to flow cytometry analysis. Cells previously gated on CD3\(^+\)CD8\(^+\) single lymphocytes. Surface expression of lentiviral constructs was monitored based on simultaneous expression of marker gene, rCD2, and TRBV 5-6 chain. Numbers on dot plots correspond to the percentage of cells in the gated population. 10,000 events recorded for each plot.
4.2.3 – Minimal differences between the recognition footprints of 868, TCR 3, and TCR 6

TCRs 3 and 6 were the first constructs I successfully cloned into the pELNS vector. Functional testing suggested there was little difference between the peptide recognition footprints of 868 and either TCR 3 or TCR 6. Reactivity to SLYNTVATL, as measured by MIP-1β production, showed no obvious difference between the mutants, although I noted an unusually high level of background MIP-1β production by 868 and TCR 3 in response to the T2 cells used as APCs, Figure 4.5A.

All three TCRs stained with the SLYNTVATL tetramer, Figure 4.5B, suggesting that the amino acid substitutions had not drastically altered TCR specificity for the HIV-1 epitope. Staining of TCR 3 appeared to be slightly reduced (14.1% Tetramer+ rCD2+, compared to 57.9% and 66.3% for 868 and TCR 6 respectively), Figure 4.5B.

The three TCRs were subject to CPL screening (Figure 4.6), but few differences were noted between their reactivity profiles, suggesting that the βT96F and βV97T substitutions were insufficient to alter the TCRs peptide preferences. These amino acid substitutions represent significant alterations in the chemical properties at the respective residue, to retain specificity despite these substitutions demonstrates the remarkable plasticity of the TCR. As such these data were not used to search the proteome databases because finding novel specificities was deemed to be unlikely. These TCRs were not subject to any further functionality testing.
Figure 4.5: TCRs 3 and 6 show little variation in their ability to recognise the cognate HLA-A2-SLYNTVAL epitope. A) 30,000 CD8+ T cells transduced with TCR constructs were cultured in the presence of 60,000 T2 cells as APC and 10^{-5} M SLYNTVAL peptide overnight before the supernatant was harvested for measurement of MIP-1β presence by ELISA analysis. B) 50,000 CD8+ T cells transduced with TCR constructs were stained with 0.5 µg HLA-A2-SLYNTVAL tetramer (relative to the pMHC component), with a PE-conjugated backbone. Cells previously gated on live CD3+CD8+ single lymphocytes. Numbers on dot plots represent percentages of tetramer-positive cells, 10,000 events measured.
Figure 4.6: TCRs 3 and 6 show little variation under CPL screening. 30,000 CD8⁺ T cells transduced with either 868 (green), TCR 3 (pink), or TCR 6 (blue) constructs were co-cultured with 60,000 T2 cells as APCs and 100 µg/mL peptide mix from a 9-mer CPL screen. Supernatants were harvested and ELISA analysis performed to measure MIP-1β presence as a proxy for T cell response. Coloured bars indicate the sequence of the index SLYNTVATL peptide. All assays performed once.
4.2.4 – TCR 22 and TCR 24 show differing specificity than the parent 868 TCR
I next moved on to assessment of TCRs 22 and 24. Both 868- and TCR 22-transduced cells responded comparably to the SLYNTVATL peptide as measured by MIP-1β production. In contrast, TCR 24-transduced cells did not respond to the SLYNTVATL peptide suggesting TCR 24 might have altered specificity (Figure 4.7A). This observation was supported by tetramer staining data showing that 868- and TCR 22-transduced cells can both be stained by HLA-A2- SLYNTVATL tetramers but TCR 24-transduced cells cannot, Figure 4.7B. The three amino acid substitutions in the TCR 24 construct (βD95I T96R V97S) appeared to have eradicated recognition of the index peptide.

As noted above, I again observed an unusually high level of background MIP-1β production by 868 TCR-transduced cells in response to T2 cells using this very sensitive readout. This background recognition was not observed for TCR 22- or TCR 24-transduced cells (Figure 4.7A). As the 868-transduced cells exhibited background activity and cells transduced with TCRs 22 and 24 did not, this provided an opportunity to potentially identify the epitope on T2 cells to which T cells with the 868 TCR appeared to be responding.
Figure 4.7: 868, TCRs 22 and 24 show differences in peptide preference. A) 30,000 CD8+ T cells transduced with TCR constructs were cultured in the presence of 60,000 T2 cells as APCs and 10^5 M SLYNTVATL peptide overnight before the supernatant was harvested for measurement of MIP-1β presence by ELISA analysis. B) 50,000 CD8+ T cells transduced with TCR constructs were stained with 0.5 µg HLA-A2-SLYNTVATL tetramer (relative to the pMHC component), with a PE-conjugated backbone. Cells were previously gated on live CD3+CD8+ single lymphocytes. Numbers on dot plots represent percentages of tetramer-positive cells, 10,000 events measured.
4.2.5 – Identifying the HLA-A2 self-epitope that the 868 TCR responds to
868, TCR 22, and TCR 24 were subject to CPL screening. Aside from a few isolated responses
at peptide positions 1 and 2, TCR 24 showed minimal reactivity at most amino acid positions
(Figure 4.8). Given that the amino acid substitutions in TCR 24 abrogated recognition of
SLYNTVATL is it possible that the substantial structural rearrangement caused by these three
amino acid substitutions (Figure 4.3E) may have interfered with HLA-A2 binding or
completely ablated the function of this TCR.

In contrast, 868 and TCR 22 showed patterns of amino acid preferences at all peptide residue
positions, Figure 4.8. Responses at each position were normalised to allow comparative
assessment of amino acid preference, Figure 4.9A. This visual analysis depicts clear
differences between the preferences of the two TCRs. For example, TCR 22 has preferences
for glutamic acid, isoleucine, and phenylalanine at positions 5, 6, and 8 respectively that are
not shared by the 868 TCR. In contrast, the 868 TCR displays a preference for leucine,
threonine, and proline residues at positions 4, 5, and 6 respectively that are not shared by
TCR 22 (Figure 4.9A).

These data were used to screen proteome databases using a webtool created by Szomolay
et al. (Szomolay et al., 2016). This tool uses CPL data from each peptide position to search
proteomic databases for epitopes that the TCR in question would likely recognise and ranks
them in order of predicted preference. Each TCR was screened against five databases
covering the proteomes of bacterial, fungal, viral, self, and tumour-associated antigens
(TAA). The top scoring peptides overall were collated and are shown, along with the database
from which each peptide originates, in Figure 4.9B. Interestingly, in a combined search of
self, pathogenic and TAA databases, 12 of the top 15 scoring peptides for the wildtype 868
TCR originated in the TAA database.
Figure 4.8: CPL screening of TCR 22 and 24 in comparison to 868. 30,000 CD8\(^+\) T cells transduced with either 868 (green), TCR 22 (orange), or TCR 24 (light blue) constructs co-cultured with 60,000 T2 cells as APC and 100 µg/mL peptide mix from a commercially available 9-mer CPL screen. Supernatants were harvested and ELISA analysis performed to measure MIP-1\(\beta\) presence as a proxy for T cell response. Coloured bars indicate the sequence of the index SLYNTVATL peptide. All assays performed once.
Figure 4.9: 868 and TCR 22 are predicted to have varying peptide preference footprints. A) a comparative overlay of the 868 and TCR 22 CPL data, showing each response as a percentage of the maximum response at that peptide position to highlight variation in amino acid preference. B) data in A were used to search proteomic databases for agonist ligands using a web tool developed by Szomolay *et al*. Said tool generated ranked lists of peptides predicted to be favoured by each TCR.
4.2.6 – 868 TCR recognises epitopes that are not seen by TCR 22
The peptides identified by CPL-based database screening of peptide databases were ordered in crude form (~40 % purity) and cells transduced with the two TCRs were assessed for reactivity against 10⁻⁵ M of each peptide. Notably, TCR 22 saw none of the peptides that were predicted from the screen data, Figure 4.10A. I therefore failed to identify a new specificity for this mutant TCR.

868 recognised two peptides from the predicted peptide lists, one of which was expected to be an 868 ligand (SQILTPPQL) and another that was predicted by the webtool to be a strong agonist ligand for TCR 22 (VSYNRIPSM), Figure 4.10A. These peptides were derived from the TAA- and self-proteomic databases respectively (Szomolay et al., 2016). The SQILTPPQL peptide comes from a protein expressed in the optical lens called lengsin. The VSYNRIPSM peptide is contained within amphoterin-induced protein 2 precursor (AMIGO2). The two peptides were ordered as pure preparations to confirm response by T cells expressing the 868 TCR but not by TCR 22 (Figure 4.10B). The lengsin-derived peptide was recognised by T cells expressing the 868 TCR. The AMIGO2-derived peptide, in contrast, did not produce a comparable response from 868-expressing T cells (Figure 4.10B).

Soluble monomers were prepared for HLA-A2-SQILTPPQL and HLA-A2-VSYNRIlimP as described in Section 2.1.7. Repeated attempts to refold HLA-A2-VSYNRIlimP were unsuccessful. This was little surprise as the 2S and 9M amino acids are suboptimal for HLA-A2 binding. Given that T cells expressing the 868 TCR also failed to respond pure VSYNRIlimP peptide in antigen titrations (Figure 4.10B), I concluded that the VSYNRIlimP sequence was unlikely to be a true epitope recognised by 868-transduced T cells. In contrast, the lengsin-derived sequence SQILTPPQL was well recognised and refolded well with HLA-A2.

Both 868- and TCR 22-transduced cells stained comparatively well with the HLA-A2-SLYNTVATL tetramer (MFI values of 2949 and 2717 respectively, blue) but only 868-transduced cells stained with the lengsin tetramer, HLA-A2-SQILTPPQL (MFI value of 911 compared to 296 for TCR 22, red) in parallel assays (Figure 4.10C). Thus, the single βD95S amino acid change in TCR 22 appears remove the response to a putative lengsin-derived epitope that can be recognised by the 868 TCR. I next aimed to validate the lengsin SQILTPPQL epitope.
Figure 4.10: Recognition of two putative self-derived epitopes by 868 TCR-expressing cells appears to have been ablated by the βD95S mutation in TCR 22. A) 30,000 CD8+ T cells transduced with TCR constructs were cultured in the presence of 60,000 T2 cells as APCs and 10^{-5} M of the crude peptide and incubated overnight before the supernatant was harvested for measurement of MIP-1β presence by ELISA analysis. Response is shown as a percentage of response to the index SLYNTVATL peptide. B) 30,000 CD8+ T cells transduced with TCR constructs were cultured overnight in the presence of 60,000 T2 cells as APC and varying concentrations (10^{-13} M to 10^{-5} M) of the three peptides recognised in A (at >95% purity) overnight before the supernatants were harvested for measurement of MIP-1β presence by ELISA analysis. The EC_{50} for each viable T cell-peptide combination is shown in the accompanying table. EC_{50} based on GraphPad Prism software analysis. Measurements performed in duplicate, error bars depict SEM. C) 50,000 CD8+ T cells transduced with 868 or TCR 22 constructs were stained with 0.5 µg tetramer (relative to the pMHC component), with a PE-conjugated backbone. Cells previously gated on live CD3+CD8+ single lymphocytes. Numbers on histograms represent MFI values of 10,000 events.
4.2.7 – Attempts to validate the lengsin epitope

Lengsin is a 57 kDa protein expressed in the ocular lens (Grassi et al., 2006; Wistow, 2006), an immune-privileged site that escapes the majority of immune scrutiny (Martin et al., 1995; Streilein, 2003; Nakatsugawa et al., 2009). The lengsin protein plays a role in cytoskeletal organisation (Wyatt et al., 2008). Lengsin expression is tightly regulated to the lens and in healthy scenarios it is expressed in no other adult tissues (Wistow, 2006). Despite this tight regulation, ectopic expression of a lengsin splice variant was described as a TAA in lung cancers in 2009 (Nakatsugawa et al., 2009). Lengsin is discussed in detail in Section 4.3.6.

Given its potential as a TAA, I next wanted to validate that the lengsin SQILTPPQL peptide as a legitimate epitope that is correctly processed and presented at the cell surface. To do so, as before, I cloned the parent protein into a lentiviral vector, pELNS, (Section 2.3.3) and transduced it into MOLT-3 cells (Section 2.4.2). Lengsin transduced well into both the wildtype HLA-A2 MOLT-3 line, and a MOLT-3 line that had previously been stably transduced with HLA-A2. Both wildtype and HLA-A2+ MOLT-3 lines were >80% positive for marker gene rCD2, confirming successful transduction (Figure 4.11A).

Response of 868- or TCR 22-transduced CD8+ T cells to transduced MOLT-3 lines was then analysed by TAPI-0 assay (Figure 4.11B). 868-transduced CD8+ T cells responded to not only lengsin-transduced MOLT-3 cells, but unexpectedly to untransduced MOLT-3 cells and MOLT-3 cells transduced with an irrelevant protein, insulin-like growth factor 2 mRNA binding protein (IMP-2). This response was HLA-A2-dependent and did not occur in T cells transduced with TCR 22 (Figure 4.11B). The observations in Figure 4.11B could be replicated (Supplementary Figure 7.2 and Figure 4.12A), in fact 868-transduced CD8+ T cells produced TNF in response to HLA-A2+ MOLT-3 cells transduced with a range of proteins (Figure 4.12B). Furthermore, 868-transduced CD8+ T cells produced TNF in response to SAR26 (an HLA-A2+ sarcoma cell line) but not MM909.27 (an HLA-A2- melanoma cell line) (Figure 4.12C). From these data, it appears that 868-transduced CD8+ T cells respond to an epitope inherently expressed in the context of HLA-A2 by MOLT-3 cells. This epitope could potentially be shared with T2 cells and SAR26, and might be lengsin-derived. The Protein Atlas contains data suggesting that MOLT-4 cells, a cell line derived from the same parental line as MOLT-3 cells, express lengsin at a low level (Protein Atlas, 2020). Confirmation that MOLT-3 cells express lengsin will require Western Blot analyses. Unfortunately, I did not have time to undertake these studies. Verification that MOLT-3 cells are being recognised via the new HLA-A2 lengsin epitope I identified could be also achieved by examining recognition of lengsin knockout
MOLT-3 cells. Alternatively, multiple epitopes could be being recognised by 868 transduced T cells, the TCR could be specific for more than one TAA.

Retaining the goal of validating the lengsin epitope, I next assessed the response of 868-transduced CD8+ T cells to a panel of three other immortalised lines, with the hope of finding a line that could be used to verify the lengsin epitope. I found that 868-transduced CD8+ T cells showed a moderate response to each immortal cell line, in an HLA-A2-dependent manner (Figure 4.13). This was not observed for CD8+ T cells expressing TCR 22 (Figure 4.13). These data were replicable using transduced T cells from two further donors (Supplementary Figure 7.3).

My data suggest that this TCR might recognise a self-antigen that is expressed in a range of immortalised cell lines. As these cell lines are cancerous, it is possible that the HIV-1-specific 868 T cells might crossreact with a common tumour antigen.
Figure 4.11: Attempts to validation the lensin-derived epitope SQILTPQPL. A) Transduced MOLT-3 HLA-A2 or MOLT-3 wt cells were cultured in 10 mL R10 suspension flasks for 1 week following spinfection. Cells previously gated on live, single lymphocytes. Surface expression of lentiviral constructs was monitored based on expression of marker gene, rCD2. Numbers on dot plots correspond to the percentage of cells in the gated population, 10,000 events recorded. B) 30,000 CD8\(^+\) T cells transduced with either 868 wt or TCR 22 were cultured in the presence of 60,000 MOLT-3 cell lines overnight before cells were harvested and assessed for surface expression of TNF. Cells previously gated on live CD3\(^+\) CD8\(^+\) single lymphocytes. Numbers on dot plots correspond to the percentage of cells in the gated population, mean of 10,000 events.
Figure 4.12: 868-transduced CD8+ T cells respond to MOLT-3 cells expressing HLA-A2. 30,000 CD8+ T cells transduced with 868 wt were cultured in the presence of 60,000 A) MOLT-3 cells lines transduced with lengsin parental protein, or an irrelevant protein (insulin-like growth factor 2 mRNA binding protein, IMP-2) B) MOLT-3 cells transduced with portions of the ZIKA virus genome, and thus expressing a range of potential viral epitopes, or C) HLA-A2+ or HLA-A2− cancer cell lines. Cells were co-cultured overnight before being harvested and assessed for surface expression of TNF. Cells previously gated on live CD3+ CD8+ single lymphocytes. Numbers on dot plots correspond to the percentage of cells in the gated population, mean of 10,000 events.
Figure 4.13: 868-transduced CD8+ T cells respond to a panel of potential APCs in an HLA-A2-dependent manner. 30,000 CD8+ T cells transduced with TCR constructs were cultured in the presence of 60,000 immortalised HLA-A2+ and HLA-A2- cell lines overnight before cells were harvested and assessed for surface expression of TNF. Cells previously gated on live CD3+ CD8+ rCD2+ single lymphocytes. Numbers on histograms represent MFI values of 10,000 events.
4.2.8 – Attempts to isolate lengsin-reactive T cells
The antigen that 868-transduced T cells are responding to in the panel of cell lines may or may not be the lengsin epitope identified by CPL-driven database screening of the TAA database. Confirmation that the recognition of T2, MOLT-3 and other HLA-A2\(^+\) cell lines is as a result of lengsin expression will require confirmation that these cells express lengsin. To confirm that the lengsin peptide is a truly processed epitope I would ideally need to find a cell that does not express lengsin and then transduce in the lengsin gene. It would also be helpful to have T cells that were primed with the lengsin peptide. As the SQILTPPQL lengsin peptide originates from a known TAA, my first thought was to look for responses to this peptide within two tumour-infiltrating lymphocyte (TIL) lines from HLA-A2\(^+\) patients that my laboratory routinely works with. However, I was unable to identify any lengsin-specific T cells within a sarcoma TIL sample (Figure 4.14A) or a melanoma TIL sample (Figure 4.14), as measured by tetramer staining.

Given the difficulty of obtaining fresh PBMC donations during the 2020 COVID-19 pandemic, I was fortunate to benefit from access to stored PBMC samples from 4 healthy donors. Each was defrosted and, after a culturing for a short period, stained with HLA-A2-SQILTPPQL tetramer (Figure 4.15). Whilst not entirely convincing, two donors (BB16 and BB18) showed some increased staining compared to the irrelevant tetramer. These two lines have been enriched for binding to HLA-A2-SQILTPPQL (Section 2.4.5) and are currently being further examined.
Figure 4.14: Lengsin-reactive T cells were not found in TIL samples. A) 50,000 MM909.42 TILs and B) 50,000 SAR26 TILs were stained with 0.5 µg (relative the pMHC component) HLA-A2-SQILTPPQL tetramer, with a PE-conjugated backbone. Cells previously gated on live CD3⁺CD8⁺ single lymphocytes. Numbers on dot plots represent MFI values of 10,000 events. Controls HLA-A2-EAAIGILTV, HLA-A2-NLVPMVATV, and HLA-A2-LLLIEGIIFFI derived from human Melan-A, cytomegalovirus pp65, and EBV BARF-1 respectively.
Figure 4.15: Potential lengsin-reactive T cells found in healthy PBMC samples. CD8+ T cells were isolated from four frozen healthy PBMC samples, 50,000 from each donor were stained with 0.5 µg (relative the pMHC component) HLA-A2-SQLTPQL tetramer, with a PE-conjugated backbone. Cells previously gated on live CD3+CD8+ single lymphocytes. Numbers on dot plots represent percentage of cells in the gated population, of 10,000 events. Red arrows indicated potentially notable staining above background. Irrelevant tetramer HLA-A24-LWMRLPLL derived from preproinsulin-signal peptide (residues 3-11).
4.2.9 – Preliminary characterisation of two further mutants of the 868 TCR
Since my return to the laboratory following the COVID-19 summer lockdown I was able to begin the analysis of two further mutants of the 868 TCR, TCRs 9 and 11 (see Table 4.2). In contrast to the TCR mutants that I have already studied, TCRs 9 and 11 both contain substitutions in their CDR3α loop (Figure 4.16A) allowing me to assess the role of the TCRα chain in peptide discrimination. TCRs 9 and 11 both have amino acid substitutions in key peptide contact residues (Table 4.1). Residue 92 in the 868 α chain is an arginine, a positively charged amino acid that makes a total of five contacts with two peptide residues – asparagine (P4) and threonine (P5) (Figure 4.2). In TCR 9 arginine 92 was substituted for glutamic acid to reverse the charge of this residue (Figure 4.16B). Residue 93 in the 868 α chain is a threonine, a polar amino acid that contacts the asparagine at position 4 in the peptide (Figure 4.2). In TCR 11 this residue has been substituted for valine, a hydrophobic amino acid (Figure 4.16C). Prior to any analysis the surface expression of the transgene was confirmed, the rCD2^+TRBV5-6^+ fraction of CD8^+ T cells transduced with the 868, TCR 9, and TCR-11 was 86.9 %, 74.7 %, and 77.4 % respectively (Figure 4.16D).

Cells transduced with either TCR 9 or TCR 11 were then subject to CPL screening (Figure 4.17, green and blue respectively). Each showed varying amino acid preferences in comparison with their wildtype parent TCR (grey bars) at a number of peptide positions (Figure 4.17). These data were then used to screen the 9-mer peptide universe (screening kindly performed by Dr Barbara Szomolay) for peptides predicted to be recognised by the mutant TCRs but not the 868 parental TCR; predicted peptides are shown in (Table 4.3).

The peptides identified by CPL-based database screening of the universe database were ordered in crude form (~40 % purity) and the three TCRs (868, TCR 9, and TCR 11) were assessed for reactivity against 10^-5 M of each peptide, Figure 4.18. The graph for each mutant TCR is overlaid with that of 868 (grey bars) in Figure 4.19 for ease of visual analysis. Figure 4.19 shows some clear incidences where peptides are seen by the respective mutant TCRs, but not by wildtype 868. These data suggest that the single amino acid substitutions in TCRs 9 and 11 have resulted in the gain of specificity for peptides that are not seen by the wildtype 868 TCR.

Further work will be needed to confirm that TCRs 9 and 11 exhibit reactivity to the peptides highlighted in Figure 4.18, and to confirm that they are not recognised by the 868 TCR. Unfortunately, I was unable to extend my PhD funding further to cover the period that I lost in the laboratory due to the COVID-19 lockdown. I am now contracted to work on a different...
project so I have very limited time in which I can undertake further experiments on the 868 TCR.

<table>
<thead>
<tr>
<th>TCR</th>
<th>Mutation</th>
<th>CDR3 Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>868</td>
<td>None</td>
<td>α: CAVRNSGYALNF β: CASSDTVSYEQY</td>
</tr>
<tr>
<td>TCR 9</td>
<td>αR92E</td>
<td>α: CAVETNSGYALNF β: CASSDTVSYEQY</td>
</tr>
<tr>
<td>TCR 11</td>
<td>αT93V</td>
<td>α: CAVRVNSGYALNF β: CASSDTVSYEQY</td>
</tr>
</tbody>
</table>

**Figure 4.16: TCR 9 and TCR 11.** A) shows the CDR3 sequences of each of TCRs 9 and 11 in comparison with the wildtype 868 sequence. Both TCRs carry amino acid substitutions in their CDR3α sequences, these substitutions are underlined. B) Model prediction of the αR92E substitution in TCR 11. C) Model prediction of the αT93V substitution in TCR 11. D) Surface expression of lentiviral constructs was confirmed based on simultaneous expression of marker gene, rCD2, and TRBV 5-6 chain. Numbers on dot plots correspond to the percentage of cells in the gated population, of 10,000 events.
Figure 4.17: TCRs 9 and 11 exhibit different peptide preferences than the 868 parental TCR. 30,000 CD8+ T cells transduced with either TCR 9 (green) or TCR 11 (blue) constructs co-cultured with 60,000 T2 cells as APC and 100 µg/mL peptide mix from a 9-mer CPL screen. Supernatants were harvested and ELISA analysis performed to measure MIP-1β presence as a proxy for T cell response. Grey bars indicate the peptide preferences of the wildtype 868 TCR for comparison. All assays performed once.
Table 4.3: CPL data was used to screen the 9-mer peptide universe for the top 20 predicted peptide agonists of each TCR. Peptides predicted to be top agonists for 868 were removed.
Figure 4.18: Recognition of putative epitopes unique to mutant TCRs. 30,000 CD8+ T cells transduced with TCR constructs were cultured in the presence of 60,000 T2 cells as APCs and 10^{-5} M of the crude peptide and incubated overnight before the supernatant was harvested for measurement of MIP-1β presence by ELISA analysis. The response of 868-, TCR 9-, and TCR 11-transduced T cells is depicted by grey, green, and blue bars respectively.
Figure 4.19: Novel specificities in mutant TCRs. Response of TCRs 9 and 11 to each peptide overlaid with that of 868 (grey) to aid visual analysis suggests some peptides produce response in T cells transduced with mutant TCRs but not with 868 wildtype.
4.3 – Discussion
In this chapter, I have demonstrated broad crossreactivity of 868-transduced CD8+ T cells. I have shown that these cells respond to peptides derived from viral, self, and TAA proteomes in the context of HLA-A2. I have further shown that a single amino acid change in the CDR3 loops of a TCR can alter the repertoire of peptides that a T cell can respond to.

Data such as that presented here provides fuel for ongoing efforts to predict what pMHC targets a given T cell will be able to respond to based on the CDR3 sequences of its TCR (Udyavar et al., 2009; Roomp and Domingues, 2011; Zoete et al., 2013; Smith et al., 2014). Greater understanding of how CDR3 loop sequences influence the peptide recognition repertoire, and how small changes to the CDR3 might alter TCR function, will provide a foundation for the rational design of TCRs against a nominal epitope. A recent model created by Jurtz et al. was able to predict a cognate antigen for a TCR based on CDR3β sequence (Jurtz et al., 2017). However, this model was limited to the training data supplied and could not be used to predict novel agonist epitopes (Jurtz et al., 2017). Uncovering further information regarding the plasticity of the TCR-pMHC interaction, and particularly how CDR sequence relates to peptide preference will lead to improvement in predictive tools.

4.3.1 – TCR transduced T cells
I chose the 868 TCR to begin investigation into how TCR sequence relates to function because it is biochemically and structurally well-characterised (Varela-Rohena et al., 2008; Cole, Fuller, et al., 2017). Repeated attempts to generate a clone from the patient 868 polyclonal HIV-1-specific line were unsuccessful (Varela-Rohena et al., 2008), I therefore turned to TCR gene transfer to deliver this TCR to a polyclonal population of primary T cells.

The ability of TCR gene transfer to redirect the specificity of T cells was first demonstrated by Dembic et al. in 1986 (Dembić et al., 1986). There have since been many studies investigating ways in which to optimise expression of the transgene in order to more effectively redirect the recipient cells (Emerman and Temin, 1984; De Felipe et al., 1999; Mizuguchi et al., 2000; Szymczak et al., 2004; Leisegang et al., 2008; Yang et al., 2008). To transduce the TCRs described in this chapter, I used a single tricistronic transfer vector to achieve stoichiometric expression of both TCR genes and a marker gene. Within the vector, the three genes were separated by 2A ‘self-cleaving’ peptide sequences which has previously been shown to be optimal for the expression of genes that function with a 1:1 stoichiometry.
The advantages and disadvantages of the various available expression systems are discussed further in Appendix Section 7.2.

4.3.2 – TCRs 3 and 6 showed little change in peptide preference

My initial aim was to study the peptide recognition repertoire of the 868 TCR and 24 mutant versions of the TCR carrying amino acid substitutions in their CDR3 loops, as described in Table 4.2. At the start of my studies, I wanted to move the 868 TCRα-2A-TCRβ-2A-rCD2 construct from the pELNS lentivirus used by Riley and colleagues (Varela-Rohena et al., 2008) to the pSF lentivirus because the pELNS virus had been sold to Novartis by its creator Jim Riley. While my laboratory retained an academic license to use this virus, we were unable to share it with other laboratories. The pSF lentivirus was purchased from Addgene and did not have this limitation. Unfortunately, despite repeated attempts at cloning, the 868 TCRα-2A-TCRβ-2A-rCD2 construct did not work in pSF. This was later tracked down to a lack of stop codon within the transduced fragment (the stop codon in pELNS lies outside of the restriction site and our historic use of this vector meant that constructs were designed without their own stop codon). Trying to troubleshoot this issue wasted ~5 months of my lab time. My supervisors then decided that I should just continue with pELNS rather than waste any more time. I have since successfully cloned 19/24 mutant TCRs into the pELNS lentivirus and expressed 12 of these constructs in T cells. I was able to screen seven TCRs by CPL.

Two of the mutant TCRs I studied, TCR 3 and TCR 6, showed little difference in their peptide specificity compared to the wildtype sequence. This was unexpected, as both TCRs bear amino acid substitutions to amino acids with differing chemical properties at key peptide contact residues, Table 4.1 and Figure 4.3. The βT96F mutation in TCR 3 involves the substitution of a small polar amino acid, threonine, to a bulky hydrophobic amino acid with a large aromatic ring, phenylalanine. As the threonine at position 96 of the wildtype 868 sequence is a key contact residue, forming two peptide contacts one of which is a hydrogen bond, I expected that this substitution would have a more drastic phenotype. The βV97T mutation in TCR 6 involves the substitution of a hydrophobic amino acid, valine, to a polar amino acid, threonine. In the wildtype TCR this residue also makes two peptide contacts, alongside an additional four contacts with HLA-A2. The retention of TCR specificity despite such amino acid substitutions demonstrates the substantial plasticity within the TCR-pMHC interaction.
I demonstrated that both TCR 3 and TCR 6 were correctly processed and expressed at the T cell surface, achieving measurable post-purification surface expression on 81% and 82% of T cells respectively (Figure 4.4). Both mutant TCRs were able to bind and respond to the cognate SLYNTVATL peptide in the context of HLA-A2, as demonstrated by MIP-1β ELISA (Figure 4.5A), and tetramer binding (Figure 4.5B). Examination of the CPL data for these TCRs in comparison to the wildtype 868 (Figure 4.6) suggested little change in peptide preference. These results confirmed that CPL screens can be used to probe the peptide recognition landscape of mutated TCRs. However, it appears that the single amino acid substitutions in TCR 3 and TCR 6 did not result in the acquisition of novel reactivities. Given the limited timescale of my project these TCRs were set aside in favour of more interesting mutants.

It is interesting to ponder why substitutions that alter chemical properties did not alter the peptide preferences of TCRs 3 and 6. It could be that the selected assay failed to detect novel peptide reactivity. CPL screens are powerful tools to assess ligand preferences with no prior knowledge of likely candidates (Houghten et al., 1991; Pinilla et al., 1992). CPL screens can rapidly assess almost the whole proteomic universe and have been used successfully in many incidences (Houghten et al., 1991; Pinilla et al., 1992; Hemmer et al., 1997, 1998; Hiemstra et al., 1997; Wooldridge et al., 2012). However, CPL screens are not without caveat, while cysteine is used as a fixed amino acid in each position it is excluded from the amino acid pool due to its propensity to form disulphide bonds. Thus, each library is devoid of peptides containing two or more cysteines. Furthermore, such libraries do not include post-translationally modified amino acids and are likely to be suboptimal for screening of highly hydrophobic peptides due to aqueous solubility issues. Alternative methods for T cell ligand discovery are discussed in Section 6.2.

It is also possible that the CPL assay is not at fault. The substitutions employed to create TCRs 3 and 6 may not have altered the peptide preference of the TCR from that of the wildtype 868 due to the natural plasticity of the TCR, and the formation of compensatory interactions. Amino acids changes in the peptide residues do not always alter TCR binding and recognition, as seen in the case of the common naturally-occurring SLYNTVATL escape mutants (Sewell et al., 1997; Cole, Fuller, et al., 2017). Furthermore, when Burrows et al. mutated key TCR contact residues on the MHC molecules they found that compensatory interactions occurred to maintain binding (Burrows et al., 2010). The TCR-pMHC interaction is extraordinarily plastic. If amino acid substitutions in both peptide and MHC molecule can be tolerated, it stands to reason, that certain single amino acid substitutions in the TCR might also retain the
peptide recognition repertoire, if TCR plasticity allows for suitable compensatory interactions.

It is also notable that the 868 TCR binds to the cognate HLA-A2-\textit{SLYNTVATL} epitope with an extremely high affinity for a naturally-occurring TCR (Cole, Fuller, \textit{et al.}, 2017). This strong binding might allow weaker binding mutants of this TCR to still appear fully functional at the T cell surface. It could be that if I had started with a weaker binding TCR, single amino acid substitutions in CDR3 may have had much greater functional effects.

\subsection*{4.3.3 – TCRs 22 and 24 have altered peptide preferences}

TCRs 3 and 6 proved useful in validating that my experimental set up was functional and demonstrated that I could successfully redirect the specificity of polyclonal primary CD8$^+$ T cells using lentiviral transfer of a TCR construct. These experiments also showed that the TCRs could still remain functional when key contact residues were mutated to residues with different chemistries. Given the limited evidence of any novel specificity with TCRs 3 and 6, I turned my focus to the next two TCRs that I was able to clone into pELNS, TCRs 22 and 24.

For TCR 22, I altered a non-contact residue, residue 95, in order to assess the impact of potential knock-on effects on the recognition footprint of the TCR. The negatively charged aspartic acid residue present in the wildtype chain was mutated to a polar, uncharged, serine - $\beta$D95S (\textbf{Figure 4.3D}). TCR 24, on the other hand, is the only mutant TCR that I was able to screen that contained more than a single amino acid substitution. TCR 24 contains three amino acid substitutions at residues 95, 96, and 97. The wildtype sequence consists of aspartic acid, threonine, and valine (DTV). These residues were mutated to isoleucine, arginine, and serine (IRS, $\beta$D95I T96R V97S) (\textbf{Figure 4.3E}). As described in \textbf{Section 4.2.1} these amino acid substitutions are based on the published structure of a HLA-A2-restricted, influenza-specific TCR, JM22 (PDB:5HHO) (Valkenburg \textit{et al.}, 2016). The JM22 TCR recognises an epitope from the M1 influenza protein (\textit{GILGFVFTL}, residues 58-66) (Valkenburg \textit{et al.}, 2016). TCR 24 did not recognise the \textit{GILGFVFTL} peptide in the context of HLA-A2 (\textbf{Supplementary Figure 7.1}); this is likely due to the remaining differences in the TCR-$\alpha$ chain.

I first assessed the ability of the mutant TCRs to bind and respond to the wildtype \textit{SLYNTVATL}. TCR 22 but not TCR 24 could respond to \textit{SLYNTVATL}, as measured by MIP-1$\beta$ ELISA, \textbf{Figure 4.7A}. Similarly, T cells expressing TCR 22 but not TCR 24 could bind HLA-A2-\textit{SLYNTVATL} tetramer, \textbf{Figure 4.7B}. T cells expressing wildtype 868 TCR responded to the T2 cells used as APCs in the assay. This background response was not observed with T cells expressing TCR
22 or TCR 24 suggesting that the amino acid substitutions between them and the wildtype TCR were sufficient to ablate this reactivity, Figure 4.7A. These results strongly suggested that the recognition repertoire differed between the 868 wildtype- and mutant TCR-transduced CD8+ T cells.

With this observation in mind, I turned to the CPL data for each TCR to see if they could be used to identify the peptide that was recognised by the 868 TCR-transduced T cells but not T cells transduced with TCRs 22 or 24, that might be responsible for background recognition of T2 cells Figure 4.8. Both the 868 wildtype TCR and TCR 22 assayed well, showing variable MIP-1β outputs to different peptide pools, reflecting variable peptide preferences. TCR 24 however, produced very low levels of MIP-1β in response to most peptide pools. At some positions, almost no MIP-1β was produced in response to any peptide pool. My work validating the surface expression of each construct, Figure 4.4, suggests this lack of response is unlikely to be due to failure of the construct expression or surface trafficking. Post-purification CD8+ T cells transduced with TCR 24 were 81.2 % positive for both the TCR β chain and the marker gene rCD2, Figure 4.4D. It is, therefore, likely that function rather than expression has been impeded by the triple CDR3 amino acid substitution in TCR 24, resulting in the minimal response observed across the 9-mer peptide universe, Figure 4.8. The triple substitution in TCR 24 may have ablated its ability to bind pMHCs and function as a TCR at all. It is further possible that the peptide length preference of the TCR might have been altered so that T cells transduced with this mutant no longer responded to 9-mer peptides. The triple substitution could have had knock on effects that were translated to other parts of the protein including the TCR α chain. In the absence of an atomic resolution structure for the βD95I T96R V97S TCR, it is impossible to know the extent to which these substitutions affected the antigen-binding site.

Fortunately, TCR 22 retained recognition of pMHC complexes but exhibited different amino acid preferences at some peptide positions compared to the 868 TCR. Peptide preference differences between TCR 22 and the 868 TCR are illustrated by the overlay graphs in Figure 4.9A. These data were input into the public webtool for analysis of CPL data created by Szomolay et al. (Szomolay et al., 2016), the output of which is shown in Figure 4.9B. The webtool ranks peptides in a predicted order of preference, whilst I did not see this predicted order manifest in terms of response (Figure 4.10A) I was able to identify a novel peptide ligand of the 868 wildtype TCR that was not recognised by TCR 22, which differs by just a single amino acid substitution, SQILTPPQL (Figure 4.10B).
I concluded that the VSYNRIPSM peptide flagged in my initial studies with crude peptides was not a true ligand for the 868 TCR as purified peptide was neither recognised nor capable of forming a stable complex with HLA-A2 (Figure 4.10). In contrast HLA-A2-SQILTPPQL refolded well and was shown to bind the wildtype 868 TCR but not TCR 22 (Figure 4.10C), demonstrating that it is possible to alter the specificity of a natural TCR with a single amino acid substitution.

4.3.4 – The 868 TCR may crossreact with a tumour-associated antigen
Throughout my studies, I noticed that T cells expressing the 868 TCR exhibited background recognition of the HLA-A2+ T2 cell line. I reasoned that this reactivity may be due to expression of lengsin in the T2 cells I used in my experiments. As such I began attempts to validate the SQILTPPQL peptide as a legitimate epitope that is processed and presented at the cell surface. To achieve this, I transduced the parental lengsin protein (along with the rCD2 marker gene) into MOLT-3 cells, both wildtype MOLT-3 cells and MOLT-3 cells that had been previously stably transduced with HLA-A2, Figure 4.11A.

I found that 868-transduced CD8+ T cells responded to MOLT-3 HLA-A2+ regardless of, or even in the absence of, a transduced protein (Figures 4.11 and 4.12). Furthermore, in an attempt to find a different APC that could be used to validate lengsin I found that 868-transduced CD8+ T cells responded to a range of cancer cell lines in an HLA-A2-dependent manner (Figures 4.12C and 4.13). This was not observed for TCR 22, which differs from the wildtype 868 by just a single amino acid substitution (Figure 4.13).

It is important to note that T cells expressing the 868 TCR have not previously been shown to respond to HLA-A2+ cell lines such as the autologous B cell line derived from patient 9300868 (Sewell et al., 1997; Varela-Rohena et al., 2008). A previous study from my laboratory demonstrated that it was not possible to transduce CD8+ T cells with a TCR that responded to a peptide presented at the T cell surface (Tan et al., 2015). High affinity HLA-A2-restricted, self-reactive TCRs could be transduced into HLA-A2+ T cells but failed to express in HLA-A2+ T cells. It was concluded that transduction with self-reactive TCRs induced T cell fratricide (Tan et al., 2015). Previous studies with the 868 TCR showed that it expressed and functioned well in HLA-A2+ T cells (Varela-Rohena et al., 2008). These findings suggest that although T cells expressing the 868 TCR recognise many HLA-A2+ cancer cell lines they do not respond to all human HLA-A2+ cells. These findings fit with the fact that there was no obvious pathology in patient 9300868 from which the HIV-1-specific 868 T cell line was grown. I
concluded that T cells expressing the 868 TCR might respond to an epitope from a TAA, Figure 4.9B.

As mentioned above, 868-transduced T cells responded to a number of cancer cell lines in a HLA-A2⁺-dependent manner. To my understanding, there are two likely explanations for these observations. The first being that endogenous expression of lengsin, a TAA, is shared between all of the cancer cell lines tested. The alternative being that multiple epitopes could be being recognised by the 868 TCR. Further work will be required to distinguish between these possibilities.

Lengsin was identified as part of the national eye institute bank (NEIBank) project in 2002, the project aimed to profile the gene expression in the human adult lens (Wistow et al., 2002). The NEIBank project found a novel protein that had sequence homology to the ancient glutamine-synthetase family, hence naming it LENs Glutamine-Synthetase-like (Wistow et al., 2002). Despite its heritage, lengsin has no enzymatic activity (Wistow, 2006; Wyatt et al., 2006, 2008). The 57kDa lengsin polypeptide forms a catalytically inactive dodecamer structure (Grassi et al., 2006; Wyatt et al., 2006) that co-ordinates the reorganisation of the cytoskeleton during terminal differentiation (Wyatt et al., 2008). Expression of lengsin is extremely tissue-specific, being tightly regulated within the lens (in terminally differentiating cells) and no other adult tissues (Wistow, 2006). Orthologs of the protein were found in mouse, rat, dog, and zebrafish (Wistow, 2006).

The tightly-regulated lens-specific expression pattern is noteworthy as in 2009 a splice variant of lengsin was described as a novel TAA expressed ectopically in lung cancers (Nakatsugawa et al., 2009). Knockdown of lengsin mRNA using RNA interference resulted in heightened cell death indicating an essential role in cancer cell survival (Nakatsugawa et al., 2009). The Human Protein Atlas suggests that lengsin mRNA is expressed by several cancer types including stomach, renal, liver, pancreatic and ovarian cancer while being most highly expressed in lung cancers (Protein Atlas, 2020). Nakatsugawa et al. later described lengsin-derived HLA-A2-presented peptide epitopes that could be recognised by specific clones grown from healthy HLA-A2⁺ donors (Nakatsugawa et al., 2011). They demonstrated that one of their peptide epitopes (FLPEFGISSA) was naturally processed and presented by HLA-A2, thus highlighting the real potential for lengsin-derived epitopes to be clinically-relevant TAAs (Nakatsugawa et al., 2011).

The lens is an immune-privileged site, meaning that – like the testis – it escapes the majority of immune scrutiny (Martin et al., 1995; Streilein, 2003; Nakatsugawa et al., 2009). Immune
privileged is achieved through a number of mechanisms including the tight junction-regulated blood-organ barriers, a lack of lymphatic drainage, low level expression of MHC molecules, and an immunosuppressive microenvironment abundant in components such as TGF-β (Martin et al., 1995; Streilein, 2003; Nakatsugawa et al., 2009). Immune privilege serves to protect vision, a sense that is essential for survival in most higher animals, from the potentially damaging effects of inflammation (Streilein, 2003). Anti-lengsin antibodies were detected in the lung cancer patients involved in the aforementioned 2009 study, with none experiencing vision disturbances (Nakatsugawa et al., 2009, 2011). This immune-protection makes lengsin a highly relevant, immunogenic cancer-antigen (Nakatsugawa et al., 2009), a relevance that has been recognised clinically – a number of lengsin peptides have been patented, as HLA-A2 complexes, for the use in anti-cancer immunotherapies. This patent covers the 8-mer lengsin-derived peptide, QILTPPQL that overlaps with the longer putative epitope I discovered (patent number W02017089777). Thus, the CPL-discovered 868 peptide ligand, SQILTPPQL, could be a clinically relevant cancer-associated immunotherapy target. I have therefore demonstrated HIV/cancer crossreactivity on the part of the 868 TCR and, moreover, showed how this crossreactivity could be eradicated with a single amino acid substitution in TCR 22.

As lengsin-derived peptides are already recognised as potentially clinically-relevant TAAs future work on this project could attempt to characterise lengsin-specific T cell clones (Figures 4.14 and 4.1). If I had had time, I would have liked to have undertaken lengsin antibody Western blots and knock-out studies to determine whether the cell lines recognised by T cells expressing the 868 TCR express lengsin endogenously.

It also remains possible that the 868 TCR binds to a different cancer-derived epitope and that lengsin expression is not responsible for the observed recognition of HLA-A2+ cancer cell lines by T cells expressing 868. Given that most of the top scoring peptides from the CPL-database screen for the 868-transduced T cells originated in the TAA database (Figure 4.9B), it could be that the 868 has capacity to interact with a range of TAAs – which could explain why T cells expressing the TCR exhibit recognition of many HLA-A2+ cell lines.

To briefly consider the original aim of this project – to use amino acid substitutions to begin assessing how TCR sequence relates to function – it is striking that the ability of T cells transduced with the 868 TCR to response to the panel of HLA-A2+ cell lines described in this chapter can be completely abrogated by a single amino acid substitution, βD95S. This demonstrates the extraordinarily precise influence of TCR sequence on TCR function.
4.3.5 – T cells expressing TCRs 9 and 11 might exhibit new specificities
During the last few months of my PhD, I was able to begin analysis of two further mutant TCRs, TCRs 9 and 11. TCRs 9 and 11 both contain substitutions in key contact residues of their CDR3α loops (Figure 4.16A and Table 4.1). TCR 9 contains the αR92E substitution, exchanging a positively charged amino acid for a negatively charged amino acid (Figure 4.16B). TCR 11 contains the αT93V substitution, exchanging a polar (hydrophilic) amino acid for a hydrophobic amino acid. Both TCR constructs were expressed well at the surface of transduced T cells (Figure 4.16D).

CPL screening of TCR 9 and TCR 11 suggested differences in amino acid preferences at various peptide positions compared to the 868 wildtype TCR and to each other (Figure 4.16). These data were used to screen a 9-mer peptide universe database for potential ligands for each TCR. Preliminary analysis with crude peptides suggests the amino acid substitutions in TCRs 9 and 11 have resulted in the gain of specificity for some peptides, meaning that T cells transduced with the mutant TCRs respond to some peptides that T cells transduced with the 868 wildtype TCR do not (Figure 4.17). These data will need confirmation with pure peptide via peptide titration and tetramer staining, as demonstrated for the HLA-A2-SQILTPQL ligand in Section 4.2.6. Ultimately any potential ligand of TCR 9 or 11 would ideally be confirmed by atomic resolution structure of the co-complex, and affinity measurements by SPR.

4.3.6 – Future optimisation to reduce polyclonal background noise
Given that the aim of this project was to investigate the influence of TCR sequence on TCR function within the context of a particular TCR, 868, it is important to consider the influence that endogenous TCRs might have on the CPL data, as – in the absence of an 868 T cell clone – the TCR was transduced into a polyclonal CD8+ T cell line.

By using a polyclonal population of CD8+ T cells in this work, I may have minimised the ‘background’ specificities from endogenous TCRs, because my transduced TCR will be present almost every cell, whereas other TCRs will only be present in single or small copy numbers. However, it is possible to enhance the expression of the introduced TCR at the expense of the endogenous TCR by maximising chain pairing between the two introduced chains, allowing the introduced TCR a competitive advantage over the endogenous TCR (Heemskerk et al., 2007).
When a TCR construct is introduced into a T cell, the transduced ‘exogenous’ α and β chains could potentially pair with the endogenous TCR chains, creating chimeric TCRs with unintentional specificities (Heemskerk et al., 2007). A single T cell could simultaneously express the introduced TCR, the endogenous TCR, and two chimeric TCRs (Heemskerk et al., 2007). Van Loenen et al. demonstrated that the chimeric TCRs formed after the introduction of an exogenous TCR can be autoreactive (Van Loenen et al., 2010) and chimeric TCRs have actually been implicated in autoimmunity (Ji, Perchellet and Goverman, 2010). Furthermore, each of these TCR pairs will compete for a limited number of CD3 subunits, providing a bottleneck to surface expression (Heemskerk et al., 2007). Heemskerk et al. demonstrated that introduction of an exogenous TCR into a T cell clone markedly reduced the tetramer staining of the cognate ligand, as the introduced TCR competed with the endogenous TCR for the CD3 subunits (Heemskerk et al., 2007). Improving the chain paring between the two exogenously introduced TCR chains can give the introduced TCR a competitive edge over the endogenous TCR. Several strategies have been utilised in an attempt to improve TCR chain pairing of transduced TCRs. These are described in Appendix Section 7.3.

Optimisations to enhance surface expression of transduced TCRs clearly have substantial potential to reduce the background noise created by residual endogenous TCRs and chimeric TCRs and are something that should be considered for future work of this kind. Even if considered surplus to requirements for initial investigations (in total, 24 mutant TCR constructs were initially designed for the work described in this chapter, Table 4.2, and not all proved interesting for full analysis) these optimisations are certainly something I would include in more in-depth investigative work on constructs that prove themselves to be interesting, such as 868, TCR 9, TCR 11, and TCR 22.

4.3.7 – Summary
In this chapter I aimed to explore the crossreactive nature of the HIV-specific 868 TCR and assess the impact of amino acid substitutions in the sequences of its CDR3 loops. I have shown that 868, and mutants thereof, can be stably transferred into polyclonal populations of CD8+ T cells and expressed at the T cell surface (Figure 4.4). Transduction of the 868 TCR could redirect primary CD8+ T cells to recognise the SLYNTVATL epitope (Figure 4.5).

I showed that some amino acid substitutions do not significantly alter the peptide recognition footprint of the 868 TCR (Figure 4.6). The interaction is highly plastic and this retention of specificity despite alteration of a key contact residues could potentially be due
to the formation of compensatory interactions. Some amino acid substitutions can ablate peptide recognition completely (Figure 4.7 A and B, and Figure 4.8), possibly by inhibiting the function of the TCR, by altering its peptide-length preference, or ablating its ability to bind pMHC targets at all. Furthermore, I showed that some amino acid substitutions may result in the gain of new specificities (Figure 4.18). The data presented in this chapter highlight plasticity within the TCR-pMHC interactions between the 868 TCR and its ligands.

In this chapter, I also demonstrated the potential for a patient-derived HIV-specific TCR (Sewell et al., 1997) to crossreact with a cancer-specific epitope, and showed that this recognition can be removed by a single amino acid substitution.

It is important to note that all of my amino acid substitutions were made to the CDR3 loops of the 868 TCR and were designed with the aim of altering peptide specificity (Garboczi et al., 1996; Garcia et al., 1996; Ding et al., 1998). Peptide specificity is, however, not exclusively determined by the CDR3 loops (Dunn, 2006; Cole et al., 2009) and according to the published 868 TCR structure, CDR1α makes two peptide contacts (Cole, Fuller, et al., 2017). It would be interesting to learn how amino acid substitutions in the CDR1 and 2 loops alter the peptide recognition footprint of the 868 TCR.

My studies provide data that increase understanding of how TCR sequence can govern TCR function and endeavour towards the ultimate aim of predicting TCR function from TCR sequence. Whilst each individual piece of experimental work can provide only information regarding the model system used, they can collectively contribute as training data for bioinformatical ventures aiming at prediction of TCR specificity. One such venture, (Jurtz et al., 2018), a webtool known as Net TCR, aims to predict, from TCR sequence alone, which HLA-A2-peptide complexes a selected TCR will recognise. The tool was able to successfully pick out the cognate antigen when challenged with TCRs that were included in the training data, but failed to do so for uncharacterised TCRs (Jurtz et al., 2018). Predictive models such as this draw on data from databases such as the VDJ database (VDJdb) (Shugay et al., 2018). The VDJdb collects data from published T cell specificity assays (Shugay et al., 2018). The tool is being continuously improved as more training data becomes available. Furthermore, compared with the publication of positive interactions there is a substantial lack of negative data for this purpose. Here I provide not only predicted ligands for the TCRs but propose a novel ligand for the 868 TCR and provide conclusive negative data. TCR 24 does not see SLYNTVATL and TCR 22 recognises neither SQILTPQQL nor VSYNRIPSM. Similarly, the preliminary data in Figure 4.19 suggests that the and amino acid substitutions in TCRs 9 and
11 respectively have resulted in the gain of specificity for peptides unseen by T cells expressing the 868 parental TCR.

To conclude, my studies provide the beginnings of investigation into the plasticity of the interaction between the 868 TCR and its peptide ligands. I have demonstrated that the specificity of the 868 TCR can be modulated by amino acid substitutions and I have shown that novel peptide specificity can both arise and be ablated by single amino acid substitutions in TCR CDR3 loops. I have further demonstrated the potential for HIV/cancer crossreactivity in a naturally-occurring, patient-derived TCR. I demonstrated that T cells expressing the 868 TCR may recognise both the index HIV-1-epitope and a TAA peptide whilst simultaneously illustrating the fragility of this crossreactivity, which can be completely eradicated by a single amino acid substitution.
5 – Peptide cargo modulates pMHC flexibility

5.1 – Introduction
5.1.1 – The flexibility of the pMHC is generally overlooked
As part of my studies I wanted to look more closely at the other side of the TCR-pMHC interaction, specifically the molecular flexibility of the pMHC-I molecule. The conditions of my Great Western 4 (GW4) Medical Research Council PhD funding required collaboration with another GW4 university, which gave me the opportunity to utilise the expertise, and technical equipment, of the Pudney laboratory. Dr Chris Pudney and colleagues regularly employ a phenomenon called the red edge excitation shift (REES, discussed below) to inspect the molecular flexibility of proteins.

Despite its well conserved binding mode (Rudolph and Wilson, 2002; Rudolph, Stanfield and Wilson, 2006) the TCR-pMHC interaction is not a rigid one, many incidences of TCR flexibility have been observed (Garcia et al., 1998; Wagner et al., 1999; Willcox et al., 1999; Reiser et al., 2002, 2003; Armstrong, Piepenbrink and Baker, 2008). Flexibility of the TCR CDR loops contributes to the crossreactive nature of the TCR, which helps provide comprehensive immune coverage to the host (Bhati et al., 2014). In contrast to the wealth of evidence concerning TCR flexibility, pMHCs have previously been viewed as rigid targets for the more dynamic TCR to interact with (Rashin and Jernigan, 2016). Yet evidence to the contrary is rapidly amounting (Hawse et al., 2013, 2014; Kass, Buckle and Borg, 2014; Van Hateren et al., 2017).

5.1.2 – The pMHC as a dynamic molecule
Two key methods that have been employed to investigate pMHC flexibility are nuclear magnetic resonance (NMR) and hydrogen/deuterium exchange (HDX).

NMR has been used study the flexible nature of proteins for decades (Snyder et al., 1975; Wüthrich and Wagner, 1978). NMR is possible because many nuclei present in organic molecules have a characteristic spin and, as a result, generate their own magnetic field. When a strong, external magnetic field is applied to the molecule the dipoles of the nuclei will align, and the different types of nuclei will change the orientation of their rotational axis, or “precess” in characteristic ways (Kwan et al., 2011). Pulses of electromagnetic radiation are then used to disrupt the dipole alignment, the disruption is measured by a detector and gives rise to the characteristic NMR spectra (Kwan et al., 2011). NMR is able to give structural
information regarding the arrangement of the nuclei in the molecule because each nucleus will be affected by the disruption to a different degree, and thus give off a different signal. The disruption that a particular nucleus experiences depends on the chemical environment in which it sits. A nucleus in an electron dense area is shielded from a larger proportion of the disruption than a nucleus in an electron sparse area (Kwan et al., 2011). Thus, NMR can provide information regarding both the nuclei that are present and their arrangement.

Importantly for studies of molecular motion, NMR is recorded over a time-scale and shifts in the spectra indicate changes in the arrangement of nuclei – in other words, conformational change. In this manner, NMR has been used to demonstrate the conformational plasticity that β2M can utilise in order to accommodate different MHC alleles (Beerbaum et al., 2013). NMR has also been used to highlight how conformational fluctuation of HLA-B*35:01 can subtly differ dependent on the peptide it is presenting (Yanaka et al., 2014).

HDX, on the other hand, monitors the exchange of hydrogen for deuterium in the amide backbone of a protein (Giladi and Khananshvili, 2020). Small quantities of protein (< 0.1 mg) are diluted in a D2O buffer and HDX occurs. After a predetermined amount of time has elapsed the reaction can be quenched by altering the pH from a neutral pH to an acidic pH and reducing the temperature, this slows the reaction to an almost complete stop. The protein is then digested with protease and the fragments analysed by mass spectrometry (Giladi and Khananshvili, 2020). HDX is particularly amenable to analysis of molecular motions because the degree of HDX that can occur within a set time is dependent on the solvent accessibility and structural rigidity of the protein. HDX of different proteins (or MHC molecules holding different peptides) can be compared as a proxy for flexibility (Giladi and Khananshvili, 2020).

HDX was used by Van Hateren et al. to demonstrate that the peptide-receptive (unbound) MHC molecule is more dynamic that the peptide-bound complex (Van Hateren et al., 2017). Using UV hydrolysable peptides, they showed that HDX increased following peptide hydrolysis, and decreased upon addition of peptide. HDX changes were observed in the peptide-binding domain, as may be expected, but also throughout the molecule - including the α3 domain and β2M (Van Hateren et al., 2017).

Hawse et al. also used HDX to analyse the peptide-dependence of MHC flexibility. They showed that varying the peptide cargo alters the HDX behaviour of HLA-A2. The variation in exchange behaviour was attributed to peptide-dependant motions within the HLA-A2 binding groove (Hawse et al., 2013). They suggested that varying the peptide cargo may
result in knock-on effects that alter the thermodynamic barriers between conformational states – facilitating or impeding conformational change. Differing peptide cargo might be responsible for an alteration in the range of available conformational states that exist and are accessible to the MHC molecule, or might influence the entropic cost for receptor binding (Hawse et al., 2013). Ultimately, varying the peptide cargo of an MHC molecule might produce consequent variations in its free energy landscape, which may have functional consequence. Like the Van Hateren study, Hawse and colleagues demonstrate that dynamic changes occur throughout the pMHC molecule (Hawse et al., 2013).

A dynamic pMHC could theoretically influence the TCR-pMHC interaction in a number of ways. A more dynamic pMHC may sample the ‘optimal’ conformation for TCR binding more frequently and therefore display improved TCR binding. In contrast, a more dynamic pMHC may confer a higher entropic cost when binding a TCR, and as a result make a more difficult target (Ayres, Corcelli and Baker, 2017). This characteristic is therefore an important consideration to the overall understanding of the TCR-pMHC interaction.

To this end, I set about furthering the work started by Van Hateren et al. and Hawse et al. in exploring pMHC flexibility. Benefiting from a collaboration with the Pudney laboratory at Bath University, I aimed to investigate whether the red edge excitation shift (REES) phenomenon and pressure/temperature (p/T) dependant fluorimetry could probe the role of the peptide cargo in modulating pMHC flexibility, thereby adding to the experimental toolkit that can be used to explore this characteristic.

5.1.3 – The REES phenomenon
The REES phenomenon, first noted in 1970 (Galley and Purkey, 1970; Weber and Shinitzky, 1970), describes the red shift in the maximum emission intensity wavelength (centre of spectral mass, CSM) that is measured when a fluorophore is excited with wavelengths towards the red edge of the spectrum (Demchenko, 2002; Catici et al., 2016). The observation did not adhere to the commonly accepted model known as Kasha’s rule (Demchenko, 2002). Kasha’s rule states that the emission spectra of a fluorophore will be independent of the excitation wavelength used provided that the excitation wavelength is within the absorption range for that fluorophore (Demchenko, 2002).

Kasha’s rule holds true when the fluorescence lifetime is longer than the time taken for solvent relaxation (Chattopadhyay and Haldar, 2014). When a fluorophore is raised to an excited state it and the surrounding solvent (the molecule and buffer) will rapidly relax into
the lowest vibrational energy level of the excited state (Chattopadhyay and Haldar, 2014). In a fully solvated environment (such as a denatured protein or a free fluorophore) solvent relaxation is much faster than the fluorescence lifetime, so emission occurs from a fully relaxed state, following Kasha’s rule. In contrast, if there are environment-induced restrictions on solvent relaxation, such as a rigid protein backbone, solvent relaxation is retarded and may be incomplete (Demchenko, 2002; Sourav and Chattopadhyay, 2007; Chattopadhyay and Haldar, 2014). If relaxation to the lowest vibrational energy does not occur within the fluorescent lifetime it results in higher energy emission following excitation at the standard wavelength (Chattopadhyay and Haldar, 2014). Different excitation wavelengths will result in different emission spectra. Increasing the excitation wavelength to longer wavelengths, and therefore using lower energy excitation, photoseselects for specific solvation states within the range of available conformational states that exists for that protein (described by its free energy landscape) (Chattopadhyay and Haldar, 2014). Using longer excitation wavelengths selects for fluorophores that are in a more solvent-relaxed state as they need less energy to become activated (Chattopadhyay and Haldar, 2014). They therefore emit at a lower energy. Experimentally, increasing the excitation wavelength results in a red shift in the CSM (Chattopadhyay and Haldar, 2014).

There are a range of conformational states available to any given protein, the REES phenomenon provides information about the dynamic behaviour of a protein by giving an approximation of the breadth of this range (Demchenko, 2002; Catici et al., 2016). Thus, in the context of this study we can use REES to investigate whether changing the peptide cargo alters the range of conformational states available to HLA-A2.

The REES effect was shown to demonstrate pressure/temperature ($p/T$) dependence in 2016 (Catici et al., 2016). The aforementioned range of available conformational states can be perturbed by non-denaturing pressures and temperatures. Temperature perturbs the range of conformational states due to differences in heat capacity between the discrete conformational states available to the protein (Akasaka, 2006). Pressure affects the range of conformational states through volume differences of the different conformational states (Akasaka, 2006). Thus, this perturbation can alter the free energy landscape of a protein (Akasaka, 2006; Catici et al., 2016). An analysis tool developed by Catici et al. (Catici et al., 2016) allows extraction of thermodynamic parameters, such as Gibb’s free energy ($\Delta G$), from fluorescence emission data taken under varying pressure and temperature conditions.
Using REES and p/T dependant fluorimetry, I probed the free energy landscapes of six pMHC complexes that can be recognised by T cells expressing the 1E6 TCR (see Table 5.1). The 1E6 TCR, which was isolated from a patient with type 1 diabetes, recognises a peptide derived from preproinsulin (PPI) (Skowera et al., 2008) in the context of the diseases risk allele HLA-A2 (Nejentsev et al., 2007). Our laboratory previously solved the structure of this TCR in complex with its cognate PPI-derived peptide (Bulek et al., 2012). This was the first ever human autoimmune TCR-pMHC class I structure. The 1E6 T cell is known to recognise well over a million different peptides as well as, or better than, the PPI ‘index’ sequence (Wooldridge et al., 2012). Structural studies showed that recognition of antigen by the 1E6 TCR is focussed around a ‘hotspot’ in the centre of the peptide (Cole et al., 2016). This hotspot was used to identify pathogen-derived peptides that could bind to HLA-A2 and act as potent activators of 1E6 T cells; and thereby provide a potential triggering mechanism of the autoimmune disease (Cole et al., 2016). This model system was selected due to this abundance of previously collected structural and biochemical data, and my particular interest in pathogenic-causes of autoimmune disease (explored further in Chapter 3). Here, I made use of these well-described ligands to examine whether the peptide cargo could tune HLA-A2 flexibility.

5.1.4 – Aims
The data presented here follows on from work by Van Hateren et al. (Van Hateren et al., 2017) and Hawse et al. (Hawse et al., 2013), using an autoimmune-relevant system. Their work demonstrated that the dynamic regions of the MHC that are not limited to the small stretch of MHC surrounding the peptide, but occur throughout the helices and beta sheet – including in areas known to interact with the TCR and its co-receptor, CD8 (Hawse et al., 2013; Van Hateren et al., 2017). These studies describe two major observations that are important to my continuation with this investigation. Firstly, that the peptide cargo – small relative to the much larger MHC molecule – can modulate the dynamic behaviour of the MHC in which it is held. Secondly, that this alteration in dynamic behaviour occurs throughout the pMHC molecule and thus may have functional consequence within the TCR-pMHC interaction.

My overall aim for this project was to investigate the potential dynamic nature of the pMHC complex using fluorescent techniques, and to assess whether the peptide would influence the dynamic behaviour of the complex. To do this I planned to compare dynamic behaviour
of the same MHC molecule (HLA-A2) when presenting different peptides. The specific aims where to:

- Investigate whether the REES phenomenon may be an applicable technique to study protein flexibility within the TCR-pMHC interaction, in particular the peptide-dependent molecular dynamics of HLA-A2. It has previously been used to investigate enzymes (Catici et al., 2016).
- Determine whether HLA-A2 complexes demonstrate a REE shift, thus acting as proxy to indicate that the structure has access to a range of conformational states.
- Investigate the influence of the peptide on the peptide-HLA-A2 complex as a whole, particularly whether altered peptide ligands (APLs) can influence the flexible nature of the HLA-A2 molecule in which they are held.
5.2 – Results

5.2.1 – Producing stable pMHC complexes
For this project six soluble pMHC complexes (Table 5.1) were refolded and purified. This involved: (a) expression of the HLA-A2 and β2M constructs separately in Escherichia coli; (b) purifying the resulting inclusion bodies; (c) refolding the two protein chains together with a peptide by slow dilution of denaturing compounds; and (d) purification of the refolded protein by anion exchange chromatography followed by gel filtration size exclusion chromatography. Figure 5.1 shows a representative example of the purification steps.

<table>
<thead>
<tr>
<th>Amino Acid Sequence</th>
<th>Affinity for 1E6 TCR (K_D, μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A L W G P D P A A A</td>
<td>208</td>
</tr>
<tr>
<td>R Q W G P D P A A V</td>
<td>7.8</td>
</tr>
<tr>
<td>Y Q F G P D F P T A</td>
<td>7.4</td>
</tr>
<tr>
<td>M V W G P D P L Y V</td>
<td>~600</td>
</tr>
<tr>
<td>R Q F G P D W I V A</td>
<td>44.4</td>
</tr>
<tr>
<td>R Q F G P D F P T I</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Table 5.1: The peptides used in Chapter 5. All were refolded with HLA-A*0201 and β2M. All can be recognised by T cells expressing the 1E6 TCR to various degrees as indicated (Cole et al., 2016). Affinity data as previously described (Cole et al., 2016). The conserved “hotspot” central GPD motif is indicated in bold text.
Figure 5.1: Production of pure pMHC complexes. A) Representative anion exchange chromatography of pMHC refolded via *E. coli* expression on an 8 mL POROS 50HQ column. After binding to the column complexes were eluted by means of a NaCl gradient. Complexes elute as a single peak, fractions corresponding to this peak are then analysed by SDS-PAGE. B) Coomassie-stained SDS-PAGE of anion exchange fractions. Samples are analysed under both reducing and non-reducing conditions. Two bands are seen if a pMHC is present, one at ~34 kDa corresponding to the HLA-A2 heavy chain and one at ~12 kDa corresponding to β2M. C) Following SDS-PAGE analysis, fractions observed to have the correct bands are concentrated and further purified by size exclusion. Gel filtration chromatography using a Superdex S200 column is used for further purification of refolded pMHC. D) The resulting gel filtration peak is analysed on Coomassie-stained SDS-PAGE to confirm the presence of the pMHC with no contaminants. Fractions are then pooled, concentrated, and frozen at -80°C.
5.2.2 – HLA-A2-ALWGPDPAAA complexes show a significant REES signal

In order to test whether: (a) this approach could be used to characterise pMHC molecules; and (b) my peptide-HLA-A2 complexes (Table 5.1) had access to multiple different conformational states, I generated soluble HLA-A2 in complex with a peptide (ALWGPDPAAA) derived from human preproinsulin. This epitope is known to be a key target on pancreatic beta cells for autoreactive CD8+ T cells in type 1 diabetes patients (Nejentsev et al., 2007; Skowera et al., 2008; Bulek et al., 2012; Wooldridge et al., 2012; Cole et al., 2016). Prior to all experiments in Bath, dynamic light scattering (Section 2.2.1) was used to confirm that freeze-thawed proteins had not aggregated or unfolded. Peaks between 1-10 nm represented refolded proteins.

For REES experiments I used tryptophan fluorescence as a molecular reporter. Tryptophan was selected as it is present in the HLA-A2 molecule (12 tryptophan in the MHC itself, with some of the selected peptides adding a 13th, see Table 5.1) and is the dominant natural fluorophore in proteins. As standard, tryptophan was excited at 295 nm to avoid energy transfer to any tyrosine residues in the protein (generally excited at 280 nm) (Moon and Fleming, 2011). For this reason, a wavelength range of 292-310 nm was selected for REES experiments.

Figure 5.2 shows the effect of increasing the excitation wavelength on HLA-A2-ALWGPDPAAA tryptophan emission. Maximum emission intensity ($\Delta \lambda_{\text{max}}^\text{Em}$) of the 13 tryptophan fluorophores within the HLA-A2-ALWGPDPAAA complex fell substantially over the 292-310 nm excitation wavelength range (Figure 5.2A). The complete fluorescence intensity spectra for all pMHC molecules are shown in Figure 5.4.

The tryptophan emission underwent a significant redshift across this excitation wavelength range (Figure 5.2B). This is demonstrated by a change in CSM, equation 1.

$$CSM = \frac{\Sigma(f_i \times \lambda_{Em})}{\Sigma(f_i)}$$  \hspace{1cm} (1)

In which $f_i$ is the measured fluorescence intensity, and $\lambda_{Em}$ is the emission wavelength.

For HLA-A2-ALWGPDPAAA, I observed a $\Delta$CSM of 11.78 nm, see Table 5.2. This indicates that the complex is unable to fully relax from its excited state during the fluorescent lifetime, this was expected as the complex is a quaternary folded protein structure. The shift observed also suggests that HLA-A2-ALWGPDPAAA can access a range of different solvation states, which can be photoselected for using different excitation wavelengths. These data therefore
demonstrated that the REES approach may provide a useful tool in investigating the flexible nature of pMHC complexes, and validated my continuation with the project.

Figure 5.2: HLA-A2 complexes show a significant REES signal. A) Relative change in emission intensity of HLA-A2-ALWGDPAAA (HLA-A2-ALW) complex across an excitation range of 292-310 nm, $\Delta \lambda_{\text{max}}^{E_{\text{m}}}$ fell significantly over the range of excitation wavelengths used. B) Change in CSM over the same excitation wavelength range. CSM calculated according to equation 1. ALW= HLA-A2-ALWGDPAAA.
5.2.3 – ΔCSM varies with different peptides suggesting the molecular flexibility of each complex is different

My initial REES results demonstrated the value of the REES technique in probing the dynamic nature of the pMHC. The data indicated that HLA-A2-ALWGPDPAAA has access to a range of conformational states. As such, I next investigated whether the same was true for other HLA-A2 complexes, and importantly for my hypothesis, whether any differences in flexibility exist between them. I used five well characterised altered peptide ligands of HLA-A2-ALWGPDPAAA which can all be recognised by the HLA-A2-ALWGPDPAAA-specific TCR, 1E6 (Cole et al., 2016) (Table 5.1). The complete fluorescence intensity spectra for each complex are shown in Figure 5.4. Each spectrum is the average of data collected in triplicate.

For each complex a significant redshift was observed in CSM, demonstrating that each complex experiences a degree of structural restriction in its ability to relax around the excited fluorophores (Figure 5.3). Furthermore, they are likely able to access a range of conformational states within their conformational equilibria. Importantly the ΔCSM of the complex varied with peptide cargo, demonstrating that peptide cargo can alter the flexibility of the whole complex. ΔCSM of 11.78, 6.69, 9.26, 8.24, 10.39, 11.77 nm were observed for HLA-A2-ALWGPDPAAA, HLA-A2-YQFGPDPAAV, HLA-A2-MVWGPDPILYV, HLA-A2-RQWGPDPAAA, HLA-A2-RQFGPDPYVA, and HLA-A2-RQFGPDGPTI complexes respectively (Figure 5.3). The CSM for each complex at each excitation wavelength is shown in Table 5.2.

The complexes demonstrated differences in their REES behaviour, indicating that they have different dynamic characteristics. As all complexes are HLA-A2, with only a differing peptide cargo, this difference in dynamic behaviour demonstrates the influence of the bound peptide on pMHC flexibility. Furthermore, the HLA-A2 complexes are structurally similar (Cole et al., 2016), and therefore the solvent accessible surface area (SASA) available for fluorescent excitation is similar, Supplementary Table 7.1. Thus, the difference in REES effect between the complexes observed here may indicate a difference in the availability of conformational states to each complex.
Figure 5.3: ΔCSM varies for pMHC complexes carrying different peptide cargo. A) Change in CSM over the same excitation wavelength range for each pMHC complex. CSM calculated according to equation 1. B) ΔCSM of 11.78, 6.69, 9.26, 8.24, 10.39, 11.77 nm were observed for ALW, YQF, MVW, RQW, RQFa, and RQFi complexes respectively. C) ΔCSM data for each pMHC complexes as a histogram allows visual depiction of the differences. D) Statistical significance for observed differences according to unpaired two-sample T-test. RQFi = HLA-A2-RQFGPDGPTI. ALW= HLA-A2-ALWGPDPAAA. RQFa = HLA-A2-RQFGPDWIVA. MVW = HLA-A2-MVWGPDPYLV. RQW = HLA-A2-RQWGDPDPAAV. YQF = HLA-A2-YQFGPDPAAV.
Table 5.2: The change in CSM over the 292-310 nm excitation range for each protein. Each peptide-HLA-A2 complex demonstrated a red-shift in the maximum emission intensity when excited with wavelengths of decreasing energy. Table shows the CSM measured at each excitation wavelength within the range of 292-310 nm. RQFi = HLA-A2-RQFGPDGPTJ, ALW= HLA-A2-ALWGPDPAAA. RQFa = HLA-A2-RQFGPDWIVA. MVW = HLA-A2-MVWGPDPYV. RQW = HLA-A2-RQWGPDPAAV. YQF = HLA-A2-YQFGPDPAAV.

<table>
<thead>
<tr>
<th>dEx(nm)</th>
<th>ALW</th>
<th>YQF</th>
<th>RQFa</th>
<th>RQFi</th>
<th>MVW</th>
<th>RQW</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0001</td>
<td>363.9714</td>
<td>362.5138</td>
<td>363.1715</td>
<td>363.8147</td>
<td>363.7469</td>
<td>363.2583</td>
</tr>
<tr>
<td>1</td>
<td>364.0236</td>
<td>362.4368</td>
<td>363.166</td>
<td>363.8664</td>
<td>363.7908</td>
<td>363.2389</td>
</tr>
<tr>
<td>2</td>
<td>364.0555</td>
<td>362.5134</td>
<td>363.2079</td>
<td>364.1293</td>
<td>363.8601</td>
<td>363.3324</td>
</tr>
<tr>
<td>3</td>
<td>364.1857</td>
<td>362.6098</td>
<td>363.4577</td>
<td>364.2485</td>
<td>363.8905</td>
<td>363.4188</td>
</tr>
<tr>
<td>4</td>
<td>364.2868</td>
<td>362.6924</td>
<td>363.4963</td>
<td>364.4822</td>
<td>364.0316</td>
<td>363.4759</td>
</tr>
<tr>
<td>5</td>
<td>364.5501</td>
<td>362.8325</td>
<td>363.5149</td>
<td>364.7304</td>
<td>364.1322</td>
<td>363.5948</td>
</tr>
<tr>
<td>6</td>
<td>364.9223</td>
<td>362.9942</td>
<td>363.7949</td>
<td>364.9785</td>
<td>364.384</td>
<td>363.6811</td>
</tr>
<tr>
<td>7</td>
<td>365.117</td>
<td>363.0174</td>
<td>364.1335</td>
<td>365.3159</td>
<td>364.6276</td>
<td>363.8299</td>
</tr>
<tr>
<td>8</td>
<td>365.436</td>
<td>363.2724</td>
<td>364.4807</td>
<td>366.0476</td>
<td>364.9234</td>
<td>364.2925</td>
</tr>
<tr>
<td>9</td>
<td>366.0682</td>
<td>363.4398</td>
<td>364.8083</td>
<td>366.6646</td>
<td>365.3219</td>
<td>364.3782</td>
</tr>
<tr>
<td>10</td>
<td>366.5828</td>
<td>363.9184</td>
<td>365.2722</td>
<td>367.2613</td>
<td>365.6973</td>
<td>364.8083</td>
</tr>
<tr>
<td>11</td>
<td>367.1234</td>
<td>364.2783</td>
<td>365.6192</td>
<td>368.1449</td>
<td>365.9969</td>
<td>365.1253</td>
</tr>
<tr>
<td>12</td>
<td>367.5902</td>
<td>364.7599</td>
<td>366.3532</td>
<td>368.9496</td>
<td>366.669</td>
<td>365.8061</td>
</tr>
<tr>
<td>13</td>
<td>368.493</td>
<td>365.2083</td>
<td>367.2738</td>
<td>369.8616</td>
<td>367.1111</td>
<td>366.3699</td>
</tr>
<tr>
<td>14</td>
<td>369.2997</td>
<td>365.6555</td>
<td>367.7869</td>
<td>370.9683</td>
<td>367.7268</td>
<td>366.7893</td>
</tr>
<tr>
<td>15</td>
<td>370.9056</td>
<td>365.9251</td>
<td>369.0839</td>
<td>372.3284</td>
<td>368.9611</td>
<td>367.8077</td>
</tr>
<tr>
<td>16</td>
<td>371.8568</td>
<td>366.9338</td>
<td>369.9415</td>
<td>373.6321</td>
<td>369.935</td>
<td>368.4971</td>
</tr>
<tr>
<td>17</td>
<td>374.094</td>
<td>368.089</td>
<td>371.9123</td>
<td>375.6312</td>
<td>371.3041</td>
<td>370.0844</td>
</tr>
<tr>
<td>18</td>
<td>375.7557</td>
<td>369.2044</td>
<td>373.5568</td>
<td>377.5811</td>
<td>373.0154</td>
<td>371.5005</td>
</tr>
</tbody>
</table>

Table 5.2: The change in CSM over the 292-310 nm excitation range for each protein. Each peptide-HLA-A2 complex demonstrated a red-shift in the maximum emission intensity when excited with wavelengths of decreasing energy. Table shows the CSM measured at each excitation wavelength within the range of 292-310 nm. RQFi = HLA-A2-RQFGPDGPTJ, ALW= HLA-A2-ALWGPDPAAA. RQFa = HLA-A2-RQFGPDWIVA. MVW = HLA-A2-MVWGPDPYV. RQW = HLA-A2-RQWGPDPAAV. YQF = HLA-A2-YQFGPDPAAV.
Figure 5.4: Fluorescence intensity spectra for each HLA-A2 complex. Protein-intrinsic tryptophan fluorescence was excited using increasing excitation wavelengths (ranging 292-310) and the emission spectra measured using a Perkin Elmer LS50B Luminescence Spectrometer (Perkin Elmer, Waltham, MA, USA). Spectra are the average of triplicate collection.
5.2.4 – Peptide cargo modulates energy barriers between conformational states

Thus far, I had shown that the REES phenomenon could be used to investigate the dynamic behaviour of the pMHC, and that this behaviour varied dependant on the peptide cargo. Each complex demonstrated a REE shift, but the degree of the shift was different for different complexes. This indicated that each complex had access to a range of conformational states, but that this range is likely different for different complexes.

Prompted by this, I hoped to further investigate the flexible nature of the pMHC complexes. I was fortunate to benefit from the extensive expertise of Dr Dragana Catici and Dr Chris Pudney from my collaborator group who not only taught me to perform the experiments described in this chapter, but employed their novel analysis tool (see Supplementary Equation 7.1, published only a few months before I began my PhD) to my data to extract critical thermodynamic information (Catici et al., 2016).

For use with the Pudney analysis tool, I next collected emission spectra from each complex, excited at the standard wavelength, over a range of non-denaturing varying pressures and temperatures. The full data is shown in Figures 7.4 to 7.9. Each spectrum, at each pressure and each temperature, was collected in triplicate.

I hoped to use this pressure/temperature (p/T)-dependant fluorimetry to probe how easily a pMHC molecule might switch between the conformational states available to it. As demonstrated by Catici et al. (Catici et al., 2016), varying the pressure and temperature of the system alters the range of the conformational states available to the molecules, and allows extraction of a number of thermodynamic parameters using the Pudney group analysis tool (Catici et al., 2016).

Fluorescence intensity data collected under varying, non-denaturing pressures and temperatures was inputted into the analysis tool. Analysis kindly performed by Dr Chris Pudney. The tool generates a number of thermodynamic parameters – some of which can be used as proxy for flexibility. Of particular interest to this study was a ΔG value for each complex, ΔG describes the energy barriers between conformational states experienced by the complex. Figure 5.5 shows the predicted ΔG values generated by this method for each pMHC complex. ΔG values ranged from 8.1 kJ mol⁻¹ (HLA-A2-RQFGPDGPI as the lowest) to 8.7 kJ mol⁻¹ (HLA-A2-RQFGPDWIVA as the highest). Each complex had a different predicted ΔG value, some of which differed outside of standard error. This indicates that the energy barriers faced by the pMHC complexes attempting to change conformational states vary
dependent on the peptide cargo. For example, it may be easier for HLA-A2-RQFGPDGPI to change conformational state, than for HLA-A2-RQFGPDWIVA. These data highlight that the relatively small string of ten amino acids sat in the binding groove can have a measurable impact on the entire ~45 kDa complex.

<table>
<thead>
<tr>
<th>pMHC complex</th>
<th>ΔG</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQFa</td>
<td>8.76726</td>
<td>0.09019</td>
</tr>
<tr>
<td>RQFi</td>
<td>8.13560</td>
<td>0.10125</td>
</tr>
<tr>
<td>ALW</td>
<td>8.71765</td>
<td>0.18125</td>
</tr>
<tr>
<td>MVW</td>
<td>8.76007</td>
<td>0.22244</td>
</tr>
<tr>
<td>RQW</td>
<td>8.70993</td>
<td>0.19335</td>
</tr>
<tr>
<td>YQF</td>
<td>8.34947</td>
<td>0.14200</td>
</tr>
</tbody>
</table>

ΔG of pHLA complexes

Figure 5.5: ΔG varies for pMHC complexes carrying different peptide cargo. ΔG values generated from p/T-dependent fluorimetry data inputted into the Pudney group analysis tool demonstrate that ΔG varies dependent on the peptide cargo carried by the HLA-A2 molecule. Some of this variation is outside of error suggesting that these pMHC complexes face true differences in the thermodynamic barriers between conformation states. RQFi = HLA-A2-RQFGPDGPI, ALW = HLA-A2-ALWGPDPAA. RQFa = HLA-A2-RQFGPDWIVA, MVW = HLA-A2-MVWGPDPDLYV. RQW = HLA-A2-RQWGPDPAAV. YQF = HLA-A2-YQFGPDPAAV. Analysis performed by Dr Chris Pudney.
5.2.5 – Peptide-MHC contacts might drive global flexibility of the complex
The data I have presented in this chapter supports previous reports (Hawse et al., 2013; Van Hateren et al., 2017) that the pMHC is a flexible structure, and that that flexibility can be tuned by the peptide cargo it carries. As such I wondered if the degree of connectivity between the peptide and the MHC molecule itself, namely the number of bonds formed between them, could be a driving force behind these observations.

CCP4 analysis performed by Aaron Wall revealed that, amongst the 6 pMHCs discussed here, the peptide sequence significantly impacted the number of bonds between the MHC and its cargo, Figure 5.6A. The cognate, PPI-derived peptide ALWGPDPAAA formed 155 bonds (defined as atoms that are less than 4 Å apart) with HLA-A2, whereas RQFGPDGPTI forms over 80 more. However, linear regression analysis of the data revealed a non-significant correlation between ΔG and the number of bonds between peptide and MHC, Figure 5.6B, (correlation according to linear regression, p=0.09, R square 0.55). Furthermore, the range of the available conformational states (as defined by ΔCSM) appears to be similarly unaffected by the number of bonds between MHC and peptide cargo, Figure 5.6B.
Figure 5.6: Peptide–MHC contacts do not correlate with flexibility of the complex. A) The number of bonds between peptide cargo and the HLA-A2 molecule for each complex was quantified using CCP4 software by Aaron Wall. B) A notable, but non-significant trend was noted between number of bonds and the ΔG (correlation according to linear regression, p=0.093, R square 0.55) but C) number of bonds did not influence ΔCSM. 

- RQFi = HLA-A2-RQFGPDGPT
- ALW = HLA-A2-ALWGPDAAA
- RQFα = HLA-A2-RQFGPDWIVA
- MVW = HLA-A2-MVWGPDPPLYV
- RQW = HLA-A2-RQWGPDPAAV
- YQF = HLA-A2-YQFGPDPAAV

<table>
<thead>
<tr>
<th>pMHC complex</th>
<th>No. bonds between peptide and MHC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALW</td>
<td>155</td>
</tr>
<tr>
<td>MVW</td>
<td>167</td>
</tr>
<tr>
<td>RQFα</td>
<td>200</td>
</tr>
<tr>
<td>RQFβ</td>
<td>237</td>
</tr>
<tr>
<td>RQW</td>
<td>200</td>
</tr>
<tr>
<td>YQF</td>
<td>199</td>
</tr>
</tbody>
</table>
5.3 – Discussion
In general, crystal structures of numerous pMHC complexes have strong similarities (Wieczorek et al., 2017). This has understandably propagated a view that the pMHC is a rigid structure (Wilson and Fremont, 1993). However, conformational plasticity within the pMHC is being increasingly noticed. Thus, it is likely that there is inconspicuous conformational plasticity within the pMHC that is undetectable by standard crystallographic methods.

5.3.1 – Evidence for a flexible pMHC
In addition to the work discussed in Section 5.1.2 (Hawse et al., 2013; Van Hateren et al., 2017), differences in conformational plasticity of the pMHC molecule have been demonstrated between different, but closely related MHCs, with the observed dynamical difference being attributed to the polymorphism (Pöhlmann et al., 2004; Bailey et al., 2014). In this study, I compared the same MHC molecule with varying peptide cargo, in an attempt to investigate what effect the peptide may have on the complex as a whole.

Peptide-dependant conformational fluctuations in the context of HLA-B*35:01 have been demonstrated using NMR spectroscopy (Yanaka et al., 2014). NMR is one of the most powerful tools for studying protein flexibility. Yet the size of the pMHC molecule can present challenges (Hawse et al., 2013). As such, we and others have sought different techniques for investigating the dynamics of the pMHC molecule.

Two key studies of pMHC dynamics by Hawse et al. and Van Hateren et al. both used HDX to demonstrate flexibility within the pMHC (Hawse et al., 2013; Van Hateren et al., 2017). Peptide-dependant effects on pMHC dynamics have the potential to influence the interaction of that pMHC with its TCR (Borbulevych et al., 2009; Hawse et al., 2013). As suggested by Hawse et al. the knock-on effects of the peptide cargo might alter the thermodynamic barriers between conformational states – facilitating or impeding conformational change. Differing peptide cargo might be responsible for an alteration in the range of available conformational states that exist and are accessible to the pMHC molecule, contributing to which pMHCs can be recognised and which cannot. Some peptides may restrict the available conformational states in such a way to limit or impede access to conformations capable of binding the TCR by raising the energy barriers between states. In addition, peptide-dependent dynamics might influence the entropic cost for receptor binding (Hawse et al., 2013). Ultimately, varying the peptide cargo of an MHC molecule might produce consequent variations in its free energy landscape.
5.3.2 – REES as a tool for exploring pMHC flexibility

With the aim of further probing the role of the peptide in MHC flexibility, I turned to the REES phenomenon. REES has been previously used to qualitatively measure changes to the free energy landscape of a protein, which describes the range of available conformational states that exists for that protein (Catici et al., 2016). Catici et al used REES to describe the ‘ruggedness’ of the free energy landscape of a metabolic enzyme (Catici et al., 2016).

REES has also been used to demonstrate rigidity within an endocrine receptor (Mishra and Jha, 2019). Having the advantage of a single-tryptophan molecule, Mishra and Jha were able to show that a very specific loop of the molecule was held in a motionally-restricted conformation, as it displayed a REE shift of just 4 nm (Mishra and Jha, 2019). The ability to describe the dynamic behaviour of such a small region of the molecule demonstrates the strength of REES as a structural biology tool.

Here, I used the REES phenomenon to probe the free energy landscape of six HLA-A2-peptide complexes. The six complexes each carried different peptide cargo. The peptides chosen were all altered peptide ligands based off a fragment of preproinsulin that is presented to the 1E6 TCR and implicated in T1D (Cole et al., 2016), Table 5.1. First, using HLA-A2-ALWGPDPAAA, I demonstrated that the REES phenomenon was applicable to the study of pMHCs. HLA-A2-ALWGPDPAAA displayed a clear red-edge shift implying that the complex cannot fully relax within the fluorescent lifetime and can access a range of different solvation states, which can be photoselected for by increasing the excitation wavelength used.

Repeating this experiment for all six HLA-A2 complexes I showed that whilst all complexes demonstrated a shift, there were differences as to the degree of the shift. As all complexes are HLA-A2, with differing peptide, this difference in dynamic behaviour may be attributed to the influence of the peptide cargo. Given that the HLA-A2 complexes are structurally similar, and therefore the SASA for fluorescent excitation is similar, the difference in REES effect between the complexes observed here may indicate a difference in the availability of conformational states to each complex. This supports the peptide-dependant effect on pMHC motion that was observed previously (Hawse et al., 2013; Van Hateren et al., 2017).

This suggestion that the free energy landscape might vary between HLA-A2 complexes prompted further investigation, for which I was directed to p/T-dependant fluorimetry. The pressure-dependence of the REES effect had recently been demonstrated for the first time by the Catici et al. (Catici et al., 2016), who showed that p/T-dependant fluorimetry is a
powerful tool for tackling issues in structural biology, derived from the unwaveringly static nature of crystallographic data (Catici et al., 2016).

Using non-denaturing pressures and temperatures it is possible to perturb the range of conformational states that is available to the protein being studied. The resulting fluorescence intensity data can be inputted into an analysis tool developed by the Pudney group (Catici et al., 2016) to extract parameters that describe the thermodynamic behaviour of proteins. Of particular interest to my study was the $\Delta G$ values, describing the energy barriers each complex experiences between the conformational states available to it. I collected emission spectra for each complex under varying non-denaturing pressures and temperatures. These data were used to calculate a singular predicted $\Delta G$ value for each complex. These values represent a proxy estimate for the ease with which a particular complex can move between its available conformational states. $\Delta G$ values differed between the six pMHC complexes, suggesting that the complexes might face different energy barriers between conformational states depending on the peptide they present.

The peptide, being a single short chain of amino acids and not a quaternary folded structure like the MHC itself, is well-documented as being a flexible component of the interaction. It can be both flattened into (Tynan et al., 2007) and pulled out of the peptide binding groove (Madura et al., 2015). I wondered whether the peptide influenced the MHC flexibility by ‘pulling’ it into a state of greater molecular motion, and whether increased contacts between peptide and MHC would result in a more flexible complex. CCP4 analysis revealed a non-significant but notable negative trend between $\Delta G$ and the quantity of bonds between the two parts of the complex. This would suggest that increased connectivity between the MHC and its peptide cargo favors a more dynamic molecule, however more data is required to verify this conjecture.

Whilst there are numerous contributing factors to the efficacy of a TCR-pMHC interaction, I have demonstrated that pMHC flexibility could potentially be one of them and that the REES phenomenon might be employed to investigate it.

5.3.3 – Summary
In summary, I showed that the REES phenomenon could be used as a tool to investigate the dynamic behaviours of immune proteins. Furthermore, these data demonstrated that pMHC complexes are subject to peptide-driven modulation of the free energy landscape. Altering the peptide cargo of the same MHC molecule can modify the free energy barriers between

195
the conformational states that are accessible to that molecule (described by \( \Delta G \)). These data would benefit from corroboration by other methodologies such as NMR and HDX.

Furthermore, the data presented in this chapter show HLA-A2-RQFGPDGPTI as having the greatest \( \Delta\text{CSM} \), 13.76 nm (see Figure 5.3), suggesting it has the widest breath of available conformational states (states in different stages of folding can be photoslected using lower energy excitation). HLA-A2-RQFGPDGPTI also had the lowest \( \Delta G \) value, 8.13 kJ mol\(^{-1}\), suggesting it experiences the lowest energy barrier between conformational states of the six complexes I studied. Together this would suggest that HLA-A2-RQFGPDGPTI both has the widest range of available conformational states, and finds it easier to switch between them. It is fairly reasonable, therefore, to say that HLA-A2-RQFGPDGPTI is the most flexible of the complexes assessed here. However, HLA-A2-YQFGPDPAV had the smallest \( \Delta\text{CSM} \), 6.69 nm, suggesting it has the narrowest breath of available conformational states. Yet, \( p/T \) analysis predicted it to have the second lowest \( \Delta G \) value, 8.35 kJ mol\(^{-1}\), suggesting it experiences the second lowest energy barrier between conformational states. This seemingly contrasting information implies that the concept of ‘molecular flexibility’ is more complicated than originally appreciated during the conception of this project. A complex can have access to a relatively limited range of conformations but move easily between them, and thus should still be considered flexible.

The data presented in this chapter suggest that, within the 1E6 TCR-PII-HLA-A2 interaction, the peptide cargo can influence the molecular dynamics of the pMHC complex. Whilst it is beyond the scope of these data to demonstrate a biologically relevant functional consequence for this observation, it is none-the-less interesting to postulate. I, and others, have shown that the peptide – small relative to the much larger MHC molecule – can have a measurable impact on the complex as a whole. This could, at least in part, help to explain why small changes in peptide sequence can alter the T cell response that can be triggered (Valmori et al., 1998; Cole et al., 2010; Madura et al., 2015)

Differences in the flexibility of the pMHC target may influence the ability of the TCR to bind in a number of different ways. A more flexible target may sample the optimal conformation for binding more frequently, and thus form higher affinity interactions. On the other hand, in other systems, a more rigid pMHC molecule – one that naturally favours the optimal conformation for binding - might provide a superior TCR target, as the entropic cost required to bind might be lower.
I have shown that REES and p/T-dependent fluorimetry can be added to the increasing number of tools useful for investigating the dynamic nature of immune proteins, supporting existing techniques such as NMR and HDX. Furthermore, these data demonstrate that the peptide can tune MHC flexibility. The full functional consequences of these observations remains to be seen.
6 – General Discussion

6.1 – Understanding the TCR-pMHC interaction
In this thesis, I have used three clinically relevant TCR-pMHC interactions to investigate the biochemical foundation underpinning the highly plastic interaction between CD8+ T cells and their targets. In Chapter 3, I profiled the peptide recognition repertoire of an auto-reactive, T1D-implicated T cell clone, InsB4. I demonstrated that the InsB4 T cell clone could strongly recognise pathogen-derived peptides and proposed three viral peptides that could potentially have broken immune tolerance and activated this T cell clone in the T1D patient from which it was isolated. In Chapter 4, I investigated the contribution of TCR CDR3 sequence to the TCR-pMHC interaction. Using the same technique as in Chapter 3, I demonstrated that the same crossreactive characteristic that enabled the autoimmune activities of InsB4, allowed for cross-disease crossreactivity in another patient-derived TCR, 868. I showed that the HIV-specific 868 TCR could recognise and respond to a TAA, potentially accounting for the observed response to a number of cancer cell lines. I subsequently highlighted the fragile and extremely specific nature of TCR-peptide recognition by demonstrating that recognition of the TAA in question could be completely eradicated by a single amino acid substitution in the CDR3β. I further demonstrated that new specificities could be conferred on the TCR using other single amino acid substitutions. Finally, in Chapter 5, I explored the potential contribution pMHC dynamics to the TCR-pMHC interaction. I outlined a novel technique for use in exploring pMHC dynamics, and found that pMHC dynamics can be altered by its peptide cargo. Full understanding of plasticity within the TCR-pMHC, to the extent where interactions can be predicted, will require detailed understanding of both complexes in the interaction. Future predictive tools will need to consider both TCR sequence, and the dynamic characteristics of potential pMHC targets.

6.2 – Investigations into the TCR-pMHC interaction
In this thesis, I utilised the strength of combinatorial peptide library (CPL) screens, to uncover TCR peptide preferences. CPL screens are discussed in detail in Section 3.1.2. Their suitability for this work lies in their almost unparalleled ability to predict novel TCR epitopes without any prior knowledge of likely candidates, almost the entire peptide universe can be scrutinised in a single assay. CPL screens provide large quantities of data to fuel antigen prediction tools such as NetTCR (Jurtz et al., 2018). The necessary exclusion of cysteine from
the non-fixed amino acid pools however, may limit discovery of peptides with more than one cysteine residue, warranting the use of other investigative methods to explore the TCR-pMHC interaction on a smaller scale, such as peptide display systems.

6.2.1 – Display systems for studying the TCR-pMHC interface

A peptide display system was used to demonstrate the impact of CDR3 amino acid substitutions in a 2005 study in which Li et al. subjected two HLA-A2-restricted TCRs to affinity maturation – the A6 TCR which is specific for the HTLV-1-derived Tax peptide (LLFGYPVYV, residues 11-19) (Yi Li et al., 2005), and the 1G4 TCR specific for NY-ESO-1 peptide (SLLMWITQC, residues 157-165) (Jäger et al., 1998). Libraries of mutant TCRs were generated containing amino acid substitutions in their CDR loops. These were displayed on the surface of M13 phage by conjugation to a phage surface protein, and subjected to repeated rounds of selection and competitive binding assays to identify variants with enhanced affinity for the pMHC target (Yi Li et al., 2005). In Chapter 4, I used amino acids substitutions at key CDR3 contact positions to alter the peptide recognition repertoire of a TCR. In 2013, the structure of a high affinity A6 variant was solved in complex with Tax-HLA-A2, it showed how just 4 altered residues in the CDR3β loop enabled enhanced contacts with the peptide that resulted in a nearly 1000-fold increase in the affinity of the interaction (Cole, 2013).

In Chapter 4, amino acid substitutions were made in the CDR3 loops of the 868 TCR. Whilst the CDR3 loops generally predominate peptide contacts, it is important to note that amino acid alterations in the other CDR loops can also generate enhanced TCRs. Dunn et al. used phage display to show that amino acid substitutions in the CDR2 loops, which predominantly contact the MHC (Rudolph, Stanfield and Wilson, 2006), could enhance the affinity within the 1G4-A2-NY-ESO-1 interaction without apparent loss of specificity (Dunn, 2006). My own laboratory also demonstrated that CDR2 amino acid substitutions could be used to enhance TCR affinity whilst maintaining exquisite peptide specificity (Cole et al., 2014). Thus, it would be interesting to evaluate the impact of amino acid substitutions in other CDR loops of the 868 TCR in future to gain a more complete understanding of how sequence relates to function within the context of the 868 TCR.

In addition to phage display systems, modified TCRs can be displayed for selection on yeast cells and mammalian cells, which offer more human-like post-translational modifications (Nguyen, Le and Maynard, 2018).
Gee et al. used yeast display to uncover the pMHC ligands for ‘orphan’ TCRs. Orphan TCRs were multimerised using MACS beads and incubated with a yeast display library featuring over 400 million unique peptides. The peptides were displayed by single chain HLA-A2-β2M complexes that were linked to yeast surface protein Aga2. Yeast displaying target pMHC were pulled out by repeated rounds of magnetic separation (Gee et al., 2018). This technique represents a powerful way to screen large peptide libraries without prior knowledge of the likely ligand. However, Gee et al. restricted their library at the P2 and C-terminal anchor positions to reflect the anchor preference of the MHC molecule they were investigating (Gee et al., 2018). Adams et al. also fixed the peptide anchors when investigating a H-2Ld-restricted TCR (Adams et al., 2016). Restricting the anchor residues during peptide display could fail to identify physiologically relevant epitopes that do not use the favoured anchor residues such as the HLA-A2-restricted Melan-A peptide EAAGIGILTV (Speiser et al., 2008; Madura et al., 2019). Indeed, my own group showed that the EAAGIGILTV peptide acted as a much stronger agonist for T cells expressing the MEL5 TCR than the position 2 heteroclitic variant peptide, ELAGIGILTV, once used in the several clinical trials (Ayyoub et al., 2003; Liénard et al., 2004; Speiser et al., 2008). Recognition of the preproinsulin (PPI)-derived epitope ALWGPDAAA by the 1E6 TCR provides a further example of how a weak TCR anchor can generate a substantially more potent peptide antigen. Substitution of the HLA-A2 P2 anchor residue from leucine to glutamine (representing a substitution from the optimal anchor residue to a poor HLA-A2 anchor residue) generated one of the best recognised singly substituted variant of ALWGPDAAA (see figure 2 here, (Bulek et al., 2012)). Indeed all the best agonists of the 1E6 TCR contained the suboptimal glutamine anchor at P2 (Cole et al., 2016) that would not have been discovered using a display system with fixed optimal MHC anchors.

Display systems have been widely and successfully used to identify TCR epitopes (Crawford et al., 2004; Birnbaum et al., 2014; Adams et al., 2016; Gee et al., 2018) and offer the advantage of being able to freely incorporate all 20 proteinogenic amino acids (Gee et al., 2018). Selection is, however, usually limited by the size of the display library, often designed around or bias towards a known epitope (Crawford et al., 2004; Birnbaum et al., 2014; Gee et al., 2018). CPL screens are not limited by the size of the peptide library so do not have to incorporate such biases.
6.2.2 – Rational design based on knowledge of the TCR-pMHC interaction

Affinity-enhanced TCRs that are used in clinical trials have not been through the rigours of thymic selection and therefore may have the capacity to recognise self-antigens. This possibility was tragically demonstrated during trials of an enhanced TCR that targeted the multi-cancer associated MAGE-A3. Adoptive transfer of patient T cells expressing a MAGE A3-specific TCR, that had been produced by phage display, resulted in undesirable off-target crossreactivity leading to fatalities (Cameron et al., 2013; Linette et al., 2013; Raman et al., 2016). These unfortunate events followed previous studies where a different MAGE A3-specific TCR had induced cancer-regression and neurological toxicity (Morgan et al., 2013). In a further example, affinity enhancement of the MART-1-specific DMF5 TCR led to a 400-fold increase in its affinity for the HLA-A2-ELAGIGILTV epitope, but brought about additional self-reactivity (Hellman et al., 2019). Epitopes that were not seen, or seen only weakly by the wild-type TCR were strong ligands for the mutant, such as the AVGIGIAVV peptide from the human CD9 protein, and the IGGIGTVPV epitope from human elongation factor 1α (Hellman et al., 2019). These studies highlight the potential dangers of T cell crossreactivity and how important it is to screen for it prior to clinical use.

The danger of off-target crossreactivity led to suggestions that, instead of random mutagenesis, more rational TCR alterations should be made. In 2010, Bennett et al. collated previously described public TCRs that recognise the HIV-1 epitope HLA-A2-SLYNTVATL, and designed CDR mutants based on motifs present in existing TCRs (Bennett et al., 2010). This more rational approach is based on two assumptions: 1) that the prevalence of public TCRs is indicative of their efficacy at recognising the desired antigen and; 2) that TCRs that are ‘public’ are less likely to encompass dangerous crossreactivities (Bennett et al., 2010). The study resulted in mutant TCRs with moderately increased affinity for the HLA-A2-SLYNTVATL epitope, and a broader recognition of SLYNTVATL escape mutants (Bennett et al., 2010).

The above described studies demonstrate the importance of understanding the TCR-pMHC interaction for the rational design of therapeutics. In Chapter 4, I used structural data published by Cole et al. (Cole, Fuller, et al., 2017) to rationally design a series of mutant TCRs based on the HIV-specific 868 TCR. This involved substituting key TCR-peptide contact residues with the hope, not of increasing the specificity of the 868-A2-SLYNTVATL interaction but of altering the specificity of the TCR completely.
6.2.3 – Further methods for peptide discovery

Understanding how TCRs contribute to disease requires knowledge of what they recognise so it is no surprise that several new techniques for discovering ligands for so-called “orphan TCRs” have arisen during the course of my PhD studies as described below.

6.2.3.1 – T-Scan

T-Scan is a method for epitope discovery that was described by Kula et al. in 2019 (Kula et al., 2019). T-Scan uses lentiviral delivery of genome-wide protein libraries, peptides are then processed endogenously and presented on MHC-I molecules at the surface of the cell, this gives the method a distinct advantage over CPL screens and peptide display methods as T-scan takes into account endogenous processing of peptides (Kula et al., 2019). In Chapter 3, I compensated for this caveat by validating any peptide of interest by lentiviral transfer of the parent gene, to determine if that antigen was genuinely processed and presented. In T-Scan, the target cell library is co-cultured with T cells. Like CPL screening and unlike peptide display systems, T-scan requires both binding and functional activation for a positive readout. The library cells carry a reporter of granzyme B activity, a protein that fluoresces following cleavage by granzyme B, allowing the target cells to be sorted from the non-target cells via FACS (Kula et al., 2019). Identified target cells are then sequenced to determine the peptide antigen they carry (Kula et al., 2019).

6.2.3.2 – Trogocytosis

In 2019, David Baltimore’s group described a method of T cell ligand discovery that draws on the phenomenon of trogocytosis. Trogocytosis describes the transfer of membrane proteins when T cells and target cells are brought into close proximity by the TCR-pMHC interaction (Joly and Hudrisier, 2003; G. Li et al., 2019). Using Jurkat cells transduced with TCRs of interest, and K562 antigen-presenting cells expressing single chain pMHC they were able to identify TCR epitopes based on the presence of T cell surface markers on the K562 cells (G. Li et al., 2019). However, the system has a number of artificialities that limit its physiological relevance. Like CPL screens, it does not rely on endogenous peptide processing. Unlike CPL screens, it uses modified, single chain pMHCs and supraphysiological levels of expression were required (G. Li et al., 2019).


6.2.3.3 – SABR
David Baltimore’s group also developed a chimeric antigen receptor called a signalling and antigen-presenting bifunctional receptor (SABR) (Joglekar et al., 2019). SABRs consist of a MHC molecule associated with CD3ζ and CD28, upon binding a TCR the CD3ζ dimerise and induce T cell-like signalling in the target cell (Joglekar et al., 2019). This signalling promotes upregulation of early activation markers such as CD69 and CD107a which can be used to identify target cells (Weiss and Littman, 1994). CD3 signalling also results in the activation of the nuclear factor of activated T cells (NFAT) transcription factor which was exploited by using NFAT-GFP in Jurkat reporter cells. In the NFAT-GFP Jurkat cell line, NFAT activation promotes the expression of green fluorescent protein (GFP) which can be used to identify target cells by FACS (Joglekar et al., 2019). Extracting the DNA from these cells and amplifying the SABR region was used to identify target epitopes (Joglekar et al., 2019).

There are many methodologies that can be used to probe T cell specificity, each with their own strengths and caveats. The three methodologies described above were published during my own studies. I selected CPL screening due to the high throughput nature of CPL-based peptide discovery, and the lack of requirement for prior knowledge or bias with regards to potential peptide ligands. Furthermore, this methodology benefited from the recently described CPL-driven database searching webtool which ranks peptide ligands (Szomolay et al., 2016). CPL screening provided me with large quantities of detailed data. It would be interesting to compare the ligand preferences of a TCR using these newly available antigen-discovery pipelines and our own CPL methodology.

6.2.4 – Peptide-MHC plasticity: The other side of the coin
It is now well recognised that the TCR-pMHC interface is dynamic and yet almost all studies to date have treated the pMHC as a fixed entity around which the more fluid TCR can mould. The plasticity of the pMHC is, however, being increasingly recognised and in Chapter 5, I turned my attention to this side of the interaction and added to the ever-increasing library of data that proves that TCR ligands are also dynamic. I showed that the pMHC is a dynamic structure whose molecular motions are fine-tuned by the peptide cargo that it carries. Understanding the dynamic behaviour of pMHC molecules is an important addition to the toolkit needed for therapeutic exploitation of the TCR-pMHC interaction. As mentioned above, Madura et al. showed that a modified melanoma peptide with enhanced binding to HLA-A2 made for a worse T cell antigen (Madura et al., 2015). The increased affinity between peptide and MHC molecule reduced the plasticity of the complex and prevented the TCR
from “pulling” the peptide out of the groove slightly, as was seen in the wildtype complex. Thus, reducing the flexibility of the pMHC made the pMHC an inferior TCR ligand (Madura et al., 2015) highlighting the importance of taking into account molecular flexibility during future endeavours to utilise TCR-pMHC interactions therapeutically.

There are a number of methods that can be employed to investigate molecular flexibility. Chiefly amongst these methods is the use of molecular dynamics (MD) simulations. This computational technique allows for high-throughput analysis of molecular flexibility, as demonstrated by the use of MD simulations to characterise the flexibility of over 50 pMHC molecules (Ayres et al., 2019). Ayres et al. found that peptide-dependent dynamic effects are transmitted throughout the MHC molecule, principally affecting the peptide-binding groove itself, but also extending out of it to influence distal regions of the molecule (Ayres et al., 2019). Despite lacking experimental supporting data, high-throughput methods such as MD simulations can be used to build large libraries of data from which future conclusions and comparisons can be drawn.

Computational data demands support in the form of experimental data. A simple way to experimentally assess molecular flexibility is by the examination of static atomic resolution crystal structures. Borbulevych et al. used crystal structures to demonstrate peptide-dependent flexibility of HLA-A2 using two peptides (Tax (LLFGYPVYV) and Tel1p (MLWGYLQYV)) that could both be recognised by the A6 TCR (Borbulevych et al., 2009). The pMHCs were almost complete molecular mimics in their unbound state, but formed a substantially different interface of contacts when bound to the A6 TCR. Comparison of the TCR-bound and -unbound structures pinpointed regions of molecular movement within the pMHC (Borbulevych et al., 2009).

Molecular flexibility of proteins in solution, including pMHC, can also be examined by nuclear magnetic resonance (NMR) and hydrogen/deuterium exchange (HDX). NMR and HDX are discussed in detail in Section 5.1.2. Beerbaum et al. used NMR to demonstrate remarkable plasticity at the MHC-heavy chain:β2M interface in complexes with different MHC molecules and the invariant β2M. Structural variation in β2M occurs at the residues involved at the β2M-MHC interface, allowing β2M to tolerate different MHC molecules without significant change to its overall structure (Beerbaum et al., 2013).

I assess pMHC dynamics in Chapter 5, benefiting from an institutional collaboration with the University of Bath, and satisfying the requirements of my GW4 PhD funding from the UK Medical Research Council. With the aid of equipment and expertise from the Bath
laboratory, I was able to use the red-edge excitation shift (REES) phenomenon to follow on from two key publications, Hawse et al. 2013 and Van Hateren et al. 2017, which use HDX to demonstrate peptide-dependent molecular motility throughout the pMHC molecule (Hawse et al., 2013; Van Hateren et al., 2017). I utilised the REES phenomenon, representing its application to pMHC molecules for the first time to my knowledge, to confirm observations of peptide-dependent MHC dynamics in an autoimmune-relevant model system (Wooldridge et al., 2012; Cole et al., 2016) as discussed in detail in Chapter 5.

6.3 –T cell therapy
Greater understanding of the TCR-pMHC interaction achieved through methods such as those described in this thesis will facilitate the development of future T cell therapies. T cell-based therapeutics have been used successfully in a variety of clinical settings as described below:

6.3.1 – Adoptive cell transfer (ACT)
One of the most prevalent forms of T cell therapy in current use is adoptive cell transfer (ACT). ACT involves the extraction and ex vivo expansion of human T cells, followed by their reinfusion into patients (Savoldo, Heslop and Rooney, 2000; Ellebaek et al., 2012). ACT has been used in a number of different clinical settings, principally to treat viral infections and tumours. For example, the infusions of donor-derived CMV-specific CD8+ T cells can be used to reconstitute cellular immunity in transplant patients following bone marrow transplant (Riddell et al., 1992). One study found that infusions of CMV-specific clones from their donors could increase CMV immunity without observable toxic side effects in 78% (11 of a cohort of 14) of patients (Walter et al., 1995).

Another complication common in transplant recipients is EBV lymphoproliferative disease. In immunocompetent individuals EBV infection is mild, but the viral genome can persist in a latent form in infected B cells (Savoldo, Heslop and Rooney, 2000). Post-transplant immunosuppression in these patients impairs the normal CD8+ T cell response that keeps the infection at bay leading to uncontrolled expansion of EBV-infected B cells (Savoldo, Heslop and Rooney, 2000). Ex vivo produced EBV-specific CD8+ T cells can be used both prophylactically and therapeutically. An early attempt to treat transplant patients with donor-derived CD3+ cells successfully achieved disease regression but two patients died from respiratory failure caused by graft vs host disease (Papadopoulos et al., 1994). Later, in an
attempt to circumvent the alloreactivity problem, Savoldo et al. used patient B cells to
generate EBV-transformed lymphoblastic cell lines (LCLs). The autologous LCLs were then
used to activate EBV-specific donor T cells which were expanded for reinfusion (Savoldo,
Heslop and Rooney, 2000). 59 prospective transplant patients were given prophylactic
treatment in this manner, none of whom developed EBV lymphoma compared with an
incidence rate of 11.5% in the control group (Savoldo, Heslop and Rooney, 2000). Such
bespoke, personalised therapy can be lengthy in development and patients may die while
awaiting treatment. To strike the balance between rejection of donor T cells and completely
autologous therapy Dorothy H. Crawford and colleagues established a bank of EBV-specific
T cells from a range of donors (Vickers et al., 2014). Patients are given EBV-specific T cell
infusions from the closest HLA-matched donor in the bank. This approach bypassed the need
for individual bespoke therapy for each parents to create an off-the-shelf therapy that has
been used successfully, with no adverse effects, to treat patients with EBV lymphoproliferative disease (a response rate of 65%) (Haque et al., 2007; Vickers et al.,
2014).

In addition to the treatment of viral infections, ACT is frequently employed to treat cancer
patients. This form of the therapy involves the infusion of tumour-specific T cells, often with
accompanying supportive cytokines such as IL-2 and IL-15 (Overwijk et al., 2003; Klebanoff
et al., 2004). Dudley et al. treated 35 metastatic melanoma patients with autologous tumour-
reactive ex vivo expanded tumour-infiltrating lymphocytes (TILs) and IL-2 achieving a 51%
clinical response rate, including three complete responses (Dudley et al., 2005). The
treatment could effectively mediate clinical regression in both the original tumour and
metastases (Dudley et al., 2005).

The above examples demonstrate the strength of ACT as a therapy to bolster endogenous
CD8+ T cell activity for clinical benefit. However, the need to generate autologous T cells for
each patient is labour-intensive and time-consuming. Furthermore, numerous studies have
shown that not all TIL respond to tumour (Scheper et al., 2019) so alternative approaches
have involved equipping T cells with receptors that do recognise tumour to ensure that all
infused T cells have the capacity to contribute to tumour clearance (Yannelli et al., 1996; Zhu
et al., 2017; Horton and Gajewski, 2018; Horton et al., 2018).
6.3.2 – TCR gene transfer

TCR gene transfer involves the transfer of \( TCR\alpha \) and \( TCR\beta \) genes from a TCR of known specificity in order to confer that specificity on a population of T cells (D. Li et al., 2019). Unlike ACT, for use in therapy, T cell genetic engineering does not rely on the presence of pre-existing disease-specific T cells. Transfer of TCR genes can instead confer that specificity on the patient’s own peripheral lymphocytes, offering it a significant advantage over ACT (Morgan et al., 2006). Disease-associated epitopes and the T cells specific for them can be identified in ongoing peptide discovery efforts such as those described in this thesis. The identified TCRs can then be cloned and ‘banked’ for use in future therapies. For example, TCRs specific for a number of TAAs have been identified, cloned, and successfully used to redirect non-specific T cells (Morgan et al., 2003; Hughes et al., 2005; Zhao et al., 2005).

T cell genetic engineering was used in Chapter 4 of this thesis to redirect a polyclonal population of donor-derived CD8\(^+\) T cells to specifically recognise the HIV-1 epitope, SLYNTVATL. I, like many others (Morgan et al., 2006; Rapoport et al., 2015), used lentiviral vectors to achieve high, stable TCR expression at the T cell surface.

Lentiviruses naturally excel at delivering genetic material to cells and boast an advantage over other types of viral vector in their heightened ability to infect non-proliferating cells (Milone and O’Doherty, 2018). Lentiviruses have also been cited as being safer than alternatives like gammaretroviruses on account of a lower propensity to cause insertional mutagenesis (Oldham, Berinstein and Medin, 2015). Lentiviral vectors first featured in clinical trials in 2003 (Levine et al., 2006), despite the replication-competency of these vectors no adverse effects were noted in an 8-year follow up (McGarrity et al., 2013). Since then, T cells transduced to express TCRs specific for tumour antigens have been used as a successful therapy (Milone and O’Doherty, 2018). In a 2015 trial, 16 of 20 myeloma patients experienced clinical responses after receiving T cells expressing an affinity-enhanced, HLA-A2-restricted NY-ESO-1-specific TCR (Rapoport et al., 2015). Methods for the optimised delivery of TCR genes are discussed in detail in Appendix Sections 7.2 and 7.3.

I believe that T cell genetic engineering offers many benefits over the use of the patient’s endogenous T cell repertoire as in ACT. As discussed above, the genetic engineering of patient T cells does not rely on the existence of disease-specific T cells within the patient, nor does it necessitate the time-consuming expansion of T cells on a patient-by-patient basis. Additionally, engineered T cells can be subjected to extensive crossreactivity analysis, without the time pressure of an awaiting patient. Large libraries of HLA-typed T cells can be
generated and stored for future ‘off-the-shelf’ allogenic T cell therapies (Mo et al., 2020; Perez, Gruber and Arber, 2020). Furthermore, Mo et al. recently described a method for off-the-shelf cellular therapy in which they designed an alloimmune defence receptor (ADR) with which a readily-available bank of disease-specific allogenic T cells can be equipped (Mo et al., 2020). The ADR recognised 4-1BB which is transiently expressed on the surface of activated lymphocytes, thus protecting against graft vs host disease (Mo et al., 2020). Such cells could be subjected to extensive safety testing, far more than bespoke therapies would have time for, and be immediately available when required (Mo et al., 2020).

In addition to not relying on the endogenous T cell repertoire in a given patient, T cell genetic engineering benefits from the ability to improve upon naturally occurring disease specific TCRs. Transduced TCRs can be subject to affinity enhancement (described in detail in Section 6.1.2) that can both enhance the effectiveness of the therapy, and broaden the specificity of the TCR to capture any potential escape mutants that may arise.

6.4 – Concluding remarks
The repertoire of TCRs on expanded, antigen-experienced T cells contains a record of all past infections and information on some potential future pathologies. There is currently much interest in using TCRs as predictors of health and disease but full realisation of this goal will require means of determining TCR function based on TCR sequence (Attaf, Huseby and Sewell, 2015). In this thesis, I present specificity data, both positive and negative, for two disease-relevant human TCRs that would provide suitable fuel to these predictive endeavours. The ultimate goal of being able to rationally design a TCR to target any given peptide ligand relies on the ability to predict positive interactions, which in turn relies on large volumes of published TCR specificity data – a small contribution to which is contained in this thesis.

I believe that the future treatment of many human diseases will rely heavily on the therapeutic application of the TCR-pMHC interaction. During my PhD project I have employed multiple techniques to investigate the interaction in disease-relevant systems. I have demonstrated that a single CDR3 amino acid substitution can provide the difference between recognition and ignorance, and I have highlighted hitherto unanticipated viral-autoimmunity and viral-cancer crossreactivity. Furthermore, I have demonstrated that extremely similar pMHC ligands can exhibit differing molecular motions.
Whilst each individual study, both here and in the literature, strictly provides information regarding the model system used, collectively they build up the framework from which future T cell therapeutics can draw the strength needed to harness one of evolutions greatest weapons against disease.
7.1 – Human immunodeficiency virus

HIVs are typically grouped into two categories: HIV type 1 (HIV-1) and HIV type 2 (HIV-2) (Coffin et al., 1986; Gelderblom, Özel and Pauli, 1989). HIV-1 is prevalent worldwide (Beloukas et al., 2016), whereas HIV-2 tends to be limited to regions in Africa (Visseaux et al., 2016). HIV primarily infects CD4+ T cells (Lasky et al., 1987; Gelderblom, Özel and Pauli, 1989; Weiss, 1993; Wilen, Tilton and Doms, 2012). Infected CD4+ T cells migrate to T cell-dense lymphoid tissues such as the lymph nodes, offering prime opportunity to spread the infection (Fanales-Belasio et al., 2010). This results in infected individuals experiencing a steady decline in the number of circulating CD4+ T cells over time, leading to life-threatening immunocompromise (Gottlieb et al., 1981; Siegal et al., 1981; Friedman-Kien et al., 1982; Mildvan et al., 1982; Schroff et al., 1983; Melbye et al., 1986; Kahn and Walker, 1998). Viral replication and chronic immune activation steadily destroy lymphoid tissue (Ford, Puronen and Sereti, 2009; Fanales-Belasio et al., 2010). This progressive immune impairment puts the patient at risk of opportunistic infections and neoplasms such as Kaposi sarcoma (Gottlieb et al., 1981; Siegal et al., 1981; Friedman-Kien et al., 1982; Schroff et al., 1983; Gupta, 1993; Hoover et al., 1993; Brooks et al., 2009; Clifford and Franceschi, 2009).

HIV has a characteristic pattern of infection consisting of an initial acute, symptomatic infection followed, after a few weeks, by a long period of asymptomatic chronic infection (Deeks and Walker, 2007; Fanales-Belasio et al., 2010). Symptoms that manifest during the acute phase include flu-like fever, rash, oral ulcers, weight loss, and lymphadenopathy (swelling of the lymph nodes) (Kahn and Walker, 1998; Deeks and Walker, 2007; Fanales-Belasio et al., 2010). Viral RNA becomes detectable in the blood approximately 10–12 days post-infection (Fiebig et al., 2003) which can be used for diagnosis (Lindbäck et al., 2000; Hecht et al., 2002). Viremia then experiences an exponential rise (Clark et al., 1991; Daar et al., 1991; Little et al., 1999) before it peaks around 21 days post infection (Little et al., 1999). Untreated, the following clinically asymptomatic period is typically 10 years and features a steady decline of CD4+ T cells culminating in death of the patient (Gupta, 1993; Deeks and Walker, 2007; Ford, Puronen and Sereti, 2009; Fanales-Belasio et al., 2010).

HIV is a retrovirus; it has two copies of its single-stranded RNA genome that must be reverse transcribed into DNA before it can be incorporated into the host genome (Fanales-Belasio et al., 2010). This genome encodes, amongst others, the gag, pol, and env genes. The gag gene
encodes core structural proteins (p24, p7, p6) and matrix protein, p17 (Gelderblom, Özel and Pauli, 1989; Fanales-Belasio et al., 2010). The env gene encodes the viral envelope glycoproteins gp120 and gp41. The pol gene encodes enzymes required for viral replication such as reverse transcriptase, which converts viral RNA into DNA, integrase which incorporates the viral DNA into the host genome, and a protease which cleaves viral precursor proteins into the proteins necessary for assembling new viral particles (Veronese et al., 1987; Gelderblom, Özel and Pauli, 1989; Fanales-Belasio et al., 2010).

The viral genome is housed, along with vital viral enzymes (reverse transcriptase, integrase, and protease) within a capsid made of polymerised p24 core antigen. This core antigen associates with the p17 matrix protein which anchors the capsid to the lipid bilayer that encases the viral particle (Gelderblom, Özel and Pauli, 1989). The virus acquires its lipid bilayer as it exits the host cell, so it often also incorporates host proteins such as intercellular adhesion molecule (ICAM)-1, MHC-I, and MHC-II (Gelderblom, Özel and Pauli, 1989). A typical viral particle measures ~110nm in diameter (Gelderblom, Özel and Pauli, 1989).

The viral life cycle of HIV can be summarised in six steps: 1) binding and entry; 2) uncoating; 3) reverse transcription; 4) provirus integration; 5) virus protein synthesis and assembly; and 6) budding (Fanales-Belasio et al., 2010). A viral particle binds to a host cell through the interaction of gp120 with CD4 on the surface of T cells (Lasky et al., 1987; Gelderblom, Özel and Pauli, 1989; Straus and Weiss, 1993; Wilen, Tilton and Doms, 2012). This adhesion triggers a conformational change in gp120 that allows it to interact with other surface receptors on the host cell such as CXCR4 and CCR5 thus stabilising the binding (Chackerian et al., 1997; Agrawal et al., 2004; Alkhatib, 2009). CXCR4/CCR5 interaction brings the virus particle into close proximity with the host cell causing the lipid bilayers to merge allowing the virus to enter the cell (Gelderblom, Özel and Pauli, 1989). Naturally occurring mutations in chemokine receptors CXCR4 and CCR5 can confer some degree of resistance to HIV-1 infection (Carmichael et al., 1993; Balotta et al., 1997; O’Brien et al., 1997; Agrawal et al., 2004; Q. Wang et al., 2017; Allen et al., 2018). Furthermore, conventional ligands for these receptors such as MIP-1α and MIP-1β can suppress HIV-1 infection in vitro by competing for the CCR5 receptor (Cocchi et al., 1995; Price et al., 1998; Garzino-Demo, 2006; Fanales-Belasio et al., 2010). CD8+ T cells are a key source of these β-chemokines (Conlon et al., 1995), and produce them in response to specific MHC-bound HIV-1 peptide epitopes (Price et al., 1998). Price et al. showed that the release of β-chemokines was epitope specific (it did not occur in response to some HIV escape mutants), and – like cytotoxic activity – mediated through activation of the TCR engaging relevant pMHC (Price et al., 1998). Thus, they
demonstrated that HIV escape mutants (discussed in detail below) not only serve to evade the cytotoxic immune response but can enhance the propagation of the infection by preventing β-chemokine-mediated inhibition (Price et al., 1998).

Once inside the cell, HIV loses its capsid—a process known as uncoating (Cosnefroy, Murray and Bishop, 2016) — and transcription of the viral RNA genome can begin. The newly produced DNA is incorporated into the host DNA by the integrase enzyme (Fanales-Belasio et al., 2010). Replication of the viral DNA is then delayed until the host cell is activated and dividing. Quiescent immune cells are a long-lived source of latent virus (Chun, Carruth, et al., 1997; Chun, Stuyver, et al., 1997; Herzig et al., 2019). This latent virus is often referred to as the latent reservoir and can be difficult to eliminate through treatment (Chun, Carruth, et al., 1997; Herzig et al., 2019).

When an infected CD4+ T cell is activated it begins the process of cell division, replicating both its own and the viral DNA (Herzig et al., 2019). Some of the earliest HIV-1 proteins to be produced include the regulatory proteins, Tat and Rev, these serve to specifically promote the transcription of HIV-1 DNA. The proteins encoded by gag and pol then begin to form the new viral particle (Fanales-Belasio et al., 2010). Env encodes a precursor protein, gp160, which is cleaved by protease into the envelope glycoproteins gp120 and gp41 (Kantanen, Leinikki and Kuismanen, 1995). Once the immature particle is assembled it migrates to the cell surface for budding and release, acquiring a new lipid bilayer coat from the host cell membrane (Fanales-Belasio et al., 2010).

There is ample evidence of CD8+ T cell involvement in the antiviral response to HIV-1. During the acute phase cytotoxic activity is detected before the appearance of anti-HIV antibodies (Borrow et al., 1994; Koup, 1994; Price et al., 1997; Allen et al., 2000; Bangham, 2009). In 1994, Burrow et al. demonstrated specific cytotoxic activity against five HIV-1 proteins (Env, Gag, Pol, Nef, and Tat) in five patients (Borrow et al., 1994). Many specific CD8+ T cell MHC-restricted epitopes have been described in the literature and are collated by the Los Alamos National Laboratory Database (T-Cell Epitope Database, 2020).

The cytotoxic immune response is sufficient to reduce viremia and bring about the chronic asymptomatic phase (Mellors et al., 1995; O'Brien, 1996). Yet despite this, the CD8+ T cell response is unable to fully clear the virus. A number of factors are responsible for this including the emergence of escape mutants (Jamieson et al., 2003; Petrova, Naumova and Gorski, 2011; Wright et al., 2012), a reduction in available CD4+ T cell help (Hoover et al., 1993; Rosenberg, Anderson and Pabst, 1998; Rodés et al., 2004), CD8+ T cell exhaustion
(Oxenius et al., 2002; Shin and Wherry, 2007; Cella et al., 2010), and activation-induced downregulation of the CD8+ TCR (Oxenius et al., 2002).

Another hurdle for the anti-HIV CD8+ T cell response and antiretroviral therapy is what is known as the latent reservoir. Mentioned briefly above, these are quiescently infected CD4+ T cells that retain the HIV-1 provirus integrated into their genome but do not express viral proteins, rendering them invisible to the immune system and difficult to remove (Chun et al., 1995; Chun, Carruth, et al., 1997; Chun, Stuyver, et al., 1997; Finzi et al., 1997). The greater the size of this reservoir of quiescently infected cells the faster the virus will rebound if antiretroviral therapy is interrupted, this means that patients must adhere strictly to lifelong treatment (Li et al., 2016). Reducing or eliminating the latent reservoir presents a significant challenge as the infected cells are effectively invisible (Herzig et al., 2019). One of the main methods of intervention is known as “shock and kill” (Archin et al., 2012). This approach involves the use of latency reversing agents (LRAs) to reactivate viral replication. Once the viral genome is actively being transcribed, viral proteins will be displayed on the surface of the infected host cell rendering it vulnerable to immune attack once again (Herzig et al., 2019; Jean et al., 2019).

Successful “shock and kill” interventions depend, therefore, on the effective killing of reactivated viral-infected CD4+ T cells by specific CD8+ T cells (Herzig et al., 2019). Whilst the specific CD8+ response to HIV-1 is well documented (Borrow et al., 1994; Koup, 1994; Price et al., 1997; Allen et al., 2000; Bangham, 2009; Kim et al., 2018), the chronic immune landscape that HIV-1 infection causes is generally associated with exhausted CD8+ T cells (Shin and Wherry, 2007; Cella et al., 2010). Attempts have been made to overcome the problem of exhausted indigenous T cells through use of adoptive T cell therapy and chimeric antigen receptors (CARs) (Deeks et al., 2002; Sung et al., 2018; Wagner, 2018; Guedan et al., 2019).

In 2019, Herzig et al. designed a 2-part CAR system to target the latent reservoir. Termed ‘convertibleCAR T cells’ they aimed to simultaneously overcome the problem of exhausted immune cells and deliver a therapy that could be turned on and off through administration (and natural clearance) of a broadly neutralising antibody (bNAb) (Herzig et al., 2019). Their CAR was specific for a modified natural molecule that was tagged to the Fc region of selected bNAbs, ensuring that their adoptively transferred T cells would only be active if and when the bNAb was also administered (Herzig et al., 2019). A further advantage of this technique lies in the wide availability of different anti-HIV-1 bNAbs, meaning that alternatives can be
selected should resistance arise, and that – theoretically – this therapy would be just as
effective at targeting other aberrant cells provided there were suitable bNAbs available (Sok
et al., 2016; Mayer et al., 2017; Dashti et al., 2019; Julg and Barouch, 2019). A limitation of
this technique is the ineffectiveness of many of the available LRAs – 100% reactivation of the
latent reservoir is rarely achieved and side-effect toxicity has been reported (Herzig et al.,
2019; Jean et al., 2019).

A further challenge for the immune response to HIV-1 infection is the emergence of escape
mutants. Viral replication is highly error-prone; this high mutational rate is beneficial to the
virus as it facilitates evasion of the host immune system (Roberts, Bebenek and Kunkel, 1988;
Goodenow et al., 1989). Rapidly introduced mutations are subject to host-applied selective
pressure that promotes viral variants that are ‘hidden’ from the host’s immune system, these
are known as escape mutants (Price et al., 1997; Menéndez-Arias, 2002; Sarafianos et al.,
2009). There are a number of ways that mutations in viral proteins can facilitate immune
evasion.

Escape mutations may result in peptides that cannot bind to MHCs (Voeten et al., 2000;
Goulder, Brander, et al., 2001; Kelleher et al., 2001; Boon et al., 2002; Feeney et al., 2004;
Rimmelzwaan et al., 2004; Casazza et al., 2005). For example, HLA-B27+ HIV-1 patients often
mount an immunodominant response against an epitope from the Gag-derived p24 protein
Immune escape can occur through mutation of the lysine or glycine residues within this
epitope, this results in a peptide that doesn’t bind HLA-B27 well (Nietfield et al., 1995;
Goulder, Phillips, et al., 1997; Kelleher et al., 2001). Occurrence of this mutation often
coincides with disease progression (Kelleher et al., 2001), and removal of the CD8-mediated
selective pressure causes reversion to the wild-type sequence (Kelleher et al., 2001).

Escape mutations may produce peptides that bind to MHCs but the epitopes are no longer
recognised by the host’s CD8+ T cell repertoire (Phillips et al., 1991). In 1997, Price et al.
showed selective pressure drove immune escape in a HLA-B8-restricted Nef peptide
(FLKEKGGL), with genomic variation being largely confined to this epitope (Price et al., 1997).
They demonstrated how positively selected variants either diminished or eradicated CD8+ T
cell recognition. In further evidence that this selective pressure was driven by the CD8+ T cell
immune response they demonstrated that the same variation did not occur in HLA-B8-
negative donors (Price et al., 1997).
Alternatively, the peptides that are produced following mutation of the viral genome might be suboptimally bound by TCRs, engaging the TCR but not activating it – this serves to prevent the TCR from encountering and potentially responding to other antigen, thus having an antagonistic effect (Bertoletti et al., 1994; Sewell et al., 1997).

In 2017, Cole et al. investigated the escape mutants that arise from a HLA-A2-restricted immunodominant Gag epitope, SLYNTVATL (Cole, Fuller, et al., 2017). A simple strategy often employed during viral escape is to mutate the positions in the epitope that anchor the peptide to its MHC, in this case positions 2 and 9 (the C-terminus). In the SLYNTVATL epitope, the Y and L residues (positions 79 and 85 in the Gag protein sequence) are crucial for the function of the Gag-derived p17 protein and thus cannot be mutated without significant fitness cost to the virus (Cannon et al., 1997; Morikawa et al., 1998; Martinez-Hackert et al., 2006). Thus, it was assumed that the escape mutants that arise from SLYNTVATL must instead abrogate T cell recognition. In fact, Cole et al. demonstrate structurally that escape mutants are still recognised by the cognate TCR. They show that the mutants achieve their escape not by preventing T cell recognition, or by fully preventing MHC binding, but by destabilising the pMHC structure at the surface of the cell (Cole, Fuller, et al., 2017).

Strategies employed for immune escape demonstrate how variations in the TCR-pMHC interaction between individuals can influence how their disease progresses. The MHC background an infectious agent faces in any individual host can dictate which escape mutants will arise, and different escape mutants will have varying impacts on viral fitness (Carlson et al., 2012) For example, both HLA-B*42:01 and HLA-B*81:01 present an immunodominant epitope from the Gag protein (Gag180–188, TPQDLNTML) (Honeyborne et al., 2006). Despite the two alleles being closely related members of the B7 superfamily, the escape mutation selected in individuals expressing either allele differs with drastic effects on viral fitness (Wright et al., 2012; Kløverpris, McGregor, et al., 2015). In HLA-B*42:01 positive individuals the most common escape mutation arises at position 3 (Q) of the peptide, relating to position 182 in the Gag protein. In contrast, in HLA-B*81:01-expressing individuals mutations more frequently arise in Gag-186 (position 7 of the peptide) (Wright et al., 2012; Kløverpris, McGregor, et al., 2015). Unlike the mutation at Gag-182, which has negligible impact on the function of the Gag protein, mutations at Gag-186 have a substantial negative impact on viral fitness. As a result of the reduced viral replicative capacity, HLA-B*81:01-positive individuals cope relatively well with the infection and experience delayed progression to acquired immunodeficiency syndrome (AIDS) (Wright et al., 2012; Kløverpris, McGregor, et al., 2015).
Even micropolymorphisms in the MHC alleles expressed by different individuals can have a significant influence on how well they control the disease. For example, single amino acid differences between HLA-B*35 alleles can influence progression to AIDS (Gao et al., 2001; Jin et al., 2002). Furthermore, of the closely related HLA-B*58:01, HLA-B*57:02, and HLA-B*57:03 alleles the latter is associated with lower viral loads than the two former. Each phenotype triggers a unique cohort of CD8+ T cells in response to infection (Kløverpris et al., 2012).

Variation in how individuals control HIV-1 infection has led to patients being classified based on their ability to deal with the chronic infection, terms such as “rapid progressors” and “long-term non-progressors” or “elite controllers” are used (Fanales-Belasio et al., 2010). Approximately 0.5% of HIV-1-infected patients are elite controllers, meaning that in the absence of treatment their own immune response suppresses viral replication to below detection limits (Deeks and Walker, 2007). Furthermore, a proportion of regularly exposed people do not contract HIV, suggesting their anti-viral immune activity can prevent the virus from establishing an infection (Rowland-Jones et al., 1995; Stranford et al., 1999). Understanding why these patients are able to respond well to the virus may aid treatment of less fortunate patients.

In 2007, Blankson and Siliciano reported that the viruses infecting a cohort of elite controllers had normal virulence (Blankson and Siliciano, 2008), and in 2008 Bailey et al. showed that elite controllers could still transmit the virus to other people (Bailey et al., 2008). This combined evidence suggests that elite controllers are not infected with an attenuated virus, but that their immune response is more effective than that of standard patients (Blankson and Siliciano, 2008). They are able to effectively control the infection despite the emergence of viral mutants, suggesting that novel CD8+ T cell responses continue to be initiated across the course of the infection (Bailey et al., 2006). Elite controllers are thought to have CD8+ T cells with a higher proliferative capacity against immunodominant epitopes (Migueles et al., 2002). They have also been shown to express higher levels of perforin, IFNγ, MIP-1-β, and TNF (Migueles et al., 2002; Betts et al., 2006).

A number of studies have suggested that certain MHC-I alleles, such as HLA-B*57 and HLA-B*27, are overrepresented in elite controllers compared to other HIV-1 patients (Lambotte et al., 2005; Bailey et al., 2006; Betts et al., 2006; Deeks and Walker, 2007). Suggesting that possession of particular MHC backgrounds enables the patients to present optimal peptide epitopes to their CD8+ T cells. Furthermore, there is evidence that selective pressure from
HIV-1 prevalence in the human population can drive the accumulation of protective MHC-I alleles (Moore et al., 2002; Leslie et al., 2005). Whole-genome association scan analysis of HIV-1 patients actually showed MHC-I alleles to be one of the strongest polymorphisms associated with protection against the disease (Fellay et al., 2007; Catano et al., 2008).

It is clear, therefore, that vital understanding can be gleaned by studying the specific TCR-pMHC interactions that occur during the course of HIV-1 infection. One TCR-pMHC interaction of particular interest to this thesis is the 868-HLA-A2-SLYNTVATL interaction. This interaction is used as a model system to investigate how TCR sequence relates to peptide preference, it has a rich history within the field of HIV-1 research, see Section 3.1.1.

7.2 – Optimised expression of transduced TCRs at the T cell surface
2A self-cleaving peptide sequences are found in the genomes of many viruses (Hahn and Palmenberg, 1996; Szymczak et al., 2004; Leisegang et al., 2008). During translation, the sequence of these short peptides causes a ribosomal ‘skip’ between the penultimate and N-terminal residues of the 2A sequence. The formation of a peptide bond is inhibited, thus breaking the continuous polypeptide chain (De Felipe et al., 1999; Szymczak et al., 2004; Leisegang et al., 2008). This mechanism results in the majority of the 2A sequence being tagged onto the end of the upstream gene, whilst the N-terminal residue, a glycine, remains attached to the downstream gene (De Felipe et al., 1999; Leisegang et al., 2008).

2A-mediated cleavage has been shown to be almost 100% efficient (Szymczak et al., 2004). In 2004, Szymczak et al. used a 2A-based vector system to restore T cell function of CD3-deficient mice. Their tetracistronic vector contained the four CD3 genes (ε, γ, δ, ζ) separated by 2A sequences (Szymczak et al., 2004). Western blot analysis of the resulting proteome revealed successful expression of all four genes. Almost no uncleaved bands were visible – indicating close to 100% cleavage (Szymczak et al., 2004).

2A peptides are, however, not the only option for multicistronic expression. Other options include the use of multiple vectors (containing a single gene each), the use of internal promotors, and the use of internal ribosomal entry sites (IRES). IRES sequences provide an additional ribosome binding site within the mRNA transcript, allowing the second polypeptide to be translated independently of the first (Mizuguchi et al., 2000; Morgan et al., 2006). Expression of the downstream genes is often poor compared to expression of the gene preceding the IRES element (Mizuguchi et al., 2000; Osti et al., 2006). Mizuguchi et al.
demonstrated that expression of downstream genes is approximately 20-50% that of the upstream gene (Mizuguchi et al., 2000). This is an undesirable caveat when expressing genes that act with 1:1 stoichiometry like an αβ TCR heterodimer at the T cell surface. Enhanced expression of downstream genes in 2A systems compared to IRES elements has been frequently noted (De Felipe et al., 1999; Klump et al., 2001). Furthermore, the efficiency of IRES elements has been shown to vary between different cell types (Borman et al., 1997), and they can be subject to competition for translation factors (Szymczak et al., 2004). IRES elements are also much larger than 2A peptides which can cause difficulties in viral vectors with limited gene carrying capacity, such as lentiviruses (De Felipe et al., 1999).

A useful comparative study of the different transfer vector options was provided by Leisegang et al. in 2008, in which they compared expression of TCRα and TCRβ when introduced by either two separate vectors, a single vector with an IRES, or a single vector with a 2A peptide (Leisegang et al., 2008). Use of two separate vectors was found to be hampered by the necessity of ensuring both vectors are taken up by the same cell (Leisegang et al., 2008; Banu et al., 2014). The 2A system was found to be less susceptible to competition from an endogenous TCR than the IRES expression method, producing enhanced surface expression and improved peptide-specific IFNγ production (Leisegang et al., 2008).

When the expression of the second gene is driven by a second, internal, promoter it invites the possibility of promoter interference, or competition between the two promoters for transcriptional machinery (Emerman and Temin, 1984; Leisegang et al., 2008). In 1984, Emerman and Temin showed that in a two-promoter vector one of the two genes was often suppressed. They suggested that the suppression was epigenetic as they could alter which gene was suppressed by pre-exposing the cells to different conditions (Emerman and Temin, 1984). For example, if cells were given two antibiotic resistance genes and cultured in the presence of antibiotic A, when subsequently exposed to antibiotic B most cells died despite possessing the resistance gene. This was also true in reverse, regardless of the order of the genes in the vector (Emerman and Temin, 1984). Thus, neither an IRES nor a two-promotor vector achieves the stoichiometric expression optimal for expression of T cell receptors.

A potential disadvantage of the 2A system is that the short peptide that remains attached to the upstream polypeptide might interfere with its biological function or contribute to immune recognition of the transduced protein. However, the system has been widely used without any noteworthy interference (Yang et al., 2008). In 1999, de Felipe et al. used 2A bicistronic vectors to confer resistance to two antibiotics, the upstream tag did not interfere
with the function of the upstream gene (De Felipe et al., 1999). In a tricistronic vector the function of the second gene was similarly unaffected, and expression of the third gene (GFP) was noted for up to three months before repurification was needed (De Felipe et al., 1999). The 2A sequence was also shown not to interfere with the expected subcellular locations of transgenes (Klump et al., 2001). Szymzak et al. suggest that the tag might actually provide an advantageous distinction between the transduced protein and endogenous protein (Szymczak et al., 2004).

The 2A autocleavage site can be further optimised by the addition of upstream spacer sequences (Yang et al., 2008). Yang et al. demonstrated that the addition of an amino acid spacer sequence (GSG or SGSG) just before the 2A sequence could enhance the expression of two different TCRs, as measured by pMHC tetramer staining (Yang et al., 2008). My construct system incorporates this spacer. There were, however, further optimisations noted by Yanh et al. that could be included and tested in future. For example, they found the addition of a furin protease recognition site prior to the 2A peptide enhanced surface expression (Yang et al., 2008).

Interestingly, a 2014 study found that the order of genes within a P2A construct could influence the expression efficiency. Using 10 virus-specific TCRs Banu et al. demonstrated that 9/10 TCRs favoured a beta-2A-alpha orientation over an alpha-2A-beta orientation, as indicated by surface expression measured by β-chain-specific antibodies and pMHC-multimers (Banu et al., 2014). This optimal expression manifested functionally as a significant increase in the target-specific release of IFNγ, IL-2, and TNF (Banu et al., 2014). This observation is not unique to the Banu study (Leisegang et al., 2008). In this chapter, I used an alpha-2A-beta orientation. It would, therefore, be insightful to perform parallel comparisons of my TCRs in both orientations in future.

7.3 – Optimising TCR gene transfer

In 2007, Kuball et al. described a method for improving chain pairing by introducing non-natural cysteine residues in the α and β TCR chains, thus enhancing chain pairing akin to the manner described for soluble TCR production (Boulter et al., 2003). This study demonstrated that the introduction of non-natural cysteine residues promoted preferential pairing between the two introduced chains, increased desirable surface expression, and reduced mispairing resulting in greater functional activity against the target antigen (Kuball et al., 2007). A paper published by the same group later the same year demonstrated that the
addition of the non-native disulphide bond was broadly applicable. Using three additional TCRs they showed that the presence of the disulphide bond improved pMHC tetramer staining, and specific killing by transduced T cells (Cohen et al., 2007).

TCR constructs can also be designed to have murine constant domains to promote interchain recognition and binding, enhancing the formation of the correct pairs (Stanislawski et al., 2001; Cohen et al., 2006; Kuball et al., 2007). Cohen et al. reported enhanced in vitro surface expression and cytokine production of murinized TCRs compared to their fully human counterparts that translated into an enhanced anti-tumour response in vivo (Cohen et al., 2006). Murinization may also enhance association with human CD3. Cohen et al. demonstrated (using detergents of different strengths) that the murinized-TCR/CD3 complex was more stable than the human TCR/CD3 complex (Cohen et al., 2006). TCR murinization has been frequently used to enhance TCR expression at the T cell surface (Leko et al., 2019; Ren et al., 2019; Tan et al., 2019) however, it does not come without caveats if such TCRs are used therapeutically as murine constant domains have the potential to be immunogenic, causing any adoptively transferred cells to be targeted by the immune system of the patient (Riddell et al., 1996).

Improving chain pairing can significantly reduce the likelihood of chimeric TCRs and resulting off-target immunity. However, the strategies employed rarely eliminate mispairing completely. For this, many have turned to downregulating the endogenous TCR. In 2009, Okamoto et al. described the down regulation of the endogenous TCR using small interfering RNA (siRNA) (Okamoto et al., 2009). The introduced TCR is subject to codon optimisation and thus differs sufficiently from the endogenous TCR to be protected from the siRNA (Okamoto et al., 2009).

Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein-9 (Cas9) techniques can also be used to remove, or ‘replace’, the endogenous TCR (Legut et al., 2018). In 2018, Legut et al. described a CRISPR/Cas9 strategy for use during simultaneous TCR gene transduction. This ‘TCR replacement’ strategy was applicable to both αβ and γδ TCRs, and produced enhanced functionality (increased tetramer staining, increased multi-cytokine responsiveness, and increased specific killing) (Legut et al., 2018). Indeed, the removal of endogenous TCR chains was observed to increase T cell function in peptide titrations by up to three orders of magnitude (Legut et al., 2018).

Roth et al. described a CRISPR-Cas9 methodology that does not rely on the use of viral vectors, citing their production as lengthy and expensive. Instead, Roth et al. introduce the
Cas9 ribonucleoprotein, the guide RNA, and the ‘replacement’ linear double-stranded DNA to the cell by electroporation (Roth et al., 2018). DNA could be inserted at precise locations within the genome without damaging cell viability or function. TCR replacement resulted in the effective redirection of T cells to specifically target a cancer antigen, producing both in vitro and in vivo responses. Importantly, they found no evidence of off-target crossreactivities (Roth et al., 2018).

Competition for the available endogenous CD3 has been shown to be another rate-limiting factor in TCR gene therapy (Ahmadi et al., 2011). Ahmadi et al. co-delivered the four genes required for expression of CD3 (CD3γ, CD3δ, CD3ε, and CD3ζ) on a single vector, separated by 2A sequences with their introduced TCR. In comparison to the delivery of TCR alone, these recipient cells expressed higher levels of the desired TCR, produced increased amounts of IFNγ in response to their target peptide, and recognised lower concentration of peptide (Ahmadi et al., 2011).

Scholten et al. used codon-optimisation of TCRs to improve surface expression. Their approach increased the portion of tetramer positive cells, and increased the specific IFNγ production in response to target cells (Scholten et al., 2006). Codon optimisation was employed in the work described in this chapter.
Supplementary Figure 7.1: TCR 24-transduced CD8+ T cells do not respond to the M1 influenza-derived GILGFVFTL epitope. 30,000 CD8+ T cells transduced with TCR constructs were cultured in the presence of 60,000 T2 cells as APCs and 10⁻⁵ M GILGFVFTL (green) or SLYNTVATL (pink) peptides overnight before the supernatant was harvested for measurement of MIP-1β presence by ELISA analysis.
Supplementary Figure 7.2: 868-transduced CD8+ T cells respond to untransduced HLA-A2+ MOLT-3 cells, repeat assay. 30,000 CD8+ T cells transduced with TCR constructs were cultured in the presence of 60,000 either MOLT-3 wildtype cells or HLA-A2+ MOLT-3 cells overnight before cells were harvested and assessed for surface expression of TNF and CD107a. Cells previously gated on live CD3+CD8+ rCD2+ single lymphocytes. Numbers on dot plots correspond to the percentage of cells (10,000 events) in the gated population.
Supplementary Figure 7.3: 868-transduced CD8⁺ T cells respond to a range of HLA-A2⁺ immortalised cell lines. 30,000 CD8⁺ T cells transduced with either 868 or TCR 22 were cultured in the presence of 60,000 of a range of cell lines (either HLA-A2⁺ or HLA-A2⁻) overnight before cells were harvested and assessed for surface expression of TNF. Cells previously gated on live CD3⁺CD8⁺rCD2⁺ single lymphocytes. Numbers on dot plots correspond to the MFI values of 10,000 events.
<table>
<thead>
<tr>
<th>pMHC complex</th>
<th>Solvent Accessible Surface Area (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALW</td>
<td>19262.674</td>
</tr>
<tr>
<td>MVW</td>
<td>19406.383</td>
</tr>
<tr>
<td>YQF</td>
<td>19044.336</td>
</tr>
<tr>
<td>RQW</td>
<td>19141.076</td>
</tr>
<tr>
<td>RQFa</td>
<td>19266.105</td>
</tr>
<tr>
<td>RQFi</td>
<td>19181.838</td>
</tr>
</tbody>
</table>

**Supplementary Table 7.1:** The SASA of the 6 pMHC complexes is similar. Calculated using PyMOL get_area function.

\[
\Delta G_{P,T} = \Delta G_0 + \Delta V_0(P - P_0) + \Delta \alpha' (P - P_0)(T - T_0) + \frac{\Delta \beta'}{2}(P - P_0)^2 - \Delta S_0 \\
(T - T_0) - T \left( \ln \left( \frac{T}{T_0} \Delta C_P \right) - 1 \right) + T_0
\]

**Supplementary Equation 7.1:** Algorithm developed by the Pudney laboratory for the extraction of thermodynamic parameters from emission spectra collected under perturbing pressures and temperatures. \(\Delta G\) was of particular use to this work.
Supplementary Figure 7.4: Pressure/temperature-dependence of HLA-A2-ALWGPDPAAA complex. Protein-intrinsic tryptophan fluorescence was excited at 292nm and the emission spectra measured using a Perkin Elmer LS50B Luminescence Spectrometer (Perkin Elmer, Waltham, MA, USA). Pressure was applied using an ISS high-pressure cell (ISS, Champaign, IL, USA) fitted with a custom fibre optic mounting connecting to the fluorimeter and a circulating water bath to control temperature (±1°C). Spectra are the average of triplicate collection.
Supplementary Figure 7.5: Pressure/temperature-dependence of HLA-A2-YQFGPDFTPA complex. Protein-intrinsic tryptophan fluorescence was excited at 292nm and the emission spectra measured using a Perkin Elmer LS50B Luminescence Spectrometer (Perkin Elmer, Waltham, MA, USA). Pressure was applied using an ISS high-pressure cell (ISS, Champaign, IL, USA) fitted with a custom fibre optic mounting connecting to the fluorimeter and a circulating water bath to control temperature (±1°C). Spectra are the average of triplicate collection.
Supplementary Figure 7.6: Pressure/temperature-dependence of HLA-A2-RQFGPDWIVA complex. Protein-intrinsic tryptophan fluorescence was excited at 292nm and the emission spectra measured using a Perkin Elmer LS50B Luminescence Spectrometer (Perkin Elmer, Waltham, MA, USA). Pressure was applied using an ISS high-pressure cell (ISS, Champaign, IL, USA) fitted with a custom fibre optic mounting connecting to the fluorimeter and a circulating water bath to control temperature (±1°C). Spectra are the average of triplicate collection.
Supplementary Figure 7.7: Pressure/temperature-dependence of HLA-A2-RQWGPDPAAV complex. Protein-intrinsic tryptophan fluorescence was excited at 292nm and the emission spectra measured using a Perkin Elmer LS50B Luminescence Spectrometer (Perkin Elmer, Waltham, MA, USA). Pressure was applied using an ISS high-pressure cell (ISS, Champaign, IL, USA) fitted with a custom fibre optic mounting connecting to the fluorimeter and a circulating water bath to control temperature (±1°C). Spectra are the average of triplicate collection.
Supplementary Figure 7.8: Pressure/temperature-dependence of HLA-A2-MVWGPDPFLYV complex. Protein-intrinsic tryptophan fluorescence was excited at 292nm and the emission spectra measured using a Perkin Elmer LS50B Luminescence Spectrometer (Perkin Elmer, Waltham, MA, USA). Pressure was applied using an ISS high-pressure cell (ISS, Champaign, IL, USA) fitted with a custom fibre optic mounting connecting to the fluorimeter and a circulating water bath to control temperature (±1°C). Spectra are the average of triplicate collection.
Supplementary Figure 7.9: Pressure/temperature-dependence of HLA-A2-RQFGDFPFI complex. Protein-intrinsic tryptophan fluorescence was excited at 292nm and the emission spectra measured using a Perkin Elmer LS50B Luminescence Spectrometer (Perkin Elmer, Waltham, MA, USA). Pressure was applied using an ISS high-pressure cell (ISS, Champaign, IL, USA) fitted with a custom fibre optic mounting connecting to the fluorimeter and a circulating water bath to control temperature (±1°C). Spectra are the average of triplicate collection.
References


Banu, N. et al. (2014) ‘Building and optimizing a virus-specific T cell receptor library for


Bentzen, A. K. et al. (2016) ‘Large-scale detection of antigen-specific T cells using peptide-


Boon, A. C. M. *et al.* (2002) ‘Sequence Variation in a Newly Identified HLA-B35-Restricted Epitope in the Influenza A Virus Nucleoprotein Associated with Escape from Cytotoxic T


Chicz, R. M. et al. (1992) ‘Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size’, *Nature*, 358(6389), pp. 764–768. doi: 10.1038/358764a0.


Cocchi, F. et al. (1995) ‘Identification of RANTES, MIP-1α, and MIP-1β as the major HIV-suppressive factors produced by CD8+ T cells’, *Science*. Science, 270(5243), p. 1811. doi:


251


977–987. doi: 10.1016/0092-8674(94)90376-X.


Ignatowicz, L. et al. (1997) ‘T cells can be activated by peptides that are unrelated in
sequence to their selecting peptide’, *Immunity*. Cell Press, 7(2), pp. 179–186. doi: 10.1016/S1074-7613(00)80521-X.


Kersh, G. J. et al. (1998) ‘High- and low-potency ligands with similar affinities for the TCR:


Lang, H. L. E. et al. (2002) ‘A functional and structural basis for TCR cross-reactivity in


Leisegang, M. *et al.* (2008) ‘Enhanced functionality of T cell receptor_redirected T cells is defined by the transgene cassette’, *Journal of Molecular Medicine*. J Mol Med (Berl), 86(5), 269


Liu, J. Q. et al. (1998) ‘Inhibition of experimental autoimmune encephalomyelitis in Lewis rats by nasal administration of encephalitogenic MBP peptides: Synergistic effects of MBP


Merkel, F. et al. (1996) ‘ Autoreactive T-cells in Goodpasture’s syndrome recognize the N-


van der Merwe, P. A. *et al.* (1994) ‘Human cell-adhesion molecule CD2 binds CD58 (LFA-3) with a very low affinity and an extremely fast dissociation rate but does not bind CD48 or CD59.’, *Biochemistry*, 33(33), pp. 10149–60. doi: 10.1021/bi00199a043.


Publishing Group, 368(6474), pp. 864–867. doi: 10.1038/368864a0.


Parkhurst, M. R. et al. (2011) ‘T cells targeting carcinoembryonic antigen can mediate


Reiser, J. B. *et al.* (2002) ‘A T cell receptor CDR3beta loop undergoes conformational changes of unprecedented magnitude upon binding to a peptide/MHC class I complex.’, *Immunity*, 16(3), pp. 345–54. Available at:


10.1038/nm0195-59.


6976. doi: 10.1073/pnas.80.22.6972.


Investigation, 118(10), pp. 3390–402. doi: 10.1172/JCI35449.


Spits, H. et al. (1982) ‘The role of T cell differentiation markers in antigen-specific and


Valitutti, S. *et al.* (1995) ‘Serial triggering of many T-cell receptors by a few peptide-MHC


