A Functional and Structural Study of HLA-B*2705 Restricted CTL Responses Associated with Delayed HIV-1 Disease Progression

Phillip Pymm

Thesis Submitted for the Degree of Doctor of Philosophy

Green Templeton College
Michaelmas Term 2012

Supervisors
Dr. Astrid K. N. Iversen
Professor Lars Fugger

Abstract

The HIV-1 Gag p24 protein contains the HLA class-1 B*2705 restricted epitope KK10, responses to which are associated with delayed progression. Data from *in vitro* proteasomal digestion studies from our group has shown the production of a number of C-terminally extended and truncated epitopes containing KK10, produced in far higher quantities during proteasomal digestion than this "optimal epitope" and that the amount of antigen made in proteasomal digestion is instrumental in determining the development of immunodominance. This work aims to characterise the contribution of these naturally processed epitope forms to the cellular immune response to this region.

Further proteasomal digestion studies have shown that the common KK10 intra-epitope escape mutant sequences R132K and L136M have major effects on epitope production by the proteasome and that a range of short peptides containing the N-terminal of the KK10 sequence are produced in large quantities by the proteasome. Recognition of the KK10 epitope forms by HLA B*2705 HIV-1 patients were characterised *ex vivo* and show recognition of KK10 epitope forms somewhat independent of the presence of KK10 recognition, we also show cross-recognition between KK10 epitope forms by CD8+ T cells, as well as recognition by CD4+ T-cells. TCR from CD8+ T-cells specific for KK10 epitope forms were found to share common features in the HLA binding CDR hyper-variable loops.

Structural studies of the HLA B*2705 molecules in complex with the KK10 epitope forms show a shared binding motif at the N-terminus, and to a lesser extent, the C-terminus of the binding groove which may facilitate cross-recognition of complexes. In addition these studies show a potentially novel binding mode for a 14mer peptide, and refolding of truncated KK10 peptides as short as a 4mer with the HLA B*2705 molecule (crystallisation with a 6mer peptide shown). This demonstrates previously unrecognised flexibility of the HLA class-1 to bind and present peptides of different lengths to T-cells.

We show that these HLA B*2705 binding-capable truncated peptides do not induce a CD8+ T-cell response in HLA B*2705 HIV-1 patients and may be able to block CD8+ T-cell responses to the KK10 epitope. This might represent a novel form of viral CTL escape. In addition we observe the presence of KK10 flanking mutations in patient sequences and significant associations between the presence of intra-epitope escape mutations, KK10 recognition and patterns of escape in flanking sequences.

Finally we note the reduction in binding of KIR3DL1 to KK10 epitope forms relative to the KK10 epitope-HLA B*2705 complex. The presence of HLA B*2705 and KIR3DL1 associate with improved disease course in HIV-1 though the mechanism through which this occurs has yet to be defined.

Acknowledgements

In the (Ahem) several years that it has taken me to reach this stage, I have relied on the help, advice and support of many people. The first of these that I would like to thank is Dr. Astrid Iversen, without whom I would not have been able to complete this project. She has given advice, knowledge and demonstrated a considerable amount of patience in guiding me through this research, as well as being instrumental in providing me with opportunities to learn and undertake a variety of techniques and develop ideas. I would also like to thank Professor Lars Fugger and his lab who have freely given assistance throughout my degree and particularly in the final stages of my project.

There are a number of collaborators, without whom, much of the work undertaken in this project could not have been performed. I would like to thank Dr. Stefan Tenzer and Professor Hansjorg Schild for undertaking the antigen processing studies for the project, Professor Peter van Endert for undertaking TAP transport assays and Dr. Simon Kollnberger for reagents, cells and advice – especially with a temperamental FPLC. Dr. Maria Harkiolaki provided invaluable guidance, teaching and work on the crystallographic studies. I would particularly like to thank her for being there when there was yet another refold to test, another Saturday of data collection and more crystals to extract in the cold room. I am also grateful to Professor Yvonne Jones for the opportunity to perform our studies at the Division of Structural Biology. I would also like to thank Professor David Price and his lab for agreeing not only to collaborate with us on the project, but also for having me under their feet for a month of the summer to learn new techniques and Professor Victor Appay for kindly providing me with reagents.

The members of the Iversen lab over the years have all been of enormous help, I would particularly like to mention Dr. Edd Wee, who has assisted with several of the experiments in this project, has reminded me of many a forgotten reagent and whose determined organisational efforts have kept the lab functioning despite my best attempts over the years.

The WIMM has been a fantastic place to work during my time here and there have been numerous colleagues who have provided help, support and friendship, whether it was a spare reagent, a coffee or a slice of cake. I can't mention everyone by name but I would like to thank my friends inside and outside of the lab including Sally, Gen, Amy, Callum, Laurie, Erie, Alba and Jon who have been invaluable in keeping me, if not entirely sane, then at least functional throughout my time here (if not always on Sunday mornings). On a project related note I'd like to thank Sally for reading through some of this thesis and trying desperately to insert some consistency and sense into my use of abbreviation and grammar.

I would not have made it to this stage without the unending support and love of all my family. My parents, Pauline and Pete have not only encouraged me to do all that I am capable of and not to give up but have led by example. I am incredibly lucky to have them and admire them enormously, as I do my sister Patti, whose selflessness, drive and sense of fun (though somehow still managing to have more common sense than I ever will) have always been an inspiration.

Finally, I would like to thank my partner Aaron for being fantastically supportive from the beginning of my degree, putting up with odd timetables, constant travel and long phone calls on obscure experiments (though not always entirely willingly), and also for making my life outside the lab the best time I have had so far.

List of Abbreviations

AIDS	**Acquired Immunodeficiency Syndrome**
APC	**Antigen Presenting Cell**
bp	**Base pairs**
CD	**Cluster of Differentiation**
CDR	**Complementarity Determining Region**
CTL	**Cytotoxic Lymphocyte**
D_β	**Diversity region of the TCR β chain**
ELISpot	**Enzyme Linked Immunospot Assay**
ER	**Endoplasmic Reticulum**
ERAP	**Endoplasmic Reticulum Amino-peptidase Associated with Antigen Processing**
FACS	**Fluorescence-activated Cell Sorter**
FITC	**Fluorescein Isothiocyanate**
FPLC	**Fast Protein Liquid Chromatography**
GALT	**Gut-associated Lymphoid Tissue**
HAART	**Highly Active Antiretroviral Therapy**
HIV	**Human Immunodeficiency Virus**
HLA	**Human Leukocyte Antigen**
HTLV	**Human T-Lymphotropic Virus**
IFN	**Interferon**
IL	**Interleukin**
J_α	**Joining region of the TCR α chain**
J_β	**Joining region of the TCR β chain**

KDa	**Kilodaltons**
KIR	**Killer Immunoglobulin-like Receptor**
MHC	**Major Histocompatibility Complex**
ml	**Millilitre**
µl	**Microlitre**
MLV	**Murine Leukaemia Virus**
mRNA	**Messenger Ribonucleic Acid**
NK	**Natural Killer**
PBMC	**Peripheral Blood Mononuclear Cells**
PBS	**Phosphate Buffered Salt**
PE	**Phycoerythrin**
SFU	**Spot-forming Unit**
SIV	**Simian Immunodeficiency Virus**
TAP	**Transporter Associated with Antigen Processing**
TCR	**T Cell Receptor**
TNF	**Tumour Necrosis Factor**
V_α	**Variable region of the TCR α chain**
V_β	**Variable region of the TCR β chain**
T_H	**T-Helper Cell**
T_{REG}	**T-Regulatory Cell**
β2M	**Beta 2 Microglobulin**

Contents

1. Introduction

1.1. The Human Immunodeficiency Virus-1 (HIV-1) and Acquired Immunodeficiency Syndrome (AIDS)

Human immunodeficiency virus-1 (HIV-1) poses one of the greatest global health challenges we face today, with an estimated 33 million people infected worldwide, and two thirds of these (22 million) in sub Saharan Africa[1]. Although deaths from AIDS have been significantly reduced following the development of, and increasing access to, highly active antiretroviral therapy (HAART) in many of the countries worst affected by the pandemic[1], there were still an estimated 2 million deaths in 2008 alone and over 25 million people have died from AIDS since 1981.

1.2. Brief outline of the HIV-1 virus, infection and transcription

HIV-1 is a retrovirus of the lentivirus subfamily. Retroviruses are named for their ability to reverse transcribe their RNA genomes into DNA and to incorporate them into the host genome through the action of a reverse transcriptase enzyme[2]. The lentivirus subfamily is a group of these viruses, which are able to insert their genome into non-dividing cells as well as those actively undergoing division. Retroviruses infect a range of animals, with several well-known mammalian retroviruses, including the murine leukaemia viruses (MLV), human T-lymphotropic viruses (HTLVs)-1-4 and the HIV-1 related simian immunodeficiency virus (SIV), and human immunodeficiency virus-2 (HIV-2)[3]. These viruses have a range of effects on their hosts, though the ability to cause chronic infection and to adversely affect the immune system is common to many viruses within this family.

The HIV-1 virus particle consists of two copies of the RNA genome of the virus with the protease, integrase and reverse transcriptase enzymes surrounded by a capsid made from

the p24 protein. The capsid is surrounded by the p17 protein that forms the matrix and outside this, the viral envelope. This consists of the host cell membrane with viral gp41 and gp120 proteins and membrane components derived from the host cell.

HIV-1 entry into the host cell is achieved through the binding of the viral gp120 protein to a co-receptor (usually CCR5 or CXCR4) along with the CD4-molecule on the cell surface[4; 5]. This binding induces a conformational change in the gp120 protein and exposes gp41, which is then able to insert its transmembrane region into the cell membrane. Which of the two co-receptors it is able to bind primarily controls the tropism of the virus: either CCR5 (R5 tropic) or CXCR4 (X4 tropic), or both (dual tropic). R5 tropic viruses are able to bind macrophages and CD4+ cells, while X4 tropic viruses preferentially infect activated CD4+ T-cells and are able to induce syncytia formation between cells. X4 tropism is linked to a rapid decline in CD4+ T-cell count and progression to AIDS[6].

Upon entry into the cell, uncoating of the virus occurs and the pre-integration complex (PIC), consisting of the contents of the viral capsid, is released directly into the cytosol. This includes Vpr, matrix and integrase, host proteins and two strands of the RNA genome. This complex remains intact in the cytosol and reverse transcription of the viral RNA into double stranded DNA occurs. Nuclear localisation signals then result in the transport of the PIC to the nucleus of the cell through the nuclear pore complex, allowing HIV to replicate in non-dividing cells with intact nuclear membranes. Following entry, the viral DNA is integrated into the host genome at sites being actively transcribed as “provirus” through the actions of the integrase and transactivator of transcription (Tat) proteins[7; 8; 9; 10]. Host cell nuclear factor kappa-light-chain-enhancer (NF-κB) binds to NF-κB binding sites in the viral DNA act to promote transcription of the sequence and begin the production of new viral particles[11; 12].

The HIV-1 RNA genome consists of nine genes with long terminal repeats at each end of the genome. The proteins of HIV are produced from these genes, some in different reading

frames. This allows the production of many proteins from a small genome. Of the nine genes present, three are structural genes common to all retroviruses: *gag, pol* and *env*. In addition there are six accessory/regulatory genes present: viral infectivity factor (*vif*), viral protein R (*vpr*), transactivator of transcription (*tat*), regulator of viral protein expression (*rev*), negative regulatory factor (*nef*) and viral protein U (*vpu*).

Gag encodes proteins that make up the capsid of the viral particle, including p17 (outer capsid, the matrix protein) and p24 (inner capsid, the capsid protein). These proteins are cleaved by the HIV protease in the forming viral particle or the immature virus after release from the cell. *Pol* encodes this protease as well as other viral proteins required for the completion of its lifecycle, including the reverse transcriptase and the integrase. *Env* encodes the proteins displayed on the viral envelope, gp41 and gp120. These two glycoproteins are also cleaved from a gp160 precursor, however, this action is not performed by the viral protease and instead these proteins proceed through the cellular endoplasmic reticulum and Golgi complexes, undergoing extensive glycosylation by the host machinery.

1.2.1. HIV-1 Reverse transcriptase and *in vivo* heterogeneity of the virus

The ability of the virus to transcribe its RNA genome to DNA and integrate it into the host genome requires the action of the viral reverse transcriptase and integrase enzymes. Once it was discovered that the virus encoded a reverse transcriptase, it became an immediate candidate for drug development, as it is vital for viral replication yet its inhibition would have no effect on normal host cell function. Indeed, two classes of antiretroviral drugs that target this enzyme, the nucleoside analogue reverse transcriptase inhibitors (NRTIs) were the first drugs to be shown to be effective and licensed for the treatment of HIV[13; 14]. However, when used alone, resistance to these drugs can occur relatively quickly.

The NRTI class of drugs exert inhibition through their incorporation into the nascent DNA chain during transcription, resulting in the termination of synthesis of the chain. The incorporation of these inhibitors is allowed due to the relatively low specificity of the reverse transcriptase for its substrates[15]. The same mechanism also results in frequent errors in substrate-usage during normal, uninhibited replication. These errors mean that the wrong DNA bases are encoded for the corresponding RNA, resulting in a mutated sequence. The high error rate of between 1.5×10^{-5} and 3×10^{-5}substitutions/site/replication[16] results in an extremely heterogeneous population of virus within an infected individual, and this, in combination with frequent viral recombination and a relatively high viral load during in most untreated patients is in fact the mechanism that allows the virus to develop resistance to antiviral therapies. It has been estimated that in untreated individuals, each possible mutation in the viral genome will occur every 1 to 2 days, with replicative fitness and selection pressure shaping the resulting population.

The high mutation rate means that there will be some viral sequences that encode mutations that will confer resistance to the effects of these drugs, for example blocking the incorporation of the NRTIs into the nascent DNA chain during synthesis. In an untreated individual these viral sequences will have no survival advantage over wild-type, however, once the patient is treated with antiretroviral drugs, viruses containing these mutations will be able to replicate successfully where others will not, and will become the majority of the population. In the same way, those viruses with mutations that allow escape from antiviral responses generated by the adaptive immune system are able to avoid elimination from the host[17; 18; 19]. The selection of viral sequences by immune responses and the fitness cost to the virus of these mutations will be discussed later in the chapter (Chapter 1.5).

With the advent of HAART, which involves the administration of a combination of drugs, usually three, targeting different processes essential for completion of the viral life cycle, viremia can be controlled for extended periods of time[20]. However, the virus is not

eliminated by this treatment and treatment breaks usually result in the return of uncontrolled viral replication[21]. Contributing to this rapid resurgence upon treatment discontinuation is the production of new virus from reservoirs in latent cells. This is predominantly a population of CD4+ memory T-cells, remaining from the initial expansion of HIV-1 specific cells formed during acute infection[22; 23]. These long-lived cells will contain the viral DNA incorporated into the host genome by the integrase enzyme, but as they are not activated will not have been producing virus and therefore may have avoided destruction by the immune system. Activation of a percentage of these cells, possibly through antigenic stimulation if they are CD4+ memory T-cells, at any given time will allow production of virus and during a treatment break successful replication can occur[23].

1.3. The course of infection of HIV-1

Initial infection with HIV-1 can occur through mucosal surfaces during sexual intercourse, or following direct intravenous inoculation. Over 80% of new infections are thought to occur at mucosal surfaces[24], with varying risks of transmission dependent on intensity and location of exposure as well as the clade and tropism of the virus to which an individual is exposed[25; 26; 27]. At these surfaces, HIV-1 is likely to be able to infect Langerhans' cells within the epithelium[28; 29] and sub-mucosal CD4+ T-cells. Dendritic cells may transport live virus on their surfaces through attachment to the DC-specific ICAM3-grabbing non-integrin (DC-SIGN) on the cell surface,[30; 31] these viruses may then infect CD4+ T-cells that contact the dendritic cell (DC)[31]. Though macrophages are able to be infected by HIV-1 and are also present in the sub-mucosa, it has been suggested that newly transmitted viruses are less able to infect these cells than CD4+ T-cells, and therefore they may not form a primary route of infection[32]. Analysis of both the first viral sequences detectable, and populations present in patients during chronic disease has shown that in the majority of mucosal infections transmission and successful infection is due to a single viral particle[32; 33; 34; 35; 36; 37]. Following

infection, timing of post exposure prophylaxis in animal models shows that a reservoir of infected cells may be set up in as little as 24 hours, though 3-6 days is the median time for this to occur[38; 39].

Following initial transmission and regardless of the location of the transmission event, one of the first sites of infection is the gut-associated lymphoid tissue (GALT), in particular CD4+ T-cells within this tissue, of which 20% can be infected in the weeks following transmission[40; 41; 42]. Up to 60% of CD4+ T-cells will commit bystander apoptosis as a result of the infection and the resultant phosphatidylserine^{+} microparticles have been shown to inhibit the immune response[43]. The widespread infection of the GALT allows the establishment of viral reservoirs and destroys much of the structure of the lymphoid tissues in the region[44]. This is not recovered even with successful HAART[40; 41; 42].

During initial infection, viral replication is poorly controlled in the absence of adaptive immunity, which has not yet been mobilised, and the virus spreads throughout the GALT and other lymphoid tissues in the body. Viral load increases rapidly and peak viremia usually occurs three to four weeks following infection[45] (Figure 1). During this phase CD4+ T-cell counts are severely reduced, and although CD8+ T-cells expand and B-cell numbers are not significantly affected by the virus, functions of these cell types that rely on CD4+ T-cells, such as CD4-dependent antibody production, as well as those that require intact germinal centres (particularly in the gastrointestinal tract) are disrupted by the loss[45]. The immune activation caused in this acute period of infection also results in production of several cytokines in large quantities[46], some of these return to normal levels following acute infection, however others remain elevated throughout infection including IL-22, IFN-γ and IL-18. Anti-inflammatory cytokines such as IL-10 exhibit a delayed response to the acute infection[47]. The rise in cytokines is higher than seen upon infection with other viruses capable of persistence such as hepatitis-B (HCB) and hepatitis-C (HCV), possibly indicating that they are not making a beneficial contribution towards effective viral clearance. It has been suggested that the

immunopathology generated by the response may be counter-productive and may worsen disease progression[46].

Following this peak in viremia, viral load in the body begins to decline and reaches a set-point, which can vary for each individual, usually by approximately 100 days post-infection[45] (Figure 1). As viral load declines, CD4+ count in the blood will return to near normal levels, though this is not mirrored in the GALT[42]. Following the establishment of viral set-point, the infection can remain under control for a number of years, although CD4+ T-cell counts slowly decline during this period. A rapid decline in both CD4+ and CD8+ T-cell counts immediately precedes the development of AIDS, a stage of the disease defined by CD4+ T-cell counts below 200 cells/µl and/or the presence of an AIDS defining condition[48].

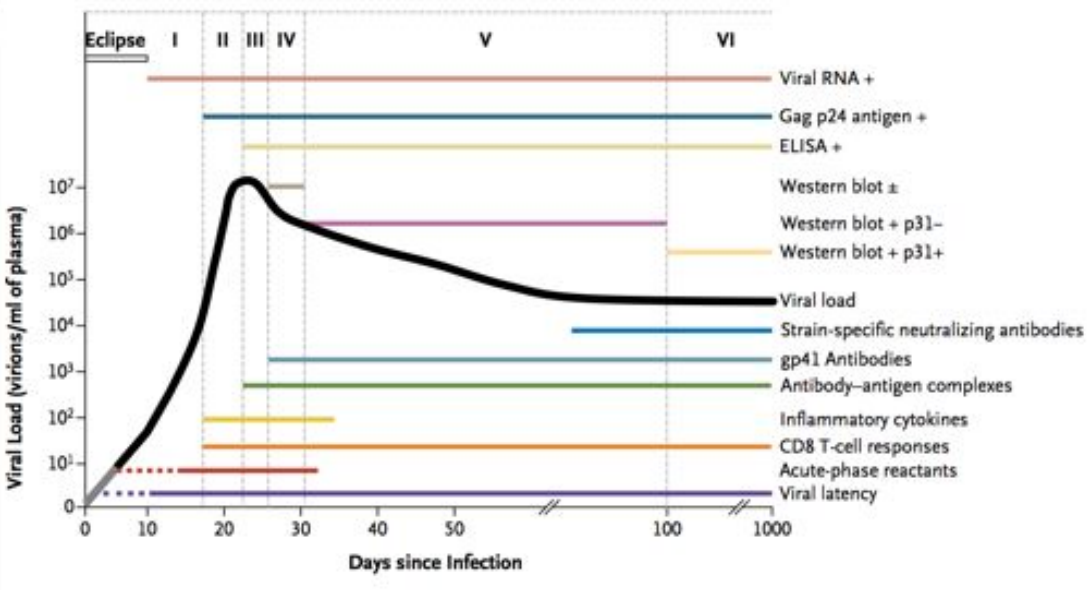


Figure 1: Acute HIV infection and the six stages of HIV infection (Roman numerals I-VI). These stages are based on the appearance of positive immune markers used for detection and staging of HIV infection in standard laboratory tests, a + sign marks a positive result and a – sign indicates a negative result. The coloured lines indicate the approximate time post infection that these positive/negative readings are valid. Reproduced with permission from[49], copyright Massachusetts Medical Society.

1.3.1. Viral Escape from the human immune system

Just prior to peak viremia, the first HIV-1 specific CD8+ T-cell responses can be detected in the blood[50; 51]. Prior to this point the viral population is largely homogeneous with few signs of escape being selected by the immune response[32; 37]. The initial CD8+ T-cell responses are usually directed towards a few peptide epitopes and often generate the rapid selection of escape mutants leading to loss of the antigen and response, although occasionally patients have displayed effective initial CD8+ T-cell responses that remain into chronic infection[51; 52; 53; 54; 55; 56].

Subsequent CD8+ T-cell responses that are seen following this initial wave induce further selection of the virus and generate greater diversity among viral sequences in the patient[51]. In some patients a minority of these responses in some patients induce changes that appear to reduce viral fitness and slow replication, providing an effective response[57], others show no or slow selection of escape mutants in their target epitope. This is sometimes due to the inability of the virus to mutate these regions without inducing significant fitness costs, and marks an effective immune response[58; 59]. Initial HIV-1 specific antibody responses are also seen at around the time of peak viremia (Figure 1). These antibodies are almost always non-neutralising and directed toward the gp41 protein[60]. The appearance of neutralising antibodies is not usually seen until several months following initial infection; at around the time that viral set-point is reached. It has further been shown that the majority of selection pressures associated with reduction of viral load following peak viremia appear to be caused by CD8+ T-cell responses to the virus[51]. Some unexplained mutations could be due to reversion of mutations transmitted from the previous host[61; 62], or could be due to selection pressure from natural killer (NK) cells which have also been shown to be able to control viral replication in combination with certain human leukocyte antigens (HLA). For example, KIR3DL1/KIR3DS1-expressing NK cells have been shown to expand during acute infection in

HLA Bw4-possessing individuals and are able to control the replication of target cells expressing these HLA molecules *in vitro*[63; 64].

The ability and requirement for CD8+ T-cells to control initial viremia and to establish a set-point for the viral load has been demonstrated in Rhesus Macaques following infection with SIV. Depletion of CD8+ T-cells in both acute and chronic infection led to the loss of control of viremia in these animals[65]. Furthermore, certain HLA class-1 types are strongly associated with altered disease progression in HIV-1. This was recently confirmed in genome wide association studies (GWAS) and related studies looking at associations for protection from HIV-1 disease progression[66; 67; 68].

1.4. The human immune system and the response to HIV infection

While neutralising antibodies do eventually appear in many patients, these are usually strain-specific and are directed towards the founder virus, therefore neutralising only a small proportion of the virus in the host at the time of their appearance[69; 70]. However, a minority of patients will develop antibodies that are broadly neutralising[71]. The patient does not benefit from induction of broadly neutralising antibodies (BNABs) as viral escape occurs easily for any one antibody[72; 73], however combinations of BNABs have been tested as prophylactics and as treatment strategies, and have so far have met with limited success[74]. While eliciting broadly neutralising antibodies following vaccination remains a significant area of research[75], and indeed has generated renewed interest following the mixed results of T-cell vaccine trials in recent years[76; 77], further examination of cellular responses that best control this disease in natural infection, and particularly of immune responses that are associated with protective effects[78; 79] may provide insight into improving immune control in all infected individuals. We are therefore concentrating on the cellular response to HIV-1 infection in this study, and in particular the two arms of the immune system that have been

associated with improved control of the disease in certain individuals, the CD8+ T-cells and the NK cells.

1.4.1. The cellular immune response

T-cells are cells of the lymphocyte lineage that leave the bone marrow during differentiation and finish development in the thymus[80]. These cells recognise antigens through a T-cell receptor on their surface. This receptor binds peptides in complex with the HLA presented on the cell surface of antigen-presenting cells (APC).

The development of T-cells is accomplished through a series of stages, from lineage commitment and thymic colonisation, to co-receptor expression and the formation and selection of functional T-cell receptor (TCR)[81] (Figure 2). Lineage commitment begins before thymic colonisation and occurs throughout the early stages of thymocyte development. Early markers of thymocyte potential and thymus-homing are the expression of CCR9 and CCR7 chemokine receptors on the double negative (DN) lymphoid progenitor cell surface[82; 83]. The activation of target genes downstream from the notch1 receptor also promotes T-lineage commitment when the notch ligand delta-like 4 (DL4) is encountered in the thymic environment[84; 85; 86]. Thus the T-cell development programme is initiated through the effects of notch1 activation along with the actions of transcription factors including Runx1 and GATA-3[87; 88; 89; 90].

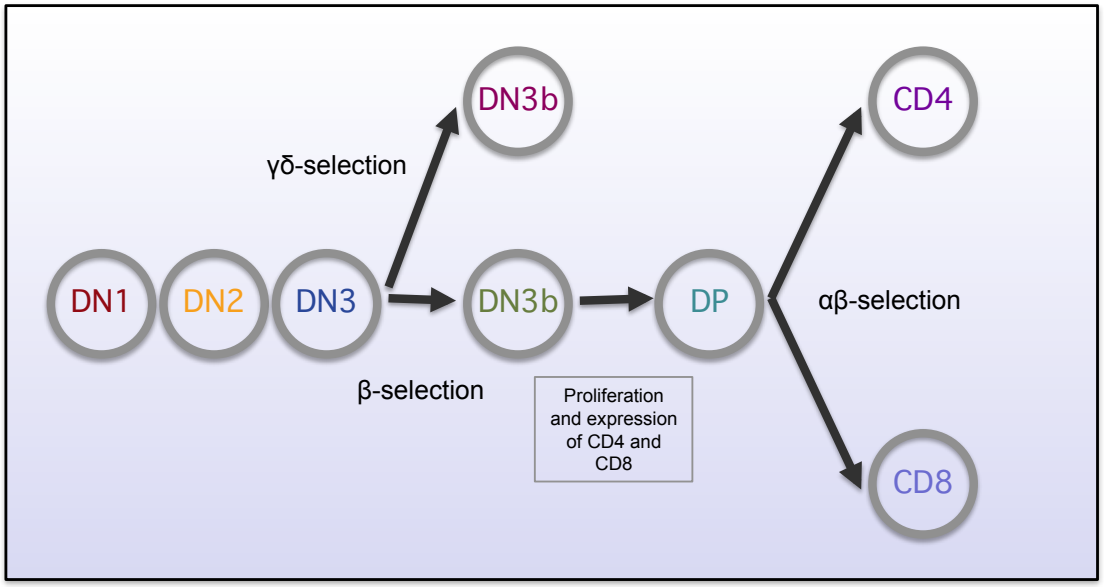


Figure 2: A basic schematic showing the stages of αβ T-cell development from the early double negative thymocyte to the single positive CD4+ and CD8+ αβ T-cells following positive and negative selection processes.

At the DN3 developmental stage of T-lymphocytes, the T-cell receptor begins rearrangement. The T-cell receptor is made up from two chains that are formed by rearrangement of several gene segments to form a single functional T-cell receptor sequence, which can then be expressed. This forms a protein containing four immunoglobulin-like domains with a transmembrane region and a short cytoplasmic tail. Three hypervariable loops on each of the TCR chain variable regions are located on the antigen-binding surface of the TCR and residues in these loops are the contact residues in the interaction between the T-cell receptor and the HLA-peptide complex[91; 92]. There are four types of chain encoded for the T-cell receptor in the genome: the α, β, δ and γ. The α and β chains form the T-cell receptor for conventional CD4+ and CD8+ T-cells, while the γ and δ chains form the TCR of a distinct set of T-cells largely recognising non-peptide, phosphorylated antigens[93; 94].

The constant (C) segment, of which there are 1-2 encoded in the germline for each chain, is joined to a variable region made up from a combination of variable (V), diversity (D), and joining (J) segments (no D segment in the α and γ chains), for which there are many

sequences encoded in the genome. At the DN3 stage, the TCR β, γ and δ chains are all able to undergo rearrangement[81; 95]. A minority of T-cells forming a functional δ and γ chain prior to a functional β chain will go on to form γδ T-cells[96], with the remainder forming αβ T-cells. This process is known as β-selection in αβ T-cells and γδ-selection in the γδ-lineage. Differential transcriptional activation is seen in these two subsets. This occurs following the successful production of either γδ or β chains at the DN3 stage and may further differentiate the populations[97; 98]. The rearrangement of these chains is controlled by the gene products RAG-1 and RAG-2 from the recombination activating genes (RAG). These proteins initiate and regulate rearrangement using non-homologous DNA end joining (NHEJ) machinery to join the TCR gene segments[99]. Joining of the J to D segments occurs prior to the joining of the V to the DJ segments. The joining of the segments in this order is partly controlled by the recombination signal sequences present at either end of each segment. These can consist of either 12 or 23 nucleotides, and joining will usually only take place between segment ends containing heterogeneous signal sequences[100], although this is not the only mechanism of control. The D segments can be joined to the J segment through either deletion or inversion of the segments, however, there is a strong preference for deletional insertion[101; 102].

Additional diversity is provided in the TCR by an enzyme known as terminal deoxynucleotide transferase (TdT), which adds diversity at the junctions of the segments through the addition of several untemplated nucleotides[103; 104; 105]. The addition of these untemplated nucleotides can result in frame shifts 3' of the addition, giving a majority of TCR sequences that are partly out of frame and unable to form a functional chain. The junctions of the V, D and J segments of the TCR chain (V and J segments in the TCRα chain) encode the region of the TCR that will become the CDR3 loop in the gene product. The sequence of this loop is therefore highly diverse in functional TCR and is the region of the receptor that forms the primary contact with the peptide antigen in the majority of TCR-HLA complexes observed to date[106; 107; 108]. Formation of a functional β chain is recognised by pairing the TCRβ gene product with a pre-TCRα chain and the expression of this pre-TCR on the cell

surface[81; 109]. Recognition of this pre-TCR inactivates the RAG genes and prevents further TCRβ formation[110]. The signalling events following the expression of the pre-TCR[111; 112] also induce proliferation of the T-cell and the expression of both the CD4+ and CD8+ co-receptors. At this stage, the cell is known as a double positive (DP) thymocyte.

Rearrangement of the TCRα chain occurs following proliferation and expression of the co-receptors with re-activation of RAG-1 and 2[99; 113]. This can occur until a successful rearrangement is obtained and is not limited by allelic exclusion. T-cells with functional TCR will then undergo positive and negative selection[114]. T-cells are positively selected for recognition of HLA molecules, but negatively selected if this recognition is strong in complexes involving presentation of self-peptides[115]. Presentation of self-peptides, found in most immunologically accessible sites in the body, occurs on the surface of medullary epithelial cells of the thymus and is partly governed by the effects of the autoimmune regulator (AIRE) transcription factor[116] and the presence of a thymus-specific proteasomal subunit β5t, which enhances the production of the HLA class-1 peptide repertoire[117]. Some T-cells possessing strongly self-peptide-HLA reactivity will, instead of undergoing cell death, become a type of T-regulatory (T_{REG}) cell[118; 119].

1.4.2. The CD8+ and CD4+ T-cells

During the stages of selection following the production of successfully rearranged TCR, cells that show weak recognition of HLA will down-regulate either the CD8+ or CD4+ receptor. The decision of which of these co-receptors is down-regulated and how the final HLA class specificity of the T-cell is decided requires further study, however, several mechanisms have been proposed as investigation of this process has progressed. The strength of signal model (Figure 3A) and the kinetic signalling model (Figure 3B) are two recent models. Current evidence suggests that of these, the kinetic signalling model is likely to be largely correct due

to the extended signalling time required for the development of CD4+ cells compared with CD8+ cells[120; 121; 122]. Though some differences in TCR binding to HLA class-1 and class-2 have been observed[123; 124] there is currently no clear distinction between TCR recognising class-1 or class-2, indeed, TCR have been shown to bind both HLA class-1 and class-2 in structural studies in the absence of the co-receptor[125]. Furthermore in mouse models where CD4+ expression is abrogated, CD8+ T-cells that recognise MHC class-2 have been shown to develop[126]. Differential requirements of some TCR for co-receptor signalling also add to the complexity of this system[127].

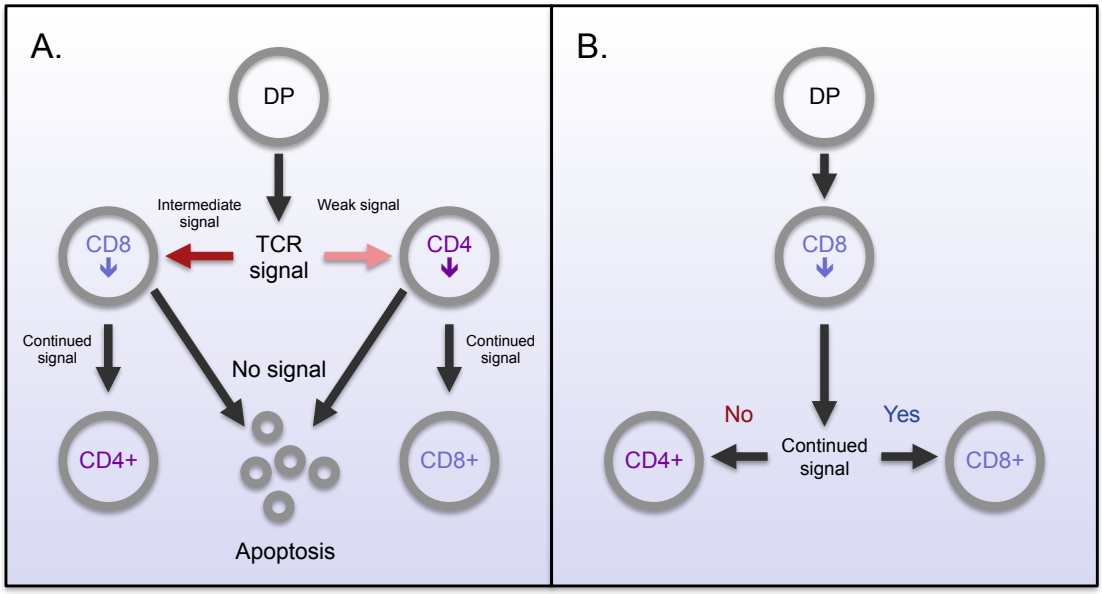


Figure 3: A basic schematic of the proposed models for co-receptor selection in thymocytes A. Strength of signal model and B. Kinetic signalling model.

Functionally, CD8+ and CD4+ cells are divided into distinct phenotypes by their cytokine secretion patterns and the effector functions that they carry out. The majority of CD8+ T-cells form a subset known as T_C1 effector cells that are characterised by the expression of genes for the production of IFN-γ and perforin following antigen recognition[128; 129]. CD8+ T-cells may produce additional cytokines such as TNF-α and IL2 following activation, in which case

they are known as polyfunctional T-cells and may have modified effector capacity compared with CD8+ T-cell producing a single cytokine[130; 131; 132].

Activation of these cells occurs in the lymph node, when they recognise and are activated by antigen presented in complex with HLA on the surface of a DC that has matured in the periphery[133; 134]. This results in the expression of the α chain of the IL2 receptor (CD25) and CD69. CD25 drives clonal expansion of the T-cell in the presence of IL-2[135], which is produced by the newly activated CD8+ T-cell[131; 136; 137]. CD69 functions to retain activated CD8+ T-cells in the lymph nodes during this initial expansion through the inhibition of S_1P_1, which is required to allow lymphocytes to leave the lymph node. This results in dramatic clonal expansion in the presence of IL-2, antigen and stimulatory receptors such as B7 (CD80/86) on the surface of DCs, and the acquisition of an effector phenotype. Following this the activated T-cells are released into the periphery to carry out effector functions[138; 139].

An important effector function of CD8+ T-cells is their cytotoxicity, which can be carried out in one of two ways. The first is the release of perforin and granzymes from granules stored within the cell; upon exposure to calcium ions in the extracellular fluid, perforin multimerises and forms pores through which the mixture of proteases including granzyme A and B can pass through and initiate apoptosis of the cell, this process takes only minutes to occur[140; 141; 142]. The second pathway involves transcriptional activation of the Fas ligand (FasL) and so requires several hours to induce. This ligand binds to Fas on the target cell surface and induces apoptosis through activation of the caspase cascade[143].

The differentiation and effector phenotypes of CD4+ T-cells are more diverse than that of the CD8+ lineage, and differ markedly in effector function. However, a minority of CD4+ T-cells has been shown to possess cytotoxic effector functions similar to that of the CD8+ T-cell[144; 145; 146; 147], though the relative use of the pathways of cytotoxicity may differ from that of classical CD8+ T-cells[147]. The majority of CD4+ T-cells are divided into subsets based on their

pattern of cytokine secretion and include T_H1, T_H2, T_H17 and T_{REG} subsets[148]. Some studies have described additional subsets including T_H3, T_R1 and T_H9 cells, each producing a signature cytokine[149; 150; 151; 152]. Differentiation of these subsets of CD4+ T-cells is mediated through modification of chromatin structure, partly dependent on the cytokine environment at the time of antigen exposure[153]. This modification allows transcriptional regulators such as T-bet in T_H1 cells and GATA-3 in T_H2 cells to access genes for production of certain sets of cytokines[154; 155; 156].

T_H1 cells are characterised by their production of IFNγ and their responsiveness to IL12[157; 158]. The production of cytokines such as the TNF family, IL-2 and GM-CSF is also characteristic of this subset[159]. T_H2 cells are characterised by the production of IL4 as well as that of IL13 and IL5 and T_H17 cells by the production of IL17A, F, IL6 and TNF[159]. The effector responses of these cells are largely mediated by the effects their cytokine production have on other cell types, both locally in the tissues and systemically and on immune responses such as antibody isotype production.

The major cytokine of the T_H1 cell subset, IFN-γ acts on a number of innate and adaptive immune cell types to promote a pro-inflammatory response. This acts to induce maturation and up-regulation of HLA class-2 on DCs, phagocytosis of pathogens and infected cells by macrophages and up-regulation of the antimicrobial oxidative systems (NADPH oxidase and iNOS) as well as promoting activation, recruitment and cytotoxicity of CD8+ T-cells and NK cells[160; 161]. In T_H2 cells the production of IL-4 and IL-13 induces the alternative activation of macrophages[162], which produce chemokines that attract eosinophils and basophils to the site of infection. The response to these cytokines is further characterised by the induction of anti-inflammatory cytokines such as IL-10[163], up-regulation of receptors for recognition of fungal antigens such as dectin, as well as enzymes that degrade chitin[164]. IL-4 and IL13 also up-regulate mucin genes and increase turnover at mucosal epithelia, while IL-5 and IL-9 promote production and survival of eosinophils, basophils and mast cells[165]. Finally, the IL-

17 and IL-6 cytokine produced by the T_H17 cell subset act on local epithelial, stromal and endothelial cells to up-regulate production of inflammatory chemokines and haematopoietic growth factors that promote the production and release of neutrophils and monocytes, initiating a local pro-inflammatory response[166; 167].

In their production of this array of cytokines and their promotion of appropriate immune response to infection, the subsets of CD4+ T-cells provide essential help to other effector arms of the immune response. Depletion of the CD4+ T-cells in HIV-1 therefore impacts the entire immune system, including the survival and effectiveness of responses of CD8+ T-cells, rendering the host vulnerable to a variety of pathogens. This produces the classic symptoms seen in the development of AIDS.

1.4.3. The antigen processing pathway

As previously mentioned, the development of the cellular responses to foreign peptide antigens relies on HLA presentation of these antigens to the TCR on the cell surface. The processes by which these peptides are produced and are transported to the HLA for binding and presentation are extremely important as they, along with the HLA, determine the parts of the pathogen that will be exposed to the immune system and so direct the immune response. The peptides presented by the HLA class-1 and HLA class-2 molecules, while having some cross-presentation, particularly by DCs in their capacity as a professional antigen-presenting cell type[168; 169], are largely from two different sources. The HLA class-1 molecules present endogenous peptides, while the HLA class-2 molecules present peptides derived from exogenous sources[170; 171; 172]. HLA class-1 are present on the majority of cell types of the body, excepting red blood cells, to enable immune recognition of cells that are infected by intracellular pathogens such as viruses and certain bacteria. HLA class-2

expression is restricted mainly to cells that are more directly involved in regular antigen presentation to T-cells, including B cells, DCs and macrophages[173].

Exogenously produced protein antigens will be engulfed by the cell through pinocytosis[174] and enter the endosomal pathway. These proteins will undergo cleavage by a variety of endosomal proteases including cathespins, aspariginal endopeptidase and IFN-γ-inducible lysosomal thiol reductase[175; 176; 177]. These proteases will continue to cleave the protein to single amino acids if peptides are not first bound by the HLA class-2 molecules which are routed to the endosomal pathway by a chaperone protein known as the invariant chain[178; 179]. The invariant chain is cleaved on entry into the endosome, leaving a peptide in the binding groove known as the class-2 invariant peptide (CLIP)[180]. The complex is then bound by a second chaperone HLA-DM (and in some cell types additional chaperones) which releases the complex only when a peptide of high enough affinity to displace CLIP and stabilise the complex binds in the binding groove[181; 182; 183; 184; 185].

Endogenous peptides are processed and presented to the class-1 HLA through a very different pathway involving proteasomal degradation and transport into the endoplasmic reticulum where loading onto the HLA class-1 occurs. These peptides begin as proteins in the cell, usually those that have been ubiquitinated after oxidative damage or have reached the end of the time they are required for (for example, cell cycle related proteins)[186; 187]. There have been some studies which suggest that newly made proteins that are not translated or folded correctly also make up a significant portion of the proteins entering the proteasome[188].

The proteasome is a multi-subunit ATP-dependent protease complex consisting of two major parts, the 20S catalytic core and 1 or 2 19S regulatory subunits[189; 190]. Together these make up the eukaryotic 26S proteasome. The regulatory subunits are responsible for recognising ubiquitin tagged proteins and mediating the unfolding of proteins and their translocation

into the 20S core for cleavage[191; 192]. This is achieved with the aid of ATP-dependent proteasomal activators such as the IFNγ induced PA28, which stabilise the open conformation of the gated ends of the central passage of the proteasome, allowing translocation[193]. The 20S core of the proteasome is formed from α-type and β-type subunits, with the α-type subunits forming a heptameric ring at each end of the structure and the β-type subunits forming two central heptameric rings, all of which are stacked into a cylinder[194; 195].

Three of the central β-type subunits are responsible for the protease function of the proteasome, each with two active sites. Two of these cleave in a chymotrypsin-like manner, two in a trypsin-like manner and two in a caspase-like manner[196; 197; 198]. This gives the proteasome the ability to cleave at a wide range of peptide motifs including after hydrophobic, basic and acidic residues, producing peptides of between 3 and 22 amino acid residues in length[199]. During infection, the production of IFN-γ induces the expression of three alternative β-type subunits with altered cleavage properties[200; 201], these replace the constitutive subunits to form the immunoproteasome, which is the major form of proteasome present in IFNγ^{high} environments[202]. The altered cleavage has been shown to favour the production of peptides with carboxy-termini that are hydrophobic or basic[203; 204; 205; 206], corresponding to the binding preferences of the HLA class-1[207]. Additionally the immunoproteasome may have a role in reducing the impact of oxidative damage on cells[208].

Proteasomal cleavage is the major determinant of the C-termini of peptides presented by HLA class-1, as evidenced by the effects of proteasome inhibitors on HLA class-1 presentation[209]. However, there are several enzymes, both cytosolic and in the endoplasmic reticulum, that are involved in the generation of HLA class-1 antigen N-termini, in particular in the trimming of the N-termini of longer HLA class-1 epitope precursors[210]. These include the aminopeptidases, leucine aminopeptidase (LAP), thimet oligopeptidase (TOP), tripeptidyl aminopeptidase II (TPPII), puromycin-sensitive aminopeptidase, nardilysin and

bleomycin hydrolase in the cytosol[211; 212; 213; 214; 215; 216; 217], as well as the endoplasmic reticulum-resident aminopeptidases 1 and 2 (ERAP1 and ERAP2) in this compartment[218; 219; 220]. Together the preferences of these peptidases create the amino terminal ends of the HLA class-1 peptides.

The action of the aminopeptidases in the cytosol will eventually degrade the peptide to its constituent amino acids. Therefore, in order to escape complete degradation, transport into the endoplasmic reticulum and binding to the HLA class-1 must occur. This transport is carried out by the transporter-associated with antigen processing (TAP), an ATP-binding cassette (ABC) transporter consisting of a dimer made up from the TAP1 and TAP2 subunits[221; 222]. Transport takes place in a canonical fashion for ABC transporters, with a central cavity formed between the two subunits providing a binding site for the substrate. This cavity alternates between the cytosolic and luminal sides of the membrane using ATP to drive the conformational change[223; 224]. Unidirectional transport and substrate release is accomplished by the central cavity having a high affinity for substrate when on one side of the membrane, with the affinity being lower in the alternate conformation. Binding affinity for the peptide is therefore the major determinant of transport by TAP[223; 225].

Despite the lack of a crystal structure for the core TAP dimer, a combination of mutational analysis and homology modelling using two related ABC transporters, Sav1866 and MsbA, and the nucleotide binding domain (NBD) of TAP1 has built up a partial picture of the TAP binding site[223; 226; 227; 228; 229], peptide transport assays have further elucidated the peptide binding preferences of TAP and their impact on the availability of HLA class-1 epitopes[225; 230]. TAP must balance specificity for its substrate with the need to bind a diverse array of peptides that will also fit the anchor preferences of the HLA class-1, allowing their presentation, thus (human) TAP is relatively permissive, but preferentially transports peptides with particular features[231]. Human TAP shows preference for peptides with basic or hydrophobic residues at the C-terminus (a preference shared by the majority of class-1 HLA

and preferentially produced by the immunoproteasome), as well as a weaker preference for basic residues at the N-terminal residues[232; 233]. TAP also shows a length preference for peptide transport, with most of the optimally transported peptides being between 8 and 13 amino acid residues in length[225; 230].

On the luminal side of the endoplasmic reticulum membrane, TAP is associated with the peptide-loading complex (PLC). This is a complex consisting of the chaperone calreticulin, the ERp57 oxidoreductase, tapasin and the empty HLA class-1 molecule[234]. Following peptide transport, the PLC, and in particular the adaptor protein, tapasin, brings the translocated peptide into proximity with the unbound HLA class-1 and observations suggest it may have a quality control function[235]. In the absence of tapasin, the loading and presentation of many peptide epitopes is strongly impacted[236; 237]. Whether this is a direct reduction in peptide loading effectiveness, or a result of free ERp57, (that usually binds and maintains the stability of tapasin) reducing the disulphide bond in the α2 domain of HLA class-1 and the destabilisation of the complex[236; 237] is unknown. The action of ERAP1 and 2 is also likely to affect the availability of peptide epitopes for binding to the HLA, as these aminopeptidases will trim the N-termini of peptide epitopes entering the endoplasmic reticulum (ER). In some cases this trimming forms an N-terminus suitable for HLA class-1 binding and so will enhance availability, while in others, optimal N-termini may be lost through trimming. The relative speed of cleavage for different peptide epitopes will affect their stability in the ER and so their availability to the HLA class-1[238]. The final step determining the repertoire of epitopes presented to CD8+ T-cells on the cell surface is specified by the binding requirements of the HLA class-1 themselves.

1.4.4. The Human Leukocyte Antigens

The HLA genes in humans are known in other organisms and generically as the major histocompatibility complex (MHC). The HLA genes are located on chromosome 6 in a gene-dense region and are the most polymorphic gene family in the genome. They were first identified as factors controlling the failure of organ transplants, while their role in the immune system was elucidated at a much later date[239].

The HLA gene loci are split into three regions in the genome, HLA class-1, HLA class-2 and HLA class-3[240]. The HLA class-1 genes include genes encoding the α-chains of the three highly polymorphic conventional class-1 molecules, HLA-A, HLA-B and HLA-C. The HLA class-1 proteins form a heterotrimer, which includes the α-chain encoded here, along with an invariant β2M domain encoded outside the HLA region and an antigenic peptide for presentation to CD8+ T-cells. All of these elements are required for the formation of stable HLA class-1 complexes. The α-chain contains the binding groove for the peptide and the transmembrane helix, anchoring the complex in the plasma membrane. In addition to the conventional HLA class-1 the region contains genes for non-conventional HLA such as HLA-E, HLA-F, HLA-G and HLA-H, these are known as class-1b molecules and share a similar overall structure to class-1[241]. These non-conventional HLA have restricted polymorphism and are not conventionally involved in presentation of peptides to CD8+ T-cells however, the HLA-E, and perhaps the HLA-G, molecules are involved in interaction with NK cell receptors (HLA-E and the NKG2A receptor and tentatively HLA-G and KIR2DL4[241; 242; 243]).

The HLA class-2 molecules share a structure analogous to that of HLA class-1, however, the composition of the complex that forms this structure is somewhat different. In place of a single α-chain making up the binding groove, the HLA class-2 molecule instead consists of two similarly sized chains, α and β[244]. Each of these contains one of the two helices forming the sides of the peptide-binding-groove; additionally each chain contains a transmembrane

domain. As with the HLA class-1 genes, there are also three major conventional class-2 loci, the HLA-DR, HLA-DP and HLA-DQ. With the exclusion of HLA-DRA, these families are also polymorphic and allow presentation of a range of peptide antigens to CD4+ T-cells[244]. The region also includes genes coding for the HLA-DM molecule, a class-2 like chaperone for HLA class-2 in the endosome and for HLA-DO, an inhibitor of DM that functions in some cell types to delay HLA class-2 presentation[181; 182].

The HLA class-3 region includes a variety of genes, some coding for proteins with immunological function such as the TNFs and components of the complement system, while others have functions unrelated to immunity.

The polymorphic nature of the HLA class-1 and 2 molecules allows the presentation of a wide variety of antigenic peptides to both the CD4+ and CD8+ T-cells. This differs markedly between individuals as each HLA can present a range of peptides but differ in their binding preferences. The high polymorphism seen in the region is of benefit as it has been shown that heterozygosity of the HLA genes improves the immune response of an individual to certain diseases, including HIV-1[245]. In addition this diversity decreases the ability of pathogens to adapt to immune responses and increases the chance that certain members of the population will survive infection with a given pathogen, so benefitting the population as a whole.

Particular HLA variants can sometimes present peptides that enable the development of especially effective or ineffective/harmful immune responses, for example there are many HLA types associated with reduced or even increased risk for certain diseases, such as malaria (HLA B*53, reduced severity), HCV (HLA B*27 increased clearance, and others) and HIV (HLA B*57, HLA B*2705, HLA B*51, HLA B*5801 improved control of disease, HLA B*3503 reduced control of disease, HLA B*5802 associated with high viral loads)[246; 247; 248]. The importance of these molecules in the development of effective immune responses is

underscored where populations have undergone long-term exposure and selection by a single disease, for example malaria. Populations in areas where malaria has been highly prevalent over an extended period can have HLA frequencies skewed towards those that are most associated with protection from severe disease[249]. It is likely that many of the differences in HLA frequencies between populations have occurred due to adaptation to local disease burdens.

In HIV-1 disease, HLA class-1 is one of the major determinants of clinical outcome[68]. Possession of certain HLA class-1 alleles (see previous paragraph) is associated with slow disease progression and improved control of viremia. Long-term non-progressors with low to moderate viral load and slowed disease progression as well as people termed elite controllers, who maintain viral loads of under 50 copies/ml and high CD4+ T-cell counts, are very likely to possess at least one of these HLA types[250; 251; 252]. Other HLA, such as HLA B*35 are associated with more rapid progression and worse control of viremia in HIV-1 infection[253]. The HLA types linked with improved disease course in HIV-1 have been found to be associated with the presentation of and response to certain HIV-1 CD8+ T-cells epitopes[254], including three epitopes for HLA B*57 (IW9, TW10 and KF11) and one epitope in HLA B*2705 (the p24 Gag KK10)[58; 251; 255; 256]. These epitopes aid the generation of effective CD8+ T-cell responses to regions of the virus where escape mutation results in reduced viral fitness or requires compensatory mutations within and outside the epitope in order to restore replicative capacity[58; 59; 257].

1.4.4.1. Structure of the HLA class-1 molecules

The HLA class-1 molecules have been extensively characterised by structural analysis, with hundreds of HLA-peptide complexes submitted to the protein data bank in recent years[258], including 5 HLA B*2705 structures complexed with peptides[259; 260; 261]. These data have revealed how HLA class-1 molecules are able to bind an impressive number of peptides of varying length and sequence in an HLA-variant-specific manner.

Peptide epitopes are bound to HLA class-1 in the binding groove, formed by two α-helices, α1 and α2, one on either side, and a β-sheet making up the base. The groove is approximately 30Å long and 12Å across at its widest point[262] and is closed at both ends, by the bulky side chains of residues at the ends of the α-helices such as W167 at the N-terminal end of the groove and Y84 at the C-terminal end (The N- and C- terminal notation refers to the peptide orientation as presented on the HLA)(Figure 29). The peptide is bound N-terminally by hydrogen bonds to conserved residues Y7, Y59, Y171 and Y159 and C-terminally by hydrogen bonds to Y84, T143 K146 and W147. Conservation of these amino acids across HLA class-1 molecules allows binding of the carboxy and amino termini of peptide ligands without reliance on specific amino acids. In addition, certain polymorphic residues along the HLA class 1 groove will also contact the backbone of candidate peptides, adding to their binding energy and influencing their conformation in the groove without dependence on specific peptide sequences.

However, HLA class 1 alleles do discriminate peptides by sequence, both at the termini and along the length of the peptide. Polymorphic residues in the binding groove ultimately determine what peptide residues are able to bind at particular positions, specifying interaction with certain types of peptide side chain. This allows HLA class-1 alleles to discriminate between preferred sequences and determine the type of amino acids an epitope must have in a given position to be able to bind to a given allele. For example, acidic residues

such as D116 and D77 located at the base of a deep hydrophobic pocket in the C-terminal end of the HLA B*27 binding groove give this allele a preference for peptides that have either a short hydrophobic residue, or a long basic residue (such as lysine or arginine) at that position[261]. This confers a degree of specificity to the epitopes presented, while still allowing a range of sequences to bind. This selection of peptide epitopes by different HLA class-1 alleles through polymorphism of amino acids along the binding groove has been shown to be important in determining levels of HIV-1 viremia in a recent study[67], and in recently published GWAS studies, amino acid residues at specific sites in the HLA class-1 molecule were found to be more strongly associated with control of viremia in HIV-1 patients than the association for whole alleles[263; 264]. Single nucleotide polymorphisms (SNPs) that corresponded to several amino acid positions along the HLA class-1 groove were associated with differential disease control.

Along the HLA class-1 binding groove, there are a series of depressions of varying size, referred to as "pockets", which are able to accommodate selected peptide side chains. There are typically 6 of these pockets along the length of the groove labelled from the A pocket at the N-terminal end of the groove to the F pocket at the C-terminal end[262](Figure 29), though the number occupied may vary with different peptide ligands. The contribution of individual pockets towards peptide binding varies, depending on the allele. The side chains of amino acids present in these pockets determine the amino acid sequences that peptides must have in order to bind in the groove and so affect the repertoire of epitopes that can be presented. For example, the C pocket of HLA A*68 is negatively charged and large, accommodating a range of basic residues, however, in HLA B*27, the same pocket is smaller and largely blocked by the hydrophobic portion of the side chain of K70, meaning it does not contribute significantly to the binding of most epitopes to this allele (Figure 29C).

1.4.5. The interaction of the T-cell Receptor with antigen

Recognition of the HLA-peptide complex by the TCR is due to contact with both the HLA and the peptide by three complementarity determining region (CDR) loops in each chain of the αβ TCR (Chapter 3). Structures of both HLA class-1 and class-2 in complex with specific TCR have been determined[258; 265; 266; 267; 268] and show how the TCR docks onto the HLA surface in each case. In addition, analysis of the kinetics of the interaction show that the on-rates of the TCR are similar whether the HLA complex presented contains a self or antigenic peptide[269]. This suggests that the docking mechanism for the T-cell receptor is based on a two-step mechanism involving weak recognition of the HLA in a concentration-dependent step, followed by more specific, concentration-independent interactions with the peptide[270]. However, the different relative contributions to HLA and peptide recognition of the CDR1 and 2 loops in some of the structures so far as well as analysis of possible transition states suggests that recognition may be more complex and/or varied than the two-step model currently indicates[106; 271].

Binding of the HLA class-1 molecule typically involves the TCR docking diagonally across the peptide-binding groove[272]. The binding of the HLA class-2 by TCR has been described as orthogonal, though in the majority of cases it is still slightly diagonally bound, but at less of an angle from perpendicular than the class-1 TCR[123; 273]. Of the TCR structures so far completed, there appears to be a single common docking mode, although there is considerably more divergence in binding angle of the TCR to HLA class-1 than to HLA class-2[124]. In each case the Vα-domain of the TCR is positioned over the α2 helix of the HLA and the Vβ domain over the α1 helix. In typical binding positions this places the Vβ CDR3 loop over the central region of the peptide-binding groove, hence the importance of this loop in peptide-dependent recognition of the HLA-epitope complex. It is not known why the β-chain is positioned on the same side of the HLA in all of the structures studied so far, but a

suggestion is that there may be undetected germline bias for recognition of particular regions of the HLA that contribute to this phenomenon.

1.4.6. Specificity and cross recognition of epitopes by the T-cell receptor

The ability of the TCR to recognise specific epitopes and the extent to which these receptors can cross-recognise epitopes similar in structure or sequence is of great interest in a number of immunological fields, from tolerance and autoimmunity, where cross-recognition between pathogen-derived peptides and host epitopes may drive disease, to cancer immunology and infectious disease.

As research into adaptive immunity has progressed it has become clear that the limited pool of T-cells circulating in a human body at a given time ($>10^8$)[274] is insufficient for each response to be generated by an individual TCR for each possible peptide epitope-HLA combination[275; 276]. Yet it is also clear that TCRs are to some degree specific in their recognition and are restricted by HLA type and the bound peptide[107; 273; 277; 278]. Cross-reactivity of TCR has been examined from both a structural and functional perspective[108; 124; 265; 267; 272; 275; 279; 280; 281; 282; 283]. Recognition of the bound peptide will, to a certain extent restrict the HLA recognition of the TCR as HLA types skew peptide presentation through type dependent sequence requirements for peptide binding. Several studies have also sought to examine determinants on the TCR surfaces for recognition of HLA itself[107; 108; 125; 277; 280; 284; 285; 286; 287; 288; 289](see chapter 5.1.1), yet a comprehensive understanding of the relative contributions made by the α and β-chain hyper-variable loops remains elusive for many interactions.

1.4.7. Killer immunoglobulin receptors

The Killer immunoglobulin receptors (KIRs) are a group of receptors encoded in humans by 14 genes present within the leukocyte receptor complex on chromosome 19q[290; 291; 292; 293]. These receptors consist of two or three extracellular C2-type immunoglobulin domains connected by a transmembrane helix to a cytoplasmic domain. The cytoplasmic domains of the receptors differ depending on whether the KIR is an activating or inhibitory receptor, with inhibitory receptors containing immunoreceptor tyrosine-based inhibitory motifs (ITIM) while activating receptors lack these motifs and contain a positively charged lysine residue in their transmembrane regions[294]. The receptors are present on NK cells as well as some types of T cells. Excepting several KIRs for which the ligand has not yet been identified, each has been found to bind a subset of HLA class-1 molecules (Table 1). Binding of these receptors to their HLA class-1 ligands on a target cell affects the likelihood of lysis by NK cells. To date, crystal structures exist for KIR2DL1, KIR2DL2, KIR2DS2, KIR2DL3, KIR2DS4 and KIR3DL1[295; 296; 297; 298; 299; 300; 301]. Of these structures KIR2DL1, 2DL2 and 3DL1 have also been shown in complex with their HLA ligand and this, along with extensive mutational analyses (selected references[259; 302; 303; 304], reviewed in[304]) particularly for the KIR3D molecules, show how the ligands are recognised by these receptors.

Gene name	Number of alleles	Recognition of HLA ligands
KIR2DL1	26	HLA-C group 2 (C*02, 04, 05, 06)
KIR2DL2	15	HLA-C group 1 (C*01, 03, 07, 08), HLA C*0501, C*0202, B*4601 and B*7301
KIR2DL3	11	HLA-C group 1 (C*01, 03, 07, 08), HLA C*0501, C*0202, B*4601 and B*7301
KIR3DL1	59	HLA B*08, B*27, B*57, A*24 (Bw4 epitope containing HLA)
KIR3DL2	25	HLA A*03, A*11
KIR3DL3	56	Unknown
KIR2DL5A	10	Unknown
KIR2DL5B	21	Unknown
KIR2DL4	27	HLA G?
KIR2DS1	15	HLA-C group 2 (C*02, 04, 05, 06)
KIR2DS2	20	HLA-C group 1 (C*01, 03, 07, 08),
KIR2DS3	12	HLA-C group 1 (C*01, 03, 07, 08),
KIR2DS4	30	HLA-C*0501, 1601, 0202 and HLA A*1102
KIR2DS5	14	Unknown
KIR3DS1	16	(Bw4 epitope containing HLA)?

Table 1: Table of KIR gene products and their HLA ligands, pink shading shows inhibitory KIR, grey shading shows KIR forms with crystal structures. Reproduced with updates from [304]. Sources [295; 296; 297; 298; 300; 301].

The KIR2D group were the first to be crystallised and all show similar patterns of binding with HLA, though the binding of the KIR2DL forms are of higher affinity than in the 2DS group. This is partly due to mutation of residues in two hydrophobic patches that also contact the HLA. Changes in packing caused by these mutations reduce the KIR2DS affinity for the ligand[301]. Despite the difference between activating and inhibitory KIR being caused by alteration to the hydrophobic interactions, the KIR/HLA interface is actually dependent on hydrogen bonding and charge complementarity to a larger extent than HLA/TCR interactions[305]. The KIR receptor was found to bind across the C-terminal end of the binding groove of the HLA ligand[298], TCRs typically bind in a more central position, but there is variation between different TCRs. In this position, the KIR D1 domain is positioned over the α1 helix of the HLA and the D2 domain is positioned over the α2 helix. Both KIR2DL1 and KIR2DL2 have been shown to provide 6 acidic residues that form salt bridges with conserved basic residues on the α1 and α2 helices of the HLA[296; 298]. In total, at least 16 residues are found to interact directly with HLA in the KIR2DL structures[306].

1.5. Interaction of HLA and HIV

1.5.1. Immune control of HIV-1 by T-cell responses

As discussed previously (Chapter 1.3), the discovery that CD8+ T-cell depletion results in loss of control of viremia and progression to AIDS in HIV-1 infected non-human primates[65] led to the identification of the CD8+ T-cell response as one of the arms of the immune system most involved in control of HIV-1 disease. Further studies have subsequently identified HLA class-1 alleles and CD8+ T-cell responses to particular regions and peptide epitopes of the virus which result in improved or reduced control of viremia[78; 79; 252]. Despite the discovery that responses to specific HLA class-1 epitopes can be protective as in the case of the HLA B*2705 restricted KK10 response[78; 79] and that viral epitopes in certain conserved regions of its genome are more likely to elicit this protection[307], a complete understanding of the mechanisms involved in protection is still missing.

This is highlighted by the fact that while these discoveries are associated with control of viremia and delayed disease progression, there are many patients with these HLA alleles and even T-cell responses to protective epitopes that nevertheless fail to control the disease[308]. An important part of this component may be the nature of the CD8+ T-cell responses generated in response to these protective epitopes in a patient. Previous studies into the generation of protective responses have shown that the CD8+ T-cells that give protection have a number of characteristics that contribute to this efficacy. A study of the CD8+ T-cell response to the HLA B*2705 restricted KK10 epitope found that the cells are more likely to be poly-functional, secreting a number of cytokines, to have greater proliferative capacity, a higher avidity for the epitope and in some cases to have a higher turnover of clonotypes[309]. This increase in poly-functionality was directly linked to the concentration of KK10 peptide used for stimulation of the cells *ex vivo*. In addition studies have shown a decreased T_{REG}

population associated with these effective responses and decreased exhaustion of CD8+ T-cells (though whether this is a contribution to effectiveness or a result of it is not known)[310; 311; 312; 313]

A recent study assessed these and other aspects of the CD8+ T-cell response to the HLA B*2705 restricted epitope KK10. This study looked at patients who were either elite controllers or who failed to control disease, but possessed a wild-type consensus sequence for the KK10 region and carried a CD8+ T-cell recognising the epitope. While failing to replicate some of the characteristics of effective T-cell control found in previous studies[308], it did find that an important correlate of protection in this group was the ability of the patient CD8+ T-cell responses to cross-react with escape variants of the epitope *ex vivo*, even when viral sequencing did not reveal these variants in the population at the time of sampling. This increased cross-recognition was associated with TCR sequences that had a decreased number of nucleotide additions in their CDR3 loops compared with the non cross-reactive TCR, possibly decreasing specificity of the receptor in this case. The cross-recognition of responses and differences in the characteristics of the CDR3 sequence in effective CD8+ T-cell populations was also associated with the ability of these T-cells to up-regulate perforin and granzyme B and to neutralize HIV infected cells *ex vivo*[308].

1.5.2. The response to the KK10 epitope

The "optimal" HLA B*2705 restricted KK10 CD8+ T-cell epitope was initially identified through the detection of a p24 Gag response associated with HLA B*2705. The location of the epitope was then refined through testing of CD8+ T-cell responses to overlapping 15-20mer peptide pools within the region. Once the peptide producing the response was identified, increasingly short peptide epitopes were tested until a maximum in the level of response by the CD8+ T-cells was seen using a 10mer version of the epitope (^{131}KRWIILGLNK^{141}, abbreviated to KK10)[314]. The HLA B*2705-restricted response to the p24 Gag region containing the KK10 epitope is detected in up to 94% of HIV-1 infected patients possessing the B*2705 allele[315], and 92% of patients in the study responded specifically to the KK10 epitope[315]. The response to this region has been observed to develop in early infection, within 50 days of the onset of symptoms[57; 310] and responses to the region can account for up to 50% of the total HIV-1 specific CD8+ T-cell response in HLA B*2705 individuals[315].

It has been shown that the response to KK10 results in an increased time taken for progression to AIDS; this is particularly evident in the stage of disease from a CD4+ count <200 to first AIDS defining illness[316]. This containment of HIV-1 infection is associated with constant KK10 specific CD8+ T-cell responses throughout the disease course and the late appearance of escape mutations in the KK10 epitope. These escape mutations are associated with a reduction in replication rate of the virus[317] until compensatory mutations occur[58]. The compensatory mutations reduce the fitness cost of escape and at this point loss of control of viral replication can occur[256; 318].

The best-known escape mutations in the region are the R132K mutation and the L136M mutation within the KK10 epitope. The L136M mutation is often seen before the R132K mutation[319; 320] and has been associated with loss of TCR recognition for some CD8+ T-cell clonotypes[321]. In addition, it has been associated with an increase in binding to the inhibitory

Ig-like transcript 4 (ILT4) receptor[321]. It has recently been observed that HLA B*2705 controllers and elite controllers more frequently possess CD8+ T-cell clonotypes that cross-recognise the L136M escape variant sequence than patients that do not control viremia[308]. The R132K escape mutation is associated with the loss of ability of the KK10 peptide to bind the HLA B*2705 groove as this residue is an anchor residue (see chapter 5) for this molecule[322; 323]. The majority of HLA B*2705 patients respond to the KK10 and when present this response is immunodominant[314] however some studies have suggested that other epitopes may also contribute to the suppression of viremia associated with the HLA B*2705 allele[324].

1.5.3. Interaction of KIR and HIV

KIRs have a number of HLA ligands, including both Bw4 and Bw6 HLA. However, there is only one KIR and HLA combination that has been found to have a significant impact on the course of HIV-1 disease to date. This is the possession of KIR3DL1/KIR3DS1 and an HLA Bw4 ligand[325]. The possession of both this KIR and its HLA ligand is associated with improved disease control and slower progression of the disease compared with individuals who have only one of these. It has furthermore been observed that the possession of KIR3DS1 protects against opportunistic infections at later stages of the disease course[326]. This association remains even when conventionally protective HLA such as HLA B*57 and HLA B*27 are removed from the analysis[325]. The possession of KIR3DS1 has been shown in some studies to have a beneficial effect, in addition to that of KIR3DL1 in individuals possessing both of these molecules[327; 328]. That both an activating and an inhibitory KIR are shown to have a beneficial effect may seem counter-intuitive, but the development of NK cells suggests that strong inhibitory responses encountered in development lead to the generation of mature NK cells that are capable of high levels of activation and efficient responses[329; 330; 331] The lack of this

inhibitory stimulus has shown to result in NK cells incapable or less capable of activation and effector function[332; 333].

The Bw4 motif is present on a number of HLA including HLA B*2705 and HLA B*57. There are two variants of this motif, Bw4-80I and Bw4-80T, of which HLA B*57 possesses the former and HLA B*2705 the latter. This is of interest as the binding of Bw4-80I to many alleles of KIR3DL1 is of higher affinity than that to Bw4-80T and the protective effect of KIR3DL1 has been observed to be particularly robust in the subset of patients with this motif. However, it has also been shown that protection is provided in Bw4-80T possessing individuals[326] and that for some alleles, binding to the Bw4-80T motif may be stronger[64]. Although the effect of KIR3DS1 is primarily observed with Bw4-80I[327] there is some debate as to whether these HLA are in fact ligands for this molecule, despite homology with KIR3DL1[334].

This raises the question as to whether the protective effect is mediated by direct interaction of KIR and HLA[335]. Further observations which call into question the exact mechanism of this protective effect are firstly that the allele of KIR3DL1 providing the strongest association with protection, KIR3DL1*004, is in fact known not to be presented on the cell surface, suggesting either an intracellular interaction or an alternative mechanism[325]. Secondly, in the same study, it was observed that while more robust protection was provided by highly expressed KIR3DL1 alleles in Bw4-80I possessing individuals than alleles with lower expression, this situation was reversed for Bw4-80T individuals; only those alleles associated with lower expression of the gene products gave significant protection in this group[325]. Despite these irregularities, the observed binding of KIR3DL1 to their HLA ligands and the correlation of this with protection suggests that for KIR3DL1, the ability of the NK cell to bind these ligands is an important feature of the mechanism of protection. Though this does not exclude the possibility of other factors/mechanisms being involved in this protection, particularly in the case of the KIR3DL1*004 allele[325].

2. Antigen processing analysis for the p24 Gag amino acid 126-150 25mer peptide and TAP transport of digestion products

Work in this chapter was a collaborative project between Dr Stefan Tenzer, Professor Peter van Endert and ourselves. Dr Tenzer carried out proteasomal digestions; Prof. van Endert performed TAP transport assays. Analysis of the proteasomal digestions was my own work.

2.1. Introduction

2.1.1. Escape from immune responses through modulation of antigen processing

The generation of effective CTL responses to viral epitopes like KK10 results in killing of infected cells and therefore exerts selective pressure on the viral sequence[319; 336; 337; 338; 339; 340; 341]. Virus with sequence mutations that reduce the anti-viral effect of these immune responses will be able to replicate better than the original virus in the presence of these immune responses and so will be selected for at the level of the viral population within an individual. The common escape mutations of the KK10 epitope demonstrate frequent mechanisms of escape from effective immune responses. The R132K and L136M changes limit the recognition of the KK10 epitope by T-cells by affecting the ability of the epitope to bind to HLA B*2705 in the case of R132K and through reduction of TCR affinity for the complex in the case of L136M[256; 318; 321; 342]. In addition to these mechanisms of viral escape, other parts of the cellular machinery that affect the ability of the immune system to recognise and respond to epitopes may also provide a means of escape.

A particularly important part of this process is the production of epitopes in the proteasome, which cleaves polypeptide chains according to their sequence. This cleavage is affected by the sequence of the polypeptide as it passes through the central channel of the proteasome where it may be cleaved multiple times, meaning that the sequence along the entire length of the polypeptide chain is involved in determining the pattern of cleavage[189; 343; 344; 345; 346; 347]. Indeed several studies have shown that selection pressure can result in viral escape through mutation of both intra—epitope and flanking mutations that reduce the production of the epitope targeted by the immune response[62; 238; 348; 349; 350; 351]. In addition to the proteasome, other cellular machinery is involved in the production and presentation of epitopes including aminopeptidases that trim epitopes and the TAP complex which transports peptides into the

ER; because both display sequence preferences for their epitope substrates and so the production and presentation of epitopes can be similarly modulated by sequence changes that affect these processes[238; 349; 352].

The maintenance of these escape mutations is of benefit to the virus while immune pressure is applied for the given region, which is usually the case within individual patients, though introduction of HAART may result in the reduction of the host responses through the reduction of antigen stimulation associated with lower viral loads[353]. The maintenance of these escape mutations following transmission to a new host will depend on several factors. Firstly, whether the donor and the recipient carry the same HLA variants (are 'HLA matched' or if not are 'HLA mismatched'). Escape mutations that carry a viral fitness cost because they hamper viral replication will only remain in the virus if they are overall advantageous to viral survival and spread. Thus, in an HLA matched recipient they may allow the virus to escape immune detection and will therefore remain, whereas in an HLA mismatched recipient they may only be disadvantageous for viral growth and will therefore likely revert. The rate of reversion is found to differ depending on the effect the mutations have on viral fitness[17; 59; 61; 257; 349; 354]. Accumulation of escape mutations over time at a population level may be reduced due to the polymorphic nature of the HLA creating very different selection pressures with different HLA allele combinations and driving reversion to the consensus sequence[17; 62]. However, a few studies have shown that for some HLA variants in some populations, adaptation by the virus may occur at this level[355; 356; 357]. If these escape mutations, while no longer advantageous, are not a disadvantage in a new host (for example due to the presence of compensatory mutations), they may be maintained or exhibit slow reversion or be lost through genetic drift. The HLA B*2705 R132K mutation when alone incurs a large viral fitness cost[317], however, when the sequence also contains compensatory mutations reducing this cost (and allowing better replication in the original host) the sequence is often maintained following transmission[58; 59; 320].

Proteasomal processing of the wild-type KK10 sequence has been analysed previously by our group[238]. We found that in addition to small amounts of the KK10 epitope, many KK10 containing epitopes with differing C-termini were produced and these were made in much larger amounts in the assay than the KK10 epitope itself. For the HLA A*02 restricted SL9 epitope and overlapping epitopes present in p17 Gag, the amount of production by the proteasome was linked to the immunodominance hierarchy of the CD8+ T-cell response in the region[238]. Further analysis of the production of epitopes by proteasomal digestion of the KK10 region and the effects of common escape mutations in the region on this production were therefore carried out in order to better understand the role that this epitope and KK10 containing epitope forms play in the immune response *in vivo*.

2.2. Proteasomal digestion of sequence and sequence variants

The HIV-B consensus sequence of the p24 Gag protein was used to create 25mer peptides containing the amino acid residues 126-150, which include the KK10 epitope. Peptides were created carrying two common escape mutations observed in the KK10 epitope in patients with a response to this peptide, R132K and L136M, either alone or in combination. Purified constitutive and immunoproteasomes (from LCL721.174 and LCL721 human Epstein-Barr virus–transformed B cell lines) were used to digest the 25mer amino acid sequence (positions 126-150 in p24 Gag, HXB2 sequence numbering). The processing results from these assays was compared with the previously published[238] proteasomal digestion of the 25mer containing the KK10 wildtype sequence. This allowed the determination of the effects of these mutations on antigen processing in the KK10 region. The epitopes present in this region are shown below (Figure 4).

Epitope	HLA restriction
GEIYKRWII	B*08
GEIYKRWIIL	A*24, B*08
EIYKRWIILGL	A*24
KRWI	B*27
KRWII	B*27
KRWIIL	B*27
KRWIILG	B*27
KRWIILGL	B*27
KRWIILGLNK	B*27
KRWIILGLNKI	B*27
IILGLNKIV	A*02
KRWIILGLNKIVR	B*27
KRWIILGLNKIVRM	B*27
IILGLNKIVRM	A*33
ILGLNKIVRMY	B*15, B*62

Figure 4: Epitopes present between amino acid residues 126-150 in HIV-1 p24 and their HLA restriction.

2.2.1. Production of KK10 and KK10 epitope forms in the wild-type P24 Gag digest

The proteasomal digestion of the KK10 wild-type containing sequence results in the production of a number of epitopes[238]. These include several KK10 containing epitope forms that are either C-terminally extended or truncated. The amounts of these naturally processed epitope forms is greater in the wild-type sequence than the KK10 epitope itself (Figure 5A. and B.). Previously published data showed one epitope form with a truncated C-terminus, the KL8 epitope. However, further investigation of the ability of HLA B*2705 to refold with and present truncated epitope forms (see Chapters 4 and 5) has revealed additional peptides which may be presented by HLA B*2705. These are the peptides KRWI (KI4), KRWII (KI5), KRWIIL (KL6) and KRWIILG (KG7). Refolding was achieved with all of these truncated epitope forms and confirmed with mass spectrometry for the KI4 and KI5 peptides, additionally we were able to crystallise and obtain structures for the KL6 and KG7 peptides

in complex with HLA B*2705. Re-analysis of the KK10 wild-type proteasomal digestion assays was therefore carried out to include these truncated epitope forms (all or part of Figure 8, Figure 9, Figure 5 and Figure 7).

Analysis of the data to include these epitope forms shows that in addition to the previously noted high levels of KL8 epitope production[238] (Table 2) the smaller truncated forms are also all produced in greater quantities than the KK10 epitope. The KL6 (or 6mer) is produced in the highest amount of any epitope form (Figure 5A and B) and at some time-points following immunoproteasomal digestion, the number of these fragments being produced is 733 times greater than the amount of the KK10 epitope (Table 2). The KL6 epitope form alone at these time-points accounts for over 50% of the total HLA B*2705 restricted epitope production. The KI5 epitope is also produced in considerably higher quantities (Table 2) and moderate amounts of the KI4 and KG7 epitopes are also seen (1.81% and 3.39% being their respective highest percentages of epitope production).

		KI4	KI5	KL6	KG7	KL8	KI11	KR13	KM14
Relative number of epitopes produced	Immunoproteasome	24.9	129	**733**	7.84	415	5.59	3.46	53.4
	Constitutive proteasome	6.86	27.6	**254**	15.6	59.4	23.8	4.41	68.3
	Mean	14.0	67.9	**445**	12.5	201	16.5	4.03	62.4
Relative molar production of epitopes	Immunoproteasome	51.3	224	**1098**	11.0	516	5.13	2.67	38.1
	Constitutive proteasome	14.1	47.8	**381**	21.9	73.8	21.8	3.40	48.7
	Mean	28.9	118	**666**	17.6	250	15.1	3.11	44.5

Table 2: Production of KK10 epitope form relative to the production of the KK10 epitope (total production observed for each epitope form over the time-course relative to total production observed for KK10). Given in number of epitopes produced and relative molar amounts. The greatest epitope production in each case is shown in bold.

The addition of these truncated epitope forms to the region means that truncated forms are the most commonly produced epitope forms at all but one time-point when the wild-type sequence is digested. There are also differences in production profiles of these epitopes between the constitutive and immunoproteasomal digests. There is greater production of the truncated epitopes by the immunoproteasome than by the constitutive proteasome, with 95.4% of B*27 epitopes produced throughout the digestion being C-terminally truncated compared with 78.9% in the constitutive digest. The relative proportions of the individual epitopes also differs between the proteasomal forms, with greater proportions of KL6 and KG7 being produced by the constitutive proteasome and greater proportions of KI4, KI5 and KL8 being produced by the immunoproteasome (Table 2).

The inclusion of these epitope forms gives rise to several new overlapping epitopes with the same C-terminal end e.g. the HLA B*08 restricted epitope GEIYKRWII shares the same C-terminus as the KRWII HLA B*2705 restricted 5mer peptide. This epitope must now be considered as both a B*08 epitope and an N-terminally extended B*2705 epitope form which may be trimmed or cleaved to the KRWII sequence. The speed at which this peptide will undergo trimming by ERAP in the ER or possible cleavage by nardilysin in the cytosol[358] is not known and will be a factor in how much of each epitope will be generated at any one time. Here I consider all of the N-terminally extended fragments produced from the proteasome as being potential HLA B*2705 epitope forms (as we have done previously[238]). The truncated KK10 epitope forms are referred to as "B27-short" epitopes in our data (Figure 8, Figure 9 panels A-D for both). However, the distinction between the truncated and extended KK10 epitope forms is not made when considering total production of B*2705 restricted epitope forms (Figure 5 and Figure 7).

In addition to the production of C-terminally truncated KK10 epitopes, there is considerable production of the C-terminally extended forms at some time-points (Figure 7). The most abundant of these in both the immunoproteasome and the constitutive proteasome is the

KM14 epitope (Table 2), followed by the KI11 epitope in the constitutive digest and the KR13 epitope in the immunoproteasomal digest. KK10 is produced in the lowest amounts of any B*2705 restricted epitope in both the immunoproteasomal digest and constitutive digests (Table 3). The C-terminally extended fragments are produced in greater percentages (of total epitopes) as a group and individually in the constitutive proteasome compared with the immunoproteasome and in both datasets the proportion of epitopes decreases over time compared with the truncated forms (Table 3).

	Constitutive proteasome					Immunoproteasome				
	1hr	2hr	4hr	6hr	Total	1hr	2hr	4hr	6hr	Total
KI4	0.48	1.02	1.77	1.98	1.49	0.30	0.50	1.96	4.41	1.81
KI5	9.20	9.26	4.69	3.98	5.98	7.81	12.6	9.71	5.33	9.38
KL6	21.8	**45.8**	**63.1**	**70.2**	**55.1**	**46.1**	42.0	**60.9**	**64.1**	**53.4**
KG7	1.06	2.41	4.53	3.94	3.39	0.30	0.43	0.66	0.82	0.57
KL8	11.0	11.1	14.9	12.6	12.9	27.6	**39.1**	25.8	25.0	30.2
KK10	0.61	0.42	0.07	0.06	0.22	0.30	0.03	0.04	0.04	0.07
KI11	14.5	6.87	2.83	1.42	5.15	1.89	0.31	0.10	0.08	0.41
KR13	3.13	1.21	0.33	0.26	0.96	0.96	0.19	0.14	0.08	0.25
KM14	**38.2**	21.8	7.81	5.59	14.8	14.8	4.87	0.73	0.16	3.89

Table 3: Production of KK10 epitope forms in the wild-type sequence at each time-point and in total for constitutive and immunoproteasomal digests. Given as a percentage of total KK10 epitope forms produced. The KK10 epitope form with greatest percentage production is shown in bold text.

2.2.2. Production of KK10 and KK10 epitope forms in the 132 R-K p24 Gag Digest

As described above (Chapter 1), the R to K mutation at position 2 of the KK10 epitope has been characterised as an HLA binding escape mutation as it greatly reduces the affinity of the peptide for HLA B*2705[256]. However, as we know that weak recognition of this mutated epitope has been seen in certain assays[256], we wanted to see whether the mutation additionally affected antigen processing in the region.

The proteasomal digests show that production of HLA B*2705 containing epitopes is changed dramatically by the inclusion of this mutation into the sequence, with the production of C-terminally extended epitope forms being completely abrogated by the mutated position (Figure 5C, D and Figure 7). The KK10 epitope is still produced by the constitutive proteasome, with total production being 0.36% (Table 4); none is produced by the immunoproteasome in this digest. The relative proportion of KK10 produced is therefore actually slightly higher than in the wild-type constitutive proteasomal digest at 0.22% (Table 3).

The proportion of C-terminally truncated KK10 epitope forms is greatly increased in this sequence, accounting for all of the HLA B*2705 epitope production by the immunoproteasome and over 99.6% of the production by the constitutive proteasome (Table 4). Despite these large differences in processing between this sequence and the wild-type, there are still some similarities in the patterns of production of the C-terminally truncated forms (Figure 5). The proportion of KL8 produced by the constitutive proteasomal digest is 16.6% of B*27 restricted epitopes, and in the immunoproteasomal digest is 31.6% (Table 4). This is very similar to the wild-type levels (Table 3). The patterns for the KG7 peptide are also similar, with a higher proportion of the epitope production being KG7 in the constitutive digest (12.3%) compared with the immunoproteasomal digest (3.88%), though the proportions produced in the wild-type are lower in both digests (Table 3).

The proportions of other epitope forms are not as closely linked to wild-type levels, a larger increase in production by the constitutive digest than the immunoproteasomal digest (Table 4) for the KL6 epitope occurs. This is possibly due to the loss of C-terminally extended epitope forms that were produced in larger proportions by the constitutive proteasome (Table 3). In addition we observed a decrease in production for KI5 in the R132K constitutive proteasomal digest (0.32% versus 5.98%) and the loss of production of KI4. However, for all of the truncated epitope forms that are seen, the pattern of increased/decreased production between the constitutive and immunoproteasomal digests remains unchanged (i.e. if relative production of an epitope form increases in the immunoproteasome for the wild-type sequence, it will also be increased compared to the constitutive digest for the RK mutant sequence).

	Constitutive proteasome					Immunoproteasome				
	1hr	2hr	4hr	6hr	Total	1hr	2hr	4hr	6hr	Total
KI4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
KI5	0.00	0.37	0.47	0.30	0.32	2.40	5.17	4.66	8.79	6.84
KL6	**75.0**	**71.5**	**65.6**	**71.0**	**70.4**	**57.1**	**47.4**	**51.3**	**62.5**	**57.7**
KG7	13.7	13.6	14.1	11.8	12.3	3.07	3.29	2.98	4.44	3.89
KL8	11.3	14.4	18.5	16.7	16.6	37.4	44.1	41.0	24.2	31.6
KK10	0.00	0.00	1.32	0.22	0.36	0.00	0.00	0.00	0.00	0.00
KI11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
KR13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
KM14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 4: Production of KK10 epitope forms in the R132K sequence at each time-point and in total for constitutive and immunoproteasomal digests. Given as a percentage of total KK10 epitope forms produced. The KK10 epitope form with greatest percentage production is shown in bold text.

2.2.3. Production of KK10 and KK10 epitope forms in the 136 L-M p24 Gag Digest

The proteasomal processing analysis performed on this sequence shows that in addition to providing TCR escape and compensation for the R132K mutation, this mutation has a significant impact on the production of KK10 epitope forms (Figure 5E and F). In particular, the production of KK10 and KK10 C-terminally extended epitope forms is completely abolished in this sequence (Figure 6)(Table 5), a more severe effect than seen with the R132K mutation. However, as with the R132K sequence, the production of C-terminally truncated KK10 epitope forms is not reduced by the mutation and they make up all of the HLA B*2705 restricted epitopes produced in the region.

Production of the truncated epitope forms is similar, but not identical to that in the R132K sequence, with the most commonly produced epitope being KL6 (up to 67.3% total)(Table 5). The proportion of KL6 produced from digestion of this sequence is slightly lower than that made after processing of the R132K sequence (up to 70.4% total)(Table 4) and the proportion of production of KI5 and KG7 is increased (Table 5). The production of the KL8 epitope is similar for the immunoproteasomal digest at 32.5% as it is for the R132K and wild-type sequences (Table 3 and Table 4), however, the proportion produced by the constitutive digest is reduced by a considerable amount when compared with these sequences at 7.18%. As with the R132K sequence, the KI4 epitope is not seen and despite the differences in proportions of some epitopes between these sequences, the relative differences in production between the constitutive and immunoproteasomal digests are again in the same direction. The same epitope is therefore always more or less abundantly produced by the same form of the proteasome.

	Constitutive proteasome					Immunoproteasome				
	1hr	2hr	4hr	6hr	Total	1hr	2hr	4hr	6hr	Total
KI4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
KI5	12.9	16.4	1.99	2.06	2.29	20.7	19.0	17.4	11.3	13.7
KL6	38.0	**41.8**	**47.7**	**69.7**	**67.3**	**49.9**	28.5	31.5	**54.0**	**46.7**
KG7	**49.1**	**41.8**	39.1	21.3	23.2	9.97	5.03	6.19	7.53	7.10
KL8	0.00	0.00	11.2	6.95	7.18	19.4	**47.5**	**45.0**	27.1	32.5
KK10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
KI11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
KR13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
KM14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 5: Production of KK10 epitope forms in the L136M sequence at each time-point and in total for constitutive and immunoproteasomal digests. Given as a percentage of total KK10 epitope forms produced. The KK10 epitope form with greatest percentage production is shown in bold text.

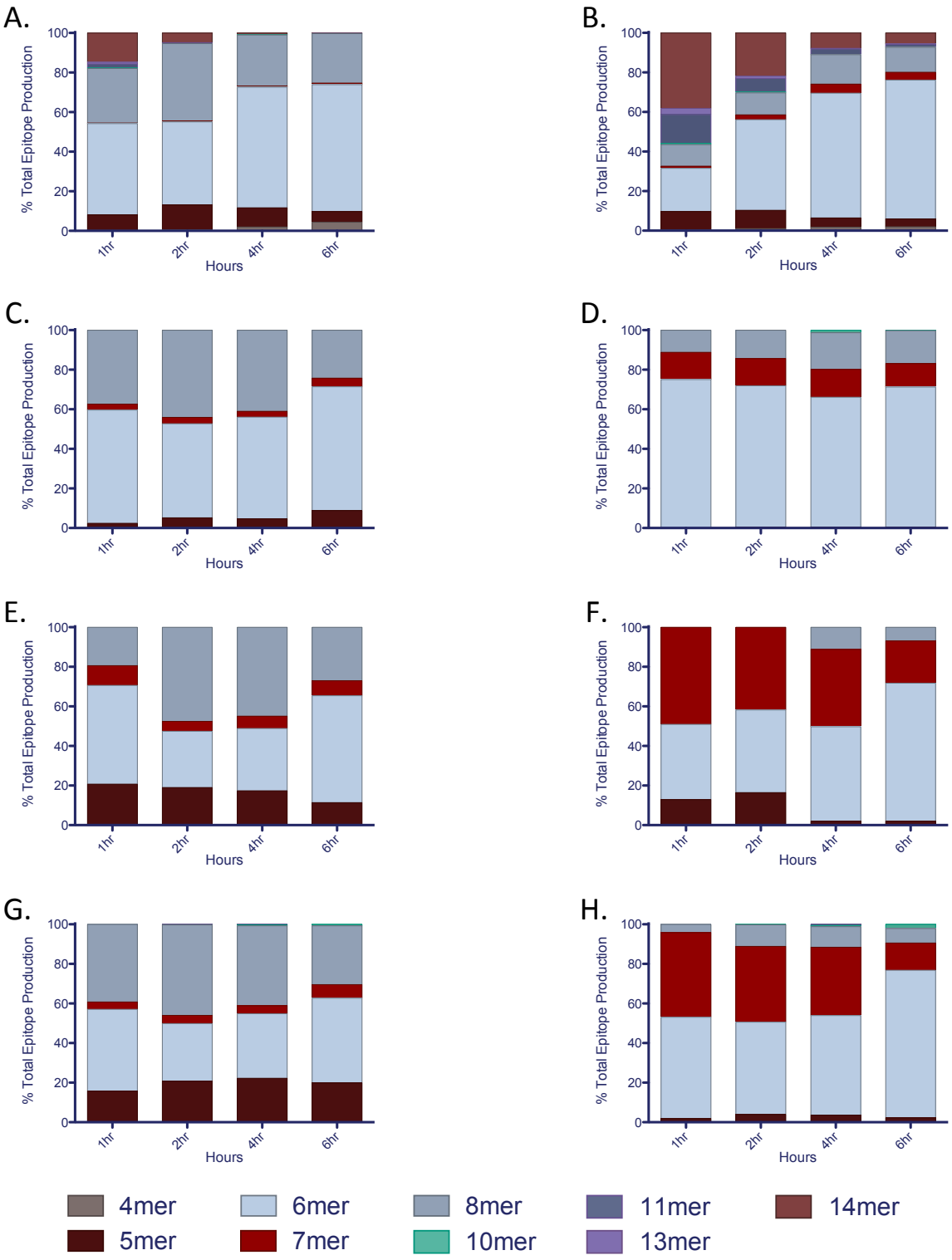


Figure 5: HLA B*2705 restricted epitope production shown in **A.** W/T immunoproteasome **B.** W/T constitutive proteasome **C.** RK mutant immunoproteasome **D.** RK mutant constitutive proteasome **E.** LM mutant immunoproteasome **F.** LM mutant constitutive proteasome **G.** RKLM mutant immunoproteasome **H.** RKLM mutant constitutive proteasome.

2.2.4. Production of KK10 and KK10 epitope forms in the 132 R-K/136 L-M p24 Gag Digest

In a large number of patients with the L136M mutation, the R132K escape mutation is also present[320]. We therefore incorporated both of these mutations into the same sequence to determine the effect they have in concert on KK10 epitope form production. As with the two mutant sequences alone, there are considerable effects on the KK10 epitope form production in this sequence (Figure 5G and H), with production of C-terminally extended epitope forms severely curtailed (Figure 6). Surprisingly, given the complete abolition of production of these epitopes by both single mutant sequences, KI11 is produced in the constitutive digest (Table 6) though this is in a very small amount compared with that made following the wild-type digests (Table 3). Of all of the digests the R132K, L136M sequence gives the greatest proportion of KK10 epitope production, with 1.72% of all production in the constitutive digest being the KK10 epitope and 0.62% in the immunoproteasomal digest (0.22% wild-type constitutive and 0.36% R132K constitutive).

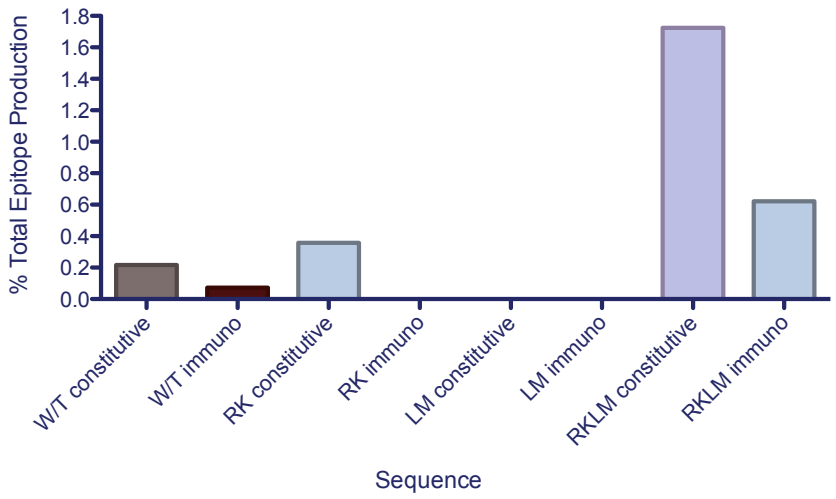


Figure 6: KK10 production as a percentage of total HLA B*2705 restricted epitope production.

Though production of C-terminally extended KK10 containing epitope forms is higher for this sequence than for either of the mutants alone, the majority of B*2705 restricted production still consists of the truncated epitope forms (Figure 5G and H). Once again, the KI4 epitope is not produced and the KL6 epitope remains the most commonly produced fragment in both digests (Table 6). The KI5 epitope is produced in similar proportions to its production in the L136M sequence, with slightly higher proportions produced by the immunoproteasome (Table 6) while the KG7 epitope is produced in proportions between those of the R132K and L136M sequences (Table 4 and Table 5). The KL8 epitope form is produced in very similar proportions to the L136M sequence (Table 5). Finally, the relative productions of both the C-terminally extended and truncated KK10 epitope forms in this sequence remain the same with respect to which proteasomal form produces the greater proportion of the epitope. Selection pressure to generate only the truncated KK10 epitope forms through sequence combinations that are adapted to proteasomal preferences may be less in case of virus with these double intra-epitopic escape mutations (R132K, L136M) due their combined effect on HLA-binding and TCR-recognition, respectively, which already decrease or abolish any HLA-B2705 restricted KK10 response.

The epitope production patterns shown by the R132K, L136M double mutant sequence are an interesting mix of the patterns seen for the R132K and L136M mutant sequences alone, with some epitope proportions more similar to one or the other of these sequences. The preserved patterns of differences in which epitopes are favoured for production by the constitutive and immunoproteasomal forms, particularly among the truncated epitope suggest that underlying cleavage preferences of the proteasomal forms are similar between the sequences, as might be expected given the sequence similarity. This is of particular interest given the large effect on production that these modest sequence changes have on the C-terminally extended KK10 containing epitope forms.

	Constitutive proteasome					Immunoproteasome				
	1hr	2hr	4hr	6hr	Total	1hr	2hr	4hr	6hr	Total
KI4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
KI5	1.82	3.91	3.58	2.20	2.50	15.7	20.7	22.2	19.9	20.1
KL6	**51.3**	**46.6**	**50.3**	**74.6**	**68.8**	**41.3**	29.1	32.7	**42.8**	**40.0**
KG7	42.7	38.3	34.5	13.7	18.9	3.72	4.16	3.98	6.82	5.98
KL8	4.22	10.9	10.4	7.34	7.92	39.2	**45.7**	**40.4**	29.7	33.2
KK10	0.00	0.25	0.64	2.10	1.72	0.00	0.00	0.47	0.77	0.62
KI11	0.00	0.00	0.56	0.00	0.08	0.00	0.32	0.25	0.00	0.07
KR13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
KM14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 6: Production of KK10 epitope forms in the R132K/L136M sequence at each time-point and in total for constitutive and immunoproteasomal digests. Given as a percentage of total KK10 epitope forms produced. The KK10 epitope form with greatest percentage production is shown in bold text.

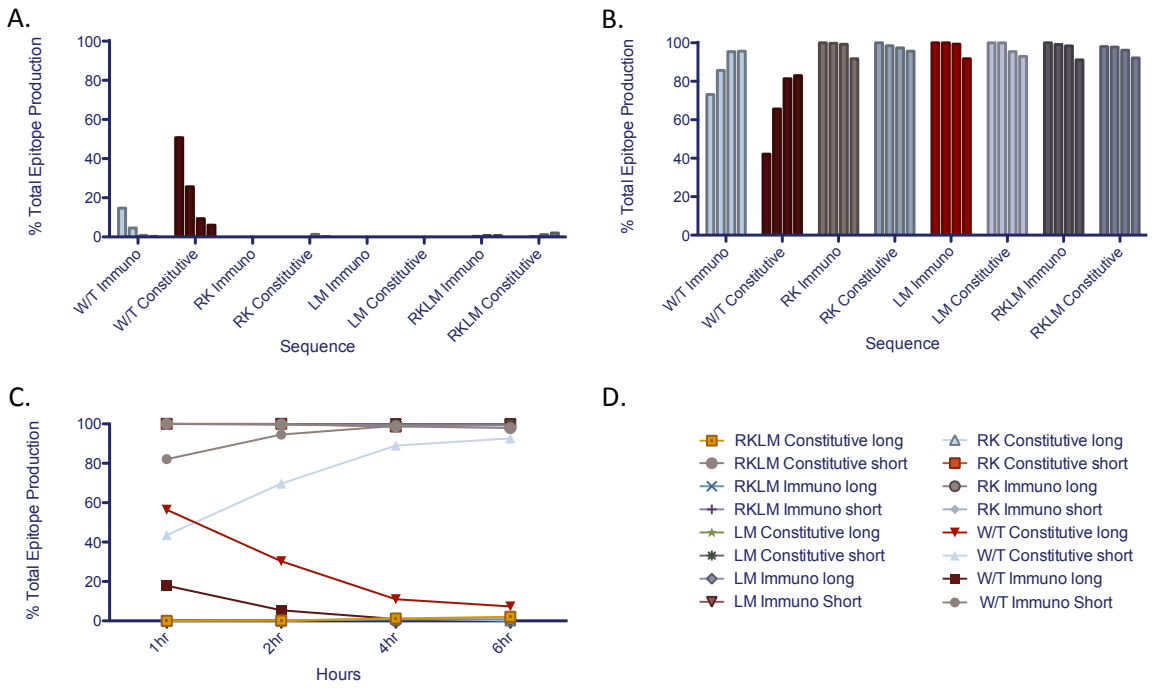


Figure 7: HLA B*2705 restricted epitope production shown by sequence used for proteasomal digestion. A. Production of KK10 and C-terminally extended epitope forms, B. Production of C-terminally truncated epitope forms and C. Production of both C-terminally extended and truncated epitope forms. D. Key for figure C. only (key for figures A and B shown in Figure 9).

2.2.5. Production of epitopes restricted by other HLA alleles in the p24 Gag digests.

In addition to the HLA B*2705 restricted KK10 epitope forms produced in this region, there are epitopes restricted by other HLA present in the region, and may partly overlap the KK10 epitope forms. Analysis of production of these epitopes and how they vary with mutations of the KK10 epitope may provide additional information on how these mutations affect processing of this 25mer region of p24 Gag. When analysing epitopes restricted by another HLA that share a C-terminus with an HLA B*2705 restricted epitope, it is important to note that, if the epitope begins N-terminal to the HLA B*2705 restricted epitope, the total production of these epitopes will also be counted as B*27 epitopes. However only those containing the N-terminus of the other epitope will be included in the total for that HLA.

The HLA B*08 restricted epitope GI9 (Figure 4) is produced in all of the sequence digests, however, the amounts vary substantially between the sequences, ranging from 0.31% in the R132K constitutive digest to 18.9% in the R132K, L136M double mutant immunoproteasomal digest (Table 7, Table 8, Table 9 and Table 10). The proportions of this B*08 restricted epitope produced (Figure 9A) account for all of the KI5 epitope seen in the three mutated sequences (Figure 10B) (Table 11). However, in the wild-type sequence the amount of the B*08 restricted epitope produced accounts for only a small amount of the total production of KI5, the majority being the KI5 fragment in this sequence (Table 11).

The A24/B8 restricted epitope GL10 (Figure 9B), shares the same C-terminus as the KL6 HLA B*2705 restricted epitope. The proportions of this epitope produced vary from 27.6% (Table 9) of total epitope production to 67.2% (Table 8). The percentage of total KL6 accounted for by production of this epitope differs between the sequences (Table 11). In the wild-type digestion up to 78.3% of the total KL6 epitope production is GL10 (Figure 10C) and in the L136M sequence, the epitope accounts for up to 94.8% of production in the constitutive digest, with the remainder being the K N-terminal form (Table 11). This trend is

increased in both the R132K and R132K, L136M double mutants, where even smaller proportions of KL6 production is accounted for by the K N-terminal form (Table 11).

In addition to the GL10 epitope, A24 can also recognise an epitope slightly C-terminal in the region, IL10. These epitopes share the same C-terminus as the HLA B*2705 restricted KL8. The proportion of total epitopes that are IL10 ranges from 6.40% in the L136M constitutive digestion to 29.5% in the immunoproteasomal digestion of the same sequence (Table 9). This pattern of production is similar in the R132K, L136M double mutant, though more of the epitope is produced in this digest, particularly by the immunoproteasome (Table 10). In the R132K sequence there is higher production in the constitutive digest and similar levels of production in the immunoproteasomal digest (Table 8) compared with the two L136M containing mutants. The wild-type digest has slightly higher levels of production in the constitutive digest than the L136M mutants but has much lower production by the immunoproteasome (Table 7).

The proportion of the KL8 epitope that is not produced from the IL10 epitope differs between the 4 sequences. Strikingly, the pattern of these differences is quite similar to the pattern seen for the KI5 epitope (Figure 10E). The largest proportions of epitope production from short, non-IL10 fragments being in the wild-type digests, the next highest proportion in the L136M immunoproteasomal digest, and the proportion is much lower in both the constitutive digest for the L136M sequence and the R132K and R132K, L136M double mutant digests (Table 11).

Though the other two truncated KK10 epitopes (KI4 and KG7) do not share restrictions with any other HLA alleles, in order to see whether this production pattern might be affected by the presence of an overlapping HLA restriction, I also looked at proportions of N-terminally extended epitope production compared with those which start with the K N-terminus. The pattern of production with more short epitopes being produced by the wild-type digest held

for both of these epitopes (the KI4 is only produced in the wild-type digest) (Figure 10A and D) showing that this effect does not depend on additional restriction by other HLA alleles.

In addition to the non-HLA B*2705 restricted epitopes of which share C-termini with HLA B*2705 restricted truncated epitope forms, there are also several epitopes produced which are not restricted by HLA B*2705. Of these epitopes, only one set is found in the digests of all of the sequences. These are the B15/B62 HLA restricted epitopes GY9 and LY8 (Figure 4). These epitopes are produced at 3.47% to 9.43% of total epitope production across the sequences (Figure 5E), with the proportion produced in the wild-type digests being higher than in the mutant sequences (Table 7, Table 8, Table 9 and Table 10).

Two other non-HLA B*2705 restricted epitopes are found in some of the sequences digested, with the HLA A*02 restricted epitope IV9 (Figure 4, Figure 5G) found following digestion of the R132K, L136M double mutant and the HLA A*33 restricted epitope MR8 (Figure 4, Figure 5F) found after processing of the L136M and the R132K, L136M double mutants. These epitopes are produced at 1% or less of total epitope production (Table 7, Table 8, Table 9 and Table 10). Their presence in only the sequences containing escape mutations demonstrate that the changes in antigen processing caused by these HLA B*2705 related mutations may produce novel epitopes for immune recognition as well as abolishing those present in wild-type sequences.

	Constitutive proteasome					Immunoproteasome				
	1hr	2hr	4hr	6hr	Total	1hr	2hr	4hr	6hr	Total
B8, B27 short	0.11	0.31	0.58	0.49	0.43	1.10	1.81	3.22	2.82	2.39
A24, B8, B27 short	16.4	30.5	**41.8**	**44.4**	36.5	30.5	26.8	33.4	27.8	29.5
A24, B27 short	9.13	8.34	10.4	8.41	9.22	18.1	25.9	15.1	4.69	16.4
B27 only	**67.3**	**52.1**	37.9	35.6	**44.4**	**38.2**	**35.7**	**44.4**	**60.6**	**44.8**
B15, B62	7.11	8.74	9.32	11.1	9.43	12.2	9.78	3.79	4.09	6.90

Table 7: Production of Epitopes by HLA restriction in the wild-type p24 Gag AA126-150 sequence at each time-point and in total for constitutive and immunoproteasomal digests. Given as a percentage of total epitopes produced. The HLA with greatest percentage production is shown in bold text.

	Constitutive proteasome					Immunoproteasome				
	1hr	2hr	4hr	6hr	Total	1hr	2hr	4hr	6hr	Total
B8, B27 short	0.00	0.37	0.47	0.29	0.31	2.40	5.16	4.62	8.06	6.49
A24, B8, B27 short	**74.5**	**69.8**	**63.9**	**67.3**	**67.2**	**56.2**	**46.1**	**49.4**	**51.1**	**50.6**
A24, B27 short	11.3	14.2	18.0	15.6	15.7	36.4	42.6	39.8	20.5	28.5
B27 only	14.3	14.1	16.3	12.6	13.3	4.99	5.89	5.42	12.0	9.37
B15, B62	0.00	1.49	1.36	4.16	3.47	0.00	0.27	0.77	8.28	5.03

Table 8: Production of Epitopes by HLA restriction in the R132K p24 Gag AA126-150 sequence at each time-point and in total for constitutive and immunoproteasomal digests. Given as a percentage of total epitopes produced. The HLA with greatest percentage production is shown in bold text.

	Constitutive proteasome					Immunoproteasome				
	1hr	2hr	4hr	6hr	Total	1hr	2hr	4hr	6hr	Total
B8, B27 short	12.9	16.4	1.90	1.92	2.13	20.7	19.0	17.2	10.4	12.9
A24, B8, B27 short	38.0	**41.8**	**44.1**	**61.2**	**59.5**	**45.9**	22.6	19.3	29.8	27.6
A24, B27 short	0.00	0.00	9.82	6.20	6.39	15.5	**40.0**	**35.4**	19.7	24.6
B27 only	**49.1**	**41.8**	39.6	23.6	25.2	17.9	18.5	27.4	**31.8**	**29.2**
B15, B62	0.00	0.00	4.02	6.22	5.93	0.00	0.00	0.68	6.69	4.63
A33	0.00	0.00	0.54	0.88	0.84	0.00	0.00	0.00	1.60	1.08

Table 9: Production of Epitopes by HLA restriction in the L136M p24 Gag AA126-150 sequence at each time-point and in total for constitutive and immunoproteasomal digests. Given as a percentage of total epitopes produced. The HLA with greatest percentage production is shown in bold text.

	Constitutive proteasome					Immunoproteasome				
	1hr	2hr	4hr	6hr	Total	1hr	2hr	4hr	6hr	Total
B8, B27 short	1.78	3.84	3.48	2.08	2.37	15.7	20.6	22.0	18.3	18.9
A24, B8, B27 short	**50.3**	**45.7**	**48.9**	**70.0**	**65.1**	**41.3**	28.4	31.9	**35.4**	**34.6**
A24, B27 short	4.14	10.7	10.1	6.91	7.51	38.8	**44.5**	**38.7**	25.4	29.5
B27 only	41.9	37.8	34.9	15.2	19.9	4.19	6.03	6.55	12.8	10.9
B15, B62	1.91	1.73	1.71	4.74	4.08	0.00	0.21	0.61	7.05	5.20
A2/B27	0.00	0.00	0.55	0.00	0.07	0.00	0.32	0.25	0.00	0.06
A33	0.00	0.22	0.37	1.09	0.91	0.00	0.00	0.00	1.12	0.81

Table 10: Production of Epitopes by HLA restriction in the R132K/L136M p24 Gag AA126-150 sequence at each time-point and in total for constitutive and immunoproteasomal digests. Given as a percentage of total epitopes produced. The HLA with greatest percentage production is shown in bold text.

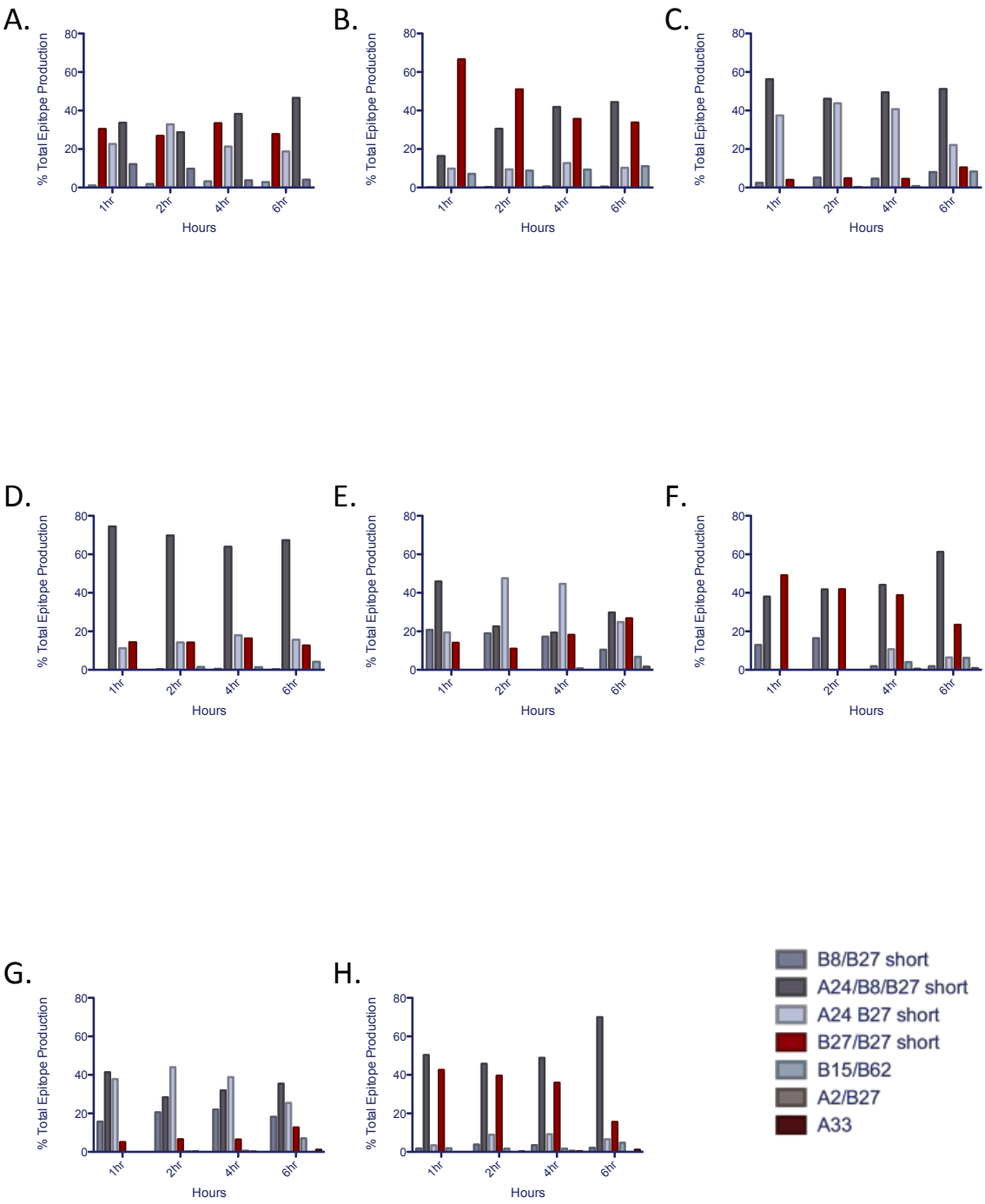


Figure 8: Total epitope production by HLA restriction shown for each sequence over time. **A**. Wild-type immunoproteasome **B.** Wild-type constitutive proteasome **C.** R132K (RK) mutant immunoproteasome **D.** RK mutant constitutive proteasome **E.** L136M (LM) mutant immunoproteasome **F.** LM mutant constitutive proteasome **G.** R132K, L136M (RKLM) double mutant immunoproteasome **H.** RKLM mutant constitutive proteasome

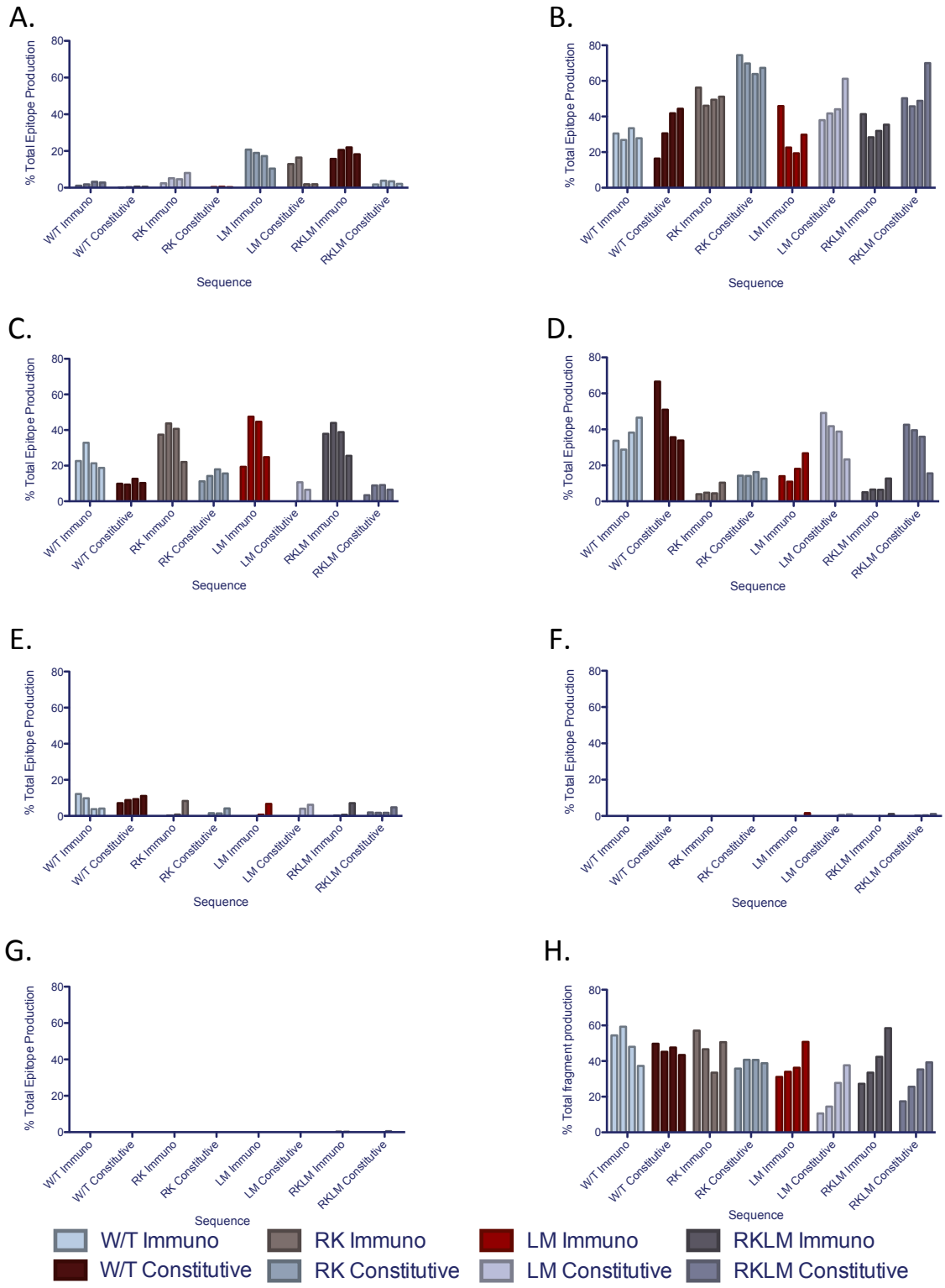


Figure 9: **A-G.** Total epitope production by HLA restriction **A.** B8/B27-short **B.** A24/B8/B27-short **C.** A24/B27-short **D.** B27/B27-short **E.** B15/B62 **F.** A33 **G.** A2. **H.** Total epitope production as a percentage of total fragment production in proteasomal digestion assays, shown by sequence.

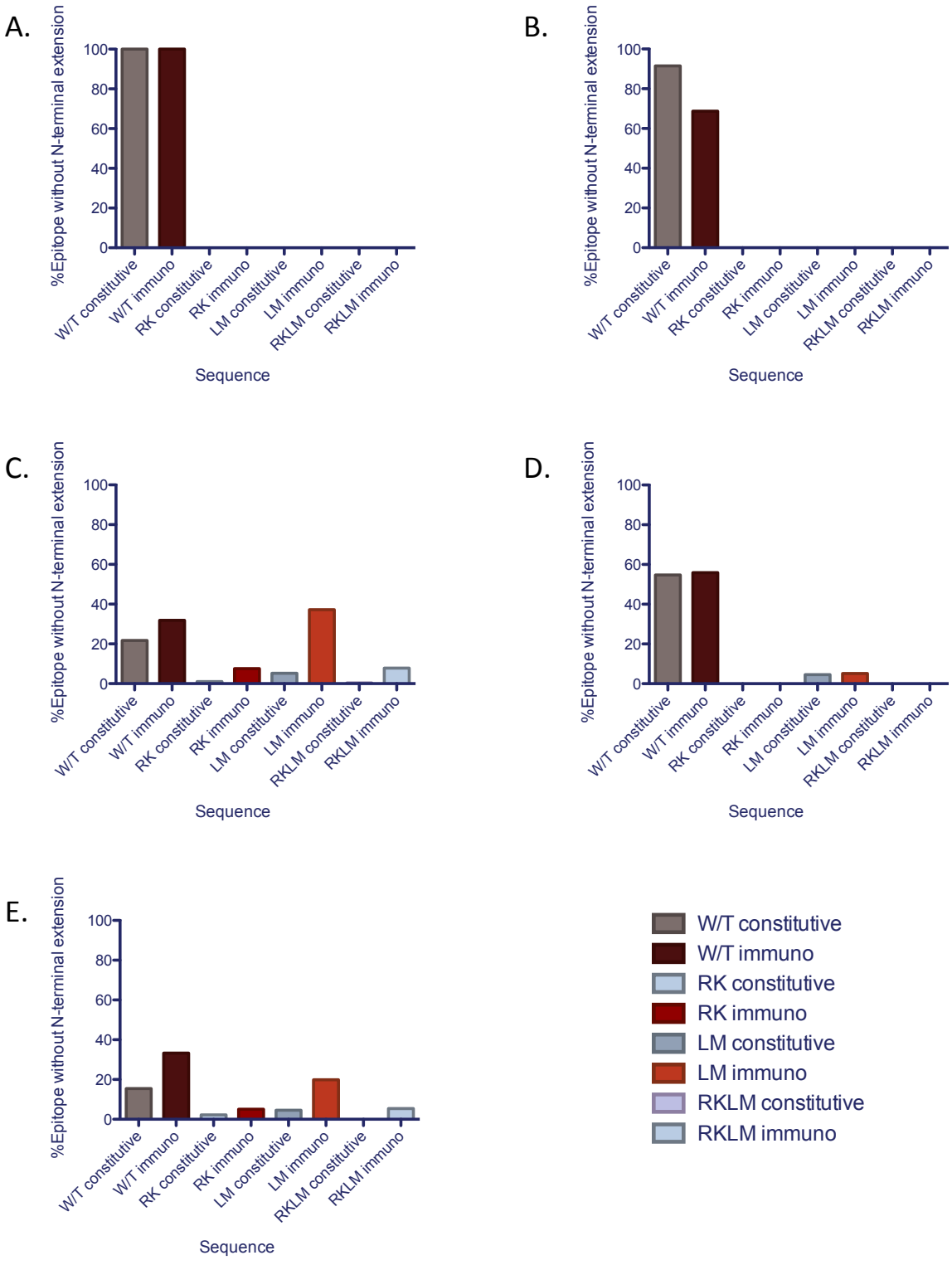


Figure 10: Proportion of truncated KK10 epitope form production consisting of peptide fragments with N-terminal K residue as a percentage of total production of that epitope form for A. KI4 B. KI5 C. KL6 D. KG7 E. KL8.

	W/T C	W/T I	R132K C	R132K I	L136M C	L136M I	RKLM C	RKLM I
KI4	**100**	**100**	0.00	0.00	0.00	0.00	0.00	0.00
KI5	**91.5**	**68.7**	0.00	0.00	0.00	0.00	0.00	0.00
KL6	**21.7**	**31.9**	**1.05**	**7.60**	**5.23**	**37.2**	**0.35**	**7.85**
KG7	**54.7**	**55.9**	0.00	0.00	**4.59**	**5.12**	0.00	0.00
KL8	**15.5**	**33.3**	**2.12**	**5.04**	**4.55**	**19.9**	0.00	**5.41**

Table 11: Proportion of truncated KK10 epitope form production consisting of peptide fragments with N-terminal K residue as a percentage of total production of that epitope form. Non-zero figures shown in bold. C=Constitutive Digest, I= Immunoproteasomal digest.

2.3. TAP transport of KK10 epitope forms

The Transporter associated with Antigen processing (TAP) is responsible for determining which peptide epitopes are taken into the ER and are then able to bind to empty HLA class-1 molecules. The transporter selects substrates at the three N-terminal residues and at the C-terminus of the peptides, it has relatively low sequence specificity, however, there are some sequence and length preferences which affect the ease of gaining access to the ER for different peptide epitopes[233]. It has been shown that of the two stages of peptide interaction with TAP, binding and translocation, the binding step determines substrate specificity and that once selected, bound peptides will be transported immediately[225]. TAP binding is therefore important to consider for our KK10 epitope forms as, transport into the ER is a prerequisite for HLA B*2705 binding and subsequent presentation on the cell surface.

2.3.1. C-terminally truncated KK10 epitope forms display ability for TAP transport

TAP binding assays[233] were carried out using synthetically produced peptides of each of the KK10 epitope forms identified during antigen processing. Binding was measured relative to a standard peptide (RL9), results were then adjusted to give binding relative to the KK10 epitope (Figure 11). Binding of TAP to the various KK10 epitope forms (Figure 11) was shown to be of moderate to high affinity, with affinity generally increasing with epitope length. Previous work has considered all peptides with an RL9 relative IC50 of less than 372 to be at least moderate affinity for TAP[230]. The lowest affinity epitope form was KI5, this peptide was transported with a RL9 relative IC50 of 51.9 (103.8 relative to KK10) and is comparable to several well-known immunodominant epitopes restricted by other HLA alleles. The KK10, KI11 and KR13 epitope forms all bind TAP with a greater affinity than the standard RL9 peptide and KR13 binds slightly better than KK10 (relative binding 0.8) so make extremely good TAP binders. This will result in efficient transport into the ER and loading of the peptides onto HLA B*2705.

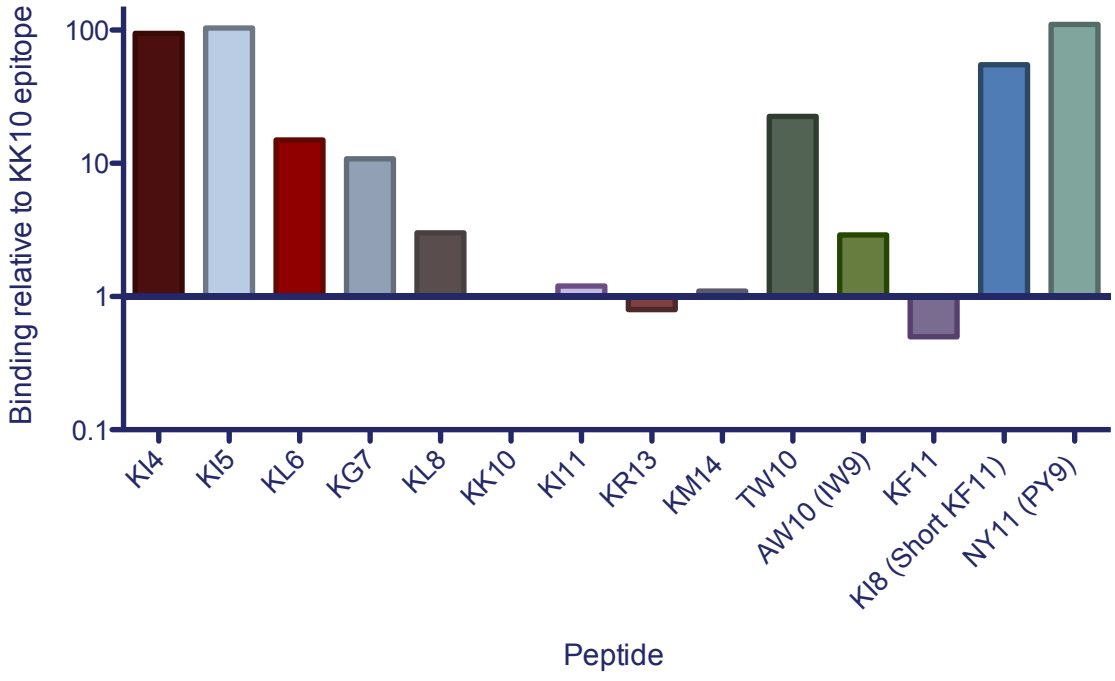


Figure 11: Binding affinity for the TAP transporter measured relative to the RL9 standard peptide as described and shown relative to the IC50 for the KK10 peptide.

2.4. Discussion

2.4.1. The impact of escape mutations R132K and L136M in the HLA B*2705 KK10 epitope on antigen processing and the production of KK10 and extended KK10 epitope forms.

Many sequence mutations have been found that affect antigen processing and the presentation of epitope, both within the epitope and in flanking regions of the sequence[62; 238; 348; 349; 350; 351]. Some of these antigen processing mutations have also been found to enhance escape through other mechanisms, such as the reduction of HLA binding or T-cell recognition, either for the same epitope as the antigen processing effect, or for another epitope[348]. However, despite the extensive study of the KK10 epitope and the effectiveness of the R132K mutation in mediating escape from the CD8+ T-cell response, its effect on antigen processing has never before been studied.

The analysis here shows that both the R132K and L136M mutations, either alone or in combination, have a profound effect on antigen processing of the region. In all three of these sequences, complete, or almost complete abolition of the production of extended KK10 epitope forms was observed (Figure 5). This abolition extended to the KK10 epitope itself in the L136M sequence, though production continued in the R132K and R132K, L136M double mutant sequences (Figure 6). Despite this depletion of KK10 and the extended epitope forms in the escape mutant sequences, the overall production of truncated KK10 epitope forms did not decrease in any of the mutant sequences, indeed it more than doubled when considering the percentage of total epitopes produced in these digests (Table 7, Table 8, Table 9 and Table 10). In particular, the KI5, KL6 and KG7 rose considerably as percentages of total KK10 epitope forms in the mutant sequences (Table 3, Table 4, Table 5 and Table 6).

The selective abolition of the KK10 epitope forms by these escape mutation containing sequences is of interest as it suggests that this may be a deliberate consequence of the escape mutation rather than an unintended side effect of escape through HLA binding and TCR recognition. That the extended epitope forms are particularly affected while the production of truncated KK10 epitope forms from the mutant sequences are maintained relative to the wild-type sequence, suggests that the combination of escape from T-cell responses to these extended epitope forms and maintenance of production of shorter epitope forms may be beneficial for viral replication.

Further to this, the production of the KK10 epitope is not abolished in either the R132K sequence or the R132K, L136M sequence, instead it increases as a percentage of KK10 epitope forms. However it is completely abolished in the L136M sequence. The L136M mutation is often seen before the R132K mutation appears in the epitope[320] and has been proposed as either a TCR escape mutation[321; 342] or as a necessary compensatory change to maintain replication with the R132K mutation[58]. However, it has also been demonstrated that many patient TCR can cross recognise the KK10 epitope containing the L136M change and furthermore that patients with this cross recognition are more likely to be long term non-progressors, which would increase selective pressure to reduce or abolish KK10 production[308]. In contrast, the later R132K mutation has been demonstrated to significantly reduce binding of the KK10 epitope to HLA B*2705 and can thereby prevent CTL recognition of the epitope[256]. Consequently, the production of the KK10 epitope with R132K, or a combination of R132K and L136M, would not be expected to invoke immune recognition of the region and selection to reduce production of the KK10 epitope form will therefore be less (or non-existing) than in patients carrying virus with only the L136M mutation.

The complete abolition of KK10 production in this particular sequence therefore introduces the compelling idea that this may be the main function of this early mutation. The cross recognition seen in some assays may be indicative of a general ability of these TCR to cross

recognise similar peptide sequences in complex with the TCR, indeed it is possible that these cross-reactive TCR are induced by the presence of the extended KK10 epitope forms produced in the region, and that it is the ability to recognise these epitope forms, rather than the L136M mutation that provides additional protection from disease progression.

However, several problems arise when considering this hypothesis. Patients generally do not lose control of viremia until after the R132K mutation and compensatory mutations have occurred[58; 256; 318]. In addition, maintenance of T-cells specific for KK10 and generation of new KK10 recognising TCR following the L136M mutation have been observed[318; 321]. Finally, the effectiveness of the L136M mutation in providing escape from recognition of this epitope is underscored by the later development of the R132K mutation, if the L136M mutation was completely effective in removing immune pressure from this sequence, development of new escape mutations would not be driven. Given that it has been demonstrated that presentation of epitopes continues to some extent in the absence of proteasomal function and that the aminopeptidase nardilysin is able to cleave at a motif that would create the correct N terminus for the epitope[358]. One possibility may be that the L136M change does prevent production of KK10 by the proteasome, but that proteasome independent production of the epitope is sufficient for some continued recognition by CTL.

The maintenance of production of the truncated KK10 epitope forms suggests that, in contrast to the extended forms, these peptides are at best not a source of selection pressure. The possibility also exists that these peptides could be beneficial for viral survival and replication. Large amounts of the KL6 and KL8 forms in particular are produced in all sequences, including the wild-type, while the largest increases in relative production occur for the KL6 and KG7 epitope forms in certain digestions (Table 7, Table 8, Table 9 and Table 10). There are also differences in how these epitope forms are produced between the wild-type sequence and in the sequences containing escape mutations. In the wild-type sequence, significant proportions of the epitopes are produced with the lysine residue forming the N-

terminus; in the escape mutant sequences, far more of the epitopes are produced in N-terminally extended forms (Figure 10). In contrast to previous studies[359], the N-terminally extended forms are preferentially produced by the constitutive proteasome for all the truncated forms apart from KI5. Production of N-terminally extended forms has been associated with improved presentation of epitopes[360] and that this occurs primarily in the constitutive proteasome with escape mutant containing sequences may be of note. This pattern of presentation does not seem to depend on whether these extended epitope forms are also presented by another class-1 HLA. However, the N-terminally extended forms of the KI7 peptide that are produced do contain an HLA class-2 DRB1*1301/1302 epitope[361] which may be presented in some patients. The epitope has been linked to strong CD4+ T-cell responses in patients and the possession of this HLA in association with DQB1*06 was associated with better control of viral replication following treatment[361]. However, there is still some uncertainty as to the contribution of CD4+ T-cell responses to the control of HIV-1 given their status as the target cell type for infection[362; 363; 364; 365; 366; 367].

The production of novel epitopes seen in the L136M and R132K, L136M sequences means that the immune escape given by these mutations may not be complete for all patients, particularly as we see an HLA A*02 restricted epitope being produced in the R132K, L136M sequence. However, the expansion of viral populations containing this sequence and continued presence of these mutations in both the patient and in HLA mismatched individuals following transmission[320], suggests that the presentation of these epitopes does not significantly affect viral replication.

3. Functional studies of HLA B2705 positive HIV-1 patient responses to naturally processed KK10 containing epitope forms.

Clonotyping work in this chapter was completed with the kind assistance of Professor David Price and his lab, particularly Dr. James McLaren and Dr. Kristin Ladell.

3.1. Introduction

In this chapter, we have used pMHC multimer technology[368] (Immudex) to analyse the recognition of KK10 epitope forms in HLA B*2705 HIV-1 positive patient PBMC. We have used dextramer combinations and TCR clonotyping to determine the extent and control of cross-recognition of these responses.

3.1.1. Studies on cross-recognition of epitopes by the T-cell receptor

Approaches taken in assessing the degeneracy of TCR have included both sequential modification of amino acid residues in a known peptide ligand (this approach was also recently used to assess the HLA-B*57/KIR3DL1 interaction[369] and the use of combinatorial peptide libraries[108; 370; 371]. This combination of approaches has shown that the TCR binds through recognition of both the HLA and the peptide, but often only a few contact residues in the peptide are bound.

In the case of HLA class-2 specific TCR, similarities in a small structural motif in the peptide epitope have been shown to be sufficient for cross-recognition[372]. Outside of these primary contact regions, substitutions are more readily accepted. In a study using a combinatorial peptide library, a type-1 diabetes derived auto-reactive HLA A*0201 restricted TCR showed ability to bind more than a million peptide epitopes within a 100 fold range of sensitivity of the index peptide[371]. This study showed that the contact residues at positions 4-6 were more highly conserved, but that outside this region, many substitutions were possible. Indeed one sequence, differing at 7 of 10 residues, gave up to 100 fold better sensitivity than the index peptide.

That the study used an auto-reactive TCR may partially explain the enormous degree of degeneracy seen in this case, as a major contributor to auto-reactivity is thought to be cross recognition of a self-epitope by a TCR generated in response to a pathogen derived peptide during infection[372; 373]. Other aspects of the immunological context of the TCR may also be important to consider when looking at T-cell cross reactivity, with one study showing variability in the extent of cross recognition depended partly on the activation state of the antigen presenting cells (APCs)[374]. A study examining 3 TCRs recognising the same HLA-peptide complex but with differing degrees of cross-recognition of other peptides showed that differences between the cross-reactive TCR and the specific TCR in this system depended on the type, location and total area of receptor-ligand contacts[288]. The more cross reactive the TCR in this study, the smaller the total area of interaction on the surface of the HLA complex, the more the contacting residues focussed on HLA residues instead of peptide interaction and the higher the percentage of the total interaction surface being mediated through Van-der-Waals (VDW) forces. The increase in the percentage of VDW type interactions is interesting as it may partly offset the decrease in interaction area, with VDW interactions being quite strong when concentrated in the same area[288].

Further insight into how the structure of the TCR modulates its ability to recognise HLA-peptide complexes has recently been given by studies examining the conformation of the CDR loops of the well-characterised A6 TCR in complex with different HLA A*02-bound epitopes[107; 282]. For this TCR, the CDR3β loop was found to adopt several different conformations when bound to the different epitopes. The range of conformations for the CDR3α loop was more restricted, while the conformations of both the CDR1 and 2 loops appeared to be fixed. Molecular dynamics simulations of the two CDR3 loops in the TCR showed that CDR3α adopts a “binding capable” or a “non-binding capable” conformation, separated by a large energy barrier. In contrast the CDR3β loop had a wide range of binding capable conformations with small energy maxima/minima. This would allow the CDR3β loop

to modify its conformation to best fit the peptide being bound, termed a conformational-selection model of TCR binding[107]

This analysis shows that, for this TCR, cross-recognition is determined largely by the ability of the CDR3β loop to exhibit conformational flexibility, however, binding studies also showed that the CDR3α loop has only one conserved interaction with peptide in the structures studied[107] and several interactions with HLA-A*02 residues. These interactions may serve to restrict the TCR specificity by HLA and to a peptide subset; however, the small number of peptide residues contacted consistently by this loop gives considerable scope for recognition of alternate peptide sequences. This study correlates well with the finding in HLA B*2705 KK10 specific TCR that increased cross recognition was associated with changes in the nature of the CDR3β loop[308].

3.2. Flow cytometry analysis of patient responses using dextramer staining

The T-cell responses of patients to the naturally processed KK10 epitope forms were assessed through the use of HLA B*2705 dextramer staining of patient PBMC (Method in Chapter 7.3.2). Patients were HLA B*2705 positive individuals identified from an HIV-1 infected Danish cohort from this cohort, we obtained 10 patients who carried HLA B*2705 and were all assessed at one or more time-points during chronic disease (Table 12, Table 13 and Table 14).

Patient Number	Time since first positive test	Time-point	CD4+ count	Viral load	Treatment	Group
101074	9 months	13/03/2007	330	112171	No	1
100967	11 years 5 months	20/11/2007	630	91258	No	1
241242	16 years 1 month	26/07/2007	540	34704	No	1
050875	4 years	11/12/2006	580	87342	No	2
170466	10 months	12/10/2006	190	1843	No	2
060473	2 years 4 months	11/10/2006	760	16625	No	1
030869	1 year 5 months	27/03/2008	540	12614	No	1
030778	1 year	22/10/2007	470	190	Viramune, Kivexa	2
081077	4 years 2 months	26/08/2008	419	39	Truvada, Efavirenz	1
111156	19 years 7 months	08/05/2006	1000	Non-detectable	Viread, Kaletra	2

Table 12: Viral load (copies/ml), CD4+ count (cells/ml), length of infection and treatment information on patients used in the study at the first time-point examined.

Patient Number	Time since first positive test	Time-point	CD4+ count	Viral load	Treatment	Group
101074	1 year	11/06/2007	320	112397	No	2
060473	5 years 11 months	03/05/2011	680	20	Truvada, Viramune	2
081077	6 years 2 months	04/08/2011	490	19	Truvada, Efavirenz	1
111156	21 years 9 months	01/07/2008	610	Non-detectable	Viread, Kaletra	2

Table 13: Viral load (copies/ml), CD4+ count (cells/ml), length of infection and treatment information on patients used in the study at the second time-point examined.

Group	Patient Number	Time-point 1	Time-point 2	Interval (Months)
1	101074	13/03/2007	11/06/2007	3
	060473	11/10/2006	03/05/2011	55
	030869	27/03/2008		
	100967	20/11/2007		
	241242	26/07/2007		
	081077	26/08/2008	04/08/2011	35
2	111156	08/05/2006	01/07/2008	26
	050875	11/12/2006		
	030778	22/10/2007		
	170466	12/10/2006		

Table 14: HLA B*2705 HIV-1 positive patient numbers and sample time-points tested for responses to KK10 epitope forms using HLA B*2705 dextramer complexes.

Dextramers consist of multiple (usually 10) HLA molecules that have been refolded around the peptide of interest linked to a dextran backbone via a biotin-streptavidin link[375]. The dextran backbone is then tagged with a fluorophore to enable analysis by flow cytometry. Dextramers were used both alone and in combination in addition to phenotypic and viability markers to assess responses to the epitopes and levels of cross-recognition.

3.2.1. CD8+ T-cell responses are seen to a variety of C-terminally extended and truncated epitope forms

Our antigen processing analysis[238] (Chapter 2) of the p24 Gag aa126-150 HIV-1 B consensus sequence demonstrated that a number of different KK10 epitope forms were being produced by both the constitutive and immune-proteasomes in addition to the optimal KK10 epitope. These epitope forms included KL8, KK10, KI11, KR13 and KM14 and these forms were incorporated into dextramers and assayed for recognition by CD8+ and CD4+ T-cells in patient PBMC. Additional shorter epitope forms identified by antigen processing were assessed by intracellular cytokine staining (ICS) or enzyme-linked immunosorbent spot (ELISpot) rather than through dextramer staining (Chapter 5.2.2).

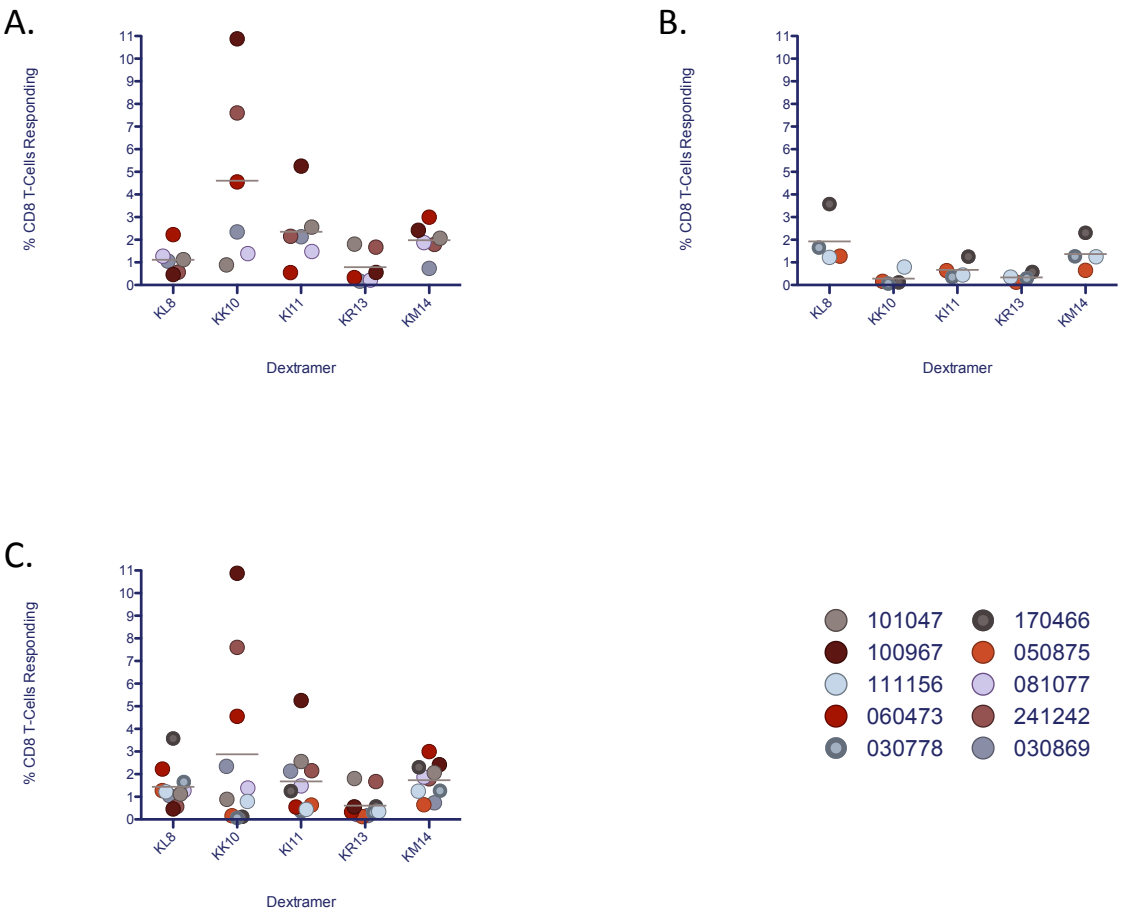


Figure 12: CD8+ T-cell responses to KK10 containing epitope forms in A. HLA B*2705 patients recognising KK10 B. Those without KK10 recognition C. All patients.

Defintion of recognition was difficult to accurately determine due to the heterogeneous nature of some populations (Figure S 12A) and differences between individual dextramers. A clear population above the gated background level of mismatched dextramer controls was therefore regarded as positive. In accordance with this definition CD8+ T-cell responses to the KK10 epitope were seen in 6 of the 10 patients tested (Figure 12C, Figure 14). The presence or absence of KK10 recognition has been used to group the patients for several analyses throughout this chapter (group 1 = KK10 responders, group 2 = KK10 non-responders). These responses varied from 1% up to nearly 11% of all CD8+ T-cells and are characterised by being highly fluorescent and clearly separated from the non-responding population. In one patient, 241242 we see both a highly fluorescent population of KK10

positive cells and a moderately fluorescent population (Figure 22) both clearly distinct from the negative population and from each other. This may indicate a difference in the number or avidity of the T-cell receptors on the surface of cells between these two populations.

Patients that we have classified as being without a KK10 response (group 2) (Figure 12B, Figure 13) show no significant positive population above levels for irrelevant dextramer controls. Any population seen at this level may represent slight differences in background binding between dextramers as a different HLA allele - HLA A*0201 was used for the irrelevant peptide control stains (patients were tested for recognition of this epitope prior to tests with HLA-B*2705 and no recognition was seen). Any slightly positive cells could also be due to slight cross recognition from TCR responding to other epitopes or possibly small, very low affinity populations. The absence of KK10 response in such a large proportion of our patients (40%) differs markedly from the 92% response rate seen in an earlier study[315], however, this could be due to the staging of patients in our cohort, which are in the chronic phase of disease and have been infected for many years at the time of sampling, as compared with the acutely infected individuals in the previous study cohort.

In addition to the recognition of the KK10 epitope in group 1, we also observed a range of responses to the naturally processed epitope forms that were tested. The two forms with the highest mean response among all patients are the KM14 epitope (1.74% CD8+ T-cells, SD 0.76) and the KI11 epitope (1.68% CD8+ T-cells, SD 1.49). The response to the KM14 epitope is more similar among the patients than that of either the KI11 or the KK10 response (SD 3.69), with the total binding ranging from 0.65% CD8+ T-cells to 3% CD8+ T-cells (Table 15). The fluorescent Intensities of the CD8+ T-cells that recognise KM14 are quite heterogeneous, with T-cells seen at all ranges of fluorescent intensity from near negative to moderate-highly fluorescent (Figure S 12). This broad spread of fluorescence suggests that there may be quite different numbers of dextramers bound to the CD8+ T cells within the KM14 positive population. This may indicate that the T cells in this population belong to a number of

different clonotypes with differing affinity or levels of TCR expression. However, distinct populations with a smaller spread of fluorescence are seen in some patients (060473, 081077 TP2). The response to KI11 has a wider range than that of KM14 in terms of the percentage of total patient PBMC recognising the dextramer (from 0.35% to 5.26% total CD8+ T-cells)(Table 15) with some patients that respond having over 5% of CD8+ T-cells binding. KI11 recognition was also variable in terms of Mean Fluorescent Intensity (MFI) with some patients displaying clearly separate positive populations with a high mean fluorescence.

The response to the KL8 epitope showed the second highest mean percentage of CD8+ T-cells binding, with a mean across all patients of 1.44%, ranging from 0.47% to 3.58% (SD 0.90) (Figure 12)(Table 15). Dextramer staining for this epitope form gave positive populations having a continuity of fluorescent intensity with the negative cells rather than being distinct (Figure S 12), it was therefore difficult to determine an exact cut-off for positive staining which was possible in most patients for the KK10 and KI11 epitope forms. The KL8 response was seen in all of the patients to some extent, but there were no patients that showed a clearly defined CD8+ T-cell KL8 positive population. Few patients recognised the KR13 epitope form and observed responses had the lowest mean CD8+ T-cell binding of any epitope form, though two patients, 101074 and 241242 had higher recognition levels, with 1.81% and 1.67% of CD8+ T-cells binding respectively (Figure 14A and E)(Table 15). Some patients recognising this epitope form, including 241242 showed clear separation from the negative cells, however for others, the staining was heterogeneous (Figure S 12).

Comparison of CD8+ T-cell recognition of epitope forms between the patients in groups 1 and 2 did not show any significant differences in responses (using a Mann-Whitney test) to the other KK10 epitope forms. However some trends were observed, with KK10 responders having a smaller percentage of CD8+ T-cells binding the KL8 dextramer (P=0.1143), and also having a greater number of CD8+ T-cells binding to the KI11 dextramer (P=0.0582)(Table

16). Though this result was not significant, it does approach significance and repeating the experiment with a greater sample number may show whether or not this is a genuine difference.

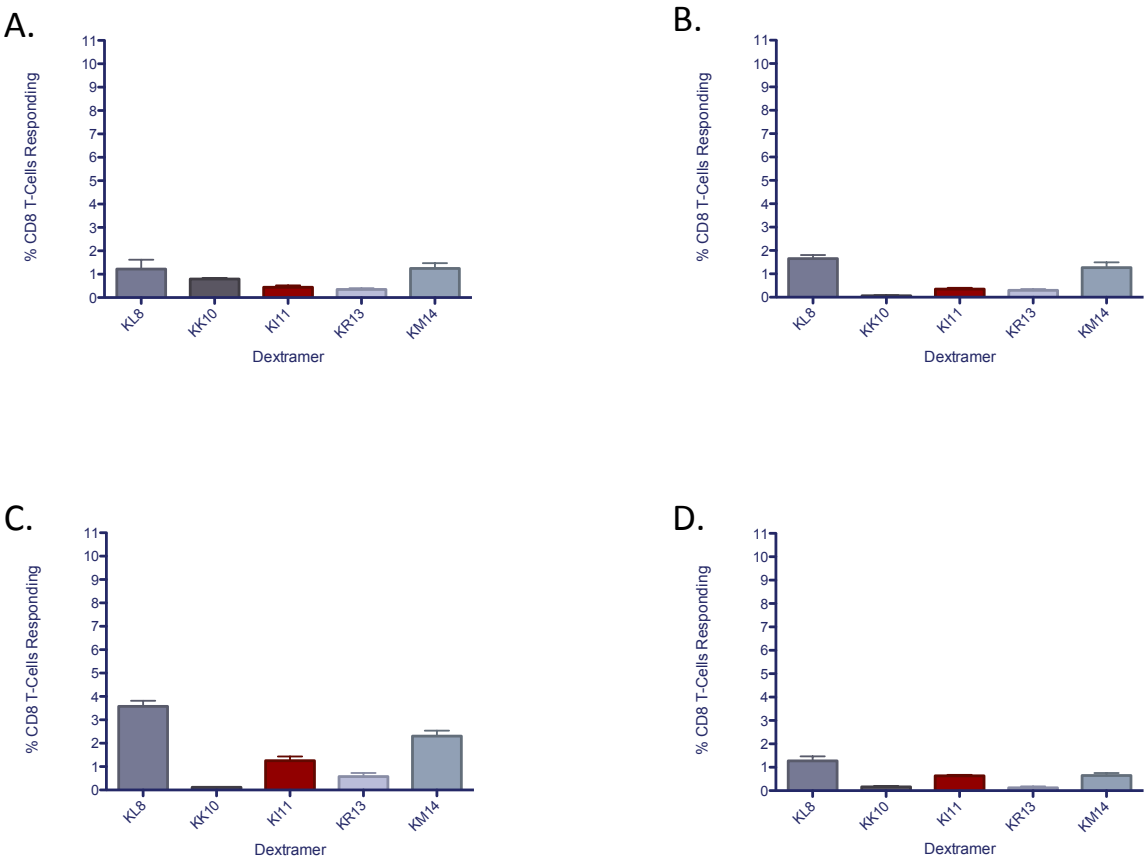


Figure 13: CD8+ T-cell responses to KK10 containing epitope forms in patients without a KK10 response A. Patient 111156 B. Patient 030778 C. Patient 170466 and D. Patient 050875.

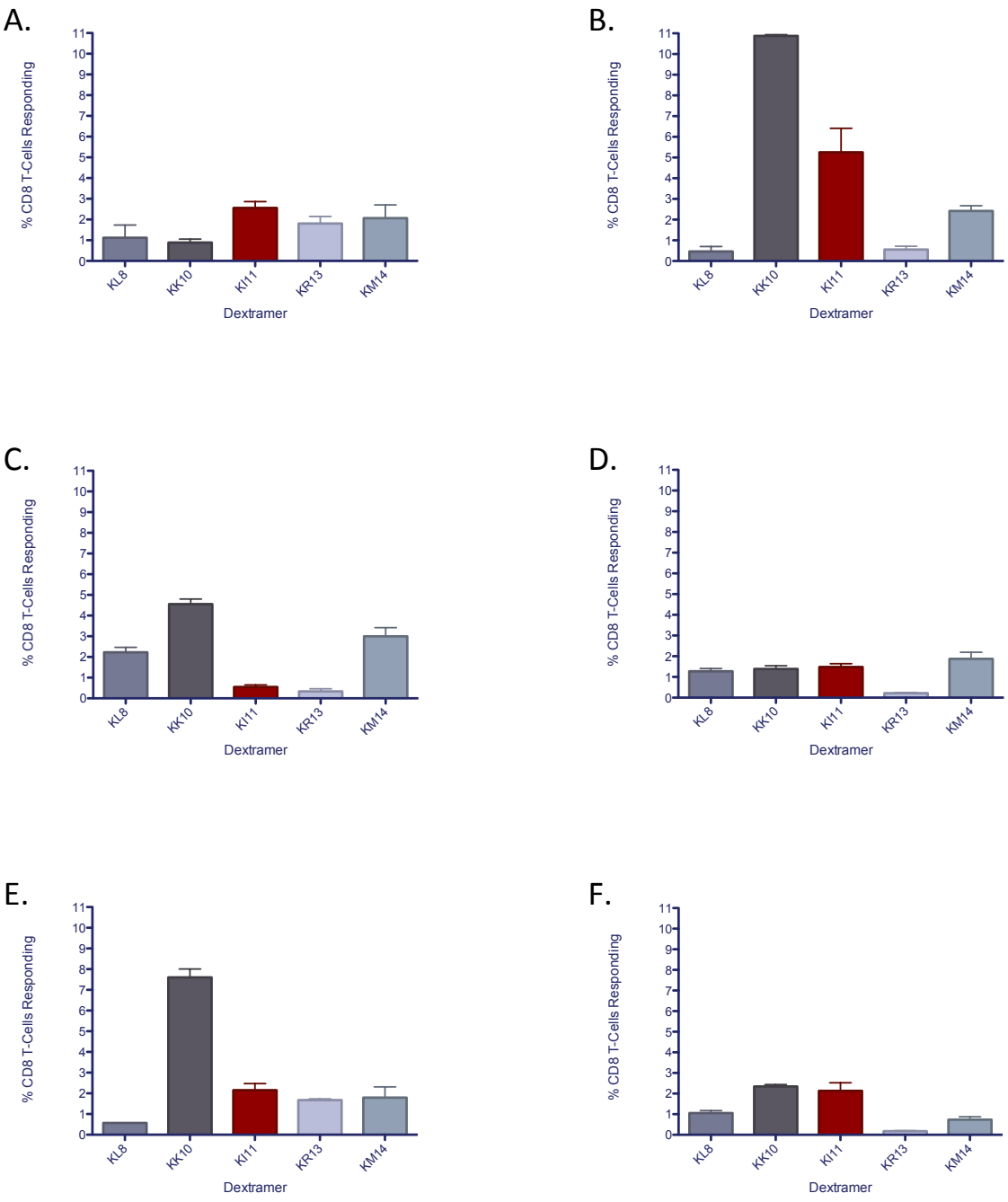


Figure 14: CD8+ T-cell responses to KK10 containing epitope forms in patients with a KK10 response A. Patient 101074 B. Patient 100967 C. Patient 030869 D. Patient 081077 E. Patient 241242 F. Patient 060473.

Group	Patient number	KL8	KK10	KI11	KR13	KM14
1	101074	1.12	0.89	2.56	1.81	2.07
	100967	0.47	10.88	5.26	0.55	2.42
	060473	2.23	4.56	0.55	0.33	3.00
	081077	1.27	1.39	1.48	0.21	1.87
	241242	0.57	7.60	2.15	1.67	1.79
	030869	1.06	2.35	2.14	0.17	0.73
	Mean	**1.12**	**4.61**	**2.36**	**0.79**	**1.98**
	Standard Deviation	**0.63**	**3.94**	**1.59**	**0.75**	**0.75**
2	111156	1.22	0.80	0.44	0.35	1.25
	030778	1.65	0.06	0.35	0.30	1.27
	170466	3.58	0.12	1.25	0.58	2.30
	050875	1.28	0.16	0.64	0.13	0.65
	Mean	**1.93**	**0.28**	**0.67**	**0.34**	**1.37**
	Standard Deviation	**1.11**	**0.34**	**0.41**	**0.19**	**0.69**
All	**Mean**	**1.44**	**2.88**	**1.68**	**0.61**	**1.74**
	Standard Deviation	**0.90**	**3.70**	**1.49**	**0.61**	**0.76**

Table 15: CD8+ T-cell responses to KK10 epitope forms in group 1 (KK10 recognition) and group 2 (no KK10 recognition) patients.

Epitope form	Group 1 mean	Group 2 mean	P value
KL8	1.1200	1.9310	0.1143
KK10	**4.6100**	**0.2848**	**0.0095**
KI11	2.1100	0.6708	0.0582
KR13	0.7913	0.3382	0.6095
KM14	1.9790	1.3690	0.2571

Table 16: Comparison of mean patient responses for each epitope form between patients in groups 1 and 2 using a Mann-Whitney test. Significant differences are shown in bold text.

3.2.2. CD4+ T-cell responses are seen to a variety of C-terminally extended and truncated epitope forms

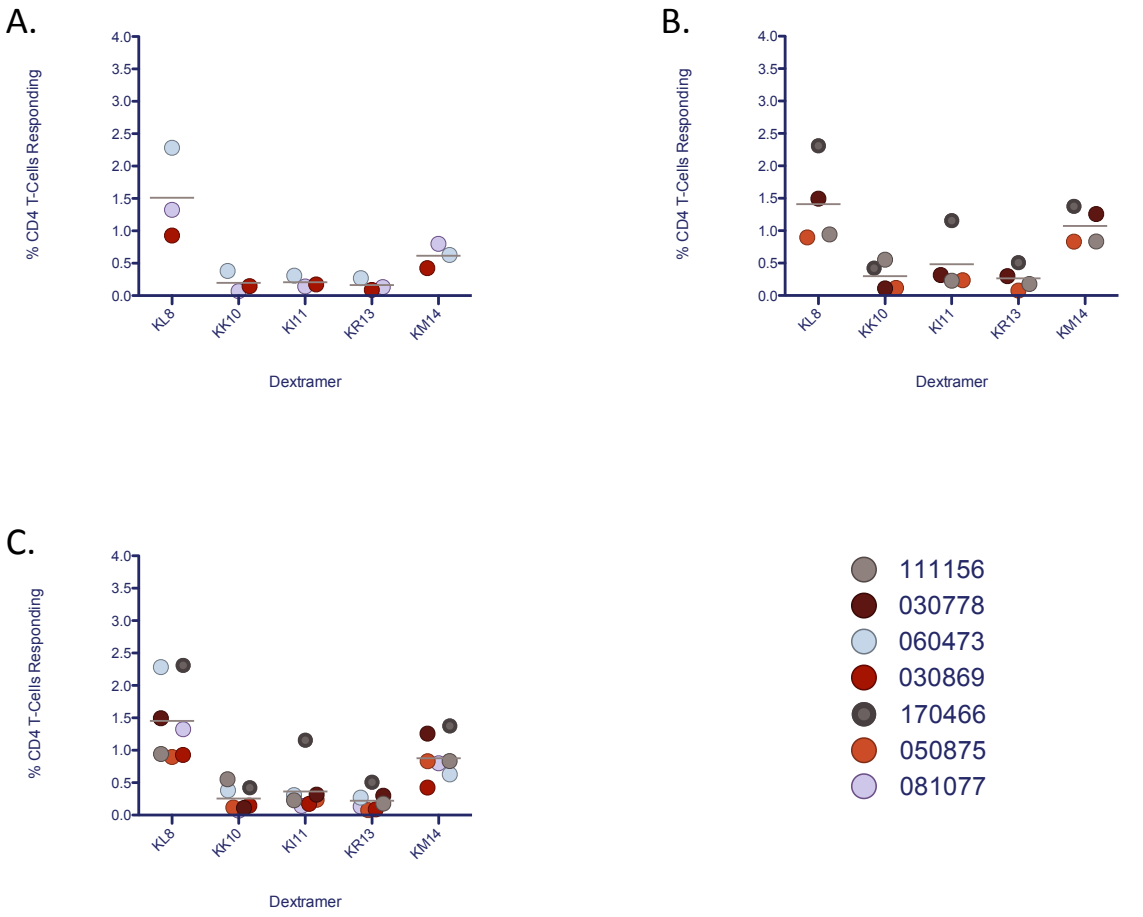


Figure 15: CD4^{+} T-cell responses to KK10 containing epitope forms in HLA B*2705 patients grouped by A. Patients with a CD8+ KK10 response B. Patients without a CD8+ KK10 response and C. All patients.

The dextramers used for the staining performed on our patient PBMC were HLA class-1 B*2705 alleles loaded with our peptide of interest. According to current understanding of recognition by TCR, these should be exclusively seen by CD8+ T-cells, even if the epitopes themselves are able to induce CD4+ T-cell responses when bound to HLA class-2[376; 377]. However, analysis of our data showed specific CD4+ T-cell responses for some epitope forms despite being bound to HLA class 1 and these responses were not seen in negative controls

with mismatched HLA-peptide complexes. It should be noted that the negative controls used in these assays were HLA class-1 complexes with irrelevant peptides that were designed for mismatching of HLA class 1 presentation and were not designed with these responses in mind. Therefore, while no recognition was seen by CD4+ of these controls and they proved suitable in this case, HLA class 2 recognition may still be possible in other patients given the recognition we have seen for the B*2705 epitope forms. An improved set of controls should be used in future studies investigating these responses.

The CD4+ T-cell recognition patterns and frequencies varied between patients. However, recognition of the short KL8 epitope and the long KM14 epitope forms were most commonly detected (Figure 15 and Figure S 14). That the largest CD4+ T-cell responses (1.45% CD4+ T-cells, range 0.47%-3.58%) were seen to the KL8 epitope is unexpected as HLA class-2 responses are usually associated with longer peptide epitopes[378; 379] relative to class-1[380; 381; 382], though recent work has suggested that cross-recognition motifs for CD4+ TCR can consist of just a small commonly structured area of two peptide epitopes[372]. The CD4+ responses to the epitope forms seen were all of a similar nature, with a heterogeneously fluorescent positive population (Figure S 14). There are no significant differences between the responses of CD8+ KK10 responders and non-responders among the CD4+ responses. Though KL8 and KM14 epitopes are the epitope forms most commonly recognised by CD4+ T-cells, one patient, 170466 (Figure 16G) also had a relatively large CD4+ response to the KI11 epitope (1.16% CD4+ T-cells)(Table 17).

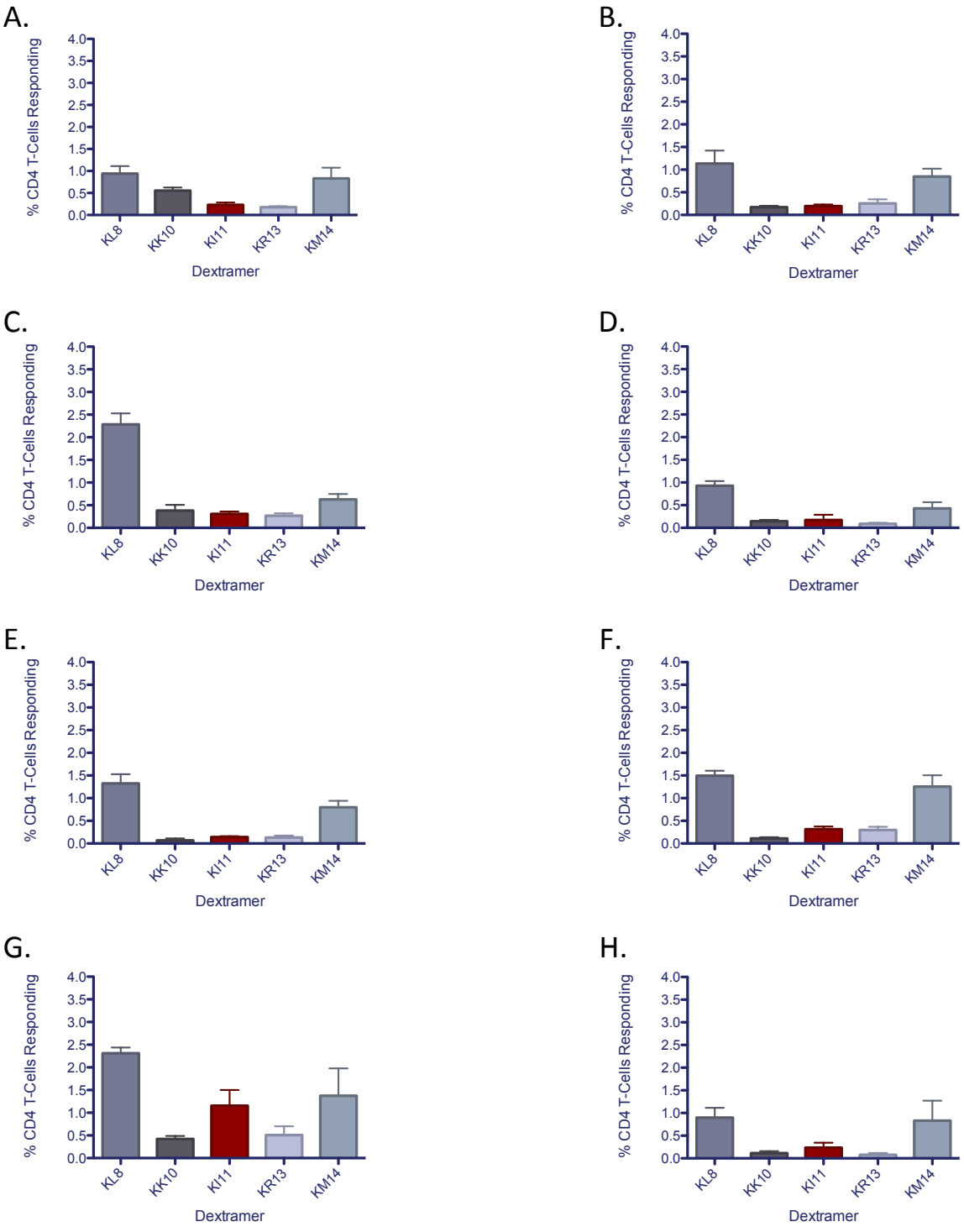


Figure 16: CD4^{+} T-cell responses to KK10 containing epitope forms in HLA B*2705 patients A. 111156 B. 111156 (later time-point) C. 060473 D. 030869 E. 081077 F. 030778 G. 170466 and H. 050875.

Group	Patient number	KL8	KK10	KI11	KR13	KM14
1	060473	2.28	0.38	0.31	0.27	0.63
	030869	0.93	0.15	0.17	0.09	0.42
	081077	1.32	0.07	0.14	0.13	0.80
	Mean	**1.51**	**0.20**	**0.21**	**0.16**	**0.62**
	Standard Deviation	**0.69**	**0.16**	**0.09**	**0.09**	**0.19**
2	111156	0.94	0.55	0.23	0.18	0.83
	030778	1.50	0.11	0.32	0.30	1.26
	170466	2.31	0.42	1.16	0.51	1.38
	050875	0.90	0.12	0.24	0.08	0.83
	Mean	**1.41**	**0.30**	**0.49**	**0.27**	**1.08**
	Standard Deviation	**0.66**	**0.22**	**0.45**	**0.19**	**0.29**
All	**Mean**	**1.45**	**0.26**	**0.37**	**0.22**	**0.88**
	Standard Deviation	**0.62**	**0.19**	**0.35**	**0.15**	**0.33**

Table 17: CD4+ T-cell responses to KK10 epitope forms in group 1 (KK10 recognition) and group 2 (no KK10 recognition) patients.

3.2.3. Cross-Reactivity of T-cell responses between subsets of KK10 epitope forms

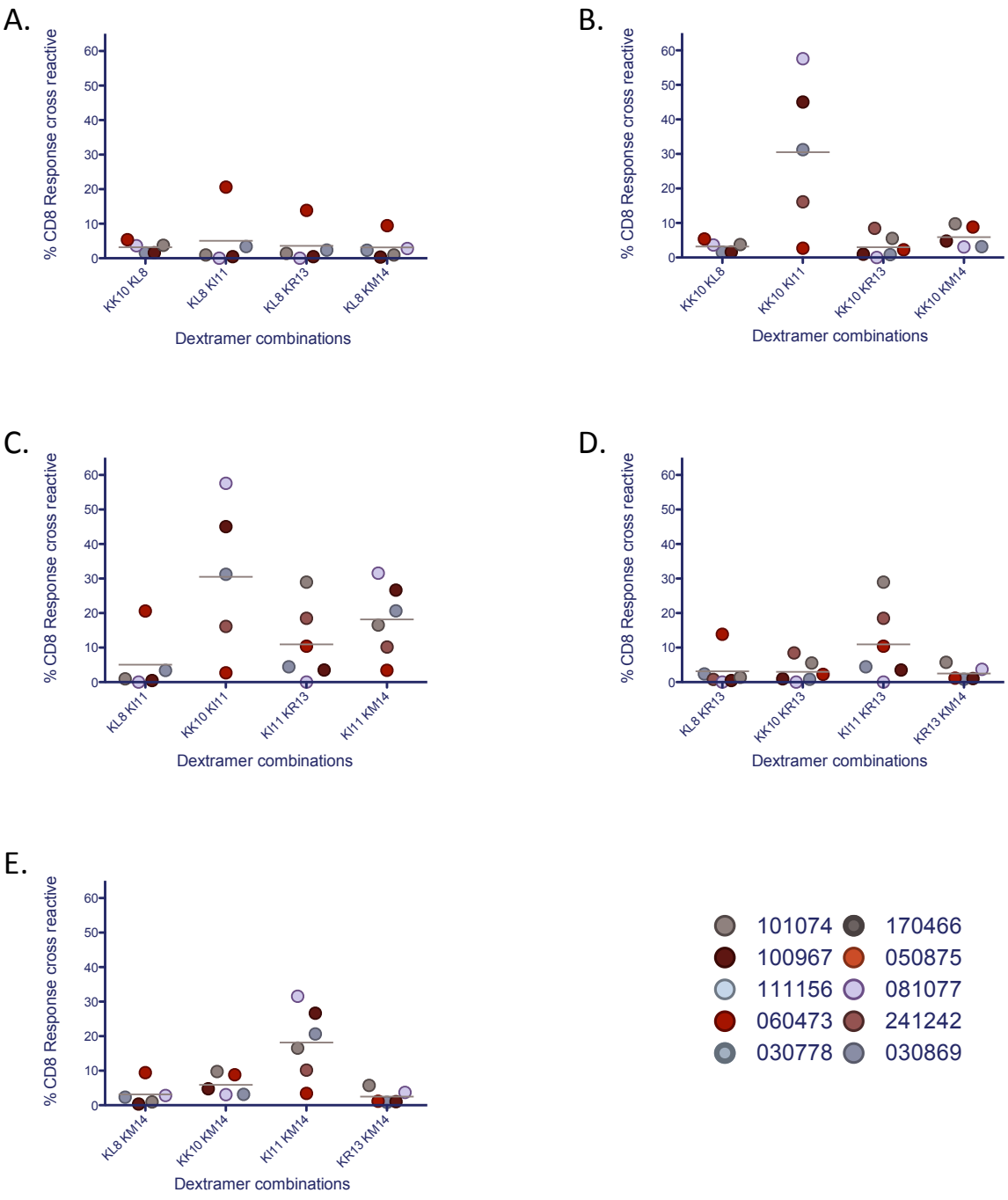


Figure 17: Cross reactive CD8+ T-cell responses between pairs of KK10 containing epitope forms in patients with a KK10 response. A. Cross reactivity with KL8 B. Cross reactivity with KK10 C. Cross reactivity with KI11 D. Cross reactivity with KR13 E. Cross reactivity with KM14.

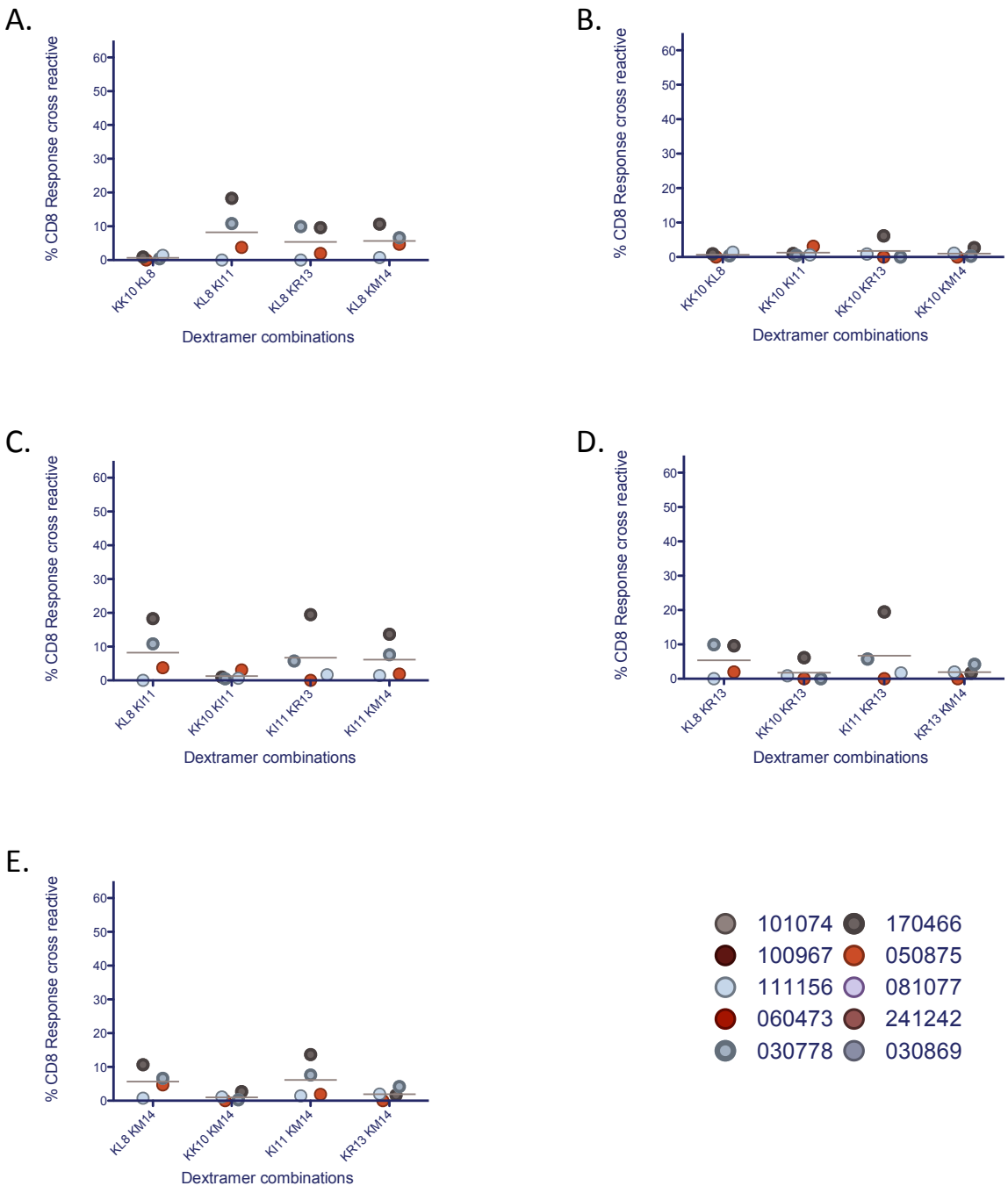


Figure 18: Cross reactive CD8+ T-cell responses between pairs of KK10 containing epitope forms in patients without a KK10 response. A. Cross reactivity with KL8 B. Cross reactivity with KK10 C. Cross reactivity with KI11 D. Cross reactivity with KR13 E. Cross reactivity with KM14.

In order to further characterise the response to the KK10 epitope forms in HLA B*2705 patients, we also assessed the cross reactivity of these responses by performing staining with two dextramers on different fluorophores. These data enabled us to assess whether responses to the KK10 epitope forms were independent of each other or relied on a shared, cross reactive set of T-cells. The levels of cross recognition were assessed for each epitope pair in all patients for both CD8+ T-cell and CD4+ T-cell populations. The levels of cross reactivity observed show marked differences depending on both the epitope pairs and the T cell subsets involved in the response.

3.2.4. CD8+ T-cell cross recognition patterns in KK10 responders and non-responders

Cross recognition of CD8+ T-cell responses to the epitope forms (Figure 17, Figure 18 and Figure S 13) differs between all patients to some extent, with individual patients occasionally displaying cross recognition which is not seen in the group as a whole (patients 081077- TP2 KK10/KM14, 060473 – TP1 KL8/KI11). However, across the patients there are several shared cross recognition patterns, particularly within the two patient groups.

The most notable difference between groups 1 and 2 is the high frequency of cross recognition between the KI11 epitope form and other epitope forms in group 1 (Figure 17C). To compare the overall levels of cross recognition in patients with responses of differing magnitudes, the sum of all dextramer positive cells in the dual-dextramer stains was taken and the cross-reactive population (i.e. those T-cells that were doubly fluorescent) was divided by this to give cross-reactivity as a percentage of the total response in each stain. When this cross-recognition is grouped by epitope form there are significant difference between the two groups for both KI11 and KK10 (though as patients are grouped by presence or absence of KK10 recognition, this difference is expected). For all group 1 patients, the mean percentage of CD8+ T-cell response that is cross-reactive is 16.0% for the

dextramer stains containing KI11. This compares to 10.3% for KK10, the next highest mean (Table 18). The difference between the levels of cross reactivity seen towards KI11 in the two groups is statistically significant at P=0.032 (Mann-Whitney test) though the KI11 epitope form still has the highest mean levels of cross recognition of all the epitope forms in the group 2 patients at 5.6% (Figure 17 and Figure 18).

Cross-recognition by T-cells of the epitope forms necessarily means that single responses to both epitope forms are needed in the patient. Therefore, when grouping a cross reactive response by epitope form as shown above, the data will also rely on the presence of the other epitope forms that give the cross-recognition.

When cross-recognition is grouped by epitope in this way, the data necessarily relies on all of the epitope forms as cross recognition of the epitope form of interest with every other form is compared. Therefore, analysis of all of the separate cross reactive responses will better allow us to understand which responses are contributing to the differences seen in cross-recognition between groups. It should also be noted that cross-recognition data cannot be considered independently from the individual recognition of the epitope forms. While there were no significant differences between the individual epitope forms (other than KK10) between the two groups (Table 16) the levels of recognition of KI11 approached significance (P=0.0582) and may similarly impact these data.

Comparison of individual cross-reactive responses (Table 19) shows that differences between the groups are significant (P=<0.05, Mann-Whitney test) for cross recognition between KK10/KL8, KK10/KI11 and KK10/KM14. That these responses all involve KK10 shows that the differences in cross recognition seen are due to a lack of KK10 recognition in group 2 patients rather than any large difference in ability of the T-cells to cross-recognise epitope forms between the two groups. However, KM14/KI11 cross-recognition (P=0.07) shows a large, but not significant difference between the groups.

In addition to the differences seen in KI11, cross-recognition of the truncated KL8 epitope form between the two groups (P=0.6343)(Table 18) shows some trends that are worth noting. In group 2, cross recognition of KL8 binding TCR is more frequent with a higher mean response (Table 18) and 2 patients display cross recognition levels at or above 10% (030778, KL8-KI11, KL8-KR13 and 170466 KL8-KI11, KL8-KM14)(Table 19). This is uncommon in group 1 where cross recognition levels with this epitope are 5% or below in five of the six patients. Only one patient (060473) in this group shows a high level of cross recognition between their KL8 response and that of other epitope forms, with 20.6% of the response to the KL8 and KI11 epitope forms being cross reactive (Table 19).

In addition to these general patterns, there are also examples of cross-reactivity between other epitope forms in individual patients (081077 TP2 KK10-KM14, 060473 TP1 KL8-KR13)(Table 19) showing that while there are significant similarities in the generation of responses to these epitope forms, they can differ in each patient.

	Group 1 mean	Group 2 mean	P value
KL8 crosses	3.6250	4.9880	0.6343
KK10 crosses	10.2900	1.1900	0.0002
KI11 crosses	16.0400	5.5990	0.0320
KR13 crosses	5.0120	3.9520	0.6671
KM14 crosses	7.9480	3.7010	0.0947

Table 18: Comparison of differences between cross reactivity for each epitope form in groups 1 and 2 (Mann-Whitney test).

Patient Group	Patient No.	KL8 KI11	KL8 KR13	KL8 KM14	KK10 KL8	KK10 KI11	KK10 KR13	KK10 KM14	KI11 KR13	KI11 KM14	KR13 KM14
1	101074	0.94	1.42	0.96	3.74	N/A	5.53	9.75	**28.98**	**16.53**	5.75
	060473	**20.62**	**13.87**	9.43	5.36	2.69	2.26	8.81	**10.40**	3.44	1.16
	030869	3.39	2.38	2.32	1.56	**31.25**	0.84	3.15	4.45	**20.63**	0.84
	100967	0.49	0.47	0.35	1.61	**45.06**	0.95	4.77	3.52	**26.65**	1.02
	241242	N/A	0.79	N/A	N/A	**16.13**	8.45	N/A	18.48	**10.19**	N/A
	081077	0.00	0.00	2.82	3.62	**57.58**	0.00	3.08	0.00	**31.55**	3.70
	Mean	**5.09**	**3.16**	**3.18**	**3.18**	**30.54**	**3.01**	**5.91**	**10.97**	**18.17**	**2.49**
	Std. Dev.	**8.78**	**5.31**	**3.64**	**1.61**	**21.94**	**3.30**	**3.17**	**10.95**	**10.40**	**2.17**
2	111156	0.00	0.00	0.74	1.41	0.62	0.88	1.13	1.70	1.43	1.95
	050875	3.77	1.96	4.70	0.00	3.12	0.00	0.00	0.00	1.88	0.00
	030778	**10.81**	9.94	6.63	0.37	0.48	0.00	0.27	5.74	7.61	4.22
	170466	**18.29**	9.61	**10.67**	0.90	1.01	6.13	2.70	**19.46**	**13.65**	1.64
	Mean	**8.22**	**5.38**	**5.69**	**0.67**	**1.31**	**1.75**	**1.03**	**6.73**	**6.14**	**1.95**
	Std. Dev.	**8.07**	**5.14**	**4.13**	**0.62**	**1.23**	**2.95**	**1.22**	**8.82**	**5.74**	**1.74**
P value		**0.62**	**0.67**	**0.41**	**0.02**	**0.03**	**0.45**	**0.02**	**0.67**	**0.07**	**1.00**

Table 19: CD8+ T-cell cross-recognition of KK10 epitope forms in group 1 and 2 patients and comparison of differences between the two groups (Mann-Whitney test) for each epitope pair. Cross-recognition above 10% of the epitope specific T-cell response is highlighted in blue bold text.

3.2.5. Cross recognition of CD4+ T-cell responses differs from that of CD8+ T-cells.

The CD4+ T-cell responses seen to individual epitopes in these assays follow a different pattern to that seen in the CD8+ T-cell response, with all of the patients tested having their largest CD4+ T-cell recognition of the KL8 epitope form (Figure 19). This is followed in magnitude by the response to the KM14 epitope form (all patients), while responses to the KK10, KI11 and KR13 epitope forms are generally low or absent. The KL8 response in the CD8+ T-cell populations was much less cross reactive than for some of the other epitope forms, such as KI11 and this is also the case in the CD4+ T-cell response, with very limited cross recognition between responses to epitope forms even where there is individual recognition of that epitope form.

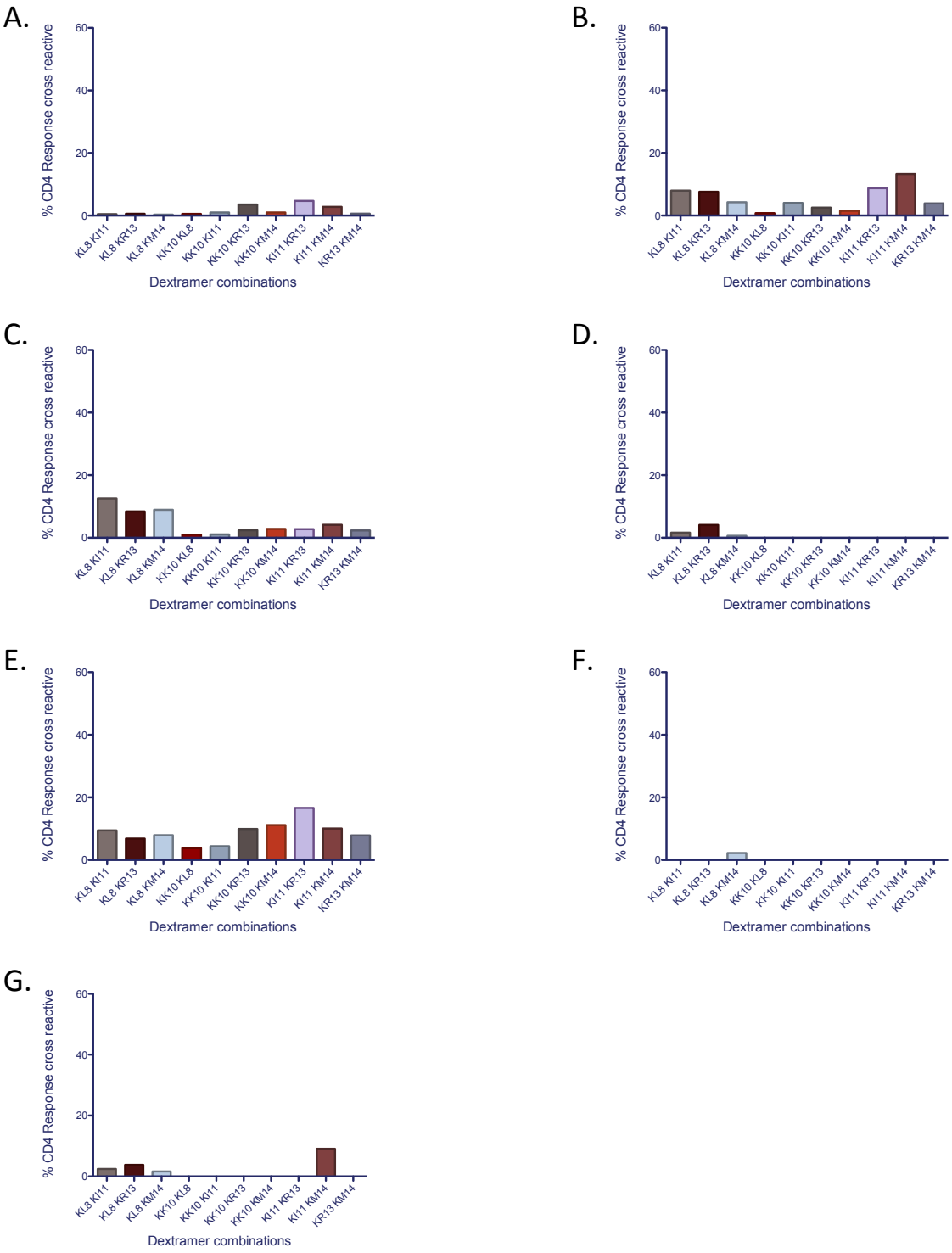


Figure 19: Cross reactive $CD4^{+}$ T-cell responses between pairs of KK10 containing epitope forms in patients A. 111156 B. 030778 C. 060473 D. 030869 E.170466 F. 050875 G. 081077

Of the small amount of cross recognition present in these responses, much of it is not between the KL8 and KM14 epitope forms but between these two responses and the KI11 or KR13 peptides, despite the low levels of response seen to the latter two forms. This is shown clearly as the patient with the highest levels of cross recognition of responses is also the patient with the highest level of individual KI11 response (170466)(Figure 16G, Figure 19E)(Table 20). In two patients (060473 and 170466) we do see a slightly higher level of cross recognition of the KL8 and KM14 epitope forms (8.92% and 7.92% respectively) that is limited or absent in the other patients. Both patients also displayed considerable cross-recognition of the KL8 and KM14 epitope forms by their CD8+ T-cells (9.43% and 10.67% respectively)(Table 20). Patient 170466 also has the highest levels of CD8+ cross recognition for many epitope form pairs in the KK10 non-responder group.

Patient Group	Patient No.	KL8 KI11	KL8 KR13	KL8 KM14	KK10 KL8	KK10 KI11	KK10 KR13	KK10 KM14	KI11 KR13	KI11 KM14	KR13 KM14
1	060473	**12.57**	8.41	8.92	0.97	1.08	2.41	2.83	2.74	4.13	2.34
	030869	1.62	4.09	0.61	0.00	0.00	0.00	0.00	0.00	N/A	0.00
	081077	2.44	3.78	1.64	0.00	0.00	0.00	0.00	0.00	9.10	0.00
	Mean	**5.54**	**5.43**	**3.72**	**0.32**	**0.36**	**0.80**	**0.94**	**0.91**	**6.62**	**0.78**
	Std. Dev.	**6.10**	**2.59**	**4.53**	**0.56**	**0.62**	**1.39**	**1.63**	**1.58**	**3.51**	**1.35**
2	111156	0.47	0.59	0.31	0.52	0.99	3.54	0.95	4.70	2.78	0.60
	030778	7.96	7.60	4.27	0.78	4.04	2.53	1.53	8.70	**13.29**	3.89
	170466	9.47	6.82	7.92	3.79	4.41	9.91	**11.14**	**16.63**	**10.07**	7.84
	050875	0.00	0.00	2.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Mean	**4.48**	**3.75**	**3.67**	**1.27**	**2.36**	**4.00**	**3.41**	**7.51**	**6.54**	**3.08**
	Std. Dev.	**4.94**	**4.01**	**3.26**	**1.71**	**2.20**	**4.22**	**5.20**	**7.04**	**6.19**	**3.60**

Table 20: CD4-T cell cross-recognition of KK10 epitope forms and means for group 1 and 2 patients. Cross-recognition above 10% of the epitope specific T-cell response is highlighted in blue, bold text.

3.2.6. Changes in responses to KK10 containing epitope forms in longitudinally assessed patients.

For 4 patients (101074, 081077, 060473, 111156 all 2 time-points)(Table 14) we were able to obtain samples from more than one time-point. Samples from all patients tested were included in the cross sectional study (Chapters 3.2.1 - 3.2.5) however, samples from the same patients from later time-points are analysed separately here to avoid skewing of the cross-sectional data.

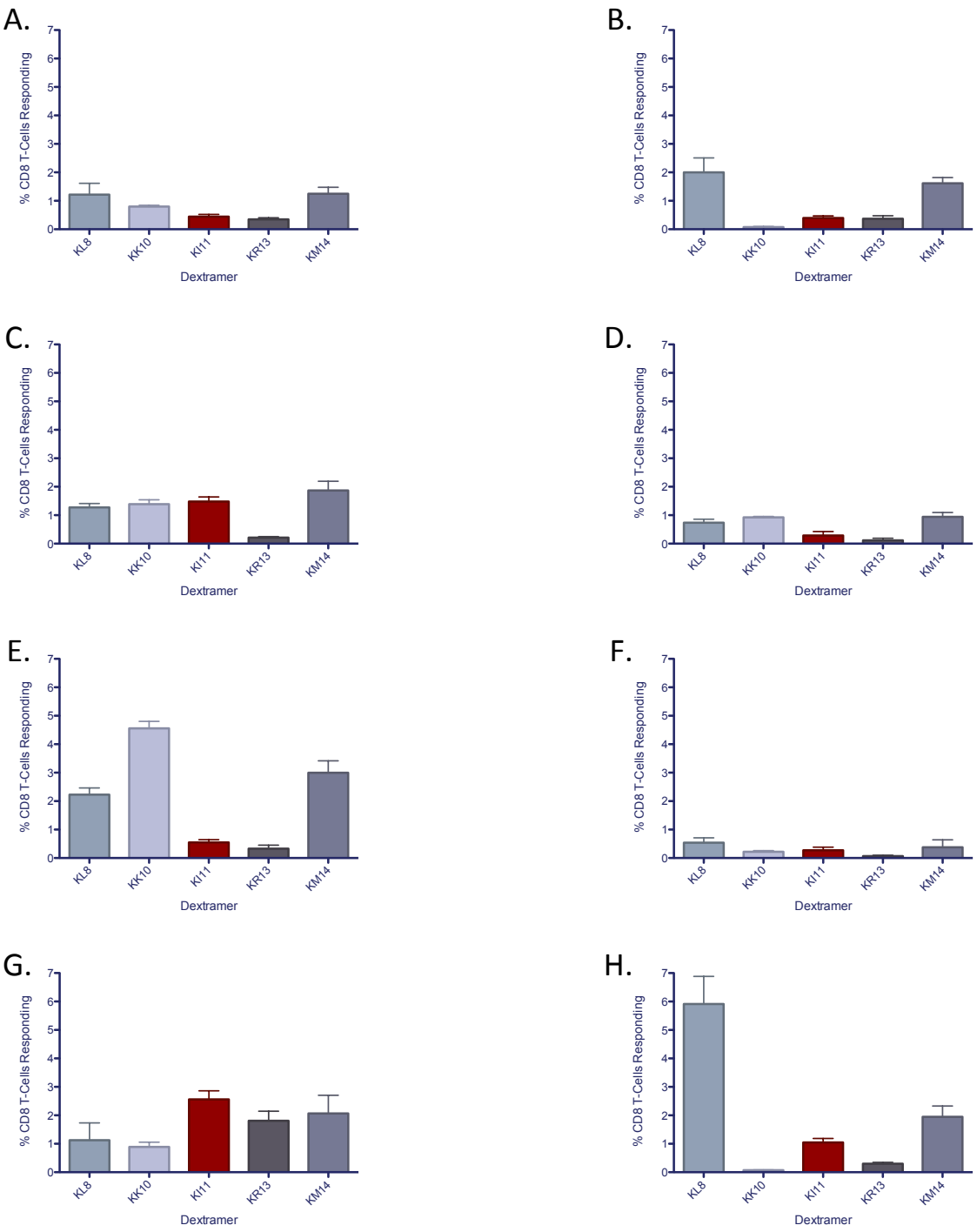


Figure 20: CD8+ T-cell responses (mean with standard error of the mean) to KK10 containing epitope forms in patients assessed at more than one time-point. A. Patient 111156 TP1 B. Patient 111156 TP2 C. Patient 081077 TP1 D. Patient 081077 TP2 E. Patient 060473 TP1 F. Patient 060473 TP2 G. 101074 TP1 and H. 101074 TP2

Patient number	Time-point 1					Time-point 2				
	KL8	KK10	KI11	KR13	KM14	KL8	KK10	KI11	KR13	KM14
111156	1.22	**0.80**	0.44	0.35	1.25	2.00	**0.08**	0.40	0.37	1.62
081077	1.27	1.39	**1.48**	0.21	1.87	0.74	0.92	**0.29**	0.12	0.94
060473	2.23	**4.56**	0.55	0.33	3.00	0.54	**0.22**	0.27	0.07	0.38
101074	1.12	**0.89**	2.56	1.81	2.07	5.92	**0.07**	1.05	0.30	1.95

Table 21: Mean CD8+ T-cell responses for KK10 epitope forms at time-points 1 and 2 for patients in the longitudinal study.

Of the 4 patients we have been able to obtain multiple time-points for, 3 are placed in group 1 in the cross sectional analysis (patients 101074, 081077 and 060473) and one in group 2 (111156). The data obtained at these time-points shows that the responses to these epitope forms can change over time (Figure 20)(Table 21). The time-points intervals are not uniform for each patient and range between months and several years (Table 14). An example of these changes is seen in the response to the KK10 epitope, at the second time-point that these patients were sampled at the KK10 response seen in the initial assay has either reduced in magnitude or is no longer present in 3 of 4 patients (111156, 060473 and 101074). Only one patient (081077) retained KK10 recognition across both time-points (Table 21). This allows us to see the responses to the KK10 epitope forms through the course of the disease in these patients and how they behave in individuals following the loss of recognition of the optimal epitope.

For patient 081077, who maintains a level of KK10 recognition (Table 21) at the second time-point, the response to individual epitope forms is slightly altered. The patient has a marked reduction in their recognition of the KI11 epitope form (Figure 20C and D). There are also slight reductions in the KL8 and KM14 populations (Table 21). The remaining two patients from group 1 (101074 and 060473) do not maintain KK10 recognition at the second time-point (Figure 20E, F, G and H)(Table 21) and recognition of other epitope forms also

differ. Patient 060473 loses recognition of the KL8 and KM14 epitope forms (Figure 20E and F), to which there was previously a robust response and no new responses to KK10 epitope forms are seen. Patient 101074 also experiences a reduction in their recognition of the KR13 and KI11 epitope forms, however KM14 recognition is maintained and the response to the KL8 form increases substantially (from 1.12%-5.92%). This is the strongest response to the KL8 epitope form seen in any of our patients. The second time-point for patient 101074 is only 3 months after the first, while over two years separate the 060473 time-points (Table 14), this difference in intervening periods may contribute to the very different changes in response seen in each patient.

The final patient for whom there is longitudinal data is patient 111156. This patient was placed in group 2 at the first time-point as the number of KK10 recognising CD8+ T-cells was >1% (Table 21) and did not form a clearly defined population. However, at time-point 2 a drop in the small number of KK10 positive cells was observed (Figure 20A and B) indicating that this initial KK10 positive CD8+ T-cells may have been a genuine response to this epitope form, perhaps remnants of an earlier response to this epitope. Apart from the change in KK10, there are increases in the level of KL8 and KM14 responses in the second time-point (Figure 20A and B), but the pattern of responses remains similar despite the two-year interval between these samples (Table 14).

Patient number	KL8 KI11	KL8 KR13	KL8 KM14	KK10 KL8	KK10 KI11	KK10 KR13	KK10 KM14	KI11 KR13	KI11 KM14	KR13 KM14
111156 TP1	0.00	0.00	0.74	1.41	0.62	0.88	1.13	1.70	1.43	1.95
111156 +3 m	1.25	0.00	1.67	0.00	0.00	7.14	0.00	0.00	0.84	0.00
081077 TP1	0.00	0.00	2.82	3.62	**57.58**	0.00	**3.08**	0.00	**31.55**	3.70
081077 +55 m	0.00	0.00	0.43	2.45	**0.82**	0.71	**42.42**	0.00	**0.66**	10.08
060473 TP1	**20.62**	**13.87**	9.43	5.36	2.69	2.26	**8.81**	**10.40**	3.44	**1.16**
060473 + 35 m	**0.00**	**0.00**	0.00	0.00	0.00	0.00	**0.00**	**0.00**	3.65	**11.73**
101074 TP1	**0.94**	1.42	**0.96**	3.74	N/A	5.53	9.75	**28.98**	16.53	5.75
101074 + 26 m	**13.66**	2.27	**10.13**	0.18	1.19	0.00	0.74	**9.20**	10.19	0.87

Table 22: CD8+ T-cell cross recognition between KK10 epitope forms for patients at time-points 1 and 2 in the longitudinal study

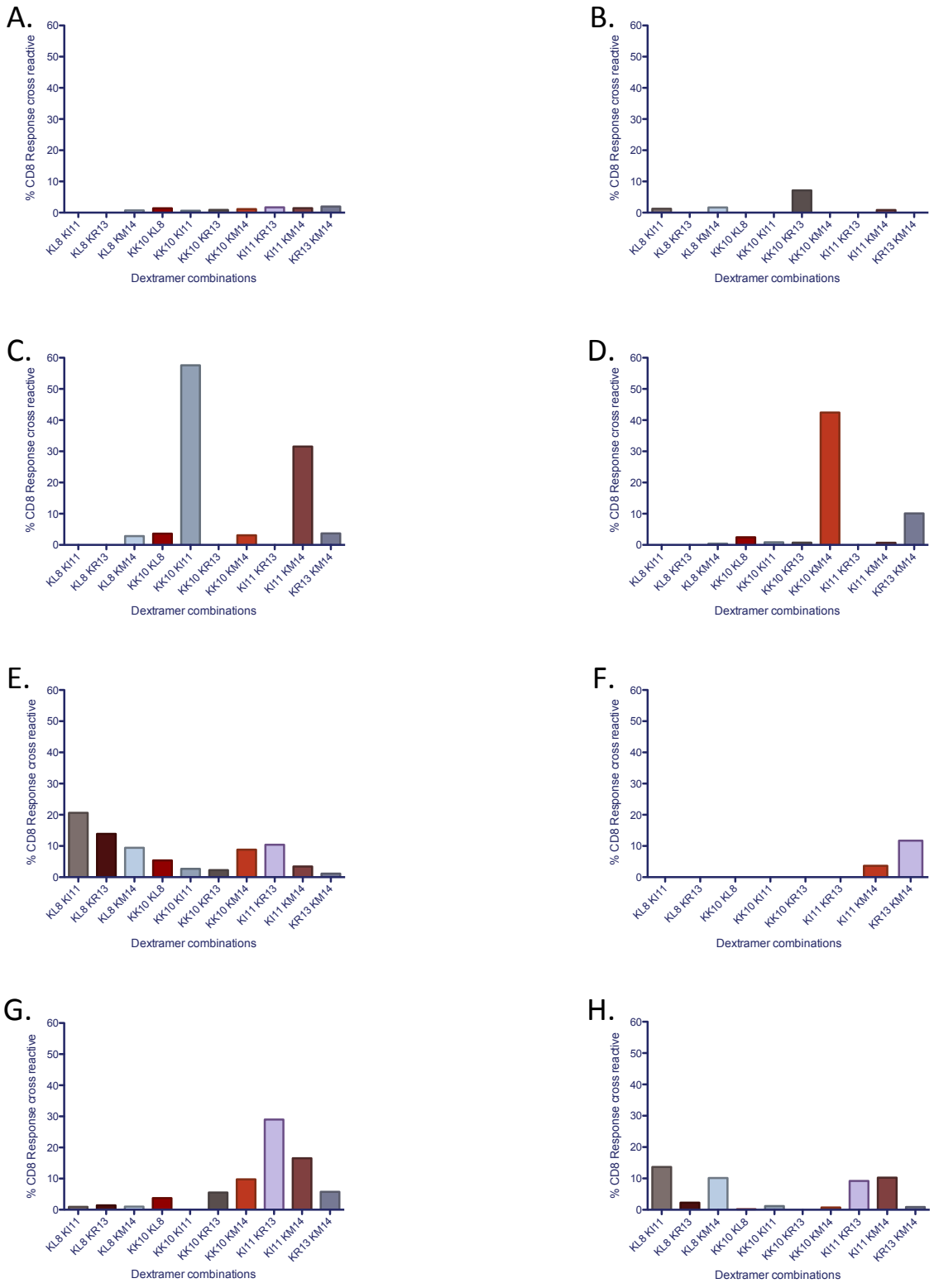


Figure 21: Cross reactive CD8+ T-cell responses to KK10 containing epitope forms as a percentage of CD8+ T-cells responding in patients assessed at more than one time-point. A. Patient 111156 TP1 B. Patient 111156 TP2 C. Patient 081077 TP1 D. Patient 081077 TP2 E. Patient 060473 TP1 F. Patient 060473 TP2.

In addition to changes in recognition of individual KK10 epitope forms, we also observed changes in cross recognition that occur between these patients over time. The three patients in group-1 for which we have this data show considerable changes in cross-recognition between the first and second time-points (Figure 21). Of these, patient 081077, who maintains recognition of KK10, nevertheless sees a dramatic shift in cross-recognition of the epitope forms (Figure 21C and D). Loss of cross-reactive responses occurs between the KK10 and KI11 epitope forms (57.6% at TP1 - 0.82% at TP2) and between the KI11 and KM14 epitope forms (31.5% at TP1 - 0.67% at TP2)(Table 22). This loss of cross-recognition is likely due to the dramatic reduction in the KI11 response seen at the second time-point in this patient (Figure 20D). In addition to the loss of the cross-reactive responses to KK10/KI11 and KM14/KI11, we also see development of a new cross reactive population in this patient between the KK10 and KM14 epitope forms (3.08% at TP1 - 42.4% at TP2)(Table 22). This may be due to the proliferation of a novel CD8+ T-cell recognising these epitope forms in the interval between the two time-points.

In patient 060473 the pattern of changes is quite different with a loss of almost all cross reactive responses that were present in the first time-point (Figure 21E and F). This reflects the loss of CD8+ T-cell recognition of the epitope forms that the patient recognised in the first time-point and the absence of any new recognition (Table 22). Patient 101074 also loses the KK10 response at the second time-point and this is seen in the cross recognition data, which shows a marked reduction in levels of cross reactivity with the KK10 epitope (Figure 21G and H). Patient 101074 does maintain both KI11 and KM14 recognition at the second time-point, but cross-recognition between these epitope forms is reduced (16.5%-10.2%)(Table 22). There is also a reduction in KI11-KR13 cross recognition (29.0%-9.20%) though this may reflect the decrease in KR13 recognition at the second time-point. The increase seen in the KL8 response at the second time-point is associated with an increase in cross recognition of this epitope form, this is particularly notable in the KL8-KI11 (0.94%-13.7%) and KL8-KM14 (0.96%-10.1%) populations. Finally, in patient 111156 there is little

change between cross recognition levels between the first and second time-points (Figure 21A and B)(Table 22), this agrees with the relatively small changes in the individual CD8+ responses to the epitope forms in the two time-points (Figure 20A and B).

3.3. Cell sorting of TCR sequences and population clonotyping

In order to better characterise the mechanisms behind the cross-recognition of the TCR recognising multiple KK10 epitope forms. Dextramer positive populations of CD8+ T-cells were sorted from patient PBMC and their TCR α and β chains sequenced. These sequences were then analysed in the context of their cross recognition patterns, structures of the HLA-KK10 epitope form complexes and current knowledge of TCR cross recognition.

3.3.1. TCR specificity and cross recognition between KK10 epitope forms

3.3.1.1. CDR3β loop sequences in TCR recognising KK10 epitope forms

Of the patients assessed by dextramer staining, we have so far been able to sort and sequence dextramer-positive populations for four patients (060473, 100967, 030869 and 24124) to some of the epitope forms that were recognised. Limitations of both time and sample material have prevented the completion of a larger study, however, we aim to extend this during future work. For patients 060473 and 100967 we were only able to obtain a response to the KK10 epitope form, and in both cases, colony PCR sequencing of a 96 well plate showed the presence of only one sequence in the KK10 recognising population.

For patients 030869 and 241242 we were able to obtain a number of CD8+ T-cell populations recognising epitope forms (some responses present in these patients were not

obtained due to limitations of the number of samples available or inability to recover cDNA of sufficient quality). These populations reveal more about the nature of cross recognition in this region for these patients and of the requirements for cross recognition of KK10 epitope forms.

We were able to see 4 TCRβ sequences in the T-cell population recognising KK10 in patient 030869 (Table 23). Two of these sequences made up 89% of the sequences obtained for this epitope form. The most prevalent sequence was only found in this population ($CDR3_{\beta}$ = CASSQGVRAHEQFF, 56%) and did not appear to be cross-reactive. The second most prevalent (CATSTGGYEQYF, 33%) was seen in both this population and that of T-cells recognising the KI11 epitope form (where it made up 21% of the sequences), indicating cross-recognition of the two epitope forms by this TCR. The two further epitope forms made up the remaining 11% of sequences and both appeared to be specific only for the KK10 epitope form.

The population recognising the KI11 epitope form was also made up of 4 TCRs in this patient, one (CATSTGGYEQYF) being the cross-reactive TCRβ seen in the KK10 population and the further 3 being unique to KI11. One of these 3, CASSERTGELFF was the most prevalent sequence in the population, making up 62% of sequences. In this patient, the KI11 recognising population was isolated from a KI11/KM14 dextramer stain due to a lack of sufficient sample to stain for individual dextramers in addition to cross reactive responses, therefore, the KI11/KM14 TCRβ sequence (CATSEGGTDTEAFF) is also likely to have been found in an individual dextramer stain for this population. This was the only TCRβ sequence found in the KI11/KM14 cross-recognising population (Table 23).

A population of CD8+ T-cells recognising KI11 and KR13 epitope forms was also isolated in a separate stain and the TCRβ sequences obtained. This population shows a single cross-reactive TCRβ sequence (CDR3β = CASSPGQFGAEAFF) that was not seen in any other

populations, including the KI11 population. There is a possibility that while recognising KI11, other, non cross-reactive clonotypes were more frequent in the single positive population and that a larger sample size would also show this TCR. Collection of a larger number of dextramer positive cells will be needed to ensure that a full picture of the TCR repertoire for these epitope forms is obtained.

The CD8+ T-cells population recognising KK10 in patient 241242 is also comprised of multiple TCR clonotypes. Uniquely in our cohort, this patient had two distinct KK10 recognising T-cell populations, one having a moderate mean fluorescent intensity and the other a much higher one. Furthermore, when assessed for cross recognition of other epitope forms, the less fluorescent population proved extremely cross reactive, making up the majority of the double positive populations in these stains (Figure 22). The more highly fluorescent population was seen to be much less cross-reactive in the double epitope form stains (Figure 22). This is reflected in the TCR sequences observed in these populations, with the $KK10_{high}$ T-cell population consisting of a single TCR sequence, while the $KK10_{low}$ population consisted of 4 TCRβ chains, all of which were found to cross react with other epitope forms. Low numbers of functional sequences obtained for this population from our sequencing mean that further TCRs may be present at lower frequencies and repeating sampling for the population would allow us to characterise it further. The most prevalent TCRβ chain (CASSRTAPDTEAFF, 57%) was found in the T-cell populations of all four epitope forms sequenced. The KK10/KI11 cross reactive population consists of 9 TCRβ chains, only one of which was found in the $KK10_{low}$ population, which may indicate that a larger number of sequences is needed to fully characterise the KK10 population. Of these forms, only one (CASSRTAPDTEAFF, 48% of sequences) was found in other epitope forms and was cross-reactive.

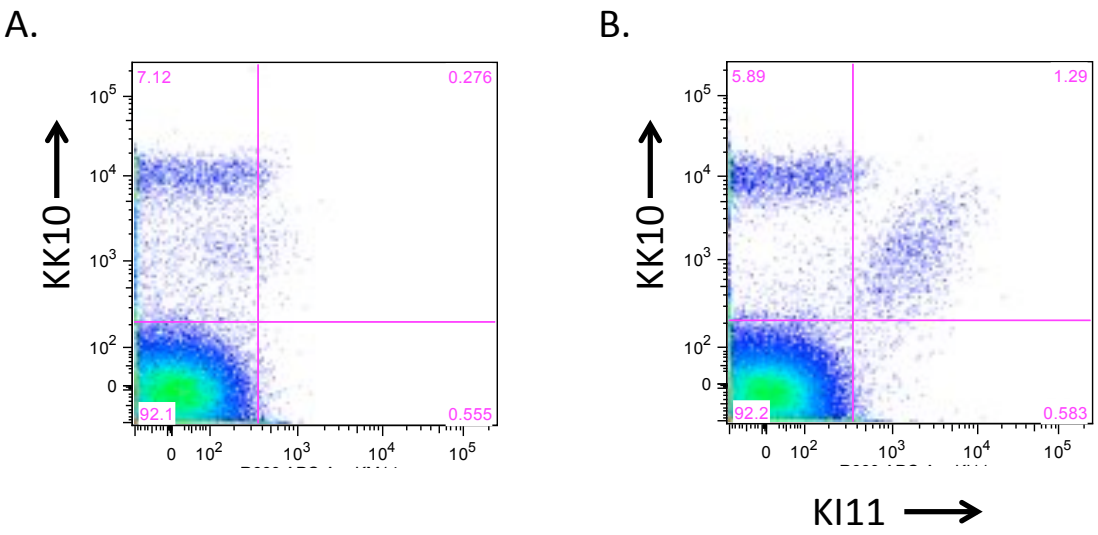


Figure 22: Plots showing the dextramer staining for A. the KK10 dextramer stain and B. KK10/KI11 double stain for patient 241242. Two KK10 positive populations can be seen for KK10, KK10-low and KK10-high. The KK10-low population is shifted in B. showing it's cross-recognition of the KI11 dextramer.

The T-cells recognising the KI11 population comprised 8 TCRβ chain, again having only the CASSRTAPDTEAFF (38%) TCRβ chain in common with the KK10/KI11 cross reactive population, the other TCRβ chain being found only in this population. In contrast, 3 of the TCRβ chain seen in the KI11 population were also found in KK10 populations and so were cross-reactive (Table 24) while the rest were seen only in the KI11 single population. That the cross-reactive sequences in the KI11 population were not found in the KK10/KI11 double positive population could be due to insufficient cells collected for complete sampling of the TCR repertoire. It could also be partly explained for some sequences by their relative affinities for each of the epitope forms. For example, the TCR containing the CDR3β CASSHLGGVTNEQFF sequence comprises 72% of the KK10/KR13 cross-reactive population, 35% of the KI11 population and only 14% of the KK10 population; this might reflect an increased affinity for the KR13 and KI11 epitope forms compared with KK10. Similarly TCRs present in the single KK10 or KI11 populations, though cross reactive, may display a preference for one epitope form, this may lower their prevalence in the cross reactive populations compared with that in the single stains of either epitope form and in some cases mean that we are not able to see these sequences in the cross-reactive populations.

The KK10/KR13 recognising population consisted of 4 TCRβ chains, 3 of which were seen in either KK10 single populations or in other epitope forms. The TCRβ chain which was not seen in another population was also the least prevalent in this case (CASSHLGGVSNAQFF, 1.5%) and the lack of a single KR13 population means we cannot determine whether this TCR would also have been present in that population. The KI11/KM14 cross-recognising population consists of the single TCRβ chain sequence CASSRTAPDTEAFF, which is cross-reactive to all 4 epitope forms tested in this patient (KK10, KI11, KR13, KM14) and is present in the single KM14 population. The KM14 population itself consists of 11 TCRβ chains, 5 of which are seen in other populations and therefore are able to cross-recognise other epitope forms. This includes the sequence seen in the KI11/KM14 population (CASSRTAPDTEAFF) as well as a sequence only found in the KM14 and KK10 populations (CAISDPPGTGEETQYF), however, it also includes 3 more sequences which were all found in the KI11 recognising population. That these sequences are not seen in the KI11/KM14 cross-reactive population implies we have an incomplete TCR profile of this recognition.

Patient	Epitope form	V_β CDR3 sequence	V_β	J_β	D_β	Frequency
060473	KK10	CASSLDMNEQYF	7-9*01	2-7*01	1*01	1.000
100967	KK10	CASSQTSGSSYEQYF	6-5*01	2-7*01	2*01	1.000
030869	KK10	CASSQGVRAHEQFF	4-3*01	2-1*01	2*02	0.564
		CASSLGTSAYEQYF	4-3*01	2-7*01	2*01	0.050
		CASSPILAIEQYF	7-2*01	2-7*01	2*01	0.063
		CATSTGGYEQYF	27*01	2-7*01	1*01	0.325
	KI11*	CASSERTGELFF	27*01	2-2*01	Unknown	0.621
		CASSERTGGLFF	27*01	2-2*01	2*01	0.035
		CASSLLGGTGELFF	7-2*01	2-2*01	2*01	0.172
		CATSTGGYEQYF	27*01	2-7*01	1*01	0.207
	KI11/KR13	CASSPGQFGAEAFF	4-3*01	1-1*01	1*01	1.000
	KI11/KM14*	CATSEGTDTEAFF	24-1*01	1-1*01	1*01	1.000

Table 23: TCR CDR3β loop sequences determined for KK10 epitope form recognising TCR from patients 100967, 060473 and 030869 along with frequency in the population and V_β, D_β and J_β, allele usage in each TCR.

Patient	Epitope form	V_β CDR3 sequence	V_β	J_β	D_β	Frequency
241242	KK10	CASSYSGTSGSPAYEQYF	6-2*01	2-7*01	2*02	1.000
	KK10 low	CASSYSGTSGSPAYEQYF	6-2*01	2-7*01	2*02	0.143
		CAISDPPGTGEETQYF	10-3*01	2-5*01	1*01	0.143
		CASSRTAPDTEAFF	4-3*01	1-1*01	1*01	0.571
		CASSHLGGVTNEQFF	4-1*01	2-1*01	1*01	0.143
	KK10/KI11	CSAPFGTSAKGQYF	20-1*01	2-7*01	2*01	0.090
		CASSRTAPDTEAFF	4-3*01	1-1*01	1*01	0.478
		CASSGLAGGALYQETQYF	7-2*01	2-5*01	2*02	0.269
		CASSSGTSGTLTDTQYF	2*01	2-3*01	2*02	0.030
		WASSLVGAGELFF	5-6*01	2-2*01	1*01	0.015
		CASRPGGETYEQYF	7-9*03	2-7*01	1*01	0.045
		CASSPGQGLYGYSF	18*01	1-2*01	1*01	0.045
		CASSYGVTYEQYF	6-5*01	2-7*01	1*01	0.015
		CSVEGGTSGLQETQYF	29-1*01	2-5*01	2*01	0.015
	KK10/KR13	CASSHLGGVTNEQFF	4-1*01	2-1*01	1*01	0.716
		CASSRTAPDTEAFF	4-3*01	1-1*01	1*01	0.075
		CASSFTSGSPENEQFF	7-9*03	2-1*01	2*02	0.194
		CTSSHLGGVSNAQFF	4-1*01	2-1*01	1*01	0.015
	KI11	CASSRTAPDTEAFF	4-3*01	1-1*01	1*01	0.378
		CASSSSGSGWTDTQYF	27*01	2-3*01	2*02	0.054
		CASSYSGTSGSPAYEQYF	6-2*01	2-7*01	2*02	0.054
		CASSFTSGSPENEQFF	7-9*03	2-1*01	2*02	0.054
		CASSLTRRSTDTQYF	13*01	2-3*01	2*02	0.027
		CASSHLGGVTNEQFF	4-1*01	2-1*01	1*01	0.351
		CASSLERPPGEKLFF	7-6*01	1-4*01	1*01	0.054
		CASSEAGTGTAPYEQYF	6-2*01	2-7*01	1*01	0.027
	KI11/KM14	CASSRTAPDTEAFF	4-3*01	1-1*01	1*01	1.000
	KM14	CASSRTAPDTEAFF	4-3*01	1-1*01	1*01	0.306
		CASTSSGDSSYEQYF	27*01	2-7*01	2*01	0.047
		CASRGGPEAFF	2*01	1-1*01	2*01	0.024
		CASSFTSGSPENEQFF	7-9*03	2-1*01	2*02	0.106
		CAISDPPGTGEETQYF	10-3*01	2-5*01	1*01	0.012
		CASSHLGGVTNEQFF	4-1*01	2-1*01	1*01	0.388
		CASSDRLAGGPNEQYF	6-5*01	2-7*01	2*02	0.024
		CATRTGLNTEAFF	24-1*01	1-1*01	1*01	0.012
		CASSILADGAYEQYF	4-1*01	2-7*01	2*01	0.024
		CASSLSIGTSGRDEQFF	27*01	2-1*01	2*02	0.012
		CASSYSGTSGSPAYEQYF	6-2*01	2-7*01	2*02	0.047

Table 24: TCR CDR3β loop sequences determined for KK10 epitope form recognising TCR from patient 241242 along with frequency in the population and V_β, D_β and J_β, allele usage in each TCR.

3.3.1.2. Features of CDR3β loop sequences involved in cross-recognition of KK10 epitope forms

TCR contact with peptide residues of the HLA-peptide complex upon binding is primarily through the interaction of the CDR3β loop of the TCR[272]. This loop often also shows the greatest divergence from germline sequences as it contains the intersection of the V, D and J segments, which make up a complete TCRβ chain. These sequence segments are modified at their joins by terminal deoxynucleotide transferase (TdT)[103; 104; 105] adding or removing a small number of nucleotides in order to create additional variation in the TCR. In typical TCR binding modes, other CDR loops contribute to recognition of the HLA complex mainly through binding of residues of the HLA rather than the peptide. However, it has been shown that for longer epitopes, which bulge out of HLA class-1 grooves, these CDR may also be involved in recognition of the peptide[383].

The sequences of the CDR3β loops responding to our epitope forms were sorted according to patient and the ability to cross-recognise KK10 epitope forms in patient 241242 (only one cross reactive TCR was found in another patient, CATSTGGYEQYF 030869)(Figure 23, Figure 24, Figure 25, Figure 26 and Figure 27). Factors that were used to characterise the sequences include presence/absence, relative proportions and location of polar, hydrophobic and charged residues in the sequences (Figure 23, Figure 24, Figure 25, Figure 26 and Figure 27) as well as length of the sequence, change in the CDR3β loop from the germline encoded sequence and number of epitope forms recognised (Table 25).

In the TCR sequences obtained from our four patients we found that differences in the CDR3β loops between cross-reactive and non cross-reactive TCRβ chain sequences were small. We did find that cross-reactive TCRβ chain CDR3 loops were slightly longer than non cross-reactive CDR3 loops (15.2 versus 14.5 residues), and that they had undergone fewer

deletions of residues encoded by the germline sequences (1.00 versus 1.83 residues). They were also slightly less likely to have a charged residue N-terminal of the common C-3/C-4 position (12/19 non cross reactive versus 4/11 cross reactive). Comparison of the sequences by epitope form recognised showed few differences between the sequences recognising different epitope forms. In part this may be due to the high degree of cross-recognition seen in the patient for which the most sequences were obtained. Differences in the percentage of TCRβ sequences being cross-reactive did differ for KM14 (50%) compared with the other two epitope forms for which we have single stained data sets (71%-75%).

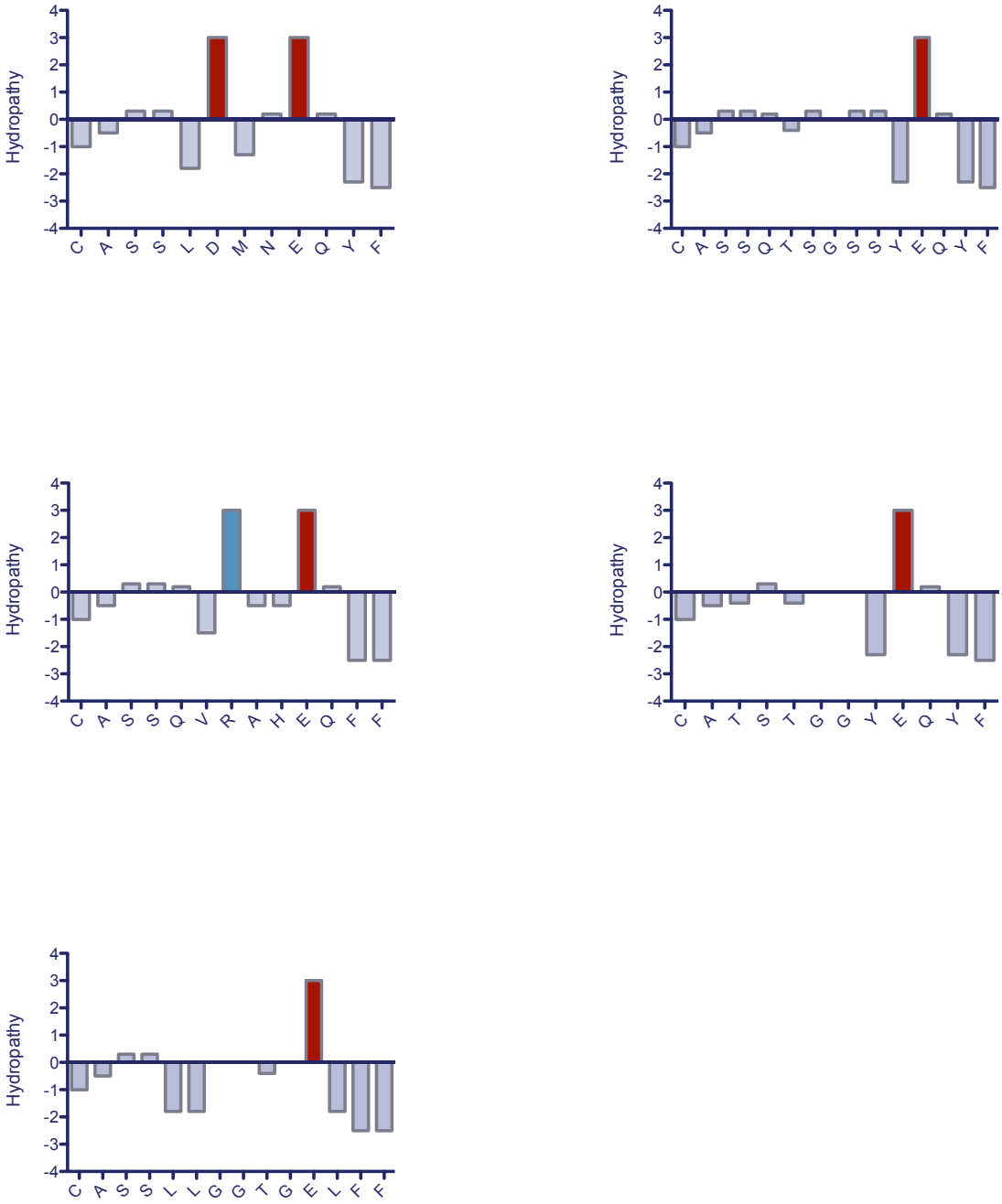


Figure 23: Hydrophobicity plots for TCR CDR3β loops taken from patients 060473 (top left), 100967 (top right) and 030869 (remainder). Hydrophobicity measured using the Hopp and Woods index of hydrophilicity[384], charged residues are shown in red (acidic) and blue (basic).

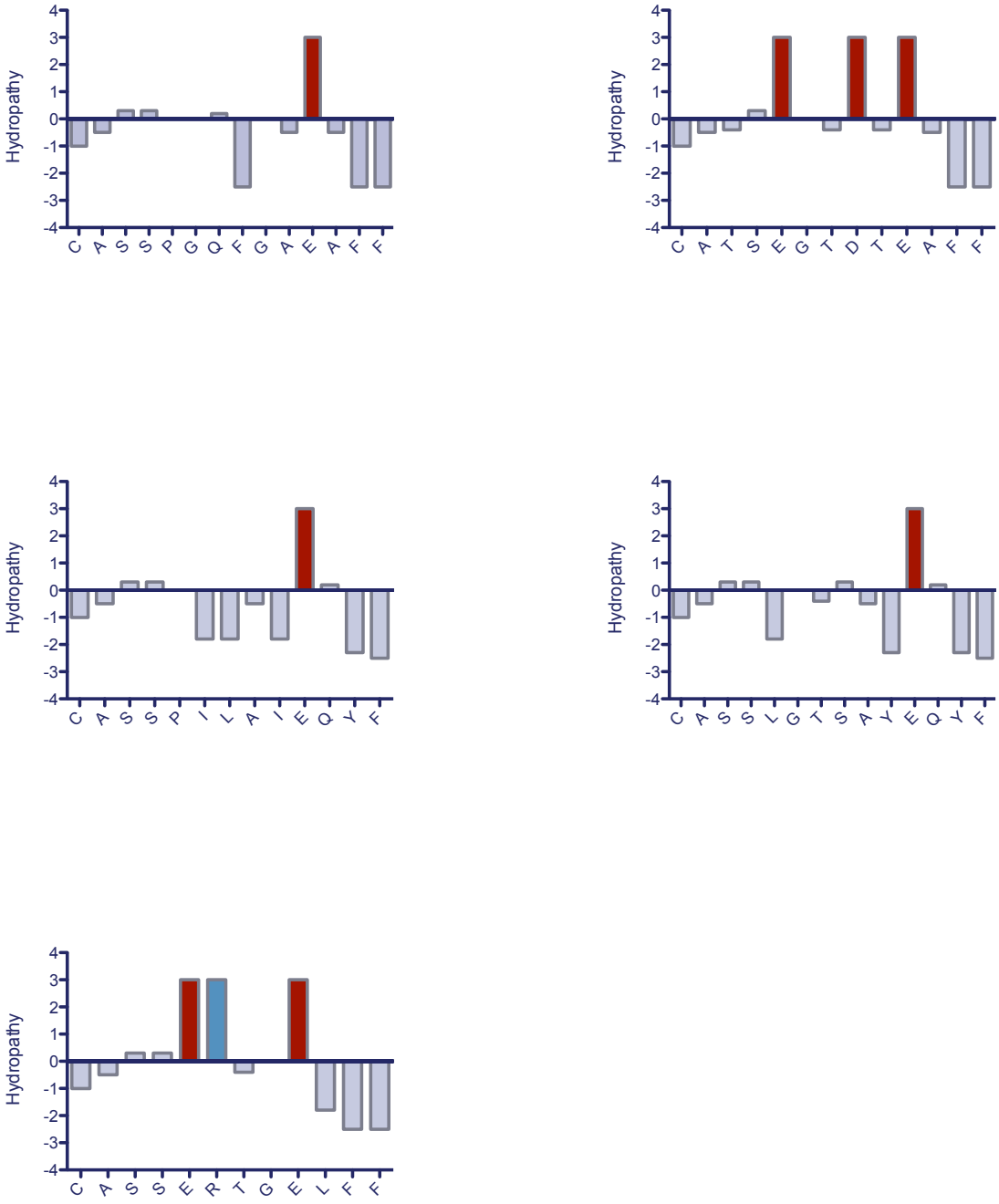


Figure 24: Hydrophobicity plots for TCR CDR3β loops taken from 030869 (plot 2). Hydrophobicity measured using the Hopp and Woods index of hydrophilicity[384], charged residues are shown in red (acidic) and blue (basic).

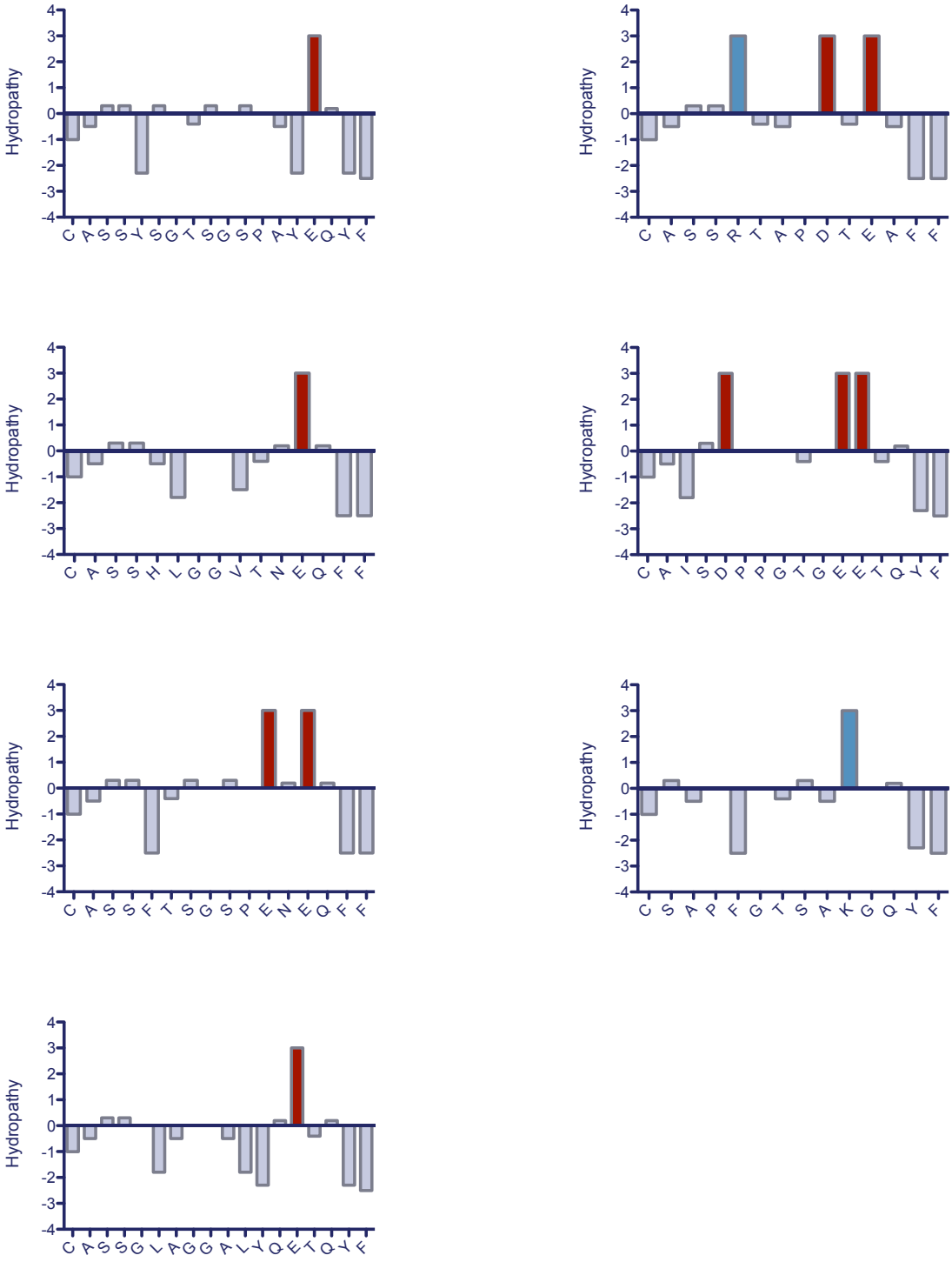


Figure 25: Hydrophobicity plots (1 of 2) for cross-reactive TCR CDR3β loops taken from patient 241242. Hydrophobicity measured using the Hopp and Woods index of hydrophilicity[384], charged residues are shown in red (acidic) and blue (basic).

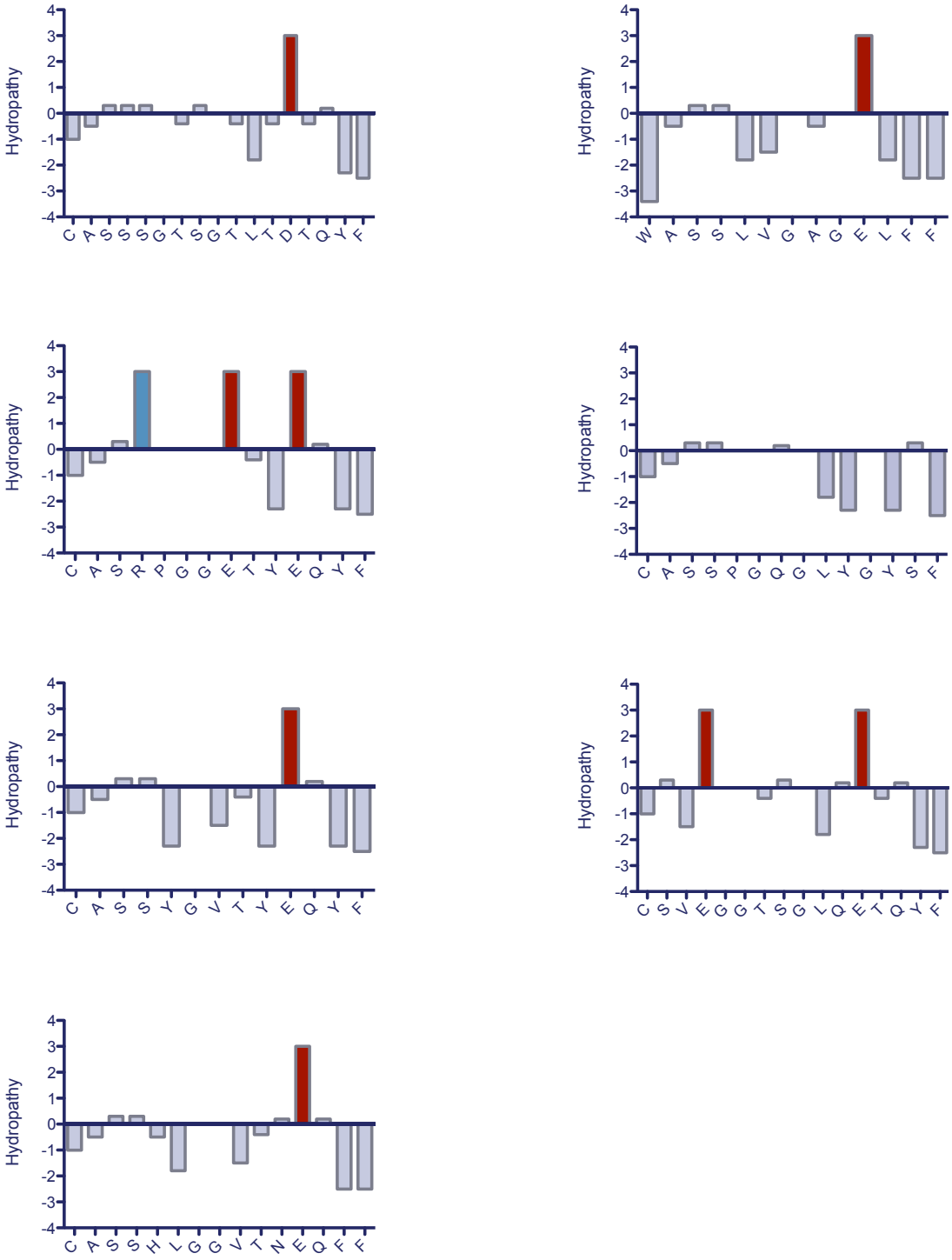


Figure 26: Hydrophobicity plots (2 of 2) for cross-reactive TCR CDR3β loops taken from patient 241242. Hydrophobicity measured using the Hopp and Woods index of hydrophilicity[384], charged residues are shown in red (acidic) and blue (basic).

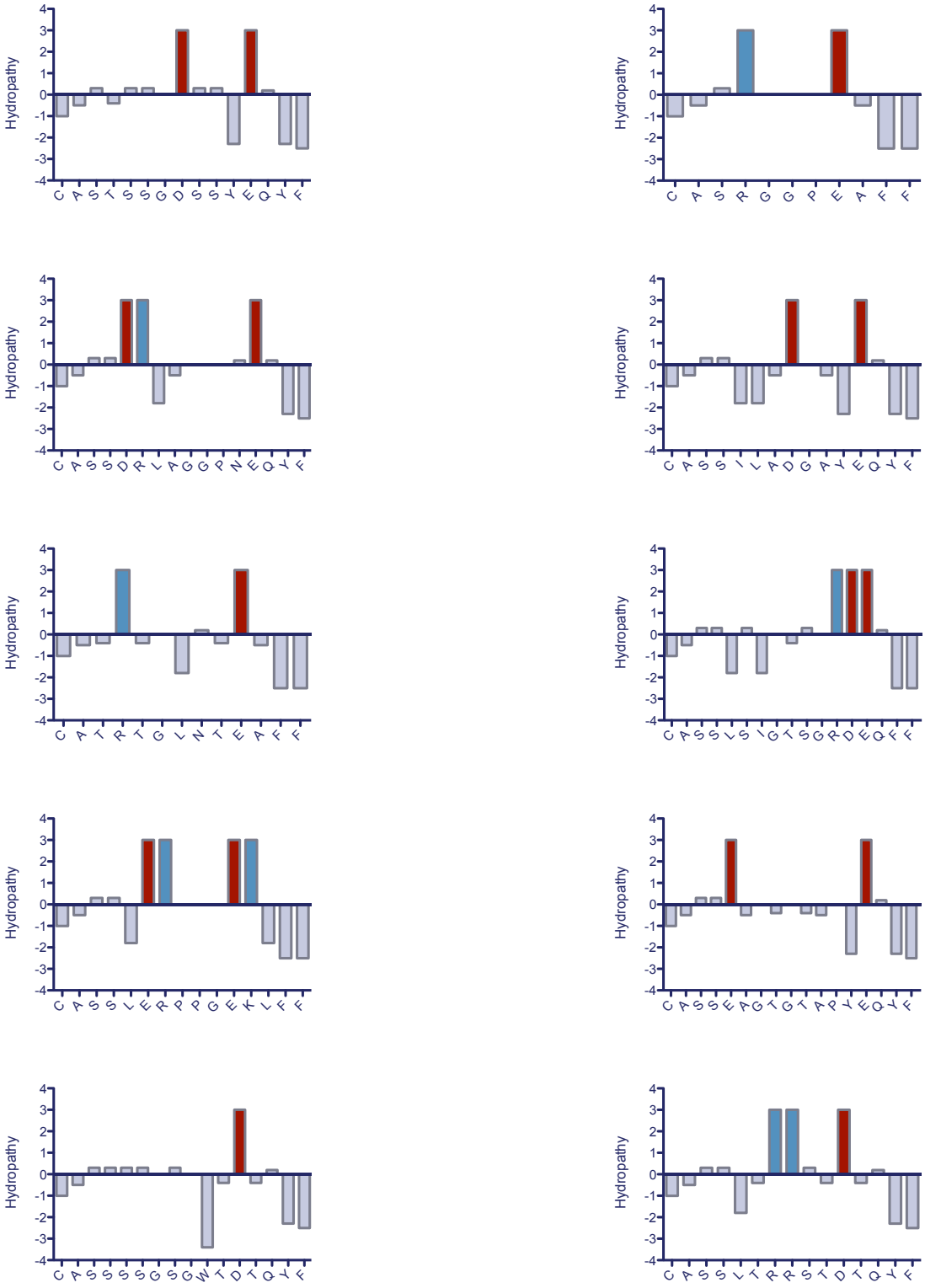


Figure 27: Hydrophobicity plots for non cross-reactive TCR CDR3β loops taken from patient 241242. Hydrophobicity measured using the Hopp and Woods index of hydrophilicity[384], charged residues are shown in red (acidic) and blue (basic).

CDR3	Germline sequence	Mutations	Inserts	Deletions	Total changes	Epitope forms recognised	CDR3 loop length
CASSLDMNEQYF	CASSLDRGSYEQYF	2	0	2	4	1	12
CASSQTSGSSYEQYF	CASSYGTSGGSYEQYF	2	0	1	3	1	15
CASSQGVRAHEQFF	CASSQGTSGRSYNEQFF	4	0	3	7	1	14
CASSLGTSAYEQYF	CASSQGTSGGSYEQYF	2	0	2	4	1	15
CASSPILAIEQYF	CASSLPPLVSYEQYF	4	0	2	6	1	13
CASSERTGELFF	CASSLDRGNTGELFF	2	0	3	5	1	12
CASSLLGGTGELFF	CASSLGTSGGNTGELFF	1	0	4	5	1	14
CATSTGGYEQYF	CASSLGTGGSYEQYF	1	0	3	4	2	12
CASSPGQFGAEAFF	CASSQGQGNTEAFF	4	0	0	4	1	14
CATSEGTDTEAFF	CATSDLGTGGNTEAFF	2	0	3	5	1	13
CASSYSGTSGSPAYEQYF	CASSYGTSGRSYEQYF	2	2	0	4	3	18
CAISDPPGTGEETQYF	CAISEPPVQETQYF	3	2	0	5	2	16
CASSRTAPDTEAFF	CASSQAPCPNTEAFF	4	0	1	5	4	14
CASSHLGGVTNEQFF	CASSQGTGGSYNEQFF	4	0	1	5	4	15
CSAPFGTSAKGQYF	CSARGTSGGSYEQYF	2	2	3	7	1	14
CASSGLAGGALYQETQYF	CASSLGLAGGQETQYF	0	3	1	4	1	18
CASSSGTSGTLTDTQYF	CASSEGTSGRSTDTQYF	3	0	0	3	1	17
WASSLVGAGELFF	CASSLGQGNTGELFF	2	1	3	6	1	13
CASRPGGETYEQYF	CASSLGQGSYEQYF	5	0	0	5	1	14
CASSPGQGLYGYSF	CASSPGQGNYGYTF	2	0	0	2	1	14
CASSYGVTYEQYF	CASSYGTGGSYEQYF	2	0	2	4	1	13
CSVEGGTSGLQETQYF	CSVEGTSGGQETQYF	1	1	0	2	1	16
CASSFTSGSPENEQFF	CASSLGTSGRSYNEQFF	4	0	1	5	4	16
CTSSHLGGVSNAQFF	CASSQGTGGSYNEQFF	6	0	1	7	1	15
CASSSSGSGWTDTQYF	CASSLGTSGRSTDTQYF	3	1	2	6	1	16
CASSLTRRSTDTQYF	CASSLGTSGRSTDTQYF	1	0	2	3	1	15
CASSLERPPGEKLFF	CASSLPPVTNEKLFF	1	2	2	5	1	15
CASSEAGTGTAPYEQYF	CASSYGTGGSYEQYF	3	2	0	5	1	17
CASTSSGDSSYEQYF	CASSLGTSGGSYEQYF	1	2	3	6	1	15
CASRGGPEAFF	CASSEGTSGGNTEAFF	2	0	5	7	1	11
CASSDRLAGGPNEQYF	CASSYGLAGGSYEQYF	4	0	0	4	1	16
CATRTGLNTEAFF	CATSDLGTGGNTEAFF	2	0	3	5	1	13
CASSILADGAYEQYF	CASSQGLAGGSYEQYF	3	0	1	4	1	15
CASSLSIGTSGRDEQFF	CASSLGTSGRSYNEQFF	1	2	2	5	1	17

Table 25: Germline V_β, D_β, and J_β sequences in the CDR3β loops of TCR recognising KK10 epitope forms. Number of changes between germline sequence and CDR3β loop found in final productive TCR shown along with CDR3 loop length and number of KK10 epitope forms recognised by the TCR.

3.3.1.3. CDR3α loop sequences in TCR recognising KK10 epitope forms

Sequences of the TCRα chains were obtained for patients 100967 and 241242, though for fewer populations than we were able to obtain TCRβ chain sequences for in this patient. Sequencing of the TCRα chain of the dextramer sorted TCR was performed using the same method as for the TCRβ chains. However, cDNA of sufficient quality was obtained for fewer populations from the mRNA isolated from the dextramer positive cells than was the case for the TCRβ chains. This could be due to differences in the efficiency of priming for the α-chain and/or due to the amplification for these chains being performed following that for the TCRβ chains, though all samples were adequately stored. Obtaining sequences for the remainder of these populations will be a priority of future work.

Nevertheless, we did obtain at least one population containing each of the KK10 extended epitope forms (although only the cross reactive forms in the case of KM14). For the KK10 population in patient 100967, there is again a single sequence for the TCRα chain, indicating this population consists of a single T-cell clone. Similarly, the KK10 high population in patient 241242 also consists of a single TCR sequence for both the α and β-chains, this is therefore another population consisting of a single clone. The remainder of the populations in patient 241242 all consist of multiple TCRα chains, as well as multiple TCRβ chains (Table 24 and Table 26), meaning that there are multiple clonotypes involved in these responses.

The KK10 low response consists of five TCRα sequences, one of which is the sequence also found in the KK10 high population (which may indicate imperfect separation of the two populations for sorting)(Table 26). Three further sequences found in this population are also present in populations for other epitope forms, indicating their ability to cross-recognise multiple KK10 epitope forms. For the TCRβ chains, we found that all of the KK10 low sequences were found in other populations, and were therefore cross reactive, the presence of a TCR sequence in the TCRα chain that is not found in any other populations may indicate

the presence of a non-cross reactive TCR not seen in the TCRβ populations (Table 26). However, more likely is that the more limited sequencing performed on the α-chain means that we have not obtained the populations that this TCR sequence is able to cross-recognise.

The KI11 population consists of seven TCR sequences, slightly fewer, but maintaining a similar pattern for the number of sequences recognizing each epitope form to the TCRβ chains. These sequences consist of both cross-recognising TCR and TCR unique to the KI11 form, as for the TCRβ chains. There are two sequences which cross-recognise the KI11 and KM14 epitope forms, one of which is found to also recognise KK10, and both of which are found in the KI11 single population. The TCRβ sequence for the same population consists entirely of a single highly cross-reactive TCR; this could mean that this TCRβ sequence is found paired to multiple TCRα chains. Finally the KR13 sequences are entirely composed of cross-recognising TCR, found in both the KK10 and KI11 populations, with one sequence additionally found in the KI11/KM14 population. This tallies with the TCRβ sequences for this epitope form, which were also dominated by widely cross-reactive TCR. An infrequent (1.5%) non-cross reactive TCR was also found in the TCRβ sequences recognizing the KR13 epitope form, that no similar sequence was picked up here may suggest that further sequencing is required to obtain a complete picture of TCRs recognition in this form.

Patient	Epitope form	V_α CDR3 sequence	V_α	J_α
100967	KK10	CAVSLLGDTGGFKTIF	1-1*01	9*01
241242	KK10	CAGTYPYSGGGADGLTF	35*02	45*01
	KK10 low	CAGTYPYSGGGADGLTF	35*02	45*01
		CAASPNNNDMRF	13-1*01	43*01
		CALGVPNDYKLSF	19*01	20*01
		CAGGSNYQLIW	1-2*01	33*01
		CAYRVNTGTASKLTF	38-2/DV8*01	44*01
	Ki11	CAGGSNYQLIW	1-2*01	33*01
		CAASPNNNDMRF	13-1*01	43*01
		CAMREGDTGGFKTIF	14/DV4*02	9*01
		CAGEAAGNKLTF	1-2*01	17*01
		CALGVPNDYKLSF	19*01	20*01
		CAASNDYKLSF	8-6*02	20*01
		CAVRDSPSGGSYIPTF	1-1*01	6*01
	KI11/KM14	CAASPNNNDMRF	13-1*01	43*01
		CAVRDSPSGGSYIPTF	1-1*01	6*01
	KR13	CALGVPNDYKLSF	19*01	20*01
		CAASPNNNDMRF	13-1*01	43*01

Table 26: TCR CDR3α loop sequences determined for KK10 epitope form recognising TCR from patients 100967 and 241242 along with frequency in the population and V_α, and J_α, allele usage in each TCR.

3.3.1.4. Features of CDR3α loop sequences involved in cross-recognition of KK10 epitope forms

We examined the features of the TCR sequences obtained for the α chain in the same way as for TCRβ (See Chapter 3.3.1.2), including hydrophobicity, the presence of charged residues (Figure 28), length and change from germline (Table 27). These characteristics were compared for sequences recognising each epitope form, as well as a comparison for those that cross-recognised multiple forms and those that were specific for one form.

As found for the TCRβ sequences, differences between the epitope forms for TCR sequence recognition was small, as were differences between cross-recognising and specific epitope forms. The KI11 and KR13 epitope forms were recognised by TCR with slightly shorter

average CDR3α loops than the KK10 and KM14 epitope forms (12.9 and 12.5, 14.0 and 14.0 respectively), the KK10 and KI11 forms also showed a higher average number of changes from the germline sequence than the KR13 and KM14 forms (2.50 and 2.43, 2 and 2 respectively). There was no change in the total number of mutations per sequence for the CDR3α loop residues from germline between the cross-reactive and non cross-reactive TCR, there was however a slight difference in the average number of residues in the CDR3α loops of the two groups (13 residues in cross-recognising TCR versus 14.3 in non cross-recognising TCR)(Table 27). There seemed to be little difference in the hydrophobicity or presence of charged residues between these two groups, with variation in each, however the relatively small sample size and limitation primarily to a single patient limits the scope of these findings.

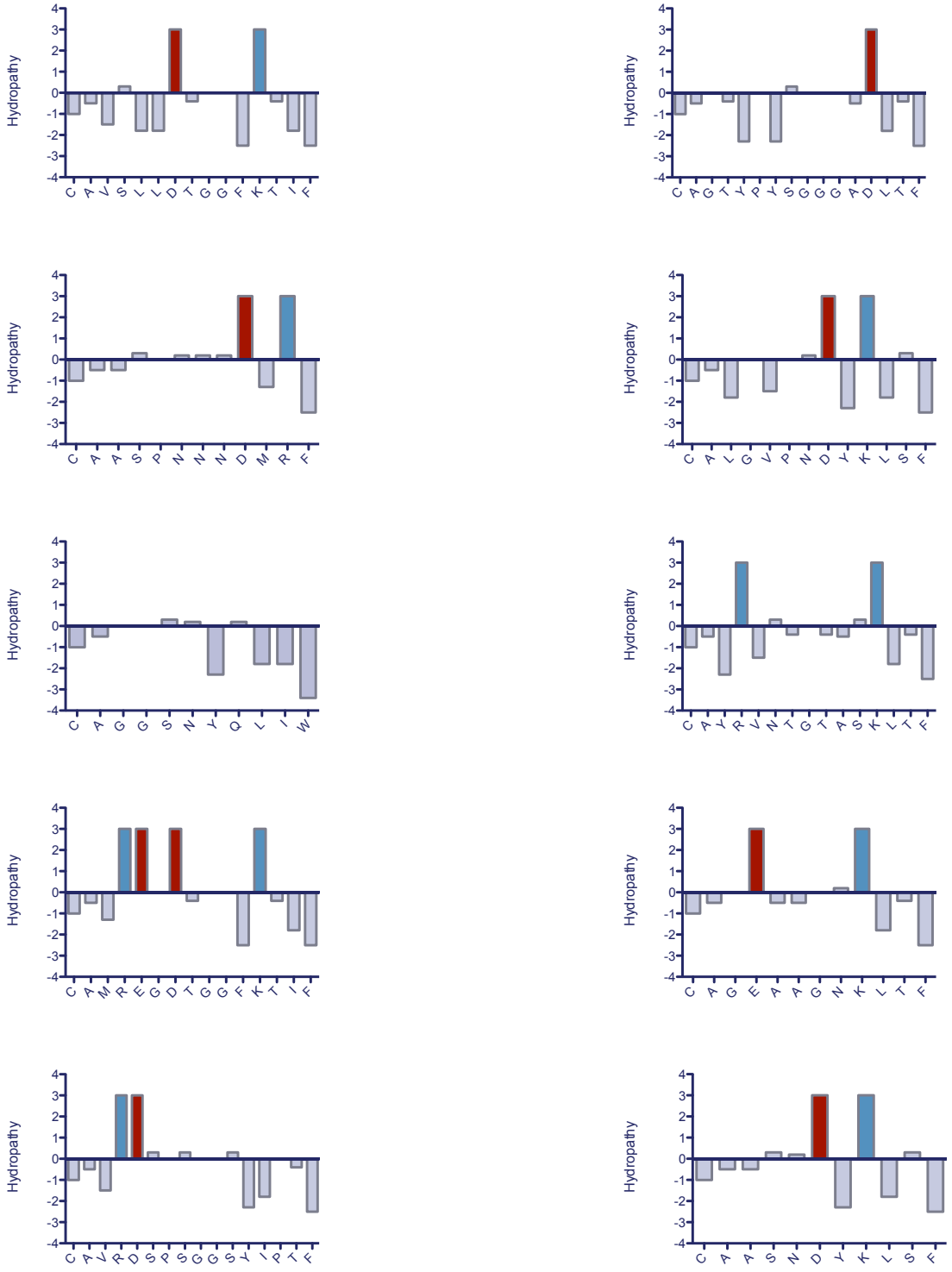


Figure 28: Hydrophobicity plots for TCR CDR3α loops taken from patients 100967 (top left) and 241242 (remainder). Hydrophobicity measured using the Hopp and Woods index of hydrophilicity [384], charged residues are shown in red (acidic) and blue (basic).

CDR3	Germline sequence	Mutations	Inserts	Deletions	Total changes	Epitope forms recognised	CDR3 loop length
CAVSLLGDTGGFKTIF	CAVRGNTGGFKTIF	2	2	0	4	1	16
CAGTYPYSGGGADGLTF	CAGQYSGGGADGLTF	1	2	0	3	1	17
CAASPNNNDMRF	CAASNNNDMRF	0	1	0	1	4	12
CALGVPNDYKLSF	CALSESNDYKLSF	2	1	0	3	3	13
CAGGSNYQLIW	CAVRDSNYQLIW	2	1	0	3	2	11
CAVRDSPSGGSYIPTF	CAVRASGGSYIPTF	1	2	0	3	2	16
CAYRVNTGTASKLTF	CAYRSNTGTASKLTF	1	0	0	1	1	15
CAMREGDTGGFKTIF	CAMREGNTGGFKTIF	1	0	0	1	1	15
CAGEAAGNKLTF	CAVRIKAAGNKLTF	2	2	0	4	1	12
CAASNDYKLSF	CAVSSNDYKLSF	1	1	0	2	1	11

Table 27: Germline V_α and J_α sequences in the CDR3α loops of TCR recognising KK10 epitope forms. Number of changes between germline sequence and CDR3α loop found in final productive TCR shown along with CDR3 loop length and number of KK10 epitope forms recognised by the TCR.

3.3.1.5. Enrichment and depletion of CDR3α and β loop amino acid residues for TCR recognising KK10 epitope forms in comparison with germline-encoded frequencies

Comparison of our patient TCR sequences with the sequences encoded by the V_β, J_β, D_β, V_α and J_α alleles in the germline reveal differences in the relative prevalence of residues at certain locations in both the TCRα and TCRβ CDR3 sequences that recognise KK10 epitope forms (Table 28). The greatest difference in prevalence between the patient and germline prevalence is the presence of a glutamic acid (E) residue at either position C-3 or C-4 of the CDR3β loop; occasionally this was substituted by an aspartic acid (D) at one of these positions. An E at position C-3 or C-4 is present in 28 of 34 sequences (82.4%) and a D is present in this position in a further 3 sequences (additive 91.2%). The difference in the number of E residues at C-3 compared with the proportion encoded in the germline (39.3%)(Table 28B) is significant, P=0.0137 Fisher's Exact Test (FET). This significance increases (P=0.0007 FET) if the comparison includes E at positions C-3 and C-4 as well as D-

4 against the number of germline sequences encoding these. Of all possible TRBJ alleles available, only 6 of 16 (37.5%) encode an E at these positions and 1 contains a D at C-4 (total 43.8%). The enrichment of these residues at this position of the CDR3β loops suggests a conserved interaction of the loop among the epitope forms; this could involve recognition of peptide or the HLA molecule by the CDR3β loop.

The second largest difference in prevalence seen between the patient sequences and the germline is the presence of asparagine (N) at the C-5 position of the J_β segment (Table 28). There is a significant decrease (P=0.0028 FET) in the presence of N at this position in the patient sequences compared with the germline. This position is quite close to the join of the J and D segments and so more likely to undergo insertion or deletion of nucleotides through the action of the non-homologous end joining machinery and terminal dideoxynuleotide transferase (TdT). This may increase the likelihood that these positions will differ from the germline encoded frequencies though significant deviations will nevertheless reveal selection of the TCR repertoire for the KK10 epitope forms.

There are several other differences that also reach statistical significance between the patient and germline V_β and J_β sequences. These are a decrease in Leucine (L) residues at position N4 (P=0.0157 FET) and an increase in glyceine residues at position N4 (P=0.0077 FET). Approaching statistical significance is the increase in prevalence of proline (P) residues at the same N4 position (P=0.0502 FET). That these significant differences from the germline sequences all occur at the final residue of the V_β segment suggests that the insertion/deletion of residues is likely to be responsible for these differences.

The central D_β sequence was not compared in a similar manner to the V_β and J_β residues in the CDR3 loop as this section can be inserted into the CDR3 sequence in both a forward orientation, through deletion, or a reverse orientation, through inversion. There is a strong bias (approximately 20:1) for this segment to be incorporated through deletion, so residues

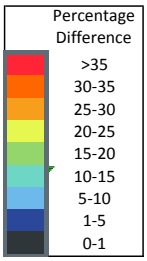
encoded by the forward and reverse sequences of the D_β segment in these positions cannot be given equal weight when comparing the possible germline sequences to those from our TCR[101; 102].

The same analyses were performed on the sequences obtained from the $CDR3_\alpha$ loops of TCRs responding to the KK10 epitope forms. Though the greatest percentage differences of the sequences from germline were lower overall than in the $CDR3_\beta$ loops, there were several differences. The two largest percentage differences in residue prevalence in the CDR3α loops were a decrease in valine (V) residues at position N2 (27.5%) in the patients and an increase in tyrosine (Y) residues at position C-4 (25.5%)(Table 28). However, neither of these differences were statistically significant, with P=0.1801 for the N2 position and P=0.0737 for the C-4 position. The lack of significance may be partly due to the small group size for the patient sequences, sequencing of a larger set of TCR sequences for this region will give a better estimate of the significance of these differences.

Despite being smaller in terms of percentage differences between the two groups, several other positions did give statistically significant differences between the patient and the germline sequences. These differences are an increase in serine residues at the C-1 position (P=0.0488 FET), an increase in the threonine residues at the C-2 position (P=0.0488 FET), an increase in the prevalence of cysteine residues at the C-5 position (P=0.0176 FET) and an increase in the prevalence of G residues at the N3 (P=0.0196 FET) and N2 positions (P=0.0353 FET). The differences from germline encoding at positions C-5 and perhaps to a slightly lesser extent N2 may be due to nucleotide insertion and deletion during rearrangement as these positions are nearer to the junction of the V_α and J_α chains. The differences seen in the C-1 and C-2 positions are more likely to be due to preferences in V_α and J_α chain selection for TCR recognising the KK10 epitope forms as these positions nearer the ends of the CDR3 region are less likely to be affected by the NHEJ machinery or TdT during rearrangement.

A.

	TRAV				TRAJ						
	0	1	2	3	-6	-5	-4	-3	-2	-1	0
A	0.00	14.1	9.90	0.00	0.32	8.71	-1.61	-6.45	0.00	-1.61	0.00
R	0.00	0.00	0.00	9.35	-3.23	-1.61	-8.06	-6.45	0.00	8.39	0.00
N	0.00	0.00	0.00	-8.70	17.1	2.26	0.65	-6.45	0.00	-1.61	0.00
D	0.00	0.00	0.00	-9.78	-1.61	-3.23	-9.68	8.39	0.00	0.00	0.00
C	0.00	0.00	0.00	0.00	0.00	20.0	10.0	0.00	0.00	0.00	-1.61
Q	0.00	0.00	0.00	-6.52	-4.84	-1.61	-9.68	1.94	0.00	-3.23	0.00
E	0.00	0.00	-3.03	-0.87	0.00	0.00	-3.23	0.00	0.00	-1.61	0.00
G	0.00	-4.04	23.9	15.7	9.35	2.58	-14.5	8.39	0.00	0.00	0.00
H	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-1.61	0.00
I	0.00	-3.03	0.00	0.00	-3.23	-1.61	0.00	8.39	-3.23	-2.26	0.00
L	0.00	-2.02	-6.16	0.00	0.00	-1.61	0.00	0.00	-11.0	-1.61	0.00
K	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.94	0.00	0.00	0.00
M	0.00	0.00	3.94	-4.35	0.00	0.00	0.00	-1.61	8.39	-1.61	0.00
F	0.00	0.00	-6.06	0.00	-1.61	-4.84	16.8	0.00	-12.9	-1.61	-5.16
P	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-1.61	8.39	0.00	0.00
S	0.00	0.00	0.00	-2.61	-11.0	3.55	-1.29	-1.61	-1.61	18.4	0.00
T	0.00	0.00	-4.04	10.00	3.55	-8.06	-4.84	-3.23	18.4	4.52	0.00
W	0.00	0.00	0.00	0.00	0.00	-3.23	0.00	0.00	0.00	0.00	6.77
Y	0.00	0.00	8.99	0.00	-3.23	-11.3	25.5	0.00	-1.61	-4.84	0.00
V	0.00	-5.05	-27.5	-2.17	-1.61	0.00	0.00	-1.61	-4.84	-9.68	0.00



B.

	TRBV					TRBJ						
	0	1	2	3	4	-6	-5	-4	-3	-2	-1	0
A	0.00	6.80	-5.27	-0.75	0.00	-6.15	14.3	-3.31	-3.31	8.46	0.00	0.00
R	-1.49	0.00	-1.49	2.11	-1.48	-3.21	2.94	0.00	0.00	0.00	-6.25	0.00
N	0.00	0.00	-0.75	0.00	0.00	-9.09	-34.6	-10.3	0.00	0.00	0.00	0.00
D	0.00	0.00	0.00	0.00	2.34	2.94	5.88	5.51	0.00	0.00	0.00	0.00
C	-0.70	0.00	0.00	0.00	-0.88	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Q	0.00	0.00	0.00	0.00	-8.28	0.00	-0.37	-6.25	0.00	26.8	0.00	0.00
E	0.00	0.00	0.00	-0.04	-0.03	2.94	5.88	-0.74	39.3	0.00	0.00	0.00
G	0.00	0.00	0.00	-2.99	8.82	17.4	-0.74	2.57	-0.37	-6.25	0.00	0.00
H	0.00	0.00	0.00	0.00	5.88	0.00	0.00	2.94	0.00	0.00	-18.8	0.00
I	0.00	0.00	1.45	0.00	0.29	0.00	0.00	2.94	-6.25	-6.25	0.00	0.00
L	0.00	0.00	0.00	0.00	-23.7	11.8	2.94	0.00	0.00	-19.5	0.00	-6.25
K	0.00	0.00	0.00	0.00	-1.77	0.00	-6.25	2.94	-3.31	0.00	0.00	0.00
M	0.00	0.00	0.00	0.00	0.00	0.00	2.94	0.00	0.00	0.00	0.00	0.00
F	0.00	0.00	0.00	0.00	5.88	2.94	0.00	0.00	0.00	0.00	16.2	12.5
P	0.00	0.00	0.00	2.94	9.11	11.8	5.88	2.94	-18.8	0.00	0.00	0.00
S	0.00	-6.06	5.49	-1.23	4.40	-21.7	-3.68	-12.5	0.00	0.00	2.94	0.00
T	0.00	-0.75	4.35	-0.04	5.88	-6.15	11.0	2.57	-1.10	0.00	-12.5	0.00
W	2.94	0.00	-3.73	0.00	0.00	2.94	0.00	0.00	0.00	0.00	0.00	0.00
Y	-0.75	0.00	0.00	0.00	-3.85	-15.2	-6.25	10.7	0.00	-3.31	18.4	0.00
V	0.00	0.00	-0.04	0.00	-2.65	8.82	0.00	0.00	-6.25	0.00	0.00	-6.25

Table 28: Percentage change in frequency of amino acid residues along the CDR3α and β loop V and J regions in KK10 epitope form recognising TCR compared with germline encoded frequency at each position.

3.4. Discussion

3.4.1. The presence of CD8+ responses to KK10 containing epitope forms indicates a broader response to the p24 Gag region *in vivo* than has previously been characterized

Our antigen processing studies of the region containing the KK10 epitope showed that in addition to the KK10 epitope itself, there are a series of overlapping epitope forms containing all or part of the KK10 sequence. The functional analysis of PBMC from HLA B*2705 HIV-1 infected individuals using dextramers for these epitope forms shows that as well as the response made to the KK10 epitope in the majority of HLA B*2705 HIV-1 infected individuals, the overlapping epitope forms may also be recognised by CD8+ T-cells in these patients (Table 15). We have additionally shown that while there are some differences between the recognition of these epitope forms dependent on the presence of KK10 recognition in the patient (KI11 recognition, Table 16), the responses can be generated in patients with and without recognition of the KK10 epitope itself (Figure 13, Figure 14). These epitope forms may form an important part of the immune response to the KK10 containing region in HLA B*2705 individuals, it has previously been found that T-cell responding to KK10 possessed more clonotypes and a more rapid turnover of dominant T-cell clones than for responses to non-protective HLA-epitope complexes[309]. The numerous epitope forms may account for some of this additional diversity in the TCR repertoire for the region if they are able to cross-recognise the KK10 epitope, our examination of this cross-recognition (described below) suggests this is indeed possible.

3.4.2. The presence of CD4+ responses to HLA class 1-peptide complexes

The presence of epitope specific responses to HLA class 1-peptide complexes from CD4+ T-cells was unexpected given previous data on the specificity of the CD8+ and CD4+ co-receptor binding for HLA class 1 and class 2 respectively. However, a number of possibilities suggest themselves to explain this phenomenon. Firstly, the nature of dextramers may present some challenges to normal binding of the co-receptor molecules. The close spacing of up to 10 HLA-peptide complexes on the dextran backbone may mean that the binding sites on the sides of the HLA molecules for the CD4+ and CD8+ co-receptors are sterically hindered, indeed an ILT2 binding assay that is used to check correct folding of HLA class 1 tetramers does not work when applied to dextramers (Figure S 8). This assay assesses refolding by the ability of ILT2 to bind to HLA class 1 and is a common method for testing newly made tetramers. However, on applying the assay to our commercially obtained dextramers, we saw no binding. It has previously been shown that ILT2 competes with CD8+ for binding to HLA class 1 and the ILT2 binding site and CD8+ binding site both involve contact with the membrane proximal α3 domain of HLA class 1 (from the structure of the CD8αα- HLA 1 complex and CD8αβ model based on Fab complex structure)[385; 386; 387; 388]. It is therefore possible that a lack of ILT2 binding will also mean that co receptors contacting membrane proximal (or in the case of dextramers, dextran proximal) regions will not be able to bind, resulting in a relaxation of specificity for T-cell receptors of a particular HLA class.

Despite the lack of ILT2 binding on dextramers compared with tetramers, initial tetramer assays with these epitope forms also indicated the presence of CD4+ T-cell responses to this region suggesting that this may not be the only (or even most likely) explanation for these responses. A more complete analysis using tetramers would confirm the presence of CD4+ T-cell responses in a system where TCR co-receptor binding has been firmly established[389]; alternatively, ability of dextramers to bind the CD8-co-receptor could be established in a similar way. Other explanations for these responses are informed by the structural data

obtained for these epitope forms complexed with HLA B*2705. These structures show that both the KL8 and KM14 epitopes likely exist for the majority of the time in conformations bound to the HLA only at the N-terminus. This means that much of the peptide is highly surface exposed and flexible and also that the C-terminus is unoccupied by the peptide (see Chapter 4.7). This presents us with at least two possible explanations for the presence of CD4+ T-cell restricted responses. Firstly, the exposed and flexible nature of these peptide epitopes mean that any TCR will be able to contact a large number of peptide residues compared with peptides bound in the groove along their entire length. The binding of a TCR to a long class 1 epitope bulging out from the centre of the groove has previously been shown to increase the proportion of the TCR residues contacting the peptide rather than the HLA molecule as it takes up more of the area on the binding surface, this results in a more antibody-antigen like interaction with less dependence on contact with conserved HLA residues[281; 390]. A TCR specific for these epitopes is therefore likely to be more dependent on peptide interactions than usual, decreasing dependence on HLA recognition. Indeed, some peptide epitopes have previously been shown to be recognised by both CD8+ and CD4+ T-cells (Astrid Iversen, unpublished data).

The use of intracellular cytokine staining (ICS) in these assays to determine the responses means that we cannot be certain whether HLA class 1 or 2 or both were used to present these epitopes, though cross recognition of these dextramers from CD4+ T-cell TCR raised against these epitopes in a class 2 complex would depend to a certain extent upon the structural similarities between the class 1 and class 2 complexes with these epitopes. A groove that can accommodate much longer epitopes in HLA class 2 may mean that similar peptide exposure would not occur when these peptides were complexed with class 2 and the resulting reactive TCR may look significantly different. However, it has been demonstrated that HLA class 2 reactive TCRs may bind peptides with quite distinct binding conformations based on small areas of conserved structural features[372] and that peptides can be presented in very similar conformations by HLA class-1 alleles and allow cross-recognition by T-cells

despite having substantial differences in polymorphic residues within their binding grooves[391]. Indeed, the mirroring of the cross-recognition patterns between the CD8+ and CD4+ responses suggests that however they are being recognised, it is likely to be through a similar or the same pathway as for the CD8+ T-cells (Figure 17, Figure 18, Figure 19).

A second possible explanation indicated by the structural data is the exposed C-terminus of the groove in both of these epitopes. This is a highly charged region and may provide a non-specific, "sticky" site for TCR binding in these two epitopes. However there is an indication that this region of the peptide-binding groove may be occupied by free amino acids (such as arginine in the refolding process) that would neutralise this charge to some extent. Additionally, though the strongest and most consistent responses are to the 8mer and 14mer dextramer complexes, there are occasional responses to other epitope forms that do not share either an N-terminal only binding mode or a highly exposed peptide (patient 170466 Figure 16G), the comparative lack of cross recognition between the CD4+ T-cell responses also makes this scenario less likely as we would expect higher levels of cross reactivity for such non-specific "sticky" responses, not lower. Cross checking of these assays with other binding or activation assays such as surface plasmon resonance (SPR) or ICS would provide dextramer independent validation of these results. Furthermore, identification, sorting and TCR sequencing of both CD8+ and CD4+ T-cells responding to these epitopes will allow us to better assess how these CD4+ responses are elicited, co-crystallisation of the TCR along with the peptide HLA complex would further allow us to see how these TCR might recognise these epitope forms.

3.4.3. Cross-recognition of KK10 epitope forms by CD8+ and CD4+ T-cells

Cross recognition between the T-cells recognising the various dextramer forms was assessed in order to better determine how they might be generated *in vivo*, and how they contribute to immune responses to the region. Cross recognition between CD8+ T-cells recognising KK10 epitope forms was particularly interesting. We were able to find both cross –recognising and singly-reactive T-cell populations for the majority of the extended KK10 epitope forms (Table 19), though responses to the KR13 epitope form when present were largely through cross recognition, confirmed in one case through the sequencing of these T-cell populations (Table 24). There were also significant differences in the levels of cross-recognition for some epitope forms between patients possessing a KK10 response and those that did not (KI11, Table 18), though this may have been mostly due to the loss of the KK10 response for which most of the cross recognising T-cell populations were seen (Table 19). Despite many of the cross-recognising T-cells showing recognition of the KK10 epitope, all but two patients also showed high levels of cross-recognition (above 10%) between two other epitope forms (Table 19, patients 050875, 111156). This was further revealed upon sequencing of the TCR sequences in patient 241242 which showed multiple cross-recognising sequences, most of these recognised KK10 but in addition, many recognised more than one KK10 extended epitope form (Table 15). A TCRα sequence was also obtained which recognised the KI11 and KM14 epitope forms, but was not present in the KK10 population (Table 27). Of all of the epitope forms, the KL8 truncated form had fewest CD8+ T-cells that were able to cross-recognise it than other epitope forms in patients recognising KK10 (Table 18), though the level of cross-recognition was higher in those patients without KK10 recognition.

The cross-recognition of the KK10 epitope and the escape mutations that occur within the epitope in later disease by CD8+ T-cells has been shown to be important for the effectiveness of CD8+ T-cell responses for control of viremia[308]. This is notable in the context of the results presented here and in the previously mentioned finding that T-cell responses to this region

involved multiple clonotypes and rapid turnover of clones[309]. The recognition of multiple KK10 epitope forms by CD8+ T-cells, and their ability to cross-recognise extended KK10 epitope forms and the KK10 epitope in the patients tested, gives a possible mechanism whereby T-cell cross reactivity may be generated to the HLA B*2705 restricted KK10 response. This increased diversity of TCR and the promiscuity of their binding amongst the epitope forms may increase the likelihood that a T-cell receptor will be present that is also able to cross-recognise the L136M mutation in the KK10 epitope sequence. This is supported by the finding that cross-reactive TCR may more heavily rely on Van-der-Waals interactions than those recognising single epitope-HLA complexes[288], these interactions are often more flexible than hydrogen bonding and salt-bridges and so may help to reduce the specificity of these TCR. The correlation of the ability to cross-recognise escape variants with control of viremia may also be due to the presence of these epitope forms, providing additional targets to raise effective T-cell responses and generate multiple clonotypes specific for the region. Indeed, patient 241242, for whom we were able to see the most diverse recognition of epitope forms and high levels of cross-reactivity began treatment 18 years following initial diagnosis.

In contrast to the cross-recognition seen between many of the KK10 epitope forms in the CD8+ T-cell responses, CD4+ T-cell recognition of the KK10 epitope forms involved far less cross recognition. This is perhaps unusual given that the binding of KK10 to HLA class-2 is quite promiscuous, being part of a supermotif that can be presented by multiple HLA types. The region also includes epitopes restricted by the HLA DRB1*13 molecule. The lack of cross-recognition may be explained by the unusual context of these epitopes for CD4+ T-cells, as they are complexed with HLA class-1. TCR have been described that recognise both class 1 and class 2 HLA, and undergo significant changes to their structure to maintain interaction in each case[125], if these CD4+ T-cell responses are generated by responses to these epitope forms in a class-2 context, recognition in class-1 may involve significant

structural change that could limit the range of peptide epitopes that the TCR is able to bind in this manner.

3.4.4. TCR Sequences show multiple clonotypes and sequence bias for recognition of KK10 epitope forms

The sequences of T-cell receptors were obtained for KK10 epitope form dextramer positive CD8+ T-cells in patients 100967, 060473, 030869 and 241242 (Table 23, Table 24, Table 26 and Table 27). For patients 100967 and 060473 we were only able to sequence the KK10 dextramer positive population, which in both cases was found to consist of a single T-cell receptor (Table 23, Table 26). In patients 030869 and 241242 we were able to obtain and sequence T-cell populations for a range of KK10 epitope forms (Table 24, Table 26). The latter two patients showed a range of TCR, some specific for particular epitope forms, while others were able to cross-recognise one or more additional forms. Several TCRβ sequences obtained from patient 241242 were present in all of the T-cell populations sequenced (Table 24).

The TCR recognising KK10 in patient 241242 consisted entirely of sequences that cross recognised at least one other KK10 epitope form, and often recognised several forms, further novel TCR sequences were obtained from double positive dextramer populations that were not seen in some of the populations responding to the single epitope form. This implies that despite detecting far more unique TCR sequences for this patient than for any other, we did not capture the entire TCR repertoire in all of the populations sequenced. This diversity and cross recognition contrasts with the sequences obtained in dextramer positive populations for patient 030869, while a range of TCR were discovered for these populations, only a single sequence was found responding to more than one KK10 epitope form. The majority of the sequences were instead unique in each response. As only four patients have been sequenced

thus far, and only two of these extensively, it is not possible to draw conclusions on how the range and cross-recognition of these T-cell populations affects the course of disease. It should also be noted that higher cell numbers in the samples for patient 241242 allowed larger populations to be collected than for patient 030869. This may be a contributing factor to the increased number of TCR clonotypes that we were able to detect in this individual.

However, it is interesting to note that of the four patients, and the two who were more comprehensively sequenced the patient with the greatest number of TCR specific for this region and most cross-recognising sequences, 241242 has been a long term non-progressor, with 18 years between diagnosis and first treatment (Table 14). This compares to the relatively rapid need for treatment of patient 030869 (within 3 years of diagnosis). For the two patient that we were only able to obtain KK10 populations for sequencing, dextramer staining for both revealed a wide range of CD8+ T cell recognition of KK10 epitope forms and cross-recognition of responses (Table 15). Of these two, patient 100967 had been infected for over 11 years without treatment at the time of sampling (Table 14), while patient 060473 required treatment four years following diagnosis.

We examined the sequences isolated from the TCR identified in our KK10 epitope form positive populations for any features that might indicate selection for recognition of HLA B*2705 and the KK10 epitope forms given that they share a significant proportion of their sequences and that many of the sequences were able to recognise multiple KK10 epitope forms. This analysis (Table 28) showed that there were differences from the frequencies encoded in germline segments. These differences help to reveal selection of specific features of the repertoire that may have been selected for their ability to recognise these epitope forms. However, we must also consider that there is a bias in gene segment usage dependent on their position in the locus, in addition, the small number of patients that these sequences were obtained from means that conclusions drawn from these results might not be representative of larger populations.

The enrichment seen for the acidic residues E and D at the C-3 and C-4 positions of these epitope forms does suggest a common site of recognition among these TCR, which is explored further with the addition of structural analyses (See Chapter 5.2) along with the changes at other positions seen in this analysis (Table 28). There is also a trend towards difference in the CDR3β loops of TCR between those that are cross reactive and those that are not. The increased number of sequences with Leucine at position N+4 in the non-cross reactive TCR suggests that either this residue is disfavoured, or that other residues are preferable at this position for cross-recognition. CASSL is a commonly encoded (44.3%) V_β sequence at the beginning of the CDR3 loop, analysis of the germline sequences of the cross-recognising TCR show that, while lower than germline frequency, the presence of CASSL in the germline sequences selected for cross-recognising TCR is not significantly lower than for the non cross-recognising TCR (29.4%, difference P value = 0.148, FET). This agrees with analysis of the germline encoded CDR1 and 2 loops which show no significant differences between cross-recognising and non cross-recognising TCR, indicating that rearrangement of the CDR3β loop is the likely mechanism behind this reduction.

4. Crystal Structures of the HLA B*2705 in complex with KK10 naturally occurring epitope forms.

Work in this chapter was a collaboration with Dr. Maria Harkiolaki who performed data collection and initial model building for all crystal structures, as well as providing extensive knowledge, expertise and assistance throughout their production.

4.1. Introduction

4.1.1. HLA Class 1 peptide binding and presentation *in vivo*

The contribution of specific pockets of the HLA class-1 binding-groove in peptide selection varies between HLAs and this is reflected in the repertoire of peptides seen to bind; in some positions, there is a clear selection for a restricted range of amino acids. Places where strict peptide sequence restrictions are commonly seen are known as "anchor positions", because the peptides are required to have specific "anchoring" residues in these positions in order to bind[207; 323; 392]. A commonly seen anchor position is at the C-terminus, with almost all HLA class-1 binding peptides showing some restriction in this position[393]. This correlates with binding to the F-pocket and it is often the primary anchor position in HLA class-1.

In HLA B*2705 in specific, (Figure 29) the F-pocket (Figure 29F) is a strong determinant of the epitope repertoire, restricting the range of amino acids able to bind as described above. However, there is another position in the B*2705 binding groove that binds an even more restricted range of amino acids and so is instrumental in determining the epitope repertoire. This anchor is determined by the B-pocket of the groove (position 2 of the peptide), the position is very tightly constrained in B*2705 and almost all peptides that bind to this allele will have an arginine present at position 2[322; 323] as this allows the formation of a hydrogen bond network between the peptide and E45, T24 and H9 of the α1 helix (Figure 30A). This therefore forms the primary anchor in HLA B*2705, indeed polymorphism at several of the residues that form the B pocket were recently identified as significantly affecting host control of HIV-1 viremia[67] presumably through determining the sequence of and conformation in which peptides can bind to the HLA class-1 alleles. The specificity of this position for arginine is illustrated particularly well by the KK10 epitope itself, the epitope is derived from a conserved region of p24 Gag in HIV-1 and mutations therein often have a significant fitness cost. However, one mutation that is commonly seen in HLA B*2705

individuals with this virus results in the substitution of an arginine at position 2 of the KK10 epitope to a lysine[322]. This is a conservative mutation giving rise to another basic residue with a long side chain, however, it can significantly abrogate binding to the HLA B*2705 B pocket as it disrupts the formation of a hydrogen bond network of the P2 arginine side chain with residues E45, T24 and H9[322 256].

The length of the peptides that HLA-class 1 binds is usually given between 8 and 10, or sometimes 8-12 amino acid residues as suggested by space availability within the class-1 groove, extended epitope conformations and range of epitopes identified through optimal epitope determination or peptide elution[322; 380]. However, some peptide elution experiments and structural studies have suggested that longer epitopes may form a significant part of the HLA class-1 repertoire[383], reviewed in[394], and some have suggested that HLA B*2705, in particular, may be able to accommodate particularly long peptides[395]. Indeed, several well known HLA class-1 epitopes are outside of the 'traditional' length range, including the HLA B*57-restricted KF11 and the HLA B*3501-restricted alt. M-CSF (14mer) epitopes[396].

Historically, an initial structural study of HLA B*2705 purified from an LG-2 B cell line[322] found that this HLA showed a strong preference for nonameric peptides, with the composite peptide density giving a clear 9mer structure with defined P2 and P9 residue specificity. However, the observed preference for nonameric peptides in that work need not be representative of an innate HLA-driven preference. Indeed, one of the most studied HLA B*2705 epitopes in the context of HIV-1 is the immunodominant KK10 10mer[314]. The range of peptides available to the HLA in an LG-2 B cell line would be strongly influenced by the peptide repertoire present as well as the presence/absence of intact antigen processing machinery in the cell line. Furthermore, longer peptide epitopes might not have been conductive to crystallisation and so were preferentially excluded during crystal formation.

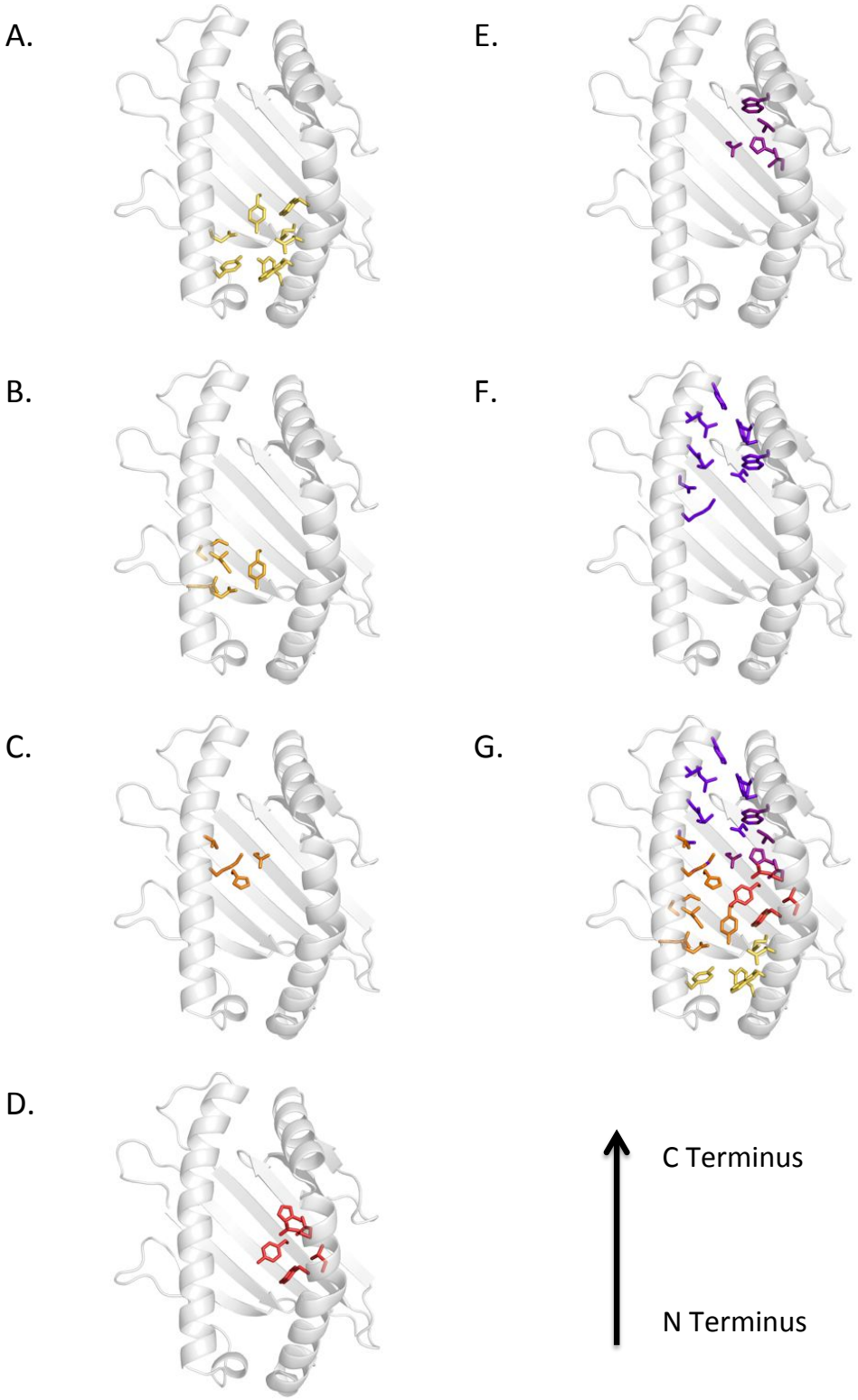


Figure 29: Adapted from[262]. The residues forming the peptide binding pockets of the HLA B*2705 binding groove. A. The A pocket, B. The B pocket, C. The C pocket, D. The D pocket, E. The E pocket, F. The F pocket and G. All pockets, where a residue takes part in formation of more than one pocket, colours are shown alternating

4.2. General Features of the HLA B*2705 structure

HLA B*2705 was crystallised in complex with naturally occurring KK10 epitope forms identified from the antigen processing of the p24 Gag 25mer (Chapter 2). To make these, purified, soluble HLA B*2705 was refolded in complex with selected epitope forms (Table 29). Screens were set up separately and crystals were produced under similar, though not always identical conditions for each of the epitope forms (Table 29). Often several high quality crystals were produced for a given epitope form (Table 29). All were of one of two space groups, $P2_12_12_1$ or $P2_1$, and the crystals of the 6 and 7mer complexes were twinned. The structures contained one (KL8, KI11, KR13, KM14) or two (KL6, KG7) HLA-B2705 peptide complexes per asymmetric unit with all diffracting to between 1.5 and 2.5Å (Table 29). The electron density maps show that while the KI11 peptide complex shows a single clearly defined peptide within the HLA binding groove, several of our structures may have areas of flexibility within the peptide, or in some cases, dual conformations.

Table 29: Crystallization conditions, data collection and refinement for KK10 epitope form structures.

	KM14	KR13	KI11	KL8	KG7	KL6
Crystallization						
Protein	15.0mg/ml in 20mM Tris pH 8.0 100mM NaCl	13.8mg/ml in 20mM Tris pH 8.0 100mM NaCl	13.4mg/ml in 20mM Tris pH 8.0 100mM NaCl	11.3mg/ml in 20mM Tris pH 8.0 100mM NaCl	12.8mg/ml in 20mM Tris pH 8.0 100mM NaCl	10.7mg/ml in 20mM Tris pH 8.0 100mM NaCl
Reservoir	20 %w/v Polyethylene Glycol 8000, 0.05 M Potassium di-Hydrogen Phosphate	20 %w/v Polyethylene Glycol 3350, 0.2 M Potassium di-Hydrogen Phosphate	20 %w/v Polyethylene Glycol 3350, 0.2 M Ammonium Chloride	20 %w/v Polyethylene Glycol 3350, 0.2 M Potassium di-Hydrogen Phosphate	30 %w/v Polyethylene Glycol Monomethyl Ether 5000, 0.2 M Ammonium Sulphate, 0.1 M MES pH 6.5	30 %w/v Polyethylene Glycol Monomethyl Ether 5000, 0.2 M Ammonium Sulphate, 0.1 M MES pH 6.5
Data collection						
Date	02/04/12	25/09/10	17/07/10	24/01/11	17/02/12	10/05/12
Beamline	Diamond I03	Diamond I24	ESRF ID23-2	Diamond I02	Diamond I03	Diamond I03
Resolution (Å)[a]	109.94-2.05 (2.11-2.05)	54.69-1.56 (1.60-1.56)	66.21-2.10 (2.16-2.10)	54.34-1.92 (1.97-1.92)	130-2.49 (2.56-2.49)	87.33-2.25 (2.31-2.25)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_1$	$P2_1$
Cell dimensions (Å)	a=51.21 b=83.02, c=109.04	a=50.95 b=82.65, c=109.39	a=51.13 b=82.80, c=110.24	a=50.90 b=82.55, c=108.68	a=45.42 b=129.87, c=90.38	a=45.51 b=130.50, c=90.15
Cell angles (°)	α=β=γ=90	α=β= γ=90	α=β= γ=90	α=β= γ=90	α=90 β=104.45 γ=90	α=90 β=104.37 γ=90
Wavelength (Å)	0.9794	0.9778	0.8726	0.9795	0.9763	1.0000
Total data	757854 (56157)	384388 (13423)	999611 (66119)	1028314 (76989)	161209 (4281)	179918 (13487)
Unique data	29865	65573	27934	76989	34488	47521
Completeness (%)	99.2 (99.9)	99.3 (94.3)	100 (100)	100 (100)	97.3 (74.3)	98.5 (97.6)
I/σ(I)	21.3 (5.5)	13.9 (2.1)	15.1 (4.8)	16.1 (4.5)	8.9 (2.1)	11.1 (2.3)
R_{merge} (%)[b]	11.8 (70.6)	6.6 (59.0)	28.3 (>100)	19.5 (93.9)	14.8 (32.5)	8.7 (61.4)
R_{pim} (%)	2.4 (14.3)	3.2 (44.7)	4.8 (20.8)	3.8 (17.8)	7.8 (30)	6.1 (41.2)
Multiplicity	25.4 (25.9)	5.9 (3.0)	35.8 (32.4)	28.7 (29.4)	4.7 (2.2)	3.8 (3.9)

Refinement						
No. of unique reflections used	2165	4517	2041	2620	1962	3476
R_{work}[c](R_{free})[d]	21.2 (23.2)	19.4 (21.4)	19.7 (22.8)	20.0 (23.1)	19.3 (22.0)	23.1 (27.4)
r.m.s.d. bonds (Å)	0.016	0.013	0.013	0.014	0.013	0.017
r.m.s.d. angles (deg)	1.72	1.67	1.72	1.73	1.89	1.91
Average B[e] (Å^2)	32.90	24.04	22.09	21.89	39.39	36.56
No. of atoms (residues) per asymmetric unit	3265 (381)	3527(386)	3354 (387)	3261 (379)	6458 (765)	6053 (734)
Program	BUSTER 2.11.2	BUSTER 2.11.2	BUSTER 2.11.2	BUSTER 2.11.2	BUSTER 2.11.2	BUSTER 2.11.2
Model quality (Ramachandran plot)[f]						
Residues in favoured regions (%)	97.1	97.7	97.4	97.6	96.7	96.7
Residues in allowed regions (%)	2.9	2.00	2.30	2.10	3.30	3.30
Residues in disallowed regions (%)	0.00	0.30	0.30	0.30	0.0	0.0
No. of domains /asymmetric unit	3	3	3	3	6	6

[a] Values in parentheses correspond to the highest resolution shell.

[b] $R_{merge}= \Sigma_j\Sigma_h (|I_{j,h}-<I_h>|)/\Sigma_j\Sigma_h(<I_h>)$, where h is the unique reflection index, $I_{j,h}$ is the intensity of the symmetry related reflection and $<I_h>$ is the mean intensity.

[c] $R=\Sigma_h||F_o|_h-/|F_c|_h|/\Sigma_h|F_o|_h$, where h defines the unique reflections.

[d] Calculated on a random 5% of the data.

[e] $B = <u^2> 8\pi^2/3$, where $<u^2>$ is the mean square atomic displacement

[f] Values from MolProbity[397]

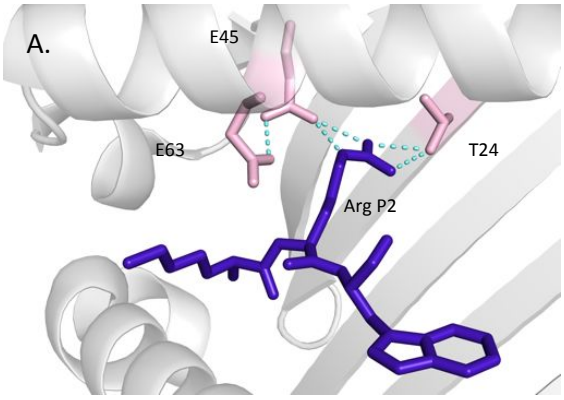
4.3. The HLA B*2705 peptide binding groove

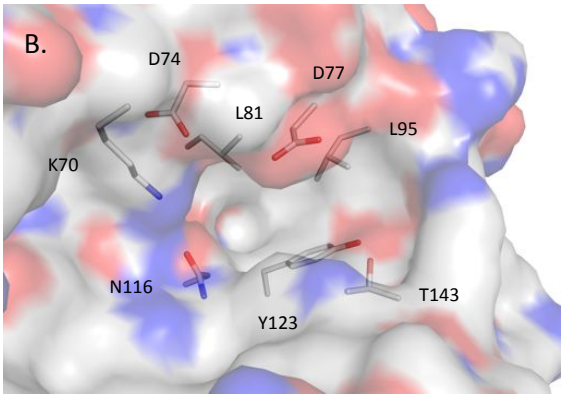
As noted previously[322] there are common motifs adopted by peptides within the HLA B*2705 binding groove. These common motifs suggest strong preferences for certain binding modes of the peptide within the groove. In our structures we are able to observe many of these common preferences despite the large difference in the lengths of the peptides bound in these complexes (from 6-14 residues).

The anchor residue provided by the P2 arginine residue binding within the B pocket of the peptide binding groove is extremely similar in all of the structures crystallised (Figure 10) and is almost identical to the previously published structure of the KK10 complex[259]. The arginine at Position 2 ($R2_{epitope}$)is stabilised, as previously described[322], through the formation of a planar hydrogen bond network between the $E45_{B27}$ $O_{\varepsilon 2}$ and the $R2_{epitope}$ $N_{\eta 2}$ and N_{ε} atoms and between the $R2_{epitope}$ $N_{\eta 1}$ and $N_{\eta 2}$ atoms and the $T24_{B27}$ O_{γ} atom. The backbone of $R2_{epitope}$ is held in place through stacking against the tyrosyl ring of $Y7_{B27}$ of the heavy chain.

In addition to the binding of the P2 anchor residue, the neighbouring residues are also bound similarly in these structures, with P1 and P3 showing high levels of conformational identity.

The HLA B*27 groove ends after the F pocket, a pocket at the base of the groove formed by the side chains of T143, D116, L81, L95, D74, D77 and Y123. This pocket seems to allow more promiscuous binding than the B pocket, with hydrophobic residues at the entrance and acidic residues at the base, thus allowing smaller hydrophobic side chains or long basic side chains (such as lysine) to bind. Nevertheless, mutations in residues within the pocket, particularly those of D116 are known to have significant effects on both the repertoire of peptides presented and the tapasin dependence of presentation[261; 398].





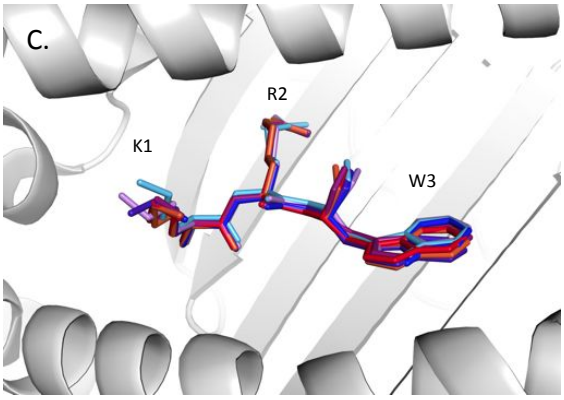


Figure 30: A. The P2 arginine of the KK10 peptide binding in the B-pocket of the HLA B*2705 groove, showing salt bridges formed between the residue and E45 and T24 of the HLA.B. The F-pocket of the HLA groove from the HLA B*27-KK10 complex, shown without the peptide. C. Overlay of peptide residues 1, 2 and 3 from the N-terminal end, showing similar positioning of peptides binding in this region.

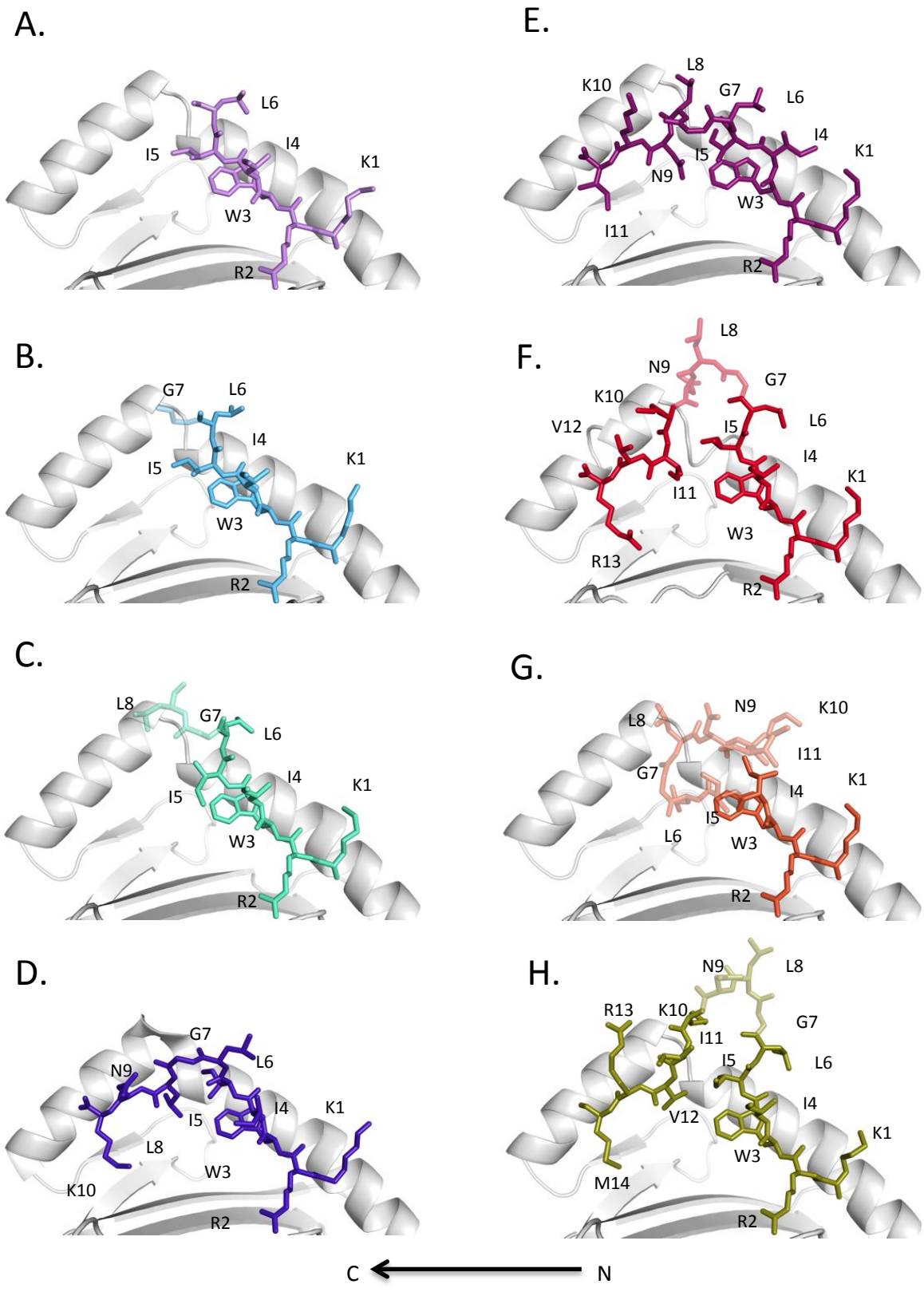


Figure 31: Panels A-H, HLA B*2705 complexed with KK10 epitope forms A. KL6 B. KG7 C. KL8 D. kk10 E. ki11 F. KR13 G. KM14 H. KM14 2nd conformation.

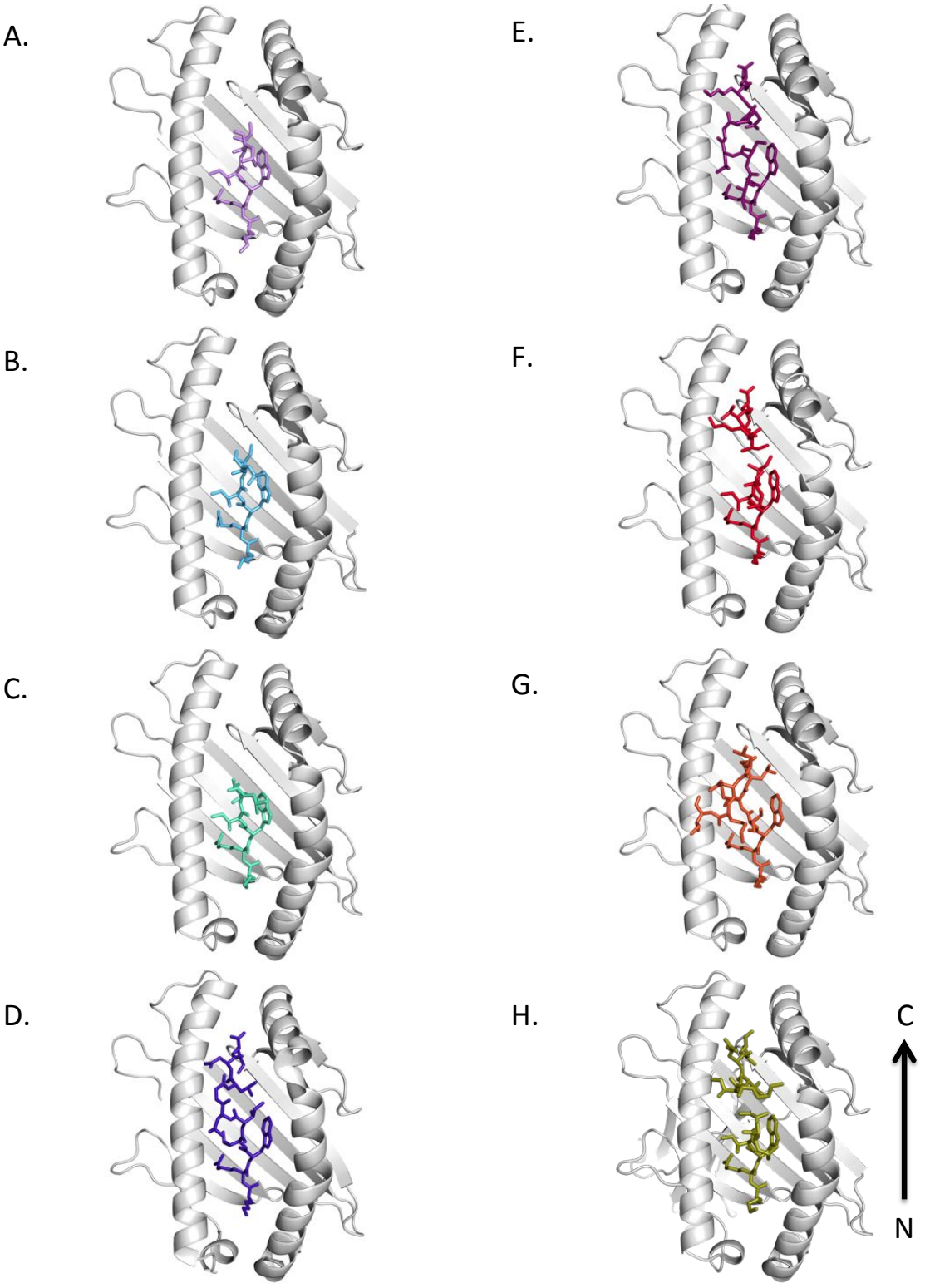


Figure 32: Panels A-H show overhead view of KK10 epitope forms complexed with HLA B*2705, A. KL6 B. KG7 C. KL8 D. kk10 E. ki11 F. KR13 G. KM14 H. KM14 2nd conformation.

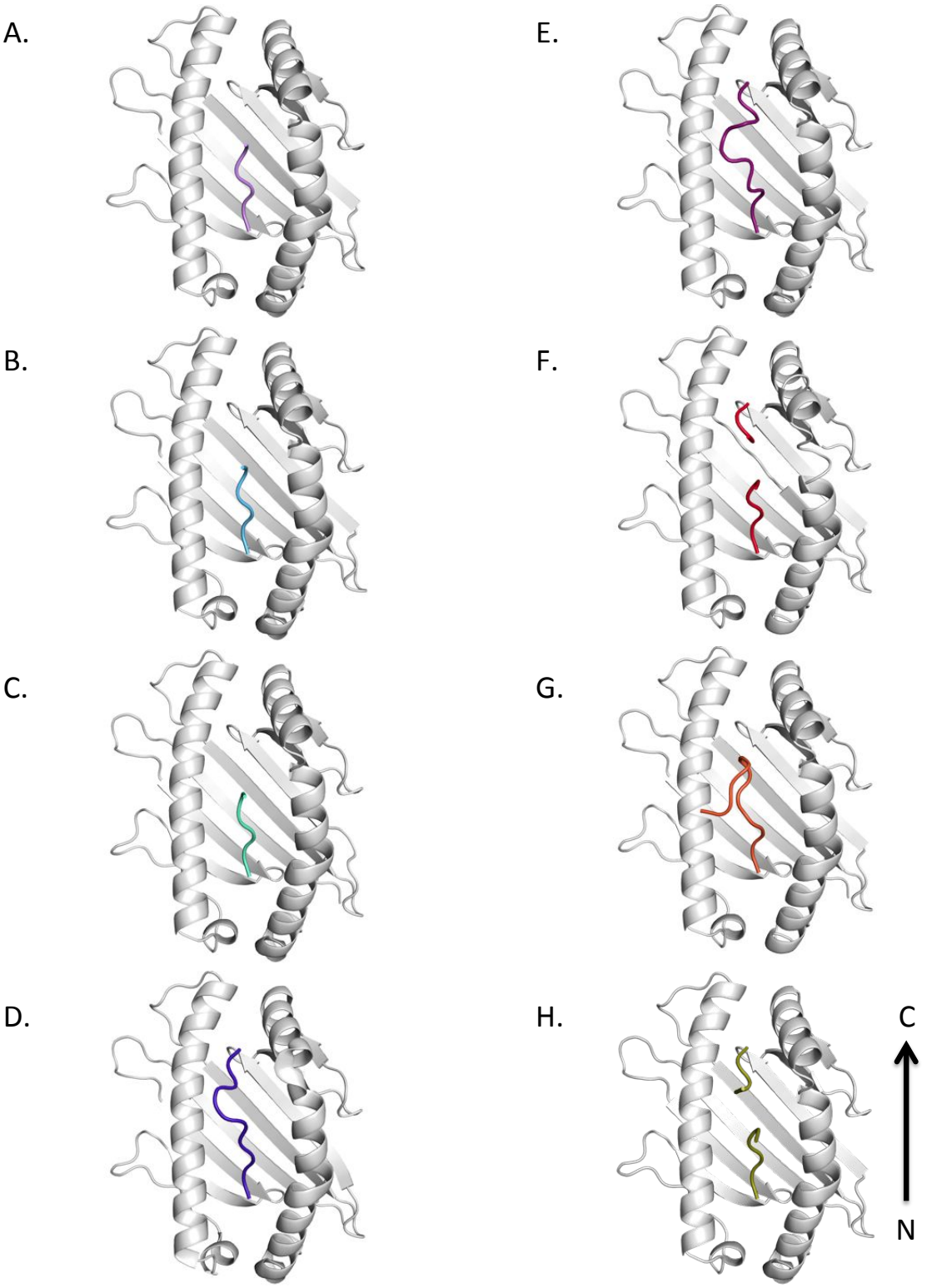


Figure 33 Panels A-H, overhead view of the backbone of KK10 epitope forms complexed with HLA B*2705 A. KL6 B. KG7 C. KL8 D. kk10 E. ki11 F. KR13 G. KM14 H. KM14 2nd conformation.

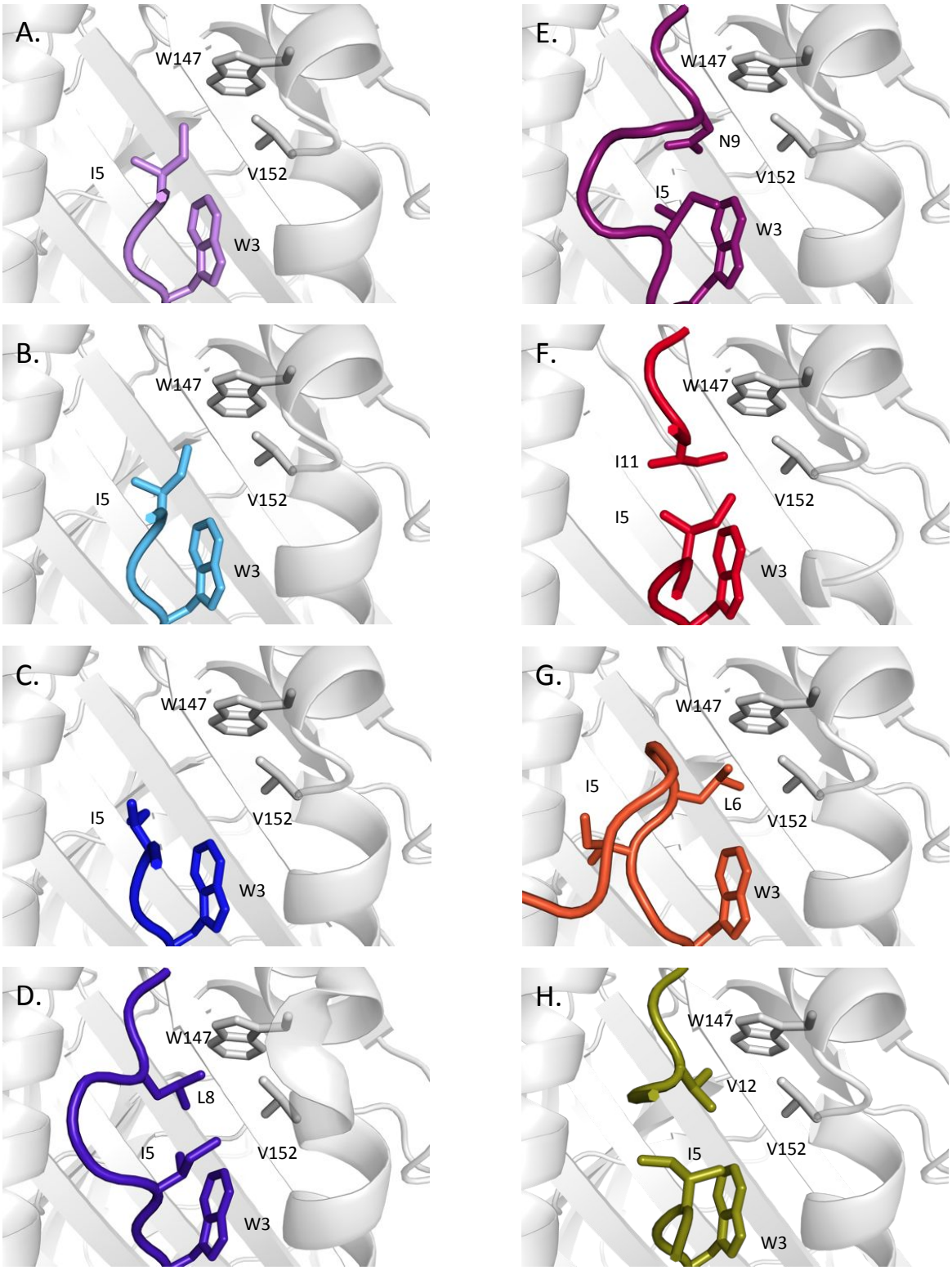


Figure 34: The E pocket shown in A. KL6 B. KG7 C. KL8 D. kk10 E. ki11 F. KR13 G. KM14 H. KM14 2nd conformation.

4.4. 10mer Structure (previously published)

The structure of the KK10 peptide in complex with the HLA B*2705 was solved prior to the start of this project [259]. The description of it here is given, as it is an extremely useful baseline structure for comparison purposes.

4.4.1. An "optimal" length epitope bound to the HLA Class-1 binding groove

As previously described in[259], the KK10 epitope is bound to the HLA B*2705 groove primarily by anchors at position 2 ($P2_{KK10}$)and position 10 ($P10_{KK10}$)of the peptide, which bind in the B and F pockets of the HLA groove respectively. The N-terminal anchor position holds residues 1-3 in a very similar conformation to that seen in all of our peptide-B*2705 complexes (Figure 30C) and is described above. The C-terminal anchor adopts a canonical position for HLA class-1, with the hydrogen bonding of the carboxy-terminus to residues K146, Y84 and T143. The side chain of the terminal lysine is accommodated within the F-pocket of the B*2705 groove (Figure 30B). The long lysine side chain allows the interaction of the lysine ζ-amino group with the acidic residues $D74_{B27}$, $D77_{B27}$ and $D116_{B27}$ through a network of ordered waters. The hydrophobic portion of the side chain can stack against $L81_{B27}$, $L95_{B27}$ and the ring of $Y123_{B27}$. The trapping of ordered waters in this pocket thus allows a degree of flexibility in the interaction of side chains with the surrounding residues and enables a range of C-terminal peptide residues to bind in this position.

The central region of the KK10 peptide between residues 4 and 7 forms a raised bulge that is supported by a layer of ordered waters between the β-strands at the base of the groove and the peptide chain (Figure 38A). This bulging allows the accommodation of 10 residues within the HLA class-1 binding groove, in a fully extended conformation as few as 8 residues may stretch across the base of the binding groove. $I4_{KK10}$ is the first residue in this central region

of the peptide, with its C_α raised approximately 3Å from the base of the groove compared with the C_α of $W3_{KK10}$. The side chain of this residue interacts hydrophobically with $I66_{B27}$ of the α1 helix. $I5_{KK10}$ continues from $I4_{KK10}$ without a significant difference in elevation, its side chain interacting with the indole ring of $W3_{KK10}$ and the side chains of $L8_{KK10}$ and $V152_{B27}$ of the α2 helix. At Position 6 the backbone of the peptide forms a tight loop consisting of residues 6 and 7 before descending back into the binding grove. The C_α of $L6_{KK10}$ is elevated by approximately 3.6Å from the binding groove floor compared with the C_α of $I5_{KK10}$ and the isobutyl side chain is surface exposed. The C_α of $G7_{KK10}$ is very slightly elevated from position 6 before kinking sharply back towards the groove and residue $L8_{KK10}$. The isobutyl side chain at this position is directed down towards the groove and sits in the E pocket, a hydrophobic pocket formed by the side chains of $I5_{KK10}$ and $V152_{B27}$ and $W147_{B27}$ of the α2 helix. The side chain of position 9 is directed away from the groove and is partly solvent exposed.

4.5. Refolding of a 4mer and 5mer peptide with HLA B*2705

Refolding of the peptides KRWI and KRWII with HLA B*2705 was carried out following their identification in proteasomal digestion assays and the successful crystallisation of the HLA B*2705 – KL6 complex. Refolding was performed as described[399] and a peak for the heterotrimer complex was seen for both the 4mer and the 5mer during purification by size exclusion chromatography (Figure 35B and C). The small size of these peaks meant that it was not possible to perform a second round of purification before concentration for crystal trials; no crystals were obtained from the trials set up. This could be due to insufficient concentration of the complex in the trial set-up or perhaps insufficient purity of the heterotrimer. It is less likely that the 4 and 5mer are insufficient to stabilise the complex as they contain all of the N-terminal residues that we see stabilising the conformation of the 6mer complex.

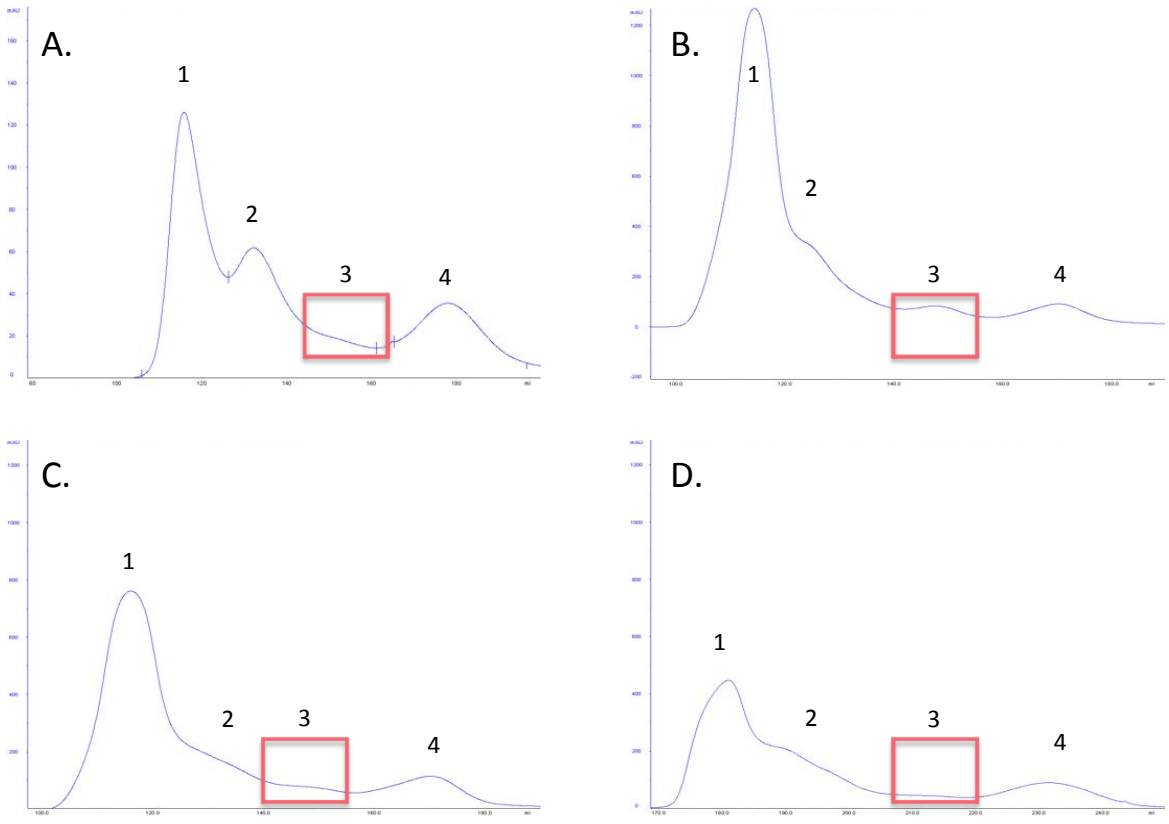


Figure 35: FPLC traces showing elution peaks of 1. Heavy aggregates 2. HLA B*2705 homodimer 3. HLA B*2705 heterotrimer and 4. B*2705 monomer/β2M homodimer in A. A control refold without peptide. B. The KI4 peptide refold. C. The KI5 peptide refold and D. The RL7 peptide refold.

4.5.1. Mass Spectrometry

To address concerns raised above, samples of the proteins were recovered from our crystallisation set-ups and sent for analysis by mass spectrometry. The results from this analysis show the presence of both the 4mer and the 5mer in their respective complexes (Table 30). Uncomplexed peptides are unlikely to be present in these wells as unbound peptide would have been separated from the complex during size exclusion chromatography, the large size difference between these two components makes it unlikely that unbound peptides could have contaminated the complex used for crystallisation. Following these results, a possible next step in crystallogenesis would be to eliminate the possibilities of insufficient concentration and/or purity. This would most easily be achieved through scaling

up, or collection of several refolds to give a larger amount of heterotrimer complex that would then be purified as normal. A wider range of crystallisation conditions could also be tried with an increased amount of product.

KK10 epitope form refold	Concentration in refold
KI4	0.2pmol/µl
KI5	10fmol/µl
KW14	1pmol/µl

Table 30: Concentrations of peptide in the KI4, KI5 and KW14 refolds. Refolds were diluted 1:100 with 1% formic acid to free the peptides prior to mass spectrometry (In collaboration with Dr. Stefan Tenzer).

4.6. Structure of the 6mer and 7mer peptides

The KL6 and KG7 epitopes, though longer than the 4 and 5mer for which refolding has been demonstrated, are nevertheless still too short to stretch across the HLA class-1 binding groove. However, crystallisation of these complexes has been possible and show that these short peptides are capable of stabilising the open conformation of the HLA B*2705 groove.

4.6.1. Short peptide epitope bound at the N-terminal anchor position only

The 6mer and 7mer peptides adopt extremely similar conformations in complex with B*27 and appear to bind to the groove solely at the N-terminus. The first 5 residues are within the binding groove, with the first 3 residues bound in a canonical fashion for the N-terminus of these KK10 epitope forms (Figure 30C, Figure 31A and B). The $I4_{KL6/KG7}$ residue C_α is shifted slightly C-terminally compared with the KK10 10mer complex and the side chain of this residue is rotated so that the $I4_{KL6/KG7}$ $C_{\delta 1}$ faces the α1 helix. However, the side chain

maintains a similar distance to the $I66_{B27}$ residue, allowing the hydrophobic interaction seen in the 10mer. The $I5_{KL6/KG7}$ C_α is shifted towards the centre of the groove away from the α2 helix. The side chain of this residue is oriented along the groove and towards the α2 helix, with the C_γ and C_δ able to engage in hydrophobic interactions with $V152_{B27}$ and $W147_{B27}$ side chains (Figure 34A and B). These side chains are in similar positions to their equivalent in the 10mer complex, forming the E pocket. Following this, the backbone of the peptide bends vertically away from the groove; the C_α of $L6_{KL6/KG7}$ is therefore nearly 3.8Å further from the groove than the $I5_{KL6/KG7}$ C_α. The side chain at this position faces N-terminally from the C_α and is raised above the groove. The residue is restrained by the interaction of the backbone amide with the carbonyl oxygen of $Q155_{B27}$ of the α2 helix. Following this residue, density for $G7_{KG7}$ in the 7mer is not sufficient for us to build it into our model and indicates that it is likely disordered.

4.6.2. Residual Density at the C-terminus of the binding groove

In both the 6mer and 7mer structures there is residual density present in the F-pocket of the binding groove (Figure 36A and B). This indicates that a molecule that has been recruited in the refolding, crystallisation or freezing processes occupies this pocket. In the 7mer structure in one of the two asymmetric units, the density strongly indicates that this molecule is an arginine amino acid (Figure 36B). It is likely this was picked up during the refolding process as the buffer used here contained arginine to suppress aggregation.

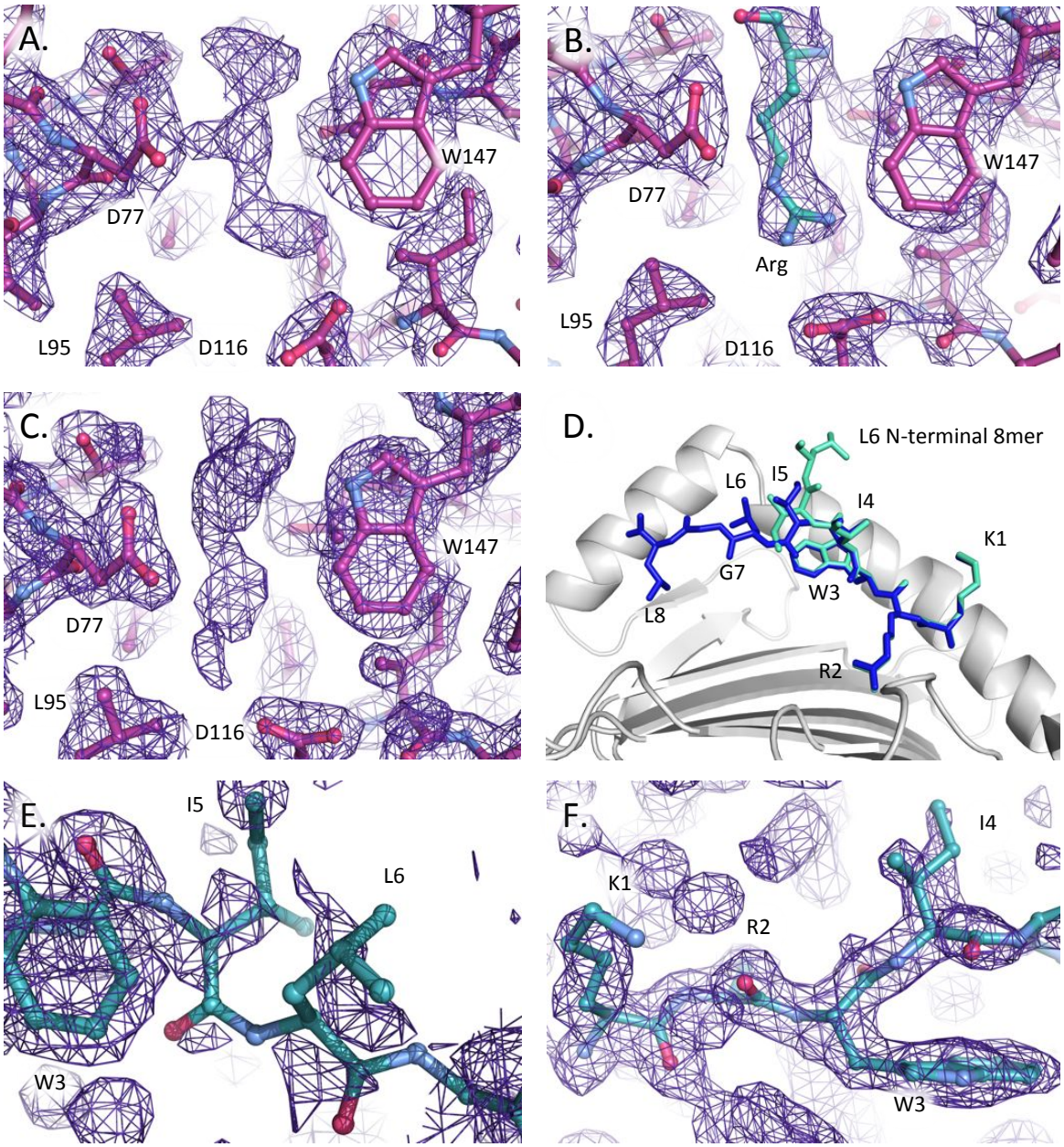


Figure 36: Residual density (map shown at 1σ) at the F-pocket of the HLA B*2705 binding groove in A. KL6 B. KG7, shown with an arginine residue modelled in the residual density and C. KL8. D. KL8 N-terminally bound and extended peptide models shown against the α2 helix of the binding groove. E. Electron density map at 1σ showing relative lack of density for central residues in the extended KL8 model and F. Map at the same σ for the N-terminal region of the KL8 peptide.

4.7. 8mer structure

The 8 residues of the KL8 peptide, if arranged in an extended conformation along the floor of the binding groove, would allow the terminal residues to bind at either end of the groove and stretch across it fully. However, our model shows that, at least for the majority of the time, we see a structure more similar to that of the 6mer and 7mer (Figure 31C). Areas of residual density found in the central region and C-terminus of the binding groove may indicate that an alternative, extended structure may occur (Figure 36D), though the strength of the density present suggests that it does so very infrequently (less than 10% of all instances). No further information could be gleaned following refinement of the data with our model of an extended conformation (density in the central and C-terminal regions is shown in Figure 36C and E).

4.7.1. The N-terminally bound form of the 8mer peptide

Following canonical binding of the residues 1-3 at the N-terminus. The density for residue 4 is very clear, and almost identical to the placement in the 6mer and 7mer structures (Figure 32A, B and C). Following this residue, the density becomes less well defined. The backbone positions of positions 5 and 6 are visible, along with an indication of the direction of the side chain in position 6; however, there is little density for the side chain of $I5_{KL8}$. The positions of the backbone for $I5_{KL8}$ and $L6_{KL8}$ are very similar to the positions in the 6mer and the 7mer and so indicate that the side chains may also occupy similar spaces. That the side chains are visible in the shorter peptides and not the 8mer may be due to the differing space groups for these crystals, and particularly the presence of crystallographic contacts of the final residues of the peptide with the adjacent asymmetric unit in the 6mer and 7mer structures. That the position of the peptide backbone is still similar in the 8mer to the position adopted in the 6 and 7mer indicates that though the crystallographic contacts with the peptide in the 6 and

7mer structures may provide some additional stability, their presence does not significantly affect the conformation of the peptide in these structures. Following the backbone of $L6_{KL8}$ there is no further visible density for $G7_{KL8}$ and $L8_{KL8}$, indicating significant flexibility of the residues, as they have no interactions with the binding groove.

4.7.2. Residual density at the C-terminus of the binding groove

As previously indicated there is significant residual density in the F-pocket of the binding groove (Figure 36C). This density may be partly attributed to the presence of an alternate, extended conformation of the peptide. However, this cannot account for all of the density seen at this position as it is significantly stronger than that seen along the centre of the groove and a peptide of this length would have little flexibility if it stretched between the B and F pockets of the groove. Two alternate possibilities exist to explain this density, firstly, that in a significant number of molecules within our crystal a second peptide is bound by it's C-terminus within the groove, and secondly, that the F-pocket is filled by a molecule picked up during refolding/crystallisation as is shown for the 7mer (Figure 36B). The latter possibility seems the more likely of the two, as we have been unable to refold significant amounts of peptide to HLA B*2705 that lacks the first and second N-terminal residues, so the C-terminus is likely insufficient to stabilise the binding groove alone. Furthermore, there is little continuation of density of this strength beyond the binding groove itself, so any peptide bound here would be held solely by one C-terminal residue and otherwise be completely flexible. For these reasons, it is likely that a molecule picked up from solution occupies it.

4.8. 11mer structure

4.8.1. A central folded loop allows a C-terminally extended peptide to fit into the HLA B*2705 binding groove

The central region of this peptide shows an elevated conformation consisting of a folded loop or "zigzag" held in place by a network of structured waters occupying the space between the peptide and the binding groove (Figure 38B). This conformation is somewhat similar to, though an extended form of, the raised bulge seen in the 10mer structure. This "zigzag" begins with $L6_{KI11}$ where the C_α is elevated and positioned towards the α1 helix by around 3Å relative to the C_α of $I5_{KI11}$. This residue is held in position only by the constraints on flexibility imposed by surrounding residues and a hydrogen bond between the backbone amide and a structured water. From this position, the C_α of the next residue, $G7_{KI11}$, moves towards the groove by about 1.5Å and remains in the same position in relation to the 2 α-helices. This contrasts with the raised loop formed in the 10mer where the C_α of this residue remains at the same elevation as $L6_{KI1}$ before kinking back toward the groove. Instead, in this longer peptide, an extra residue is accommodated by extending this loop, with $L8_{KI11}$ completing the "zigzag" by shifting towards the α2 helix by almost 3.8Å and remaining at the same elevation as the $G7_{KI11}$ C_α. Following this residue, the backbone drops by 3.4Å back towards the groove and enters the C-terminal region.

4.8.2. The extension of the central loop affects binding positions throughout the peptide

The extension of this loop to include $L8_{KI11}$ means that this residue is no longer placed in the hydrophobic E pocket formed by $W3_{KK10}$, $I5_{KK10}$ and $W147_{B27}$ in the 10mer. The residue that

follows $L8_{KI11}$ and is directed toward the groove is in this case $N9_{KI11}$, which is not hydrophobic. Consequently, there are several alterations in the structure compensating for this change. The C_α of $I4_{KI11}$ in the 11mer structure is pushed upward and N-terminally compared with the 10mer, this displaces the side chain and increases the distance between it and residue $I66_{B27}$, reducing the strength of this interaction (Figure 34H). The side chain is slightly shifted away from the groove resulting in an increase in solvent exposure for the C_β and $C_{\gamma 2}$ atoms. The change in position of this residue may also affect the position of $R62_{B27}$ compared with the 10mer (Figure 39F). The C_α of $I5_{KI11}$ is again slightly N-terminal of its equivalent position in the 10mer and the side chain is further toward the α1 helix. The interaction with $W3_{KI11}$ is maintained, though the distance between $I5_{KI11}$ and $V152_{B27}$ is increased, in this position the side chain of $N9_{KI11}$ can be accommodated without hindrance from $I5_{KI11}$ $C_{\gamma 2}$. The $N9_{KI11}$ C_β is oriented toward the hydrophobic pocket occupied by $L8_{KK10}$ in the 10mer, but then kinks away to allow the δ-oxygen and δ-amino groups to interact with ordered water molecules between the peptide and the α1 helix (Figure 34E).

4.8.3. The residues of the C-terminal region

The F-pocket of the 11mer is occupied by the C-terminal residue $I11_{KI11}$ that is a shorter and more hydrophobic residue than the lysine at the C-terminus of the 10mer peptide. However, the conformation of the backbone is identical to that of the 10mer, with the carboxylates hydrogen bonding to $K146_{B27}$, $Y84_{B27}$ and $T143_{B27}$. As the side chain of $I11_{KI11}$ is shorter than that of the 10mer, it is not able to reach the base of the F-pocket and does not interact with the acidic residues there. Instead it is stabilised by hydrophobic interactions with $Y123_{B27}$, $L81_{B27}$ and $L95_{B27}$ of the heavy chain, with the Leucine side chains moving slightly from their 10mer positions into the pocket to interact with the isoleucine. Residues $D116_{B27}$ and $K70_{B27}$ are slightly shifted N-terminally where they interact with ordered waters.

4.9. 13mer structure

The complex of a 13mer peptide with an HLA class-1 molecule has been previously shown in [383]. In this paper, a "super-bulged" yet apparently rigid conformation was seen and explained how a peptide of this length was accommodated within the class 1 binding groove. Our KR13 peptide complexed with HLA B*2705 loosely follows the general binding conformation of this structure, with a central bulge of the peptide stabilised by intra-peptide interactions. However, the loop formed by this bulge in our structure appears to be disordered and is not rigid which does not allow us to see the apex of the central bulge. This could be due to the absence of proline residues in the KR13 peptide, which may have increased the rigidity of the conformation in the B*3501-EBV complex[383].

4.9.1. A tight loop of peptide outside of the groove permits binding of extended epitopes within the restricted class 1 groove

While the N-terminal region of this peptide is bound in the canonical manner for our HLA B*2705 complexes, the central region of the peptide shows a different binding mode to that seen in the 10 or 11mer allowing the accommodation of two extra residues within the confines of the groove. $I4_{KR13}$ is in a similar position to its equivalent in the 10mer complex, with only a 0.5Å shift from the 10mer position and a rotation of the isoleucine side chain, the interactions of both side chain and backbone with the HLA molecule are similar. $I5_{KR13}$ is shifted slightly C-terminally and is more elevated than its equivalent in the 10mer complex, with a shift of 1Å between C_α's of the 10mer and the 13mer. The side chain of the isoleucine is oriented slightly more towards the centre of the binding groove and engages in hydrophobic interaction with the side chain of $I11_{KR13}$ of the peptide, this being in a similar position to $L8_{KK10}$ in the E pocket (Figure 34F). This interaction links the N-terminal and C-terminal sides of the peptide and between these residues, the peptide bulges out of the

groove in a partly flexible loop. This bulge out of the groove enables the long 13mer peptide to fit into the B*2705 groove.

From position 5, the backbone of the peptide turns sharply upward away from the groove, rising vertically by 3.6Å to the C_α of $L6_{KR13}$. The backbone is held in this position through the interaction of the isoleucine residues and by a hydrogen bond between an ordered water molecule and the carboxylate of the residue. The side chain of L6 faces towards the N-terminal end of the peptide groove, it is solvent exposed and poorly ordered, with a temperature factor (a measure of the relative vibrational motion of atoms within the structure) of 185.5 (compared with 11.18 for $I5_{KR13}$). The residue is shifted by 4.9Å from it's 10mer equivalent and the C_α is 4.6Å further towards the α1-helix compared with the 10mer. The backbone appears to continue to move upwards from this point, away from the groove. Though there is some density continuing beyond $L6_{KR13}$, the features are not defined and we are unable to build a model for the residues from $G7_{KR13}$ to $N9_{KR13}$ inclusive; possibly due to a relative lack of restraints and so increased flexibility of these residues.

The backbone of the $L10_{KR13}$ residue begins to return towards the groove from the flexible loop formed by the central residues. The side chain of this residue is oriented towards the α1 helix and the terminal amine forms a salt bridge with $E76_{B27}$ of the HLA heavy chain, stabilising the residue (Figure 37B). The side chains of the equivalent residues in the 10 and 11mers are in significantly different positions and neither interacts with $G76_{B27}$ (though $L10_{B27}$ and $E76_{B27}$ do interact in the 11mer, this residue is in a position equivalent to $V12_{KR13}$ in the 13mer, Figure 37A and C). Following this residue, the C_α of the $I11_{KR13}$ residue is 3.6Å closer to the groove. The isoleucine side chain faces N-terminally and parallel to the groove, engaging in hydrophobic interaction with $I5_{KR13}$ and completing the C-terminal base of the loop.

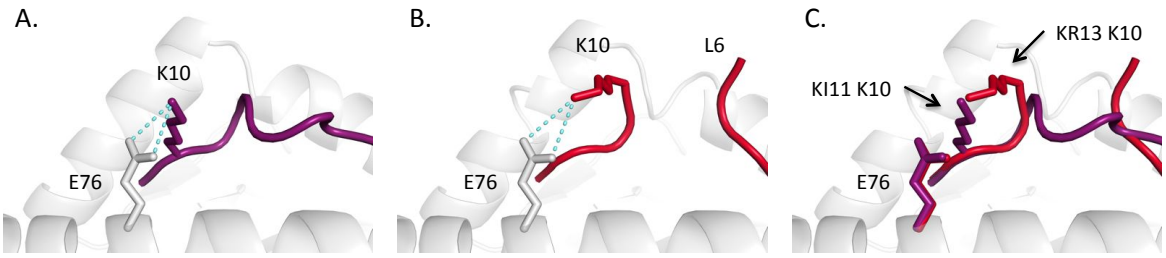


Figure 37: The interaction of K10 and E76 in A. the KI11 complex and B. the KR13 complex, C. KI11 and KR13 peptides showing K10 side chain position and E76 positions.

4.9.2. The C-terminal region

Following the central loop, the C-terminal region of the peptide is bound in a similar manner to the KK10 complex. The side chain of $V12_{KR13}$ faces out of the groove and the $R13_{KR13}$ backbone carboxylates are bound in the same manner as the 10mer complex (described in 5.4.1). The side chain of $R13_{KR13}$ occupies the F pocket, and the length of this side chain means that it reaches to the base of the pocket. The arginine terminal amine groups form hydrogen bonds with the side chain carbonyl groups of $D77_{B27}$ and $D116_{B27}$ as well as with $L70_{B27}$ through an ordered water. This differs from a previously published HLA B*2705-flu peptide complex which also ended with a C-terminal arginine. In this case $L70_{B27}$ was seen to move away from the F pocket and did not interact with the side chain[259]. A comparison of the two structures suggests that this may be due to a smaller side chain at residue 3 in this peptide (Y rather than W) that may prevent the adoption of this conformation in the KR13-HLA B*2705 complex.

4.10. 14mer structure

The KM14 epitope in complex with HLA B*2705 is an example of a peptide which is longer than the usual given range for HLA class-1 of 8-12 (or sometimes 8-10, or 9-12) amino acid residues. A previous study has identified several 14mer epitopes for HLA B*3501 [396] and shows these long peptides bound at their N and C termini to the ends of the class-1 groove with large protruding loops characterising the central regions. Therefore, we might have expected a similar conformation for the KM14-HLA B*2705 complex, particularly considering the similarities between the previously crystallised HLA B*3501-13mer complex[383] and our HLA B*2705-KR13 complex. We were able to obtain data from two crystal structures for KM14 under different crystallisation conditions. Our first model shows a peptide that may exist in multiple conformations and may indicate a novel mode of binding to HLA class-1, however, the data for the peptide in the binding groove in this model is unclear in places. Our second model shows a 14mer peptide clearly bound at both the N and C termini of the groove with a large central bulge in a similar manner to the previously described 14mer complexes mentioned above.

4.10.1. The N-terminal region of the 14mer complex

The N-termini of both models is clear and shows a similar structure up to the $I4_{KM14}$ residue. Statistics shown are for a model of this complex up to the $I4_{KM14}$ residue (Table 29). These N-terminal residues are well ordered in the groove and adopt very similar conformations to that seen in the other complexes, with the first three amino acids adopting almost identical conformations. The position of $I4_{KM14}$ in this model is similar to that of the 11mer structure, with the side-chain of the residue oriented away from the groove and partially solvent exposed. The side chain is flipped from the position in the 11mer, with the $C_{\delta 1}$ group towards the C-terminus of the residue. Following this, in the first structure obtained the residual

electron density was insufficient for us to be highly confident of producing an accurate model of the peptide. However, the residual electron density within and above the groove has allowed us to build a longer, if rather speculative structure for this epitope for discussion and to aid future studies. This model will be discussed below followed by the structure seen in the second of our crystals.

4.10.2. Leucine at position 6 acts as an anchor residue for the 14mer peptide

The central region of the KM14 peptide is somewhat different to the complexes we have so far discussed. There is a significant amount of disorder within the peptide groove that may indicate the existence of multiple conformations of the peptide. The $I4_{KM14}$ residue is the first of the central region in this model and is well defined in the 14mer. The C_α is shifted by 0.9Å towards the N-terminus compared with the KK10 peptide. The $I4_{KM14}$ side chain faces away from the groove and is close to the α2 helix. The side-chain, additionally, engages in hydrophobic interactions with the $T3_{KM14}$ side chain. Following this residue the geometry of the backbone is flipped from that in the KK10 peptide and is different to any of the other epitope forms. The $I5_{KM14}$ C_α is at a similar elevation to that of $I4_{KM14}$ and its side-chain, which is oriented toward the α1 helix, is surrounded by numerous residues from both the α1 helix and the peptide. These include the hydrophobic $A69_{B27}$ and $I66_{B27}$ along with the C_β and C_γ of $L70_{B27}$. These, along with some of the backbone C_αs of the α1 helix may help to stabilise the $I5_{KM14}$ side chain in this position through hydrophobic interaction. In the 10mer complex, the side chain of $I5_{KK10}$ faces the α2 helix and engages in hydrophobic stacking with the $T3_{KK10}$, forming part of the E pocket.

The backbone from the C_α of $I5_{KM14}$ drops slightly towards the groove and α2 helix by 0.9Å to the $L6_{KM14}$ C_α. The side chain of this residue is oriented towards the α2 helix and the binding groove. The side chain is stabilised in the hydrophobic E pocket formed from some of the

same residues seen in the 10mer E pocket. The pocket is formed by $W147_{B27}$ and $V152_{B27}$ on the α2 helix and $W3_{KM14}$ of the peptide, these interact with the $L6_{KM14}$ side chain through hydrophobic interactions (Figure 34G). The position of this residue is extremely different from that seen in any of the other HLA-B*27 complexes that we have solved for KK10 epitope forms as in these complexes, L6 usually forms the apex (or near apex) of a central loop. In this case the peptide continues close to the base of the binding groove up to this position and $L6_{KM14}$ forms an anchor for the peptide. Following this, the peptide backbone kinks up and out of the binding groove, rising almost perpendicularly by 4.8Å. This vertical rise includes the whole of the $G7_{KM14}$ residue.

4.10.3. A folded conformation may exist in addition to the loop conformation of the 14mer.

Beyond $G7_{KM14}$, the electron density loses clear definition as the peptide is elevated away from the binding groove. There is a partial density present near the F-pocket of the groove. This could indicate that at this point a population of 14mers in the crystal forms a tight loop, containing residues $L8_{KM14}$ to $V12_{KM14}$, leading to $R13_{KM14}$ and $M14_{KM14}$ in position at the end of the groove. However, there is also residual density indicating another plausible, but very different, path. This density represents a peptide folds back towards the N-terminus beyond $G7_{KM14}$ and does not anchor C-terminally in the binding groove.

While the density present for this conformation is continuous following from the N-terminal portion of the peptide and is visible at reasonably low levels of noise (0.7-1 standard deviation from mean electron density), it lacks distinct features and only the peptide backbone is discernible. However, in response to this peptide conformation, a shift in the position of the side chain of the α1 helix $R62_{B27}$ to a position above the groove is observed. Next to this side chain is an area of better-defined electron density for the 14mer backbone

at the K10 position which indicates a possible interaction (hydrogen bonding) between the carboxylate of the peptide backbone for residue $K10_{KM14}$ and the terminal amino groups of $R62_{B27}$. This interaction likely helps stabilise the peptide in this folded position and restrain the C-terminal end of the peptide in an atypical position. However, the $R62_{B27}$ side chain is also present in this position in several of our structures where there is no indication of interaction with another peptide. The possibility that this residue is flexible and freely adopts this conformation, rather than forming a hydrogen bond must therefore be considered (Figure 39A, B and C). In addition to this, we also refolded the HLA B*2705 molecule with a KW14 peptide (KRWIILGLNKIVRW), this was designed to have a bulky hydrophobic residue at the C-terminus that would not bind in the C-terminal binding groove and might serve to stabilise a folded back conformation. We were unable to obtain crystals for this complex, but successful refolding evidenced by mass spectrometry (Table 30) was observed.

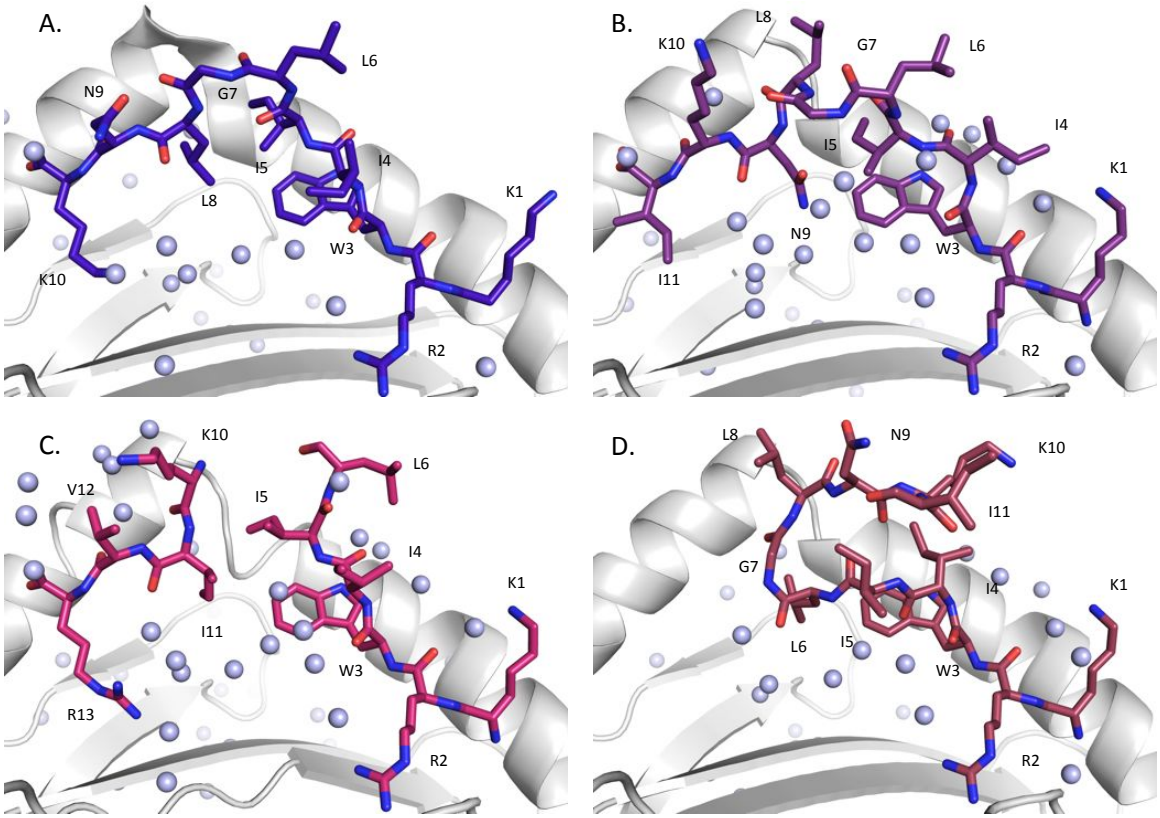


Figure 38: Water structure surrounding the peptide-binding groove in the A. KK10 B. KI11 C. KR13 and D. KM14 complexes.

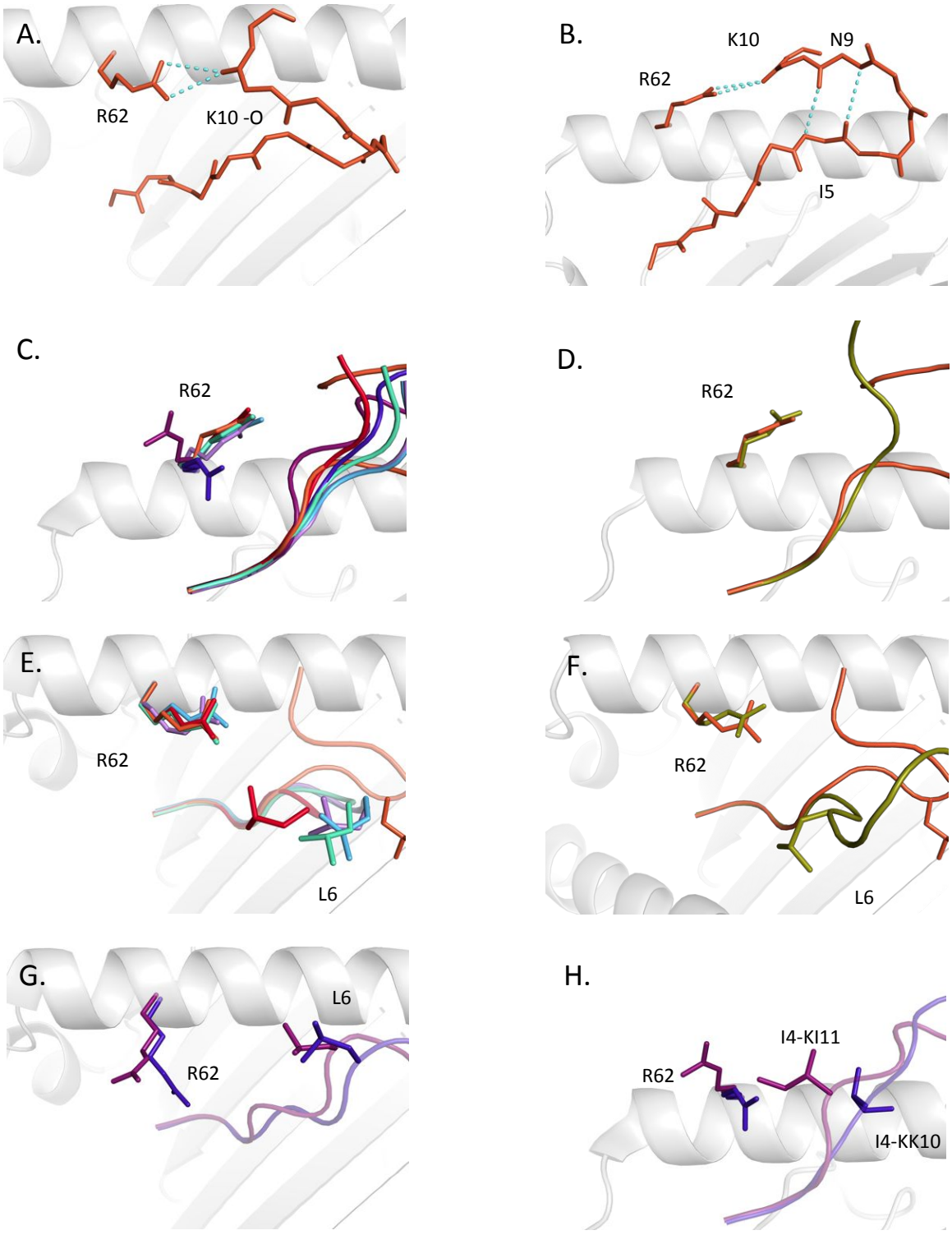


Figure 39: The positions and interactions of residue arginine 62 A. In the KM14 peptide B. In the KM14 peptide, showing intra-peptide hydrogen bonds between the backbone carbonyl-amides of residues 5 and 9. C. In all peptide forms. D. In both conformations of the KM14 epitope E. In it's position relative to leucine 6 In C. KL6, KG7, KL8, KR13 and KM14 F. In KM14 in both conformations of the KM14 epitope and G. KK10 and KI11. H. In it's position relative to isoleucine 4 In KK10 and KI11.

4.10.4. A second structure determined for the HLA B*2705 KM14 complex

The second structure we were able to obtain for the KM14 complex was derived from a crystal appearing 186 days following crystallization (Figure 31H), compared with 6 for the first crystal and in different conditions (20 %w/v Polyethylene Glycol 6000, 0.100 M HEPES pH 7.0), this means that although the density for the peptide is clearer in this structure, we are not able to comment on their relative validity as *in vivo* models since the different crystallization conditions make a direct comparison of the structures impossible.

This structure shows a peptide that is bound at both the N and C terminal ends of the groove, with a large central bulge. The N-terminal region is similar to that described for the first KM14 structure, differing slightly in the position of $I4_{KM14}$, the C_α of this residue remains in the same position, but the C_β shifts 2.4Å towards the α1 helix, and the side-chain faces this helix rather than the α2 facing position of the first 14mer. This is the same position that the $I4_{KR13}$ adopts. The structure of this model of the 14mer closely follows that of the KR13 complex, with the C_α of $I5_{KM14}$ being shifted by only 0.6Å out of the groove compared to the C_α of the $I5_{KR13}$. As in the 13mer complex, the peptide backbone kinks upwards to the C_α of $L6_{KM14}$. This atom is shifted N-terminally by 1.3Å compared with the $L6_{KR13}$. Following this, peptide density is insufficient to build an accurate model of the peptide chain until the $K10_{KM14}$ residue; however, an approximation of the loop based on areas of residual density was created for electrostatic modelling (area of 50% transparency, Figure 31H). The $K10_{KM14}$ C_α is located 6.73Å out of the groove from the base of the central loop at $V12_{KM14}$. The side-chain of $K10_{KM14}$ faces the α2 helix and hydrogen bonds with the backbone carbonyl oxygen of $A150_{B27}$. The peptide backbone drops towards the binding groove to the $I11_{KM14}$ C_α and then to $V12_{KM14}$, which forms the anchor at the C-terminus of the central loop. The side-chain of the $I11_{KM14}$ faces the α2 helix and may form a weak hydrophobic interaction with the $T73_{B27}$ C_γ on this helix, though this is 4.8Å from the $I11_{KM14}$ side-chain.

The $V12_{KM14}$ anchor is in a position equivalent to that of the $I11_{KR13}$, with the C_α shifted only 1.0Å out of the groove. The side-chain of this residue similarly forms hydrophobic interactions with $W147_{B27}$, $V152_{B27}$, and $W3_{KM14}$ in the E pocket. The peptide backbone adopts the canonical binding position at the C-terminus, with very similar positions to the 10mer, 11mer and 13mer complexes. The side-chain of the C-1 residue $R13_{KM14}$ faces out of the groove and is solvent exposed, which may serve as a feature for T-cell recognition. The R13 residue shares a similar basic character to the $K10_{KI11}$ residue that shares a similar orientation and may facilitate cross-recognition in TCR recognising these peptide side-chains. The $M14_{KM14}$ residue side-chain is bound in the F pocket and forms the C-terminal anchor.

4.11. Discussion

The structural models we have made showing binding of these epitope forms in the HLA B*2705 complex is, to our knowledge, the first study of a series of physiologically relevant viral peptide epitope forms differing only in the length of the sequence bound. This study, as well as being important for the study of long-term non progression and the generation of host immune responses, also informs us how epitope length directly affects conformation of the peptide and it's ability to bind to HLA B*2705, with implications for all HLA class-1. The systematic crystallisation of increasing length epitope forms has shown general features of binding in the HLA B*2705 groove and how these requirements are met by each epitope form. The study also adds to information on the accommodation of longer peptide forms in the HLA class-1 groove and the similarities and differences between the binding of these forms in different alleles[383; 396].

4.11.1. The N-terminal anchor of the KK10 epitope forms adopts a conserved conformation and is sufficient to stabilize the HLA B*2705 binding groove

In all of the epitope forms that were crystallised we see a common binding motif at the N-terminus of the peptide that includes the anchor residue at position 2. This motif seems to be common not just to our epitope forms, but is seen across a range of epitopes that bind to HLA B*2705[322] with the side-chains of residues 1-3 adopting very similar positions within the binding groove. The structures obtained by our short epitopes, including the 6mer, 7mer and 8mer peptides further show that the N-terminal region of the peptide is sufficient to stabilise the HLA B*2705 complex as in these structures we see residues from 1-5 in contact with the HLA binding groove, residues C-terminal of this are above the groove and solvent exposed (Figure 31 A-C). This indicates that the major contribution to the stabilisation of the complex is from the N-terminal peptide residues in these structures. Residual density at the C-terminus suggesting that a molecule may be picked up from solution to fill the F-pocket in these structures shows that filling this pocket may also be necessary to maintain a stable complex (Figure 36).

However, our data shows that occupation of the C-terminus alone is unlikely to be sufficient for stabilisation as a peptide that lacked residue 1 at the N-terminus failed to stabilise the complex as well as the 4mer and 5mer N-terminal peptides (Figure 35). This is in contrast to a previous study showing stabilisation of the murine H-2D(b) MHC by a C-terminal pentapeptide[400] and to a further study on the areas of the peptide groove involved in determining binding in HLA B*15 through assessment of relative conservation of residues at positions in HLA B*15 peptide epitopes[401]. This study showed that the C-termini of peptides binding HLA B*15 were more conserved than the N-terminal anchors and may have had a larger effect on determining epitope binding.

4.11.2. Common binding motifs of KK10 epitope forms within the B*2705 groove

In addition to a common binding motif at the N-terminus of the peptide, there are further areas along the groove that show preferences for certain binding modes and where binding of differing residues in our epitope forms can lead to significant differences in the conformation of the peptide. The first of these is the E pocket, this pocket, consisting of $W147_{B27}$ and $V152_{B27}$ on the α2 helix and $W3_{peptide}$ forms a hydrophobic patch midway along the groove. In the KK10 optimal epitope, this is occupied by the L8 residue (the I5 residue also forms part of the pocket), and so stabilises the peptide following the central bulging region. In several other peptide forms the pocket is bound by part of an isoleucine side-chain, being occupied by I5 in the 6mer and 7mer (Figure 34A and B) and I5 and I11 in the 13mer (Figure 34F). In the 13mer epitope form these residues are positioned at the N and C-terminal ends of the central peptide loop and as for $I5_{KK10}$ and $L8_{KK10}$ their interaction of their side-chains with each other and the E pocket helps to stabilize the peptide at either end of the central loop. This anchoring of the central peptide loop involving the E pocket is a common feature of the 10mer, 11mer and 13mer peptides, and the differences in residues involved and their relative orientations helps to accommodate the differently sized peptide loops/bulges seen in each complex.

This pocket is bound by L6 in the long model of the 14mer (Figure 34G) and the side chain of this residue is clearly visible in the electron density, indicating its stability in the pocket. We suggest in our model for the 14mer peptide that this interaction with the E pocket in the 14mer acts as a mid-way anchor to the HLA and stabilises a possible folded conformation.

However, in two of our epitope forms, the pocket is bound less optimally, with N9 in this position in the KI11 peptide and I5 being positioned much further away from the pocket in the 8mer structure (Figure 34C and E). In the 11mer this results in a shift of the polar asparagine side-chain away from the hydrophobic pocket. This shift affects the conformation

of the peptide N-terminal to N9, with $I4_{KI11}$ being displaced and shifted further out of the groove and $I5_{KI11}$ shifting to accommodate the N9 side-chain, this results in I4 being partly solvent exposed (Figure 31E) and so may increase its availability for TCR binding (Chapter 5.2.1). It also results in a slight increase in the distance between $I5_{KI11}$ and $N9_{KI11}$ compared with $I5_{KK10}$ and $L8_{KK10}$, and the shift in the side-chain of N9 away from the α2 helix actually allows the backbone to extend further towards it, meaning the C_{α} of N9 is actually closer to the α2 helix than the equivalent in $L8_{KK10}$. These changes allow room for a larger central loop in the 11mer and the incorporation of an extra residue in this loop through a more pronounced "zigzagging" of the peptide backbone between the α1 and α2 helices.

In the 8mer peptide, the side chain of the I5 residue which interacts with the E pocket in the 6mer and 7mer peptides is rotated so that the side-chain is oriented perpendicular to the base of the groove, increasing the distance between the E pocket and the $C_{\delta 1}$ group (Figure 34C). The electron density of this residue is not as clear for this position in the 8mer peptide as in the 6 and 7mer peptides and asymmetric crystal contacts may have helped to stabilise the I5 in the 6mer and 7mer structures so that it's true flexibility is not represented. Alternatively, the increased flexibility of this side-chain in the 8mer peptide could be due to the reduction in its interaction with the E pocket, which may have helped to stabilise the peptide.

The second region of the groove where we see differing residues binding at the same position of the HLA is the F pocket at the C-terminus of the groove. Unlike the E pocket, the different residues binding here all do so in similar conformations and all of our longer epitope forms have residues at their C-termini which match the binding preferences of the HLA B*2705 F-pocket (long basic residues or short hydrophobic residues), although the quite different side-chains of these residues do mean that there are some differences in the residues contacted within the F-pocket by the different epitope forms, with the $I11_{KI11}$ residue not reaching the base of the groove and having no groups available for hydrogen

bonding it is unable to interact with $D77_{B27}$, $D116_{B27}$ and $L70_{B27}$ as $K10_{KK10}$ and $R13_{KR13}$ do in this position. This difference does not affect the backbone of the peptide N-terminal to the F-pocket in the 11mer compared with the 10mer or 13mer and the residues are in similar positions along the HLA groove in each complex, the C-1 residue is slightly raised with the side chain facing out of the groove and the side-chain of the C-2 residue is involved in binding at the E pocket.

4.11.3. Longer epitopes can be accommodated by HLA B*2705 through novel peptide conformations

The structures that we have shown in this system include several epitopes that are longer than the "standard" range of 8-10 amino acids that have historically been considered optimal for HLA class-1 binding. Nevertheless, these epitope forms have been able to bind within the B*2705 groove and our structures show how these elongated epitope forms are accommodated through the adoption of certain conformations within the groove, some of which are partly or wholly novel.

To our knowledge there are no previous structures of HLA B*2705, or indeed of any HLA B*27 subtype complexed with peptide epitopes longer than 10 amino acid residues. Two complexes of HLA B*27 have been shown with decameric peptides, one of these is the KK10-HLA B*2705 complex used as a reference structure here[259], the other is HLA B*2709 complexed with a synthetic peptide (s10R) modified from part of the HLA B*2707 subtype sequence[402](Figure 40A). This structure shares several features of the KK10 conformation, including a canonically bound N-terminal region, binding within the F pocket at the C-terminus and a central bulge stabilised partly by intra-peptide contacts between residues at positions 5 and 8. However, differences in this peptide mean that quite a different surface will be presented to TCR, with both positions 4 and 9 significantly more solvent exposed

than in KK10 and the central bulge being shallower and longer at around 1.4-1.9Å closer to the groove at its apex than in the KK10 structure.

These common elements are also maintained to some extent in our extended epitope forms KI11, KR13 and the second structure obtained for KM14. The intra peptide interaction between the residues at the termini of the central bulge (I5 and N9) is much reduced in the 11mer compared with the other forms as N9 is shifted away from the E pocket and interacts with ordered waters between the side-chain and the base of the groove. Though there is an asparagine ($N8_{s10R}$) in this position in the s10R peptide it is able to interact through hydrogen bonding with the $R5_{s10R}$ side-chain and so adopts a different conformation (Figure 40A). The backbone of residue I4 in this complex is in an almost identical position to that of the s10R peptide, and this may be due to the reduced interaction between residues 5 and 9, mimicking the effect of fitting in the long $R5_{s10R}$ side-chain. However, the central bulge in the 11mer, while not quite as raised as for KK10, is still slightly higher than s10R at its apex and adopts an exaggerated form of the zigzag of the KK10 peptide backbone between the α1 and α2 helices rather than the more extended conformation of s10R in this region (Figure 41A). The conformation of the 11mer in complex with the HLA B*2705 groove therefore contains elements of both the KK10 and s10R conformations as well as a unique conformation at the E pocket. Comparison of the KI11 structure with that of 11mer peptides crystallised in complex with other HLA class-1 [255; 403] show that all of these structures contain a central bulge with a bed of ordered waters between this bulge and the base of the groove. However, this is stabilised differently in each structure, with the Bade-Doding B*4104/HEEAVSVDRVL (HL11) complex showing intra-peptide interactions between the residues at each end of the bulge, in this case at positions 3 and 9, as well as an α-helical turn towards the end of the bulge (Figure 40B). The Guillame Stewart-Jones B*5703/ KAFSPEVIPMF (KF11) structure contains proline residues at positions 5 and 9 which add rigidity to the peptide conformation (Figure 40C). Neither of these adopts quite the same zigzag conformation that is seen in the KI11 complex.

In the 13mer and 14mer complexes, the intra-peptide interaction between the anchoring residues either side of the central bulge ($I5_{KR13}$ and $I11_{KR13}$, $I5_{KM14}$ and $V12_{KM14}$) returns and is further stabilised through interaction with the E pocket. The shorter side-chains of these residues relative to the s10R peptide means that, as in KK10, the central loop is less spread out in this complex. Both the N and C-terminal regions are more similar to the KK10 structure in these complexes, with the backbone of the peptides in different positions to the s10R complex. Another structure of a 13meric peptide with HLA B*3501 has also been completed[383](Figure 40D), comparison of this with the 13mer structure shows that again both have a central bulge/loop of peptide, however, the central loop is rigid in the B*3501/LPEPLPQGQLTAY (LY13) complex and is raised away from the groove more N-terminally than the KR13 peptide, beginning at position 4 (Figure 40D). The 14mer backbone is also slightly raised away from the groove at either side of the central bulge in comparison to the 14mer, but far less than for the LY13 complex. The rigidity of this peptide is provided by proline residues at positions 2, 4 and 6 as well as the intra-peptide interaction between $P6_{LY13}$ and $L10_{LY13}$; though we cannot clearly see the apex of the KR13 peptide loop, it is also likely to be raised higher above the groove (though being flexible could probably bend) than the LY13 peptide in order to fit the 3 residues that we are unable to see into the tighter loop (Figure 41C).

Our first model for the 14mer in complex with HLA B*2705 is, to our knowledge a completely novel conformation for a peptide bound to HLA class-1. While there are no 14meric complexes for HLA B*2705, there is a structure of a peptide of this length in the HLA B*3501/LPAVVGLSPGEQEY (LY14) complex[396], there is also a structure of a 16meric peptide in complex with HLA B*4103[403]. The 16mer structure (AL16) shows a central region that is unstructured, presumably forming a flexible loop of peptide bulging out from the HLA class-1 groove (Figure 40F). The 14mer structure uploaded to the PDB is complete and shows a central bulged peptide with the top of the loop leaning towards the α1 helix (Figure 40E

Figure 41D). However, in the publication, the structure of the central region of this peptide is referred to as flexible[396]. I must therefore assume that they have been able to complete their model by looking at data with higher levels of noise, so although the structure is given for the loop region of the peptide, it is likely to be only a possible conformation. Both of these structures therefore show wholly or partly flexible central loops with the termini bound in canonical conformations.

A feature of the structure of the KM14-HLA B*2705 complex that may lend additional weight to the model we propose is the shift of the residue R62 $_{B27}$, this residue moves from its position over the peptide in the KK10 complex, where it interacts with Q163 $_{B27}$ in a clamp over the peptide[260; 404; 405]. This clamp has been shown to be flexible[406] and upon TCR binding may adopt an exposed conformation to serve as a contact residue for the TCR[406]. Our structures show that the position of R62$_{B27}$ may also be influenced by the position of L6 $_{peptide}$ (Figure 39) and that the conformation of the peptide chain and side chains in this region may "encourage" R62$_{B27}$ to adopt this more exposed conformation. However, considering the increased distance between the L6 $_{KM14}$ residue and R62$_{B27}$ compared with the 6,7,8 and 13mer complexes, it may be that the "encouragement" to switch position in this case is due to the ability to serve as a contact residue for the extended peptide chain of KM14 at position 10. The presence of considerable residual electron density in the structure at this location indicates that it is likely that R62$_{B27}$ is interacting with something in this position and that it may be the KM14 peptide chain.

This proposed model for the binding of the KM14 peptide is significantly different to previously crystallised HLA class-1 complexes presenting long epitopes[396; 403] and shows a novel peptide conformation within the HLA class-1 groove. Though we acknowledge that this data is insufficient to conclude that this model is in any way definitive, we suggest that it is a possible scenario for binding of the KM14 peptide.

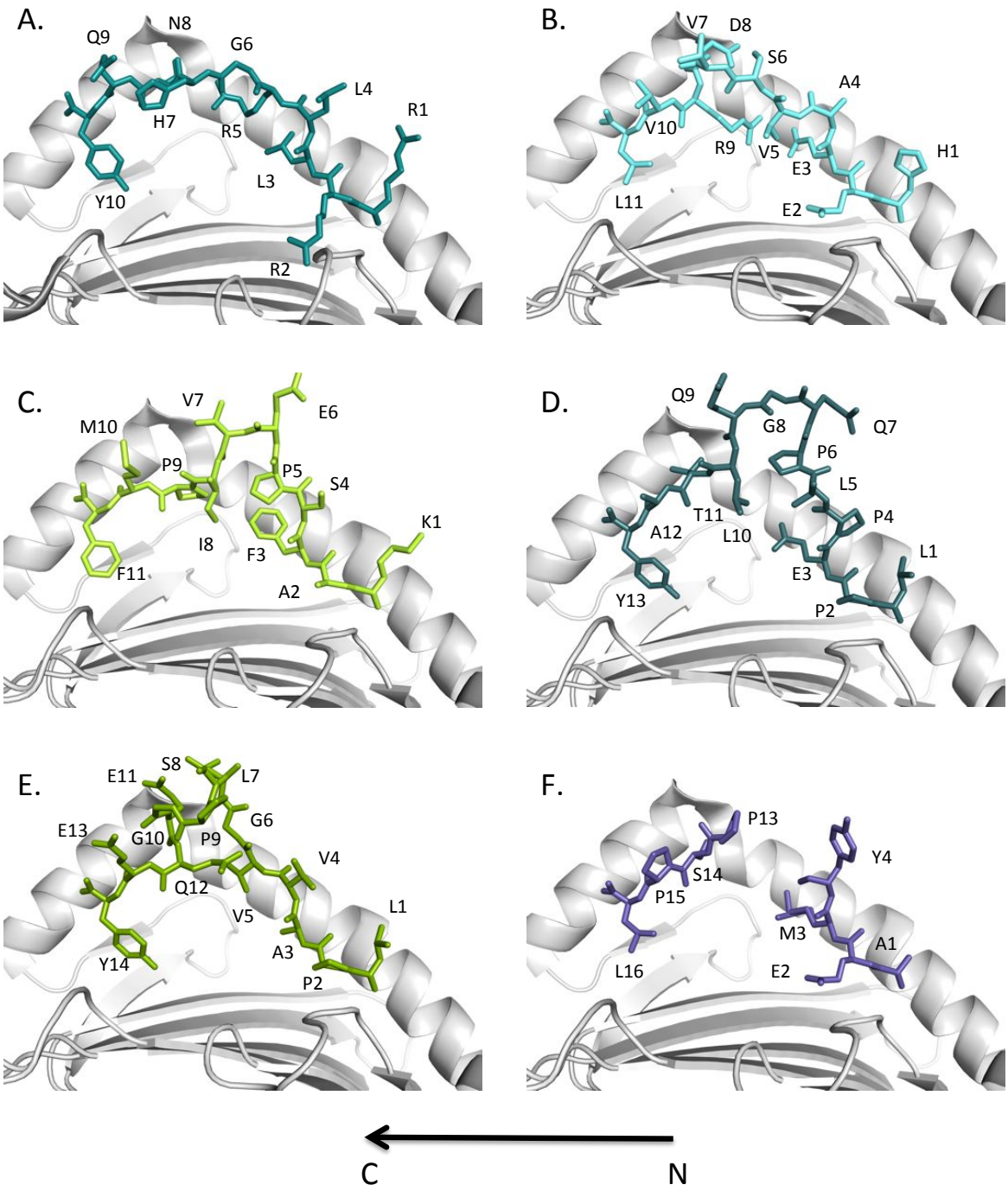


Figure 40: Side view of conformations of other long peptide forms that have been previously crystallised against an HLA class-1 groove **A.** s10R **B.** HL11 **C.** KF11 **D.** LY13 **E.** LY14 **F.** AL16

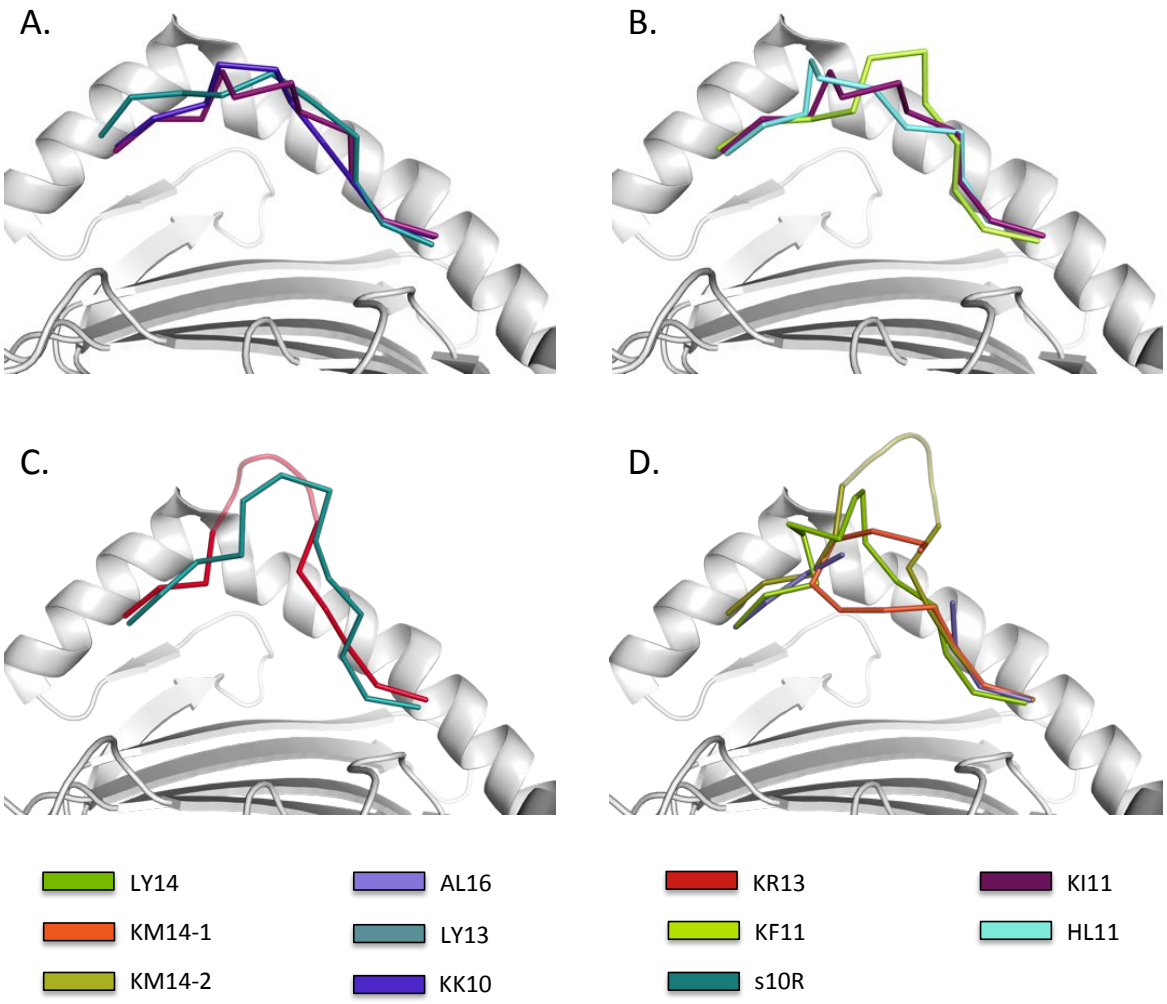


Figure 41: Backbones of previously crystallised long HLA class 1 epitopes shown grouped by epitope length with the appropriate KK10 epitope forms against an HLA class-1 molecule **A.** KK10, KI11 and s10R **B.** KI11, HL11 and KF11 **C.** KR13 and LY13 **D.** KM14, LY14 and AL16.

5. Structural and sequence dependency of immune cell interactions with KK10 epitope form-HLA complexes

I would like to thank Dr. Edmund Wee of the Iversen Lab for his advice and assistance with the viral sequencing and ELISpot assays carried out in this chapter.

5.1. Introduction

The interaction of the HLA class-1 complex with the immune system occurs through a number of receptors. The peptide-dependent interaction with αβ T-cell receptors has been the most studied of these and is vital for the development of immune responses to pathogens including HIV-1. Yet, it has also become clear that the HLA molecule interacts in a peptide dependent manner with another class of receptors, the KIR and that these have a significant impact on the host response to infection. In addition there are peptide dependent[321] and independent interactions[127; 385; 386] with a range of receptors that modify the development of the immune response. While receptors may engage with the molecule at different points of contact, the structure and characteristics of the surface of the HLA determine the interaction. For peptide dependent interactions such as those with TCR and KIR, this surface includes the peptide epitope and is modified by the sequence and conformation of the epitope bound. Using our structural data, TCR sequencing and prior knowledge of these interactions, we wanted to understand how the KK10 epitope forms change this surface and affect how the immune system sees and responds to these complexes.

5.1.1. The recognition and binding of HLA class-1 complexes by the T-cell receptor (TCR)

Recognition of the HLA-peptide complex by the TCR is due to contact with both the HLA and the peptide by three CDR loops in each chain of the αβ TCR (Chapter 3.1.1). Structures of both HLA class-1 and class-2 in complex with specific TCR have been determined[258; 265; 266; 267; 268] and show how the TCR docks onto the HLA surface in each case. Binding of the HLA class-1 molecule typically involves the TCR docking diagonally across the peptide-binding groove[272]. The binding of the HLA class-2 by TCR has been described as orthogonal, though in the majority of cases it is still slightly diagonally bound[123; 273]. Of the TCR structures so far

completed, there appears to be more divergence in binding angle of the TCR to HLA class-1 than to HLA class-2[124] however, in each case the Vα-domain of the TCR is positioned over the α2 helix of the HLA and the Vβ domain over the α1 helix. In typical binding positions which places the Vβ CDR3 loop over the central region of the peptide binding groove, hence the importance of this loop in peptide dependent recognition of the HLA-epitope complex.

Several characteristics of the HLA could explain these differences in docking between the two classes. Firstly, the peptide bound in the HLA class-1 groove is bound at both the N and C termini, in order to be accommodated in the class-1 groove; many peptide epitopes therefore bulge in the centre of the groove. As the HLA class-2 peptides are not bound at the termini, the majority of peptides adopt an extended conformation in the class-2 binding groove[407]. Analysis of TCR contact sites show that while TCR specific for HLA class-1 complexes recognise both conformation of the backbone and peptide side-chains, the TCR specific for class 2 are more limited to side-chain interactions, though recognition and binding of small structural motifs has been shown in this interaction[372]. The differences in the structure of peptides between the two classes could affect the angle at which the TCR is best able to contact both the peptide and the HLA, and the more variable conformation of the class-1 epitopes within the groove could lead to a more diverse range of binding modes.

A second possible reason for this difference is the angle at which the HLA is presented on the surface of the cell[124] The class-1 HLA is tethered to the cell through a single trans-membrane helix on the heavy chain, meaning that it is able to adopt a number of orientations on the cell surface, presumably on docking of the TCR, the HLA class-1 molecule would be extended away from the surface and outward facing. The tethering of HLA class-2 is through trans-membrane helices on both the α and β chains of the molecule. This somewhat fixes the orientation of class-2 relative to the surface, it has been proposed that the binding groove in class 2 would be at a tilted angle to the surface rather than facing directly outward[124] The differences in the orientation of the two HLA classes may affect the binding angle that the

TCR must adopt in order to dock at contact sites and further helps to explain the greater variability of HLA class-1 TCR binding.

The orientation of the TCR on the HLA surface has been demonstrated in one case to be relevant for TCR activation[278], in this study the crystal structure of the agonist peptide- HLA complex and the TCR showed binding of the TCR in a typical, diagonal manner on the HLA surface. An artificial peptide that the TCR bound with higher affinity than the agonist, but did not generate a cytokine response to, was also crystallised and showed a markedly different binding orientation, in this case centred over the α1 helix and nearly parallel to the HLA binding groove. This suggests that how a TCR binds HLA can be as important as if the TCR binds in terms of generating an immune response, and may lend support to the idea of the TCR as a mechanotransductor[408]. While this was done using an artificial peptide, it may be of relevance in our work as we show a number of TCR are present that cross-recognise our KK10 epitope-forms (Chapters 3.2.3 and 3.3.1) and introduces the possibility that not all forms that are recognised by dextramer binding need necessarily be agonists of the T-cell response.

5.1.2. KIR3DL1 recognition of HLA-B

The binding of KIR3DL1 to the HLA B*5701 complex occurs at two discontinuous sites across the three KIR3DL1 domains. D0 binds an area of the HLA close to β2-microglobulin which is largely conserved in HLA-A and B, this binding predominantly involved contact with the backbones of residues in the region, giving "innate" recognition of HLA class-1. This interaction site is not seen in previously crystallised HLA/KIR complexes and the binding of KIR3DL1 is diminished without it[369]. The D1 and D2 domains form a second binding site, with the D1 domain docked over the α1 helix of the HLA and the D2 domain docked over the α2 helix close to the C-terminal end of the peptide binding groove. The HLA B*5701 peptide

is contacted through a water mediated interaction involving residues Y200 and E282 in the D1 and D2 domains, as well as a Van-der-Waals interaction with the side-chain L166 of the D1 domain, all contacting the C-1 position of the peptide epitope.

HLA-B*2705 contains a Bw4 motif which is a known conserved KIR3DL1 recognition motif[64; 409] and interacts with the D1 domain in the HLA B*5701-KIR3DL1 structure. In this interaction HLA B*5701 residues E76, R79 and R83 form a network of salt-bridge interactions with each other and the D1 domain, this creates a small hydrophobic pocket in which $I80_{B57}$ is positioned, it then makes a single Van-der-Waals contact with $L166_{KIR3DL1}$. The mutation of the I80 residue to T in HLA B*2705 was shown to slightly reduce binding affinity, possibly due to the disruption of the salt-bridge interactions by the T80 side-chain[369]. The interactions of the D2 domain with HLA B*5701 showed the importance of three residues on the α2 helix for interaction with KIR3DL1, I142, K146 and A149[369]. These residues are all conserved in the HLA B*2705 sequence and should enable recognition by KIR3DL1.

5.2. Impact of structural conformation of the TCR binding faces of KK10 epitope form-HLA complexes on selection and recognition by TCR

In the previous chapters of this thesis, I have described the structures of the KK10 epitope forms in complex with HLA B*2705 as well as the sequences of the TCR that recognise these epitope forms. In order to better understand these data, I have conducted further analyses of the KK10 epitope form structures. These analyses include the production of electrostatic models of the surface of the complexes as well as determination of the solvent exposure of the amino acid residues of the peptides when bound in the HLA B*2705 groove. In addition, the functional effects of the truncated KK10 epitope forms were explored.

5.2.1. Solvent exposure of the HLA bound peptides and structural features accessible for TCR recognition

The electrostatic models of the surfaces of the KK10 epitope form-HLA complexes show the spread of net charge over the surface of the molecule, allowing us to observe it's surface conformation and properties. The electrostatic potential of the surfaces is dependent upon the relative positions and sequence of the epitope residues. The shared HLA sequence and folding present in these complexes means that for regions removed from the peptide binding groove, the electrostatic surfaces are highly similar, while surrounding the peptides, there are significant changes between the epitope forms. Despite this, the sequence identity and shared conformation of particular areas of the peptide in these epitope forms does provide some areas of similarity between some or all of the structures.

One area of similarity in the complexes is that surrounding the N-terminal region of the peptide-binding groove, because all of the peptides share similar binding conformations in

this region and share a surface exposed $K1_{peptide}$ (Figure 42). This results in a small positive charge over the area surrounding this residue, the degree of charge varies between the structures, with the HLA B*2705-KK10 complex having a higher net charge in this area than the other complexes. This may be due to the $R62_{B27}$ residue adopting an alternative conformation in the KK10 structure compared with the other epitope forms. It is likely that in solution this arginine would be partly flexible[406] and the difference in the degree of positive charge in this region between the epitope forms may be less pronounced.

This area is the only peptide-surrounding region to have a similar electrostatic surface potential in all of the epitope forms. However, several forms share features in other parts of the peptide surface. The KL6, KG7 and KL8 complexes (Figure 42A, B and C) all have peptides that do not stretch across the peptide binding groove, resulting in an F-pocket without bound peptide in all three of the complexes. The empty F-pocket forms a highly negatively charged region exposed to the surface in these epitope forms; this constitutes a considerable difference between the surface of the complexes with epitopes that do not stretch across the binding groove and the surface of complexes with epitope forms that fill the F-pocket and so may have a marked on the ability of T-cells to cross recognise the two groups. However, in the crystal structures we obtained for these epitope forms, residual electron density is present in the F-pockets of these structures, suggesting a molecule was picked up from solution and fills the F-pocket. It is possible that this also occurs *in vivo* and this might alter the surface potential of the complex considerably from that shown in our models.

In these three complexes, the C-terminal residues of the peptides are not constrained by binding to the HLA and so are likely flexible. Therefore, while shown in a fixed position in the structures, these flexible residues may adopt several conformations and their electrostatic potential will be diffused across these conformations. Residues 7-9 in the KR13 and KM14 complexes (Figure 42F and G) are also poorly defined in the crystal structures and are likely

to be partly flexible, meaning that their contribution to the electrostatic potential in these complexes will also be more diffuse than indicated in our models of the surface.

Despite the sequence similarities between the epitope forms in our complexes, the remainder of the surface potential surrounding the peptides and binding groove varies substantially. This is likely due to the differences in conformation and relative position of the residues in each epitope form induced by fitting the different C-terminal residues into the F-pocket in each complex in the extended epitope forms.

In the KK10 structure (Figure 42D), the positive charge over the N-terminus of the peptide-binding groove extends up to the C_α of $W3_{KK10}$. Following this position, the surface shows little net charge in the central region up to the $L8_{KK10}$ residue, though there is a patch of positive charge under the $L8_{KK10}$ position on the surface of the HLA. At residue $N9_{KK10}$ a net negative charge is present on the $\alpha1_{B27}$ helix side of the binding groove, near the position of the $N9_{KK10}$ side-chain. It is likely that the $E76_{B27}$ side-chain is a contributor to the negative charge in this region. The change in the position of this residue in the KI11 and KR13 complexes (Figure 42E and F) along with possible interactions with these peptides may help to explain a lack of negative charge at the equivalent positions of these peptides. The $\alpha2_{B27}$ helix side of the KK10 peptide has largely no net charge. At the C-terminus of the KK10 peptide, the $\alpha1_{B27}$ helix side of the binding groove continues the region of negative charge, while the surface on the side of the peptide chain facing the $\alpha2_{B27}$ helix is somewhat positively charged.

In the KI11 complex (Figure 42E), the N-terminal positive charge appears slightly reduced compared to KK10 (though scales are based on approximate models for solvation, discretization and iterative solving for the calculated energies, these processes introduce individual errors into the model and small deviations in atom positions in individual structures can occasionally have disproportionate effects, so direct comparison of charge

intensity may not be realistic), and a slight negative charge is present on the $\alpha2_{B27}$ helix side of the groove around the $I5_{KI11}$ residue. The central region of the KI11 peptide (Figure 42E), between residues $L6_{KI11}$ and $N9_{KI11}$ shows little net charge on the surface, as for the central residues in the KK10 peptide. However, the shift of the KI11 peptide towards the $\alpha1_{B27}$ helix and movement of the $N9_{KI11}$ residue away from the E pocket leaves a small pocket on the $\alpha2_{B27}$ helix side of the groove in this central region that is negatively charged. The $K10_{KI11}$ residue gives a slight positive charge on the $\alpha1_{B27}$ helix side of the peptide groove at this position, compared with a negative charge being present at the equivalent position in the KK10 complex (at $N9_{KK10}$). On the $\alpha2_{B27}$ helix side of the groove at this position, a small, negatively charged pocket is present between the helix and the peptide chain. Finally at the C-terminus of the groove, there is a slight positive charge towards the $\alpha2_{B27}$ helical side of the groove and a patch of negative charge towards the $\alpha1_{B27}$ helix, as for the KK10 complex.

The KR13 complex (Figure 42F) is quite different from the previous two complexes in that it has very little negative charge on the surface of the peptide and binding groove, the exception being a slight negative charge on the $\alpha2_{B27}$ helix side of the binding groove at the N-terminus (though the peptide chain itself remains positively charged as for the other structures). This may occur as the exposed residues in this structure are mainly uncharged or positively charged, additionally interactions of the peptide neutralise negative charges present on the surface of the binding groove in this complex and there is a shift in the exposure of pockets of the binding groove compared with the KK10 and KI11 structures. Following the positively charged N-terminal region of the peptide, a slight positive charge is retained up to the $I5_{KR13}$ residue and into part of the central bulge. This region is largely uncharged on the surface of the peptide, however, the shift of the central bulge towards the $\alpha2_{B27}$ helix in this complex exposes a pocket between the peptide and the $\alpha1_{B27}$ helix that is positively charged. This is similar, but slightly larger than at the equivalent position of the KK10 complex, reflecting a slightly larger shift of the peptide chain towards the $\alpha2_{B27}$ helix in this structure. Following this region, the $K10_{KR13}$ residue gives a slight negative charge

towards the $\alpha1_{B27}$ helix side of the groove, the salt bridge of this residue with $E76_{B27}$, although over an increased distance compared to $K10_{KI11}$ serves to neutralise the negative charge this residue contributes to the region in the KK10 complex. The C-terminus of the peptide and binding groove are positively charged on the $\alpha2_{B27}$ helical side of the peptide chain as in the KK10 and KI11 complexes, however, the $\alpha1_{B27}$ helical side is uncharged in this structure, differing from the negative charges present in both the KK10 and KI11 structures.

The final C-terminally extended KK10 epitope form, the KM14 peptide, may exist in multiple conformations in the binding groove as indicated by two different structures obtained from crystallisation of the complex in two different buffer conditions (Table 29). However, we only describe one structure (Figure 42G) in this chapter as a complete peptide gives a better model of the electrostatic surface present, and for the first KM14 conformation (Figure 31) we are unable to model the peptide following $I11_{KM14}$ as no density can be seen. The N-terminal region of the peptide is again slightly positively charged, however, as in the KR13 complex, there is a region of negative charge present on the $\alpha2_{B27}$ helix side of the groove. The slight positive charge continues along the peptide chain up to the beginning of the central bulge. From residues $L6_{KM14}$ to $N9_{KM14}$ the top surface of the peptide chain is largely uncharged, however a positive charge is present on the sides of the peptide bulge and on the $\alpha2_{B27}$ helix side of the peptide-binding groove. On the $\alpha1_{B27}$ helix side of the central bulge, there is a small, negatively charged pocket exposed underneath the peptide chain. This is in a slightly different (but somewhat overlapping) position to the pocket exposed in the KR13 complex, however appears to have the opposite charge to the pocket in the 13mer.

The $K10_{KM14}$ side-chain continues the positive charge present on the $\alpha2_{B27}$ helical side of the peptide chain at the end of the central bulge, while the $I11_{KM14}$ residue facing the $\alpha1_{B27}$ helix gives an uncharged surface on this side of the chain, further towards the $\alpha1_{B27}$ helix and slightly C-terminal to the $I11_{KM14}$ residue, a region of negative charge is present, most likely contributed by the $E76_{B27}$ side-chain at this position. On the $\alpha2_{B27}$ helix side of the groove, the

positive charge is extended further C-terminally by the side-chain of the $R13_{KM14}$ residue, a small pocket is present on this side of the peptide groove at this position and also gives a net positive charge (as opposed to the positive net charge of the pocket at this position in the KI11 structure). Finally the C-terminus of the peptide is largely positively charged in the KM14 complex, however there is a patch of negative charge on the $\alpha1_{B27}$ helix side of the groove, similar to that in the KK10 and KI11 structures.

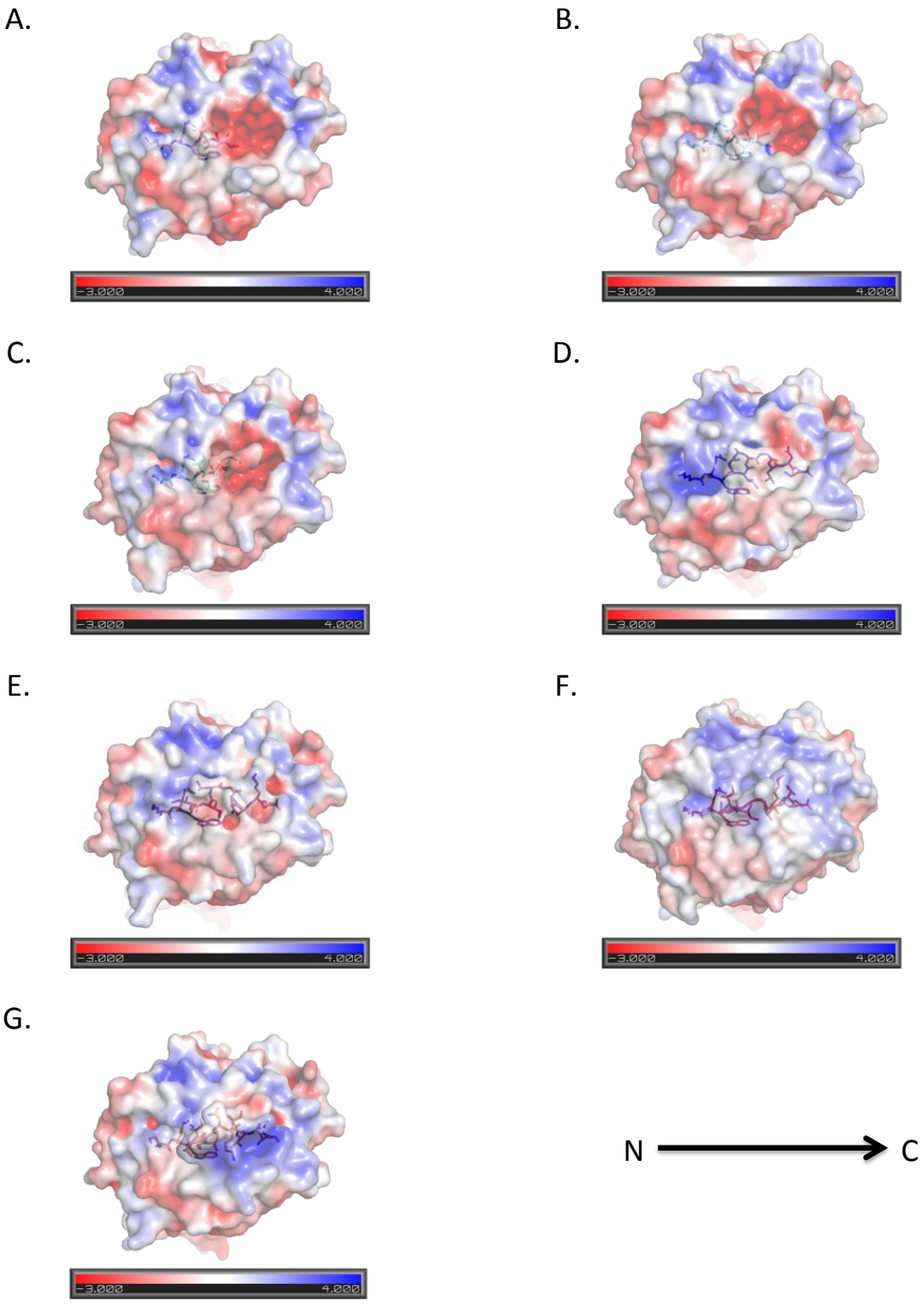


Figure 42: Electrostatic surface views of peptide-HLA B*2705 complexes showing the TCR contacting face of the complex and bound peptide **A.** KL6 **B.** KG7 **C.** KL8 **D.** KK10 **E.** KI11 **F.** KR13 **G.** KM14

The areas of similarity found in the electrostatic surfaces of the KK10 epitope form complexes may allow cross recognition of some of the epitope forms, while areas of difference may elicit clonotypes specific for certain forms. The number of unique and shared structural motifs visible on the surface of the HLA-peptide complexes along with the degree of steric hindrance to TCR binding that may arise from protruding and/or flexible loops will also contribute to the ability of the complexes to be recognised by TCR and levels of cross recognition.

The surface area of the HLA-bound peptide that is exposed to solvent (Table 32) gives an indication of the percentage of peptide exposed, and in combination with the surrounding HLA, suggests what unique features are likely to be available for TCR recognition. It is unclear whether the TCR will be able to bind and recognise unique features in these regions if parts of the peptide chain appear to be flexible in the structure. It has previously been suggested that such TCR recognition may involve the peptide being "pushed" further into the groove or over part of the HLA surface upon TCR docking[410]. A prior structure of a TCR in complex with a 13mer peptide epitope bound to HLA class-1 showed an increased percentage of the TCR CDR loops were involved in peptide recognition relative to structures using shorter epitopes, and that this is likely a result of the large central bulge of the peptide reducing the accessibility to parts of the HLA surface[383]. However, when crystallised without the TCR, this peptide epitope was not observed to have flexibility in the central bulge, and so different patterns of binding might be seen in these cases.

Percentage solvent exposed surface area for the peptide was calculated using the Protein interfaces, surfaces and assemblies service at European Bioinformatics Institute (PDBePISA)[411]. This was performed for the entire peptide chain (Table 32) and for individual residues (Table 31). The results show that the C-terminally extended KK10 epitope forms increase in the amount of surface area exposed as length of the epitope form increases, as well as in the percentage of total surface area exposed to solution. The increase is larger in

the KR13 and KM14 complexes (from 28.4% in the KK10 to 37.2% and 44.6% in the KR13 and KM14 complexes respectively) compared with the KI11 complex (28.4% to 29.0% in the KI11 complex) indicating that accommodating the extra residue in the KI11 epitope form does not greatly increase the amount of the peptide that bulges out of the groove, perhaps due to the shift of the central region of the peptide toward the $\alpha 1_{B27}$ helix in this complex.

The truncated KL6 epitope form displays the least surface area exposed to solution of any epitope form, though the percentage of total surface area exposed is higher than in both the KK10 and KI11 complexes (32.4%). Similarly, the KG7 complex shows the next lowest surface area exposed, though it is only 7 $Å^2$ smaller than the KK10 complex (Table 32) and is a higher percentage of the total surface area of the peptide (36.0%). The surface area exposed in the KL8 complex is greater than in both the KK10 and KI11 complexes (at 602 $Å^2$ versus 459 $Å^2$ and 484 $Å^2$ in the KK10 and KI11 complexes) and is second only to the KM14 complex in terms of the percentage of the epitope exposed to solution. The relatively high percentages of peptide surface area exposed to solvent and in the KL8 complex, total area exposed compared with the KK10 complex reflects the lack of binding at the C-terminus of these peptides (Table 31). This results in the C-terminal region being almost entirely solvent exposed, however, this may not result in an increased number of features available for TCR recognition as the peptides are flexible and may prevent TCR docking by obscuring contact residues on the surface of the HLA molecule. Alternatively, docking of a TCR may force these small peptides into the HLA binding groove and dramatically reduce the amount of exposed peptide available for binding.

Comparison of the percentage solvent exposure by residue clearly shows the increase in solvent exposure of the C-terminal region of the peptide chains in the truncated epitope forms (Table 31). It also gives a simple comparison of conformation of the peptides and the relative exposure of different residues between epitope forms that may be important in TCR recognition. The solvent exposure of the N-terminal 4 residues is similar in all the epitope

forms (excepting a relative decrease in exposure for $K1_{KI11}$) reflecting the similarity in structure of this region in all of the crystal structures (Figure 31). This is followed by an increase in solvent exposure at position 5 for all but the KK10 and KI11 epitope forms (which show a slight decrease) followed by a rise in all epitope forms towards the central bulge of the peptide in the extended epitope forms, or an unbound C-terminus in the truncated forms, the L6 residue is highly (over 75%) exposed in all epitope forms. In the KK10 epitope form, the $G7_{KK10}$ residue is also highly exposed (86.9%), this exposure then decreases in position 8 (28.7%) at the C-terminal anchor of the central bulge, rises slightly as the side-chain of $N9_{KK10}$ faces out of the groove and decreases towards the C-terminal anchor.

This is slightly different in the KI11 complex, where the exposure of $G7_{KI11}$ is considerably lower (21.7%) as it is stacked next to the $\alpha1_{B27}$ helix. The exposure then increases again for $L8_{KI11}$ as it now forms part of the central bulge, before decreasing at the $N9_{KI11}$ anchor. Unlike for the KK10 epitope, the exposure of the $K10_{KI11}$ (49.5%) is increased in the complex as the side-chain is now exposed to solution instead of forming the anchor. This exposure is maintained in the KR13 and KM14 complexes. Both the KR13 and KM14 complexes follow a very similar pattern of surface exposure through the peptide, the main differences between them being an extra residue included in the central bulge region ($I11_{KM14}$) and the increase in solvent exposure at the C-1 position of the KM14 complex due to the presence of the long, exposed R13 side-chain.

Peptide in complex	Total peptide surface area Å^2	Solvent exposed surface area Å^2	% Peptide exposed
KL6	1191	386	32.4
KG7	1256	452	36.0
KL8	1389	602	43.3
KK10	1619	459	28.4
KI11	1668	484	29.0
KR13	2031	754	37.2
KM14	2177	972	44.6

Table 32: Total and solvent exposed peptide surface area for HLA-KK10 epitope form complexes.

Residue	% Exposed						
	KL6	KG7	KL8	KK10	KI11	KR13	KM14
LYS 1	17.9	20.0	18.3	18.1	5.66	18.4	13.5
ARG 2	4.80	4.81	3.86	1.72	1.32	2.35	3.54
TRP 3	16.2	15.4	14.7	14.6	8.17	15.2	14.9
ILE 4	36.4	42.3	34.5	42.1	50.5	34.7	34.2
ILE 5	46.5	40.0	58.1	29.7	35.0	45.6	56.0
LEU 6	82.8	80.3	76.2	100	92.5	80.6	86.6
GLY 7		91.0	100	86.9	21.7	99.1	92.5
LEU 8			99.1	28.7	73.4	100	100
ASN 9				34.6	12.5	51.4	100
LYS 10				2.20	49.5	74.2	62.4
ILE 11					1.31	28.2	72.8
VAL 12						22.2	28.2
ARG 13						2.60	63.6
MET 14							1.87

Table 31: Percentage solvent exposure by peptide residue for KK10 epitope forms.

5.2.2. Effects of truncated, partly flexible KK10 epitope forms on T-cell recognition and responses.

In order to assess the CD8+ T-cell response to the truncated epitope forms, we measured the production of cytokines by patient CD8+ T-cells following incubation with the KK10 epitope forms using intracellular cytokine staining. This was done for 4 cytokines, CD107a, interleukin 2 (IL2), interferon gamma (IFNγ) and tumour necrosis factor alpha (TNFα) and has to date has been completed for the group 1 patient 081077, this patient was previously found to recognise KK10 and the extended KK10 epitope forms in our flow cytometry assays (Chapter 3.2). Patient 081077, despite receiving treatment throughout the study, maintained the response to KK10 at all time-points tested and developed a response to KM14 in later time-points. We were therefore able to compare responses to epitopes recognised in dextramer staining with those to the truncated epitope forms in this patient (Figure 43, Figure S 15, Figure S 16).

In this patient we were able to show production of CD107a, IFNγ and TNFα in CD8+ T-cells following incubation with KK10 and extended KK10 epitope forms. Both CD107a and IFNγ were produced in a dose dependent manner in the assay, reaching saturation as peptide concentration increased (Figure S 16). However, cytokine production by CD8+ T-cells was not observed for the truncated epitope forms (Figure 43. This includes the KL8 epitope form, which was recognised by other patients' CD8+ T-cells in the dextramer-binding assay, Chapter 3.2). This indicates that no T-cell activation occurs following presentation of these epitope forms and could be due to absence or the quality of recognition by T-cells of some or all of the epitope forms, or due to a lack of presentation of these forms on the cell surface.

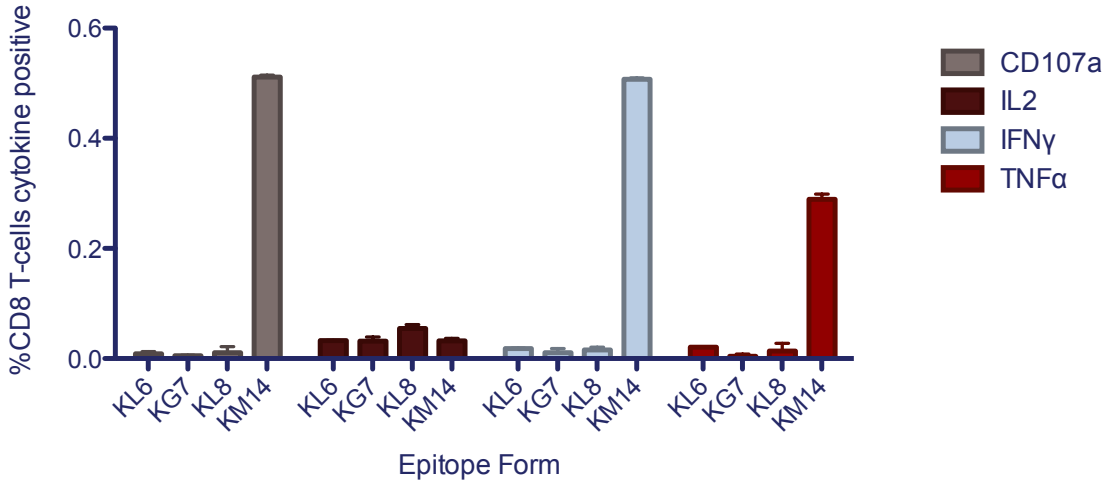


Figure 43: Percentage of CD8+ T-cells producing cytokines at maximum peptide concentrations following incubation with KK10 epitope forms in patient 081077.

In the light of processing, trimming, refolding and structural data (Chapters 2.2, 2.3, 4.6 and 4.7) the possibility that this negative result is due to lack of presentation on the cell surface seems unlikely. However, an IFNγ ELISpot assay was designed in order to test for the presentation of these epitope forms on the cell surface and also to assess their effect on immune responses. This assay used T2 cells as target cells for presentation of the KK10 epitope forms, which were selected due to presentation shown in a previous assay and the lack of variability, limitations in number and presence of other HLA in patient samples.

The previous assay showed that epitope forms from KL8 to KM14 were presented by HLA B*2705 in this context. T2 cells are a TAP deficient, HLA class-1 deficient murine B-cell line. These cells do not display HLA or present HLA class-1 peptide antigens on their surface unless transfected with specific HLA molecules. We used T2 cells transfected with HLA B*2705 (kindly provided by Dr Simon Kollnberger) to assess the ability of KK10 epitope forms to stabilise HLA B*2705 in these cells and be presented on the cell surface. Binding was shown for all HLA B*2705 epitope forms tested (Figure 44), showing *in vivo* presentation of the truncated KL8 epitope form and extended KK10 epitope forms. Unusually

we saw some binding for the HLA B*57 TW10 epitope which was used as a control, though the KL8 epitope containing an R132K mutation showed no binding.

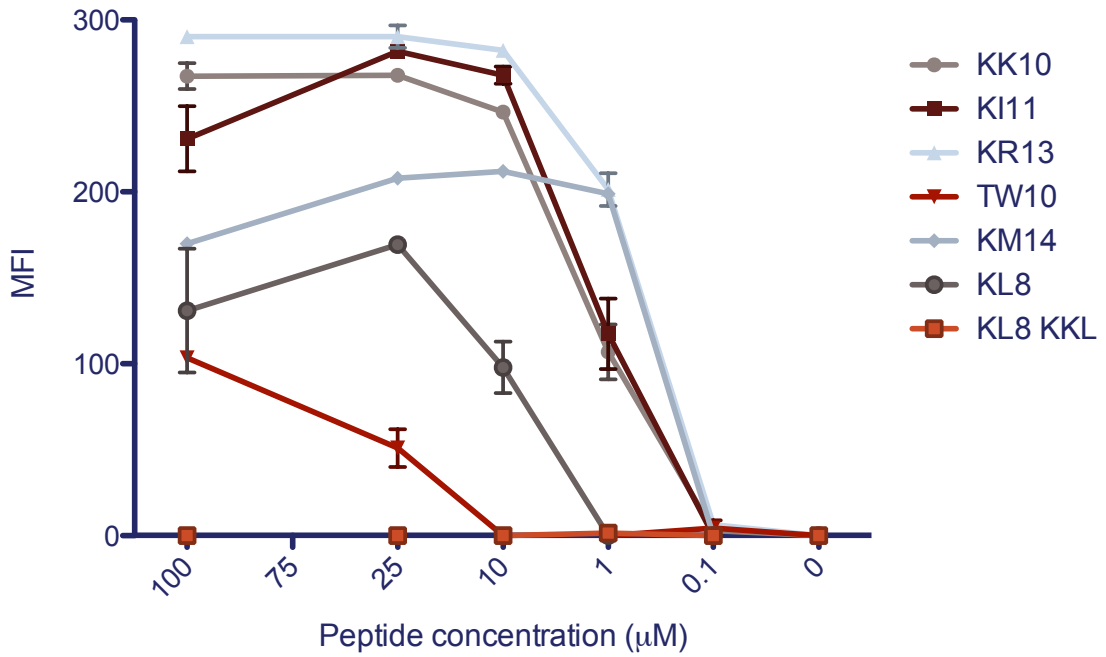


Figure 44: T2 HLA B*2705 stabilisation assay. Fluorescence indicates binding of refolded HLA B*2705 on the cell surface by the HLA B*2705 specific ME-1 antibody following incubation with KK10 epitope form peptides.

Given the demonstrated presentation of the KK10 epitope complexes using the T2 cells, the IFNγ ELISpot assay was designed using the T2 cells for antigen presentation. This assay used the KK10 epitope and truncated epitope form peptides (either alone or in combination). These peptides were first incubated with T2 cells transfected with B*2705, a KK10 specific CD8+ T-cell clone G12C (kind gift from Professor Victor Appay) was then mixed with the T2 cells and incubated overnight on a 96 well ELISpot plate.

Titration of the KK10 peptide using this experimental set-up allowed us to determine the half maximal inhibitory (saturating in this case) concentration (IC50) for this peptide (Figure 45F) and so calculate amounts needed for the KK10 truncated epitope peptide forms according to ratios obtained from proteasomal processing of the region. The IC50 KK10 was

then used in a competition assay with the truncated epitope forms. Given that, in our previous assay, we observed no CD8+ T-cell cytokine response to the truncated epitope forms (and Figure S 17), we hypothesized that presentation of these epitope forms on the cell surface in competition with the KK10 epitope should give a reduction in IFNγ production with increasing concentration of the truncated epitope forms.

An initial assay has shown that for the KG7 and KL8 epitope forms, there is a decrease in the production of IFNγ as the concentration of the truncated epitope form increases (Figure 45B and C). The assay also shows a small decrease when all of the truncated epitope forms (KI4, KI5, KL6, KG7 and KL8) are combined (Figure 45E). However, data for the KL6 epitope form and for the combination of truncated forms up to the KG7 peptide in length are inconsistent (Figure 45A and D). Replication of this assay using increased numbers of replicates as well as a higher affinity KK10 clone should allow us to obtain more accurate measurements of cytokine production (Repeat data added prior to final submission can be found in appendix Figure S 18). In this initial assay, low numbers of cells per well (200), adopted from a recommended protocol, meant that small variations in cell number between repeats may have considerably affected the data. In addition, for some peptides and pools we were not able to test the higher ratios indicated by the proteasomal processing data as the DMSO concentration would have caused significant toxicity.

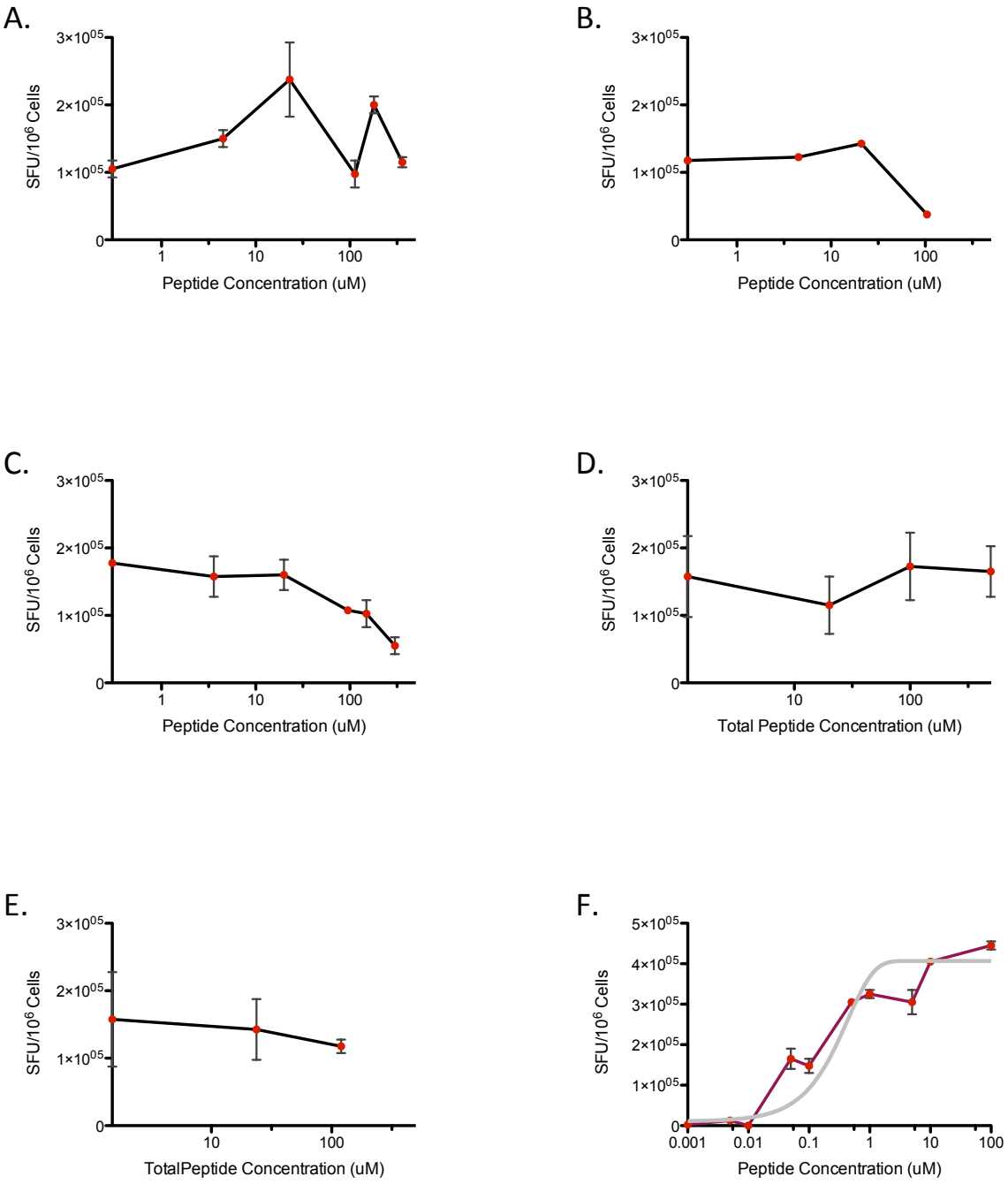


Figure 45: IFNγ secretion given in spot forming units per million cells in ELISpot KK10 competition assays for **A.** KL6 **B.** KG7 **C.** KL8 **D.** Truncated epitope forms from KI4-KG7 **E.** Truncated epitope forms from KI4-KL8 and **F.** The titration of IFNγ secretion in response to the KK10 peptide to obtain an IC50 value, overlaid on data is the fitted one-site binding curve giving the IC50 concentration.

5.3. Footprints of immune recognition on the viral p24 Gag sequence

Recognition of viral epitopes by the cellular arm of the immune system and CD8+ T-cell responses targeted against these epitopes exert selection pressure on the replication of the viral population within the host[17; 51; 52; 58; 257]. This pressure acts against viral sequences that continue to contain and present the targeted epitope to the immune system, as cells infected with these viral sequences are destroyed, preventing viral replication. Any HIV that either does not contain or does not present targeted epitopes will therefore have a survival advantage relative to those that do and selection will result in outgrowth of such viruses. These sequences will differ from the consensus sequence at residues which have allowed the virus to escape from the immune response, whether this is through loss of immune cell recognition[17; 412], HLA binding[256] or alteration of proteasomal processing[238; 343; 350; 413; 414]. These changes are known as escape mutations and allow us to see the "footprints" of immune responses that have exerted selection pressure on the virus.

Viral sequencing was carried out for p17 and p24 gag regions of the virus on HLA B*2705 HIV-1 patient samples used for flow cytometry and cell sorting using nested PCR. These sequences were then proofread, aligned and the region of p24 surrounding the KK10 epitope analysed for mutation from the clade B consensus sequence. Known or proposed escape mutations were compared between patients from groups 1 and 2 (Figure 46). The two common escape mutations in the KK10 sequence R132K and L136M were found in a large percentage of the sequences of the group 1 patients (50.4% and 43.6% respectively)(Figure 46C) but were present in significantly fewer sequences in the group 2 patients (11.6% and 2.33%, P= <0.0001 for both mutations–Fisher's exact test). The R132K mutation was associated with the L136M mutation in most cases excepting 9 sequences in group 1 where the R132K mutation was associated with an L136I mutation (patients 101074, 060473, 100967, 030869 and 081077) and 1 sequence in group 1 where the L136M mutation was

found with the wild-type R132 residue (patient 081077). The L136Q mutation was also observed in both groups in a total of 3 sequences (patients 111156 and 060473).

This result confirms that the KK10 recognition detected through dextramer staining reflects the presence of an effective KK10 response *in vivo*, as selection pressure in the group with the KK10 recognition is significantly higher than in patients without this recognition. However, HIV from some patients in group-2 did show these mutations (050875, 170466 and 111156), though this was at lower frequencies than for the group 1 sequences (maximum 50% of viral population versus 100% maximum observed in group 1 patients).

One mutation within the KK10 epitope sequence occurs at a position that we propose will impact recognition by the KIR3DL1 receptor in the light of previous studies[259; 300; 415] and our own results (Figure 47), the N139D mutation, was also observed in both patient groups (1.71% group 1, 14.0% group 2). The difference in distribution between the total sequences in the two groups was significant (P=0.0051, Fisher's exact test). However, it should be noted that the number of patients having the mutation was the same in both groups (2 in each). The differences in frequency and total sequences between the groups give a significant difference in distribution of this mutation, but a larger sample size may indicate whether this translates into a difference in numbers of patients with the mutation in each group.

Two flanking mutations associated with changes in proteasomal processing of the KK10 epitope, N120S and N120H (Astrid Iversen, unpublished data), were also observed in the HLA B*2705 patients (Figure 46). The distribution of the N120S sequence was significantly different between groups 1 and 2 (12% versus 25.6%, P=0.0486 Fisher's exact test)(Figure 46C), while the N120H mutation showed no difference between the two patient groups (29.1% versus 41.9%, P=0.1327 Fisher's exact test). All sequences containing the N120S mutation were observed to have the consensus KK10 epitope sequence, while the N120H mutation was observed in KK10 consensus containing sequences in 22 of 34 sequences in

group-1 and in all 18 N120H containing sequences in group-2 (Figure 46D). The association of both the N120S mutation and the N120H mutation was significantly associated with maintenance of a consensus KK10 sequence (P= <0.0001 N120S, P=0.0012 N120H, Fisher's exact test).

Mutations from consensus were observed at almost every position in the region in at least one viral sequence (Figure 46A), however, there were a number of mutations within or just outside the region shown that, while having no prior known impact on KK10, did occur in multiple sequences and/or patients and were sometimes associated with uneven distribution between the two patient groups. The first of these is the L111P mutation, present in a total of 56 sequences, was associated with group-2 patients (P=<0.0001, Fisher's exact test) and in all but 3 sequences was found with a consensus KK10 epitope sequence (P=<0.0001, Fisher's exact test)(Figure 46D). Just downstream of this at position 115, the less frequent I115L mutation (20 sequences total) was also significantly associated with group-2 patients (P=0.0009, Fisher's exact test) though it was found in 9 of the time-points sequenced in group-1 at lower frequencies. It was often found in sequences containing the L111P mutation and the N120S mutation (18/20 sequences). A mutation downstream of this position, E128D (21 sequences) was also associated with the L111P change, being found only in sequences containing the L111P mutation and frequently in those containing the L111P and I115L mutations (18/21) (19/21 also contained the N120S mutation). The association of these sequences may indicate a shared selection pressure or a network of compensatory changes in this region, Indeed, the E128D mutation has previously been described in the context of an R132G mutation as a compensatory change[58]. Its occurrence in this group, exclusively with R132 containing sequences suggests that it may be compensatory for other changes or be selected for another reason.

Downstream of the KK10 epitope two common mutations occur at the T148 residue, these are T148V (67 sequences) and T148S (26 sequences). Neither mutation shows significant

association with groups 1 or 2, T148V shows an association with the R132K and L136M mutations in the KK10 epitope (P=<0.0001, Fisher's exact test)(Figure 46F). All 19 T148S sequences in group-1 and 6 of 7 T148S sequences in group-2 contain the consensus residues at positions 111, 115, 120, 132 and 136. Thus the T148S mutation is significantly associated with the absence of mutations affecting T-cell escape, HLA binding and proteasomal processing as well as several with no previously identified function (111 and 115) (P=<0.0001, Fisher's exact test comparing sequences with and without consensus residues with presence of T148S)(Figure 46E). The only common mutation that appears in sequences containing the T148S substitution is N139D, a proposed KIR3DL1 escape sequence. All 8 of the sequences containing the N139D mutation also contain the T148S change.

A.

B.

	Number of sequences mutated				
	N120S	N120H	R132K	L136M	N139D	**Total Sequences**
Group 1	14	34	59	51	2	**117**
Group 2	11	18	5	1	6	**43**
P value	**0.0486**	**0.1327**	**<0.0001**	**<0.0001**	**0.0051**	

C.

	% Total sequences mutated				
	N120S	N120H	R132K	L136M	N139D
Group 1	12.0	29.1	50.4	43.6	1.71
Group 2	25.6	41.9	11.6	2.33	14.0

D.

	Number of sequences				**Total Sequences**
	L112P	L116I	T148V	T148S	
Group 1	26	8	50	19	**117**
Group 2	30	12	17	7	**43**
P value	**<0.0001**	**0.0009**	**0.8567**	**1.000**	

E.

		Number	**Total**
		T148S	**Sequences**
Consensus 112, 116, 120, 132, 136		25	**30**
Change from consensus		1	**130**
P value		**<0.0001**	

F.

	Number of sequences				**Total Sequences**
	N120S	N120H	L112P	I116L	T148V	
Consensus KK10	25	40	53	18	22	**93**
R132K/L136M	0	12	3	2	45	**67**
P value	**<0.0001**	**0.0012**	**<0.0001**	**0.0015**	**<0.0001**	

Figure 46: Sequencing of the region of p24 Gag containing the KK10 sequence, showing **A.** HXB2 clade B consensus sequence for amino acid residues 119-152 on top line and all observed mutations at the amino acid level in our HLA B*2705 patient sequences. Sequence mutations affecting the KK10 epitope shown boxed. **B.** Number and **C.** Percentage of sequences containing KK10 affecting mutations in group 1 and group 2 patients (Table 14). P values for significance of difference in frequencies between the groups (Fisher's exact test). **D.** As C for mutations of unknown effect. **E.** and **F.** Difference in frequency of mutations between (KK10) consensus sequences and sequences with KK10 or other specified changes.

5.4. Recognition of KK10 epitope form-HLA B*2705 complexes by the KIR3DL1 and KIR3DL2 receptors

In addition to the interactions with and recognition of the KK10 epitope forms by CD8+ T-cells and the effects of their response to this region on disease, it has also been shown that the interaction of KIR and HLA can be important in HIV-1 infection and disease[63; 328; 416; 417]. In particular, KIR3DL1 has been shown to bind HLA B*2705 and the KK10 epitope in a peptide dependent manner[259]. KIR3DL2 is known to bind HLA A*03 and HLA A*11 complexes in a peptide dependent manner[418] as well as HLA B*2705 homodimers independent of peptide[419] but has previously shown either no or weak binding of HLA B*2705 heterotrimeric complexes. It therefore acted partly as a control for KIR3DL1 binding, ensuring there were no homodimers present in the KK10 epitope form dextramers. We therefore assessed recognition of the KK10 epitope forms by the KIR3DL2 as well as the KIR3DL1 receptors.

5.4.1. Binding assay for KIR3DL1 and KIR3DL2 receptors

In order to test recognition of the KK10 epitope forms by the KIR3DL1 and KIR3DL2 receptors, we used two transfected BAF cell lines, one displaying the KIR3DL1 receptor and the other displaying the KIR3DL2 receptor (kind gift from Dr Simon Kollnberger). These were used in combination with dextramers of the KK10 epitope forms as previously described[259], replacing tetrameric complexes with our larger multimers. Binding of the dextramer to the transfected BAF cells (assessed by mean fluorescence) indicates recognition of the HLA-peptide complex by the KIR3DL1 or KIR3DL2 receptors.

Binding of the KK10 epitope forms by KIR3DL1 (Figure 47A) shows that the KK10 peptide is capable of binding to the KIR3DL1 receptor (MFI = 50.5) as described previously[259]. Some

binding is also seen to the KI11 epitope form though the mean fluorescence is less than half of the value for the KK10 peptide (MFI = 22.6). The KM14 peptide appears to have a mean fluorescence a little above the unstained KIR3DL1 BAF cells (MFI = 10.4 compared to MFI = 1.05 unstained) that may indicate some weak binding. The KL8 (MFI = 6.41) and KR13 (MFI = 3.38) epitope forms appear to have little interaction with KIR3DL1. When tested for KIR3DL2 binding (Figure 47B), none of the dextramers showed strong binding. The highest mean fluorescence was for the KI11 epitope form (MFI = 13.37, unstained MFI = 3.86). This may indicate weak recognition of the complex, stronger recognition would be expected if homodimeric B*2705 heavy chain was present in significant amounts, but it is also possible that a small amount was present in the dextramer.

Three escape mutant sequences for KL8 were also tested, the KKWIILGL (KL8 KKL) mutant, the KRWIIMGL (KL8 KRM) sequence and the doubly mutated KKWIIMGL (KL8 KKM) sequence (Figure 47E and F). Of these three sequences the two forms containing the R-K mutation showed no binding to the KIR3DL1 or KIR3DL2 transfected BAF cells. The KL8 KRM epitope form showed slightly raised mean fluorescence for KIR3DL1 (MFI = 10.1), however, this epitope form also showed a slightly raised mean fluorescence in the KIR3DL2 binding test (MFI = 10.7). While neither of these are strong interactions, it might indicate that a peptide independent mechanism of recognition is responsible for the binding, whether due to presence of homodimer in the dextramer or other slight aberration.

Following previous work done by other groups on the peptide dependence of the interaction of KIR3DL1 with HLA-B alleles[259; 369; 415] and the structural and sequencing results obtained in the project, we also wanted to examine the effect of peptide mutations on KIR recognition of the KK10 epitope. Our patient sequences showed two mutations present at position 9 (C terminal -1 position) of the sequence, N-D and N-E. This position is equivalent to the peptide at position 8 of the 9mer in the KIR3DL1-HLA B*5701 complex that is shown to interact with the KIR receptor. We therefore tested the sequences KRWIILGLDK (KK10 D) and

KRWIILGLEK (KK10 E) in our KIR3DL1 binding assay. Additionally, as binding was reduced or abrogated in other KK10 epitope forms, we tested the replacement of the C-1 position of the KI11 and KR13 epitope forms with an N residue as present in the KK10 peptide in order to determine whether binding was restored. This gave the sequences KRWIILGLNNI (KI11 N) and KRWIILGLNKINR (KR13 N).

Binding of the KK10 mutant sequences KK10 D and KK10 E was shown to abrogate binding to the KIR3DL1 receptor (MFI = 2.59 KK10 D, MFI = 2.42 KK10 E, compared with MFI = 50.5 KK10) (Figure 47C) indicating that this residue is necessary for the binding of HLA B*2705 KK10 to the KIR3DL1 receptor. However, the KI11 N and KR13 N sequences failed to restore binding to KK10 like levels (MFI = 9.49 KI11 N and KR13 N) and actually reduced binding in the case of KI11 (KI11 MFI = 22.6), showing that other factors in these epitope forms contribute to the reduction in the ability of KIR3DL1 to recognise them. This could be due to the conformation the mutated sequences adopt in the groove failing to replicate that of KK10. In the case of KR13, the larger central bulge may prevent KIR3DL1 recognising the complex. Testing with KIR3DL2 showed no increase in binding of any of the mutated epitope forms.

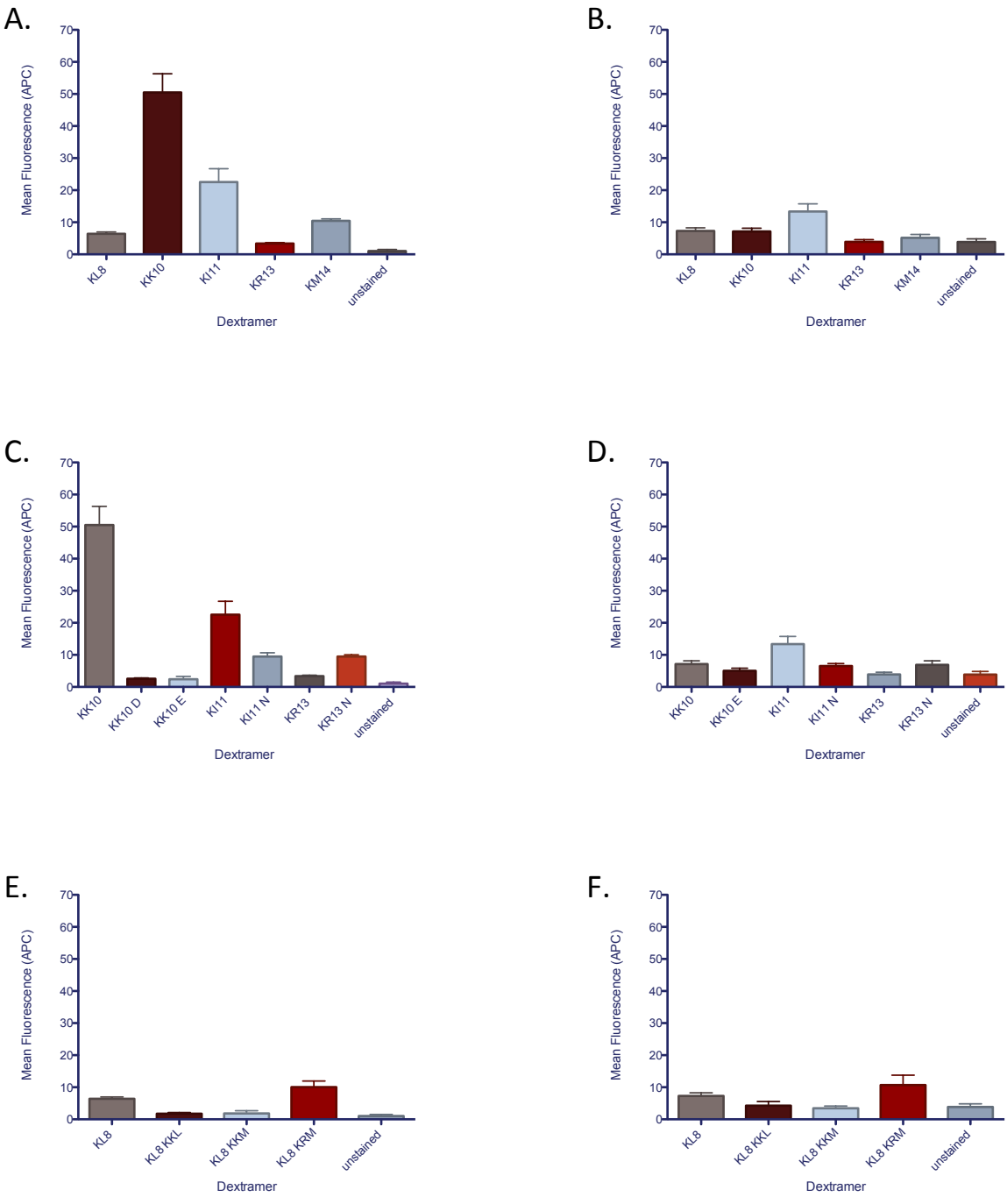


Figure 47: Binding of HLA B*2705-KK10 epitope form dextramer complexes measured by mean fluorescent intensity to **A.** KIR3DL1 and **B.** KIR3DL2. **C**. Binding of KK10 epitope form KIR3DL1 binding mutant variants to KIR3DL1. **D**. Binding of KK10 epitope form KIR3DL1 binding mutant variants to KIR3DL2**. E**. Binding of KL8 and KL8 escape variant sequences to KIR3DL1. **F**. Binding of KL8 and KL8 escape variant sequences to KIR3DL2. Mean fluorescence of untransfected BAF cells stained with the relevant KK10 epitope form dextramer were used as baseline measurements and subtracted from the above values (unstained untransfected BAF was subtracted from unstained KIR3DL1 and KIR3DL2 transfected cells to give the unstained values).

5.4.2. Electrostatic surfaces of the KK10 epitope form-HLA complexes and their interaction with the KIR3DL1 receptor.

Previous studies and structural work have shown that KIR/HLA interactions are typically more dependent more on hydrogen bonding and charge complementarity than those of HLA/TCR, which include a more significant contribution from Van-der-Waals forces[305]. Good charge complementarity was also shown for the D2 domain of KIR3DL1 bound to HLA B*5701[369]. Therefore to better understand the abrogation of binding seen for the KK10 epitope forms (compared with KK10 itself) to the KIR3DL1 receptor, we looked at the electrostatic surfaces of the KK10 epitope form complexes in combination with their structures for the area of the peptide-HLA surface highlighted for interaction by the structure of the KIR3DL1-HLA B*5701 complex[369] (Figure 48, Figure 49).

The structure of the HLA B*2705 KK10 complex shows that the $N9_{KK10}$ residue is a possible point of interaction for the KIR3DL1 complex, in a comparable position to the S8 residue of the HL B*5701 self-peptide in the KIR3DL1 structure (Figure 48E and F). Asparagine at this position could allow either water-mediated interaction with the E282 and Y200 residues as in the HLA B*5701 structure or direct interaction with the δ2-amino group. The electrostatic surface surrounding this residue and that of KIR3DL1 at the binding site, when combined with the structural data help to give an indication of how the differences in the peptide-HLA structure may affect the interaction of KIR3DL1 and the HLA (Figure 48E and F and FigureD). The surface of KIR3DL1 between the E282 and Y200 residues forms a negatively charged pocket, the negative charge continues along the surface of the molecule from this pocket towards the end of the D2 domain. An area of positive charge is present on the E282 side of the pocket and a relatively uncharged area on the Y200 side, which becomes positively charged moving away from the pocket.

The surface electrostatics of the B*5701 complex taken from the same structure as the KIR3DL1 (Figure 49H)[369] show that the area surrounding the serine residue is not highly charged, however, to the sides of this residue are two small pockets each of which carries a charge. That on the side of the α2 helix is positively charged, while that on the side of the α1 helix is negatively charged. KIR3DL1 is docked to the HLA such that an area of positive charge is in proximity to the α1 helix and a surface containing areas of negative charge sits over the α2 helix. Thus there is a degree of charge complementarity to the interaction surfaces, and although the binding may not be completely optimal[369] it is clear that it plays an important role, along with non-polar and water mediated interactions, in the binding of these two molecules.

The surface of the KK10 HLA B*2705 (Figure 49D) shows that while the N9 residue may interact with the $E282_{KIR3DL1}$ and $Y200_{KIR3DL1}$ in the negatively charged pocket, interactions with the α1 and α2 helices are likely to be similar to that in the HLA B*5701 complex. The distribution of charged areas on the electrostatic surface model shares key features with that of HLA B*5701 in the area bound by the KIR molecule (Figure 49H). This includes a patch of negative charge over the α1 helix near the C-1 terminal position of the peptide chain, as well as positive charge at the C-terminus of the peptide chain. The patch of positive charge over the α2 helix present in the HLA B*5701 structure is absent here and the area is uncharged, this is likely due to the replacement of a lysine residue in this position in the HLA B*5701 peptide with a Leucine in KK10. However, the impact on binding of this change should not be large as the side chains of these residues are not exposed and so no direct interactions would be affected.

In the KK10 epitope forms, significant differences are found from the KK10 complex in both structure and the resulting electrostatics of the molecule that may be responsible for the impact seen on KIR3DL1 recognition. The truncated epitope forms (Figure 49A, B and C) that have been crystallised show obvious structural changes likely to inhibit recognition and

binding of KIR3DL1. All three structures show a peptide that is flexible at the C-terminus and does not reach the C-1 position within the groove. Though the peptide may be pushed down if KIR3DL1 began to bind through recognition of conserved residues, in the KL6 and KG7 complexes, the peptide chains are too short to reach across the groove and in the 8mer, the residue at the C-1 position would be a glyceine, unlikely to be able to bind the $E282_{KIR3DL1}$ and $Y200_{KIR3DL1}$ side-chains, as observed in mutational analysis using the HLA B*5701 – KIR3DL1 interaction[369]. The potentially unfilled C-terminus also has a strongly negative charge and would not compliment the surface of the KIR3DL1 molecule in a binding position similar to that seen for HLA B*5701.

The KI11 epitope form (Figure 49E) is only extended by a single amino acid compared with the KIR3DL1 recognised KK10 epitope, nevertheless, binding is reduced by more than half to this epitope form (Figure 47A). There are several changes in the structure that could account for this reduction. At the C-1 position of this peptide, asparagine is replaced by a lysine residue, this could interact through formation of a salt bridge with the $E282_{KIR3DL1}$ and $Y200_{KIR3DL1}$ side-chains, though it is a much longer side chain than that of the asparagine and could prove difficult to accommodate. Indeed replacement of S8 in the HLA B*5701-peptide complex tetramers with larger charged residues such as arginine still showed binding by KIR3DL1, though with reduced affinity[369]. The change of this residue to lysine also enables it to interact with the $E76_{B27}$ residue on the α1 helix, it is unclear whether binding of KIR3DL1 would break this interaction and allow the $K10_{KI11}$ residue to bind $E282_{KIR3DL1}$ and $Y200_{KIR3DL1}$, if not this interaction may provide a further barrier to binding. On the surface of the KI11 complex, the interaction of these residues results in far less negative charge in this area than in the KK10 complex, the pocket on the side of the α2 helix at this position is now also negatively charged rather than positive, altering the charge complementarity of the two structures.

The abrogation of binding of KIR3DL1 to the KR13 epitope form (Figure 47A) correlates with significant deviation of the binding surface of this complex from that of KK10. The C-1 position in the KR13 complex is occupied by the V12 residue (Figure 49F), additionally the side-chain of the $K10_{KR13}$ residue ends at a similar position to the end of the side-chain of K10 in the KI11 complex and also interacts with the $E76_{B27}$ side-chain. In this complex the stretching of the K10 side-chain across the C-1 position and presence of a valine is likely to make KIR3DL1 binding to this region difficult, as interaction of $E282_{KIR3DL1}$ and $Y200_{KIR3DL1}$ with valine is unlikely[369], and though they could interact with the terminal group of the lysine side-chain, this may be blocked from adopting a suitable orientation by it's interaction with $E76_{B27}$ and steric hindrance from $V12_{KR13}$. Furthermore, alignment of the two structures shows that the C-terminal end of the central bulge in this peptide may clash with KIR3DL1 in the binding position adopted by HLA B*5701, it must therefore be forced to adopt an altered conformation to accommodate KIR3DL1 binding. The effects of these changes to the HLA B*2705 surface can be seen on the electrostatic model (Figure 49F), this shows the absence of negative charge over the α1 helix at this position compared with the KK10 complex due to the replacement of asparagine with valine and interaction of the $E76_{B27}$ and $K10_{KR13}$ residues.

The C-terminal region of the HLA B*2705-peptide complex is again different in the KM14 structure, and also seems to abrogate recognition of the complex by KIR3DL1, although perhaps not as completely as for the KR13 epitope form (Figure 47A). In this complex the C-1 position is occupied by the R13 residue, the side chain of this residue adopts an extended, solvent exposed conformation and does not appear to interact with the $E76_{B27}$ residue. The central bulge is also present in this epitope form as for the KR13 complex, and although it is shifted slightly away in some areas, there are still clashes with the KIR3DL1 structure that would require a shift in conformation in order to bind. These two factors could account for the reduction in binding affinity observed in comparison to KK10. The surface of the complex shows that the $R13_{KM14}$ and free $K10_{KM14}$ side-chains contribute to an increase in

positive charge over the α2 helix at this position compared with the KK10 and HLA-B*5701 complexes. An area of negative charge is still present over the α1 helix at this position in the KM14 complex and so the surface electrostatics, while altered in intensity, are not dissimilar from complexes recognised by KIR3DL1. This might indicate that steric constraints and clashes, rather than the electrostatics of the surface are of greater importance in the reduced binding of this epitope form by KIR3DL1.

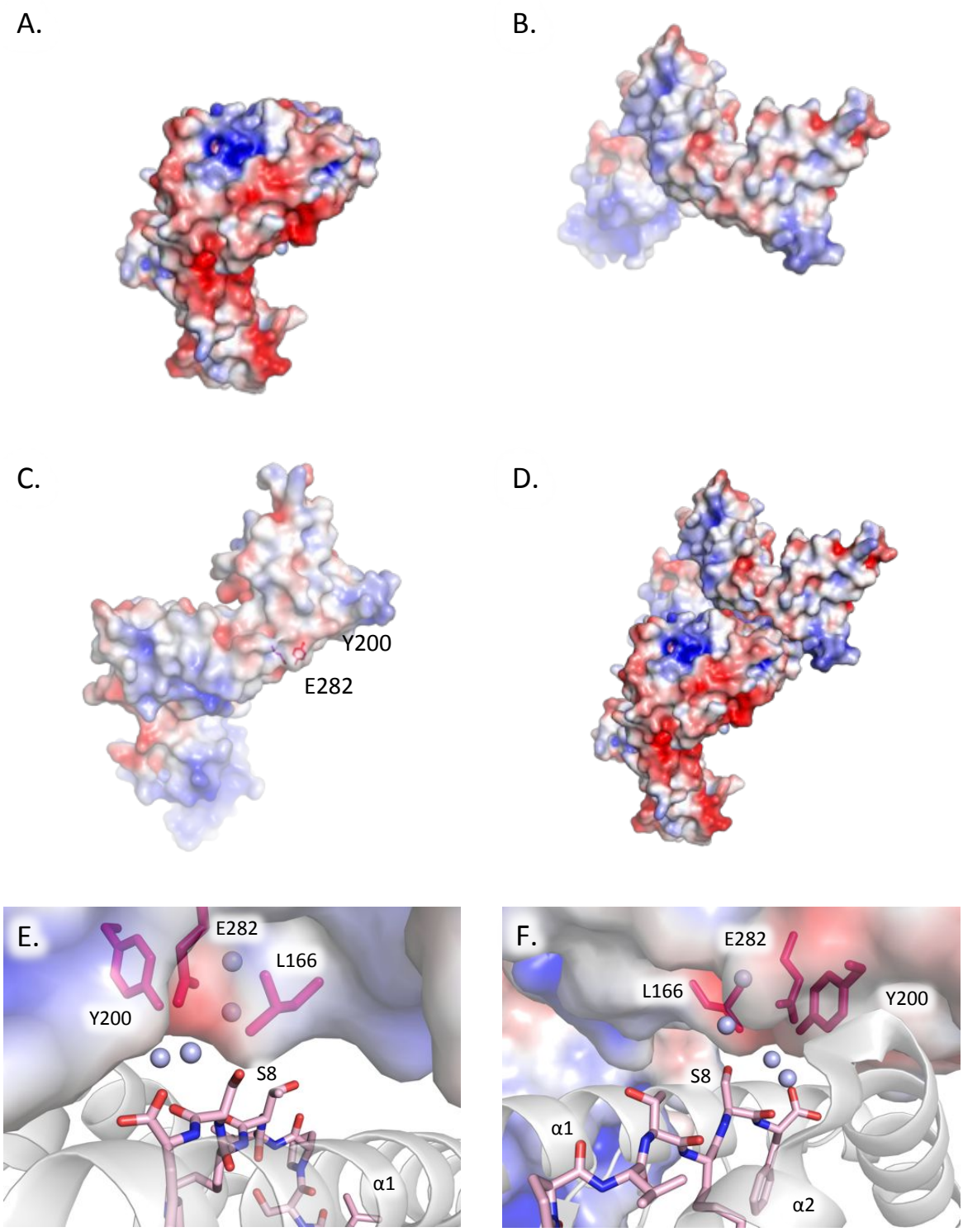


Figure 48: Surface Electrostatic models for **A.** The HLA B*2705 KK10 complex shown with peptide binding groove. **B.** KIR3DL1 **C.** KIR3DL1 showing the peptide contacting residues Y200 and E282. **D**. KIR3DL1 and HLA B*2705 models superimposed in an approximate binding configuration**. E.** Interaction of HLA B*5701-self peptide complex with KIR3DL1 shown from the C-terminal end of the peptide groove and **F.** From the N-terminal end. KIR3DL1 and HLA B*5701 structures from PDB published in[369]

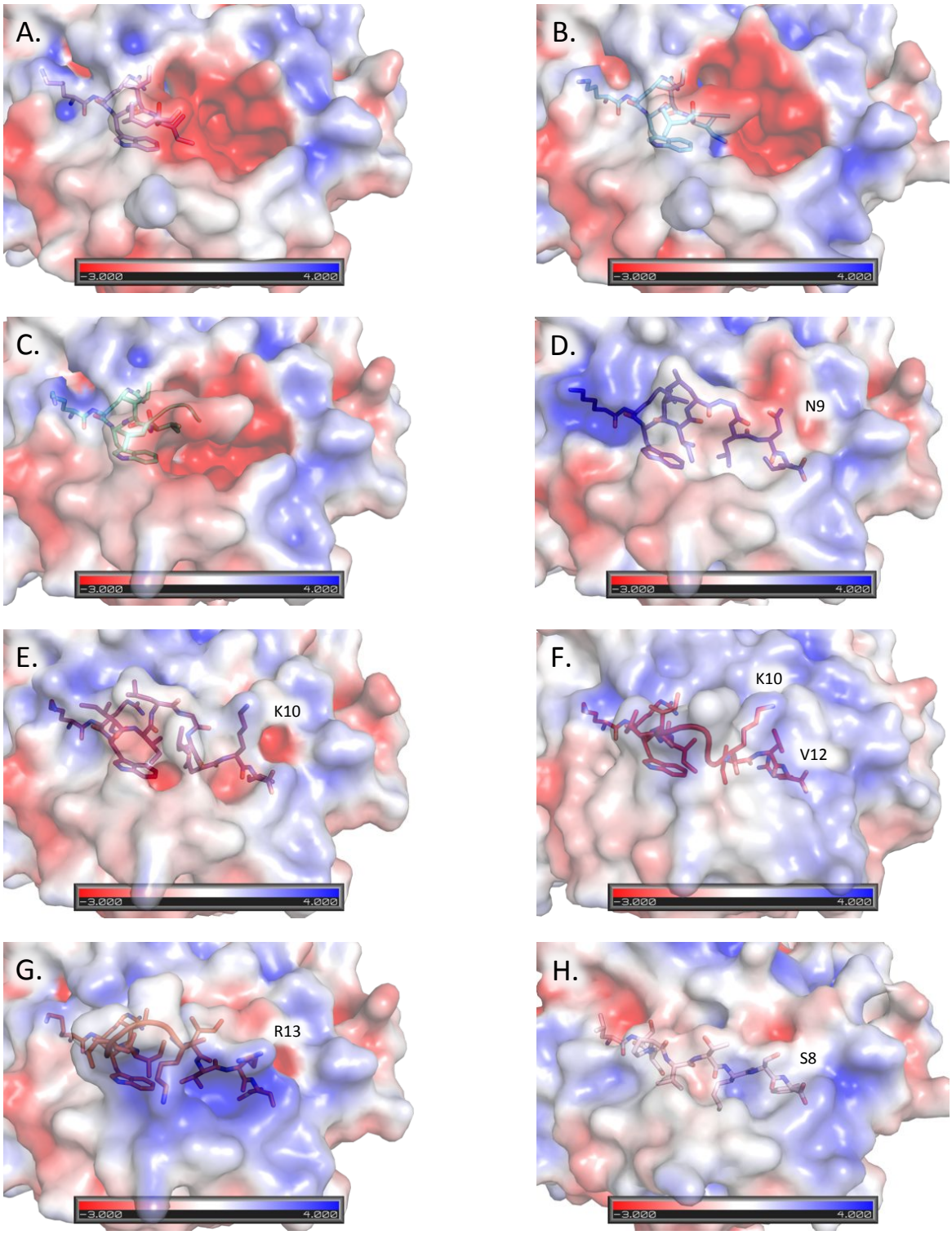


Figure 49: Electrostatic surface models for the peptide binding surfaces of the HLA B*2705-KK10 epitope form complexes, possible KIR3DL1 interacting residues are labelled where present in **A.** HLA B*2705-KL6 **B.** HLA B*2705-KG7 **C.** HLA B*2705-KL8 **D.** HLA B*2705-KK10 **E.** HLA B*2705-KI11 **F**. HLA B*2705-KR13 **G.** HLA B*2705-KM14 and **H.** The HLA B*5701-self peptide complex as determined in the KIR3DL1- HLA B*57 structure[369].

5.5. Discussion

In this chapter we have looked at the integration of data from multiple techniques in order to better understand the how the KK10 epitope forms interact with immune receptors and the consequences of these interactions for the immune response. Further extensions to this work are necessary in order to obtain a complete picture, nevertheless, this work underlines the importance of consideration of all of the KK10 epitope forms when assessing the HLA B*2705-restricted immune response to this region.

5.5.1. The structural basis for TCR recognition and cross-reactivity between a subset of KK10 epitope forms

The structural data and the electrostatic surfaces generated from this data show that there are a number of differences between the surfaces of the KK10 epitope forms; this is particularly true in the central and C-terminal regions of the peptide binding groove. Though some TCR do not bind in a canonical fashion to HLA class-1[267; 420] a large proportion of the structures of these HLA-peptide-TCR complexes show that the TCR CDR3β loop contacts the HLA at the central/C-terminal end of the peptide binding groove[272], where the KK10 epitope forms differ from each other relative to the more similar conformation of the N-terminal end of the peptide (Figure 31). Indeed, in this chapter we have also shown that the difference in the epitope forms at the C-terminus of the peptide groove is sufficient to significantly affect the binding of KIR3DL1 to the epitope forms. Nevertheless, despite the differences between the different HLA-KK10 epitope structures, we have also found substantial levels of

CD8+ T cell cross-recognition between KK10 epitope forms and have been able to isolate and sequence cross-recognising TCR (Table 23 and Table 24).

This may be due to a general difference between the recognition of HLA by TCR and KIR. It has been observed that TCR binding of HLA is more reliant on Van-der-Waals interactions than KIR[305], allowing a greater flexibility in binding orientations compared with direction and distance sensitive hydrogen bonds and salt bridges. Furthermore, while TCR recognition in general occurs over a wider area of the peptide and stretches more N-terminally than that of KIR, the same TCR has also been observed adopting altered conformations when binding to different ligands[125; 267]. This has been particularly true of the conformations of the CDR loops (and the CDR3β loop in most cases) and has also been shown to affect even the core organisation of the molecule. The increase in flexibility compared with KIR and the ease of conformational change makes it difficult to assess likely areas of binding and interaction between the TCR and the HLA. Nevertheless, we discuss some of the possibilities below.

Identification of conserved residues and sequences among TCR recognising KK10 epitope forms was performed by comparing the repertoire of CDR3α and β loops that we obtained from dextramer positive TCR populations with the abundance of these residues in the same position in the germline encoded sequences. Any large differences in the proportions of particular amino acids at a given position would then reveal positive or negative selection for particular residues among the KK10 epitope form reactive T-cells (Table 28). From this comparison we observed that the KK10 epitope form specific T-cells are enriched for glutamic acid and aspartic acid residues at the C-3 and C-4 positions of their CDR3β loops relative to the germline. In addition, there is a marked reduction in the proportion of asparagine present at the C-5 position. Assuming a conventional or near conventional docking mode of the TCR to the HLA complexes, the glutamic and aspartic acid residues present may be able to interact with the C-1 position of the peptide. This is an asparagine in

KK10, a lysine in KI11 and an arginine in KM14. In KR13 the C-1 position is a valine, but the end of the side-chain of the C-3 lysine is present at a similar position to the C-1 lysine in KI11 and so this might preserve the interaction for some TCR. Alternatively, the absence of cross recognition of KR13 in some patients (Table 19) could reflect the loss of the polar/basic residue at this position.

The central regions of the CDR3β loops contain a high proportion of small non-polar and hydrophobic residues (53/89 residues in the region). The high proportions of glyceine in this region reflect the germline propensity for glyceine in the TR-D_β segments, presumably to give flexibility and the ability of the loop to turn. Proline is also encoded at high frequencies in this region, presumably also in order to allow turning of the peptide chain at the apex of the loop, but this residue is reduced in our sequences, and glyceine increased in frequency. This difference may show a tendency in these TCR for increased flexibility, provided by glyceine, that allow the adaptation of the CDR3β loop to a number of KK10 epitope forms. The low frequency of large and polar or charged amino acids in this central region (9/89) may also improve flexibility and the ability to cross recognise peptides. The presence of relatively small hydrophobic residues across the centre of the groove would allow a large number of Van-der-Waals contacts to be made with the bound peptide, affording binding without the need for a high degree of specificity. The addition of small polar residues such as threonine and serine, that are present at considerable frequencies in the central region of the loop (27/89), would add the opportunity for some hydrogen bonding in addition to Van der-Waals contacts, while maintaining the flexibility of the chain. This pattern extends to the C-6 and C-5 positions of the loop, where a significant reduction in the number of asparagine residues at C-5 occurs compared to germline, this coincides with an increase from germline in the frequency of alanine and threonine at C-5 and glyceine, proline, Leucine and valine at C-6.

The surface of the KK10 epitope and the extended epitope forms in the central bulge region of the peptide is largely uncharged and consists of non-polar residues that are highly solvent exposed (Figure 42). This shows a complementarity with the central regions of the CDR3β loop and there is a possibility this will provide a suitable area for binding and recognition. In four of the most abundant and most cross-reactive TCR identified in patient 241242, three (Figure 50A, B and C) contain relatively long central regions containing these small hydrophobic and polar residues, with occasional large hydrophobic residues towards the termini of the chain and an acidic residue at the C-terminus. One cross-reactive TCR (Figure 50D) does not entirely follow this pattern and has a short central TAP motif between charged residues. Extension of this analysis and structural studies would confirm or refute these suggestions.

A.	C**ASSYSGTSGSPAY**EQYF
B.	C**ASSHLGGVT**NEQFF
C.	C**ASSFTSGSP**ENEQFF
D.	CASSR**TAP**DTEAFF

Figure 50: CDR3β sequences of four highly cross-reactive TCR isolated from patient 241242. Charged residues showed coloured blue (basic) or red (acidic). Aromatic or residues with a ring showed in larger font. Central regions consisting largely of small hydrophobic and small polar residues shown in bold.

5.5.2. KK10 truncated epitope forms do not elicit CD8+ T-cell responses

The failure of the truncated KK10 epitope forms to induce cytokine production from CD8+ T-cells shows that these epitope forms are not recognised by the host immune system. Given the typical binding orientation of the TCR to HLA class-1, this is not completely unexpected, as the recognition of peptide side chains and backbone conformation occurs primarily through the CDR3 loops, particularly the CDR3β loop contacting the central to C-terminal end of the peptide[265; 272; 287; 288]. At this position, the truncated epitope forms will have either no peptide residues (in the shortest forms) or no fixed structure (in the longer 7 and 8mer

forms). The peptide may be pushed into the groove on TCR binding[410], but docking may be unable to occur without at least some recognition of the peptide, equally, the flexible nature of the peptide at the unbound C-terminus may be able to sterically hinder binding. However, even given these reasons, it was surprising that this lack of cytokine production extended to the KL8 truncated epitope form, as we had previously shown recognition of this epitope form by CD8+ T-cells in dextramer staining experiments (Table 15). This may be explained by the docking of TCR in an unusual conformation to this epitope form as they are not able to dock in a typical conformation at the centre-C-terminal end of the peptide groove, in some cases, unusual docking conformations have been associated with a lack of T-cell activation, despite high affinity binding[278].

Though preliminary, our data also indicates that presentation of the KG7 and KL8 epitope forms on the cell surface does occur, though repeated measurements must be carried out in order to confirm this finding (Figure 45). If confirmed, this suggests a role for the truncated epitope forms in shaping the immune response to the region. The presentation of these epitope forms will prevent the HLA from binding more immunogenic KK10 epitope forms and the abundant production will ensure that the short epitope HLA complexes greatly outnumber the amount of KK10 and extended epitope forms that are found on the cell-surface. This may reduce CD8+ T-cell responses to KK10 and the extended epitope forms and might increase the number of cells that escape CD8+ T-cell mediated cell death. We therefore propose that the production and presentation of these truncated epitope forms is a novel CTL-escape mechanism because the production and presentation of these short epitope-forms likely increase viral survival and spread. This hypothesis is reinforced by our antigen processing data, which shows the continued production of these truncated epitope forms following escape mutations in the KK10 epitope, when production of KK10 and KK10 extended epitope forms is severely reduced or abolished (Figure 5).

5.5.3. Footprints of immune recognition of KK10 epitope forms on the viral sequence

Sequencing of the viral genome at p17 and p24 Gag in our HLA B*2705 patient cohort allowed us to analyse the immune footprints present in these genomes (for viral sequences present at approximately 10% or higher of total population) in context of their recognition of KK10 epitope forms in dextramer staining. Those patients without recognition of the KK10 epitope showed a markedly different pattern of mutation in the region surrounding the KK10 epitope (positions 109-166 included in analysis) including significant differences between the presence of recognised escape mutations within the KK10 epitope (Figure 46). However, it is striking that several patients (050875, 111156 and 170466) displayed low frequencies of these mutations despite no obvious KK10 recognition.

There are several reasons why these mutations might be present in patients without recognition of KK10 at the time-point tested. Patient 050875 does not respond to KK10 but does have a small but distinct population of CD8+ T-cells which recognise the KI11 epitope form, this may have been able to drive selection in the sequence. The second patient, 111156 shows relatively low responses to other epitope forms (though does seem to respond to KM14) but has been on HAART at all time-points tested in this study and is the only patient showing undetectable viral load throughout. Antiretroviral treatment and the reduction in viral load often reduces or abrogates immune responses as very little antigen continues to be presented[353]. Finally patient 170466, as for 050875 also has considerable levels of CD8+ T-cell KI11 recognition. In addition, though first testing positive only a short time before the time-point tested (Table 12) this patient is likely to have first been infected 15-20 years ago through exposure to a partner with the virus. The viral load of this patient is also the lowest in our group among those not receiving treatment (1843 copies/ml) (Table 12). This could have a significant effect on the immune response compared with patients infected for shorter periods, and sequence changes may be evidence of a previous response to the KK10 epitope.

The differences in frequencies of other mutations between the groups that are thought to affect presentation or immune response to the KK10 epitope (Astrid Iversen, unpublished data) is also of interest as it reveals possible reasons for the development, or lack thereof of CD8+ T-cell recognition of the epitope. While the difference in frequency of the N120S mutation between the two groups is sufficient for statistical significance, more interesting is the increased difference when looking in all patients at the association between this mutation, the N120H mutation and the presence of escape mutations within the KK10 epitope. The N120S associates almost exclusively with sequences lacking the R132K and L136M KK10 escape mutations, and the association is statistically significant, though not complete between the lack of these mutations and the presence of the N120H substitution. Both of these changes are associated with severely reduced production of the KK10 epitope during proteasomal processing and so limit the exposure of this epitope to the immune system. That these changes are associated with absence of KK10 escape mutations within the epitope indicates that they can occur earlier than the 132 and 136 escape mutations. It also suggests that once these mutations are gained, selection pressures decrease or do not continue, as this would likely result in sequences containing both the upstream N120 changes and the intra-epitope mutations. The further association of subset of sequences containing the N120S mutation with mutations at positions 111, 115 and 128 (L111P, I115L and E128D) suggests that these mutations may be compensatory changes occurring as a result of the N120S change or may serve to enhance the effect of the mutation.

The intra-epitope mutation N139D was shown in our work to abrogate binding of the KK10 HLA B*2705 complex to KIR3DL1 (Figure 47C). This mutation was present in only a small number of sequences (8/160 and 4 patients) but thus far shows a significant correlation with group-2 patients and with sequences lacking other KK10 escape mutations (Figure 46A). The association of this epitope with wild-type sequences may indicate that these viral populations are also driven by selection pressure from KIR3DL1^{+} Natural Killer (NK) cells.

Though escape from an inhibitory KIR may seem counter-intuitive, this may depend on the balance of activating versus inhibitory signals that arise from an NK cell binding an infected target cell such as one displaying the KK10 epitope. It may also indicate that licensing of NK cells through inhibitory KIR is important in increasing their effectiveness and cytotoxicity[330; 332; 333; 421]. It is intriguing that this mutation occurs preferentially in patients that show no T-cell recognition of the KK10 epitope, suggesting that selection pressure by CD8+ T-cells is greater than from NK cells and drives escape where a T-cell response is present. In addition, it indicates that when T-cell pressure is absent or reduced, NK selection pressure may capable of driving escape in the viral population. A further investigation into recognition of this epitope would provide further insight into its role, testing the ability of KK10 epitope form specific T-cells to cross recognise this sequence would show whether this change also affects the T-cell response.

In addition to the analysis of previously described mutations affecting KK10 in this region, and the linkage of these with several mutations of unknown function upstream of KK10, there were two other changes in the region downstream of the epitope found across sequences from multiple patients in our study, T148V and T148S. We assessed whether these changes were associated with either our patient groups, or the presence of previously identified mutations in the region. The T148V mutation was present in both patient groups and in both consensus and KK10 escape containing sequences, however, in our sample, it's distribution favoured sequences containing the KK10 escape mutations (P=<0.0001, FET) and it may therefore be associated with compensatory changes to the region, there are no known HLA B*2705 CD8+ T-cell epitopes covering this position of p24[422] so it is unlikely to be a direct escape mutation. The T148S mutation was also present in a considerable number of our sequences (26/160), this mutation was found in both patient groups, but interestingly was strongly associated with the lack not just of the R132K HLA binding and L136M CD8+ T-cell KK10 escape mutations, but also of the N120S/H upstream mutations affecting proteasomal processing of KK10 (P=<0.0001), with just one sequence of the 26 containing

any of these. The 26 T148S sequences without mutations at positions 120, 132 and 136 account for all but 4 of the sequences without these mutations in our study, of the remaining 4, only one has a complete consensus sequence, with another containing L111P and T148V and the other two containing G117A and T148V changes. One mutation that is associated with the T148S mutation is N139D; all 8 sequences containing this change also have the T148S mutation. These changes are conservative and whether the T148 mutations are of functional significance is unknown, however, while the V148 is the consensus amino acid in most other HIV subtypes, S148 is not, suggesting that these substitutions are not neutral.

5.5.4. The KIR-HLA interface in KK10 epitope form-HLA B*2705 complexes

Unlike the HLA-TCR interface, the binding of KIR3DL1 to HLA class-1 is dependent upon charge complementarity[305] and on groups of specific, germline encoded residues on the HLA, in addition to peptide recognition. This results in a much more restricted mode of binding than that of the TCR, where evidence of germline encoded recognition of fixed HLA class-1 residues is limited and for those that are identified, direction-independent, hydrophobic interactions predominate[108; 125; 287; 288]. This much more restricted, charge based binding mode allows the comparison of the structurally determined KIR3DL1 binding of HLA B*5701 with potential binding sites in our KK10 epitope form-HLA B*2705 complexes with a higher degree of confidence than the identification of possible TCR binding sites. This comparison is further aided by our data on KIR3DL1 binding ability of the KK10 epitope forms.

The analysis of the surfaces and structure of the KIR3DL1 and HLA B*2705-KK10 epitope form complexes at the proposed binding site combined with data on the ability of KIR3DL1 to bind these epitope forms shows us the importance of the peptide sequence and conformation at the binding site, and the charge complementarity of the surrounding HLA

for KIR3DL1 binding (Figure 47,Figure 48Figure 49). Relatively small changes in the amino acid at the binding position, from asparagine in KK10 to lysine in KI11 give a reduction in binding of KIR3DL1 by more than half, showing the dependence of this interaction on the bound peptide as well as on matched surfaces. The further reduction in interaction of KIR3DL1 with other KK10 epitope forms with larger differences at these positions and changes to the conformation of the bound peptide re-enforce this view.

While not unexpected, given prior data[259; 415; 423], the lack of KIR3DL1 binding to our epitope forms may be relevant in HIV-1 infection, as lack of binding of the inhibitory receptor to these complexes may affect killing by NK cells. That these epitope forms, both extended and truncated, are produced in far higher quantities than the KK10 epitope itself may mean that they have a significant presence on the cell surface and give a different response than might be expected from examination of KK10 alone.

The association of KIR3DL1 with slower disease progression in HLA B*2705 patients[325; 327; 328; 416] has been intriguing since its discovery, as the binding of inhibitory receptors to HLA ligands has been shown to prevent target cell lysis *in vitro*[64]. However, our data show that escape from KIR3DL1 recognition may be driven by selection during HIV-1 infection (Figure 46). Licensing of NK cells has been a proposed mechanism where recognition of cognate HLA by these inhibitory receptors results in an activated phenotype and more efficient NK cytotoxicity[333]. However, the mechanism of any direct link between recognition of virus infected cells by these inhibitory receptors and improvement in the immune response remains tentative[424; 425] and requires a more thorough understanding of the outcomes of KIR recognition and the contribution of NK cells to the host immune response.

6. Concluding remarks

The studies presented here give a further insight into some of the mechanisms behind an immune response to the KK10 epitope region in HIV-1 p24 gag and the viral modulation of this response. This response is associated with control of HIV-1 disease in a subset of individuals with HLA-B*2705 and so provides a rare example of an effective immune response to HIV. The understanding of effective immune responses and mechanisms of immune escape are of paramount importance in the development of an HIV vaccine as exploitation of the mechanisms behind these responses and the development of strategies that reduce the ability of virus to escape from the immune system will be required to provide competent immunity in the general population.

Briefly the studies here show that HLA B*2705 positive patients develop CD8+ T-cells responses specific for naturally processed KK10 epitope-peptide forms identified by proteasomal digestion[238] and that some T-cells are able to cross recognise multiple epitope forms and are particularly able to cross-recognise the KK10 epitope form. These KK10 epitope form specific TCR have been sequenced and characterised for shared features that may enable cross-recognition. Crystal structures of the various KK10 epitope forms in complex with HLA B*2705 show shared and unique structural features, corresponding well with recognition patterns seen in our functional data. Also shown is that a large number of the truncated KK10 epitope forms are produced during proteasomal digestion and that the presence of KK10 escape mutations in the epitope sequence, while abrogating production of KK10 extended epitope forms, does not reduce production of these truncated forms. In contrast, extended KK10 epitope forms are not produced following digestion of long peptides with KK10-associated CTL escape mutations (R132K and L136M) and in peptides with the L136M escape mutation, no KK10 epitope forms are produced. These truncated epitope forms are capable of binding the HLA B*2705 molecule and may be able to block the presentation of KK10 and extended epitope forms at the ratios seen in proteasomal digestion. Finally we show that KIR3DL1 recognition of the KK10 epitope is considerably

weaker or absent for the KK10 epitope forms, analysis of structural features and surface electrostatics provides a basis for this loss of binding.

The ability of patients to recognise KK10 epitope forms, both with and without the recognition of the KK10 epitope form itself in the patient, as well as the high level of cross-recognition seen for some epitope forms with KK10 (and the presence of cross recognition between other epitope forms in some patients) may be an important part of the effective immune response seen to this epitope. The cross recognition of CD8+ T-cell clones among these epitope forms may increase the number of T-cell clonotypes able to recognise the epitope relative to regions where fewer epitope forms are present. This could possibly sustain the immune response and recognition towards the epitope region even when escape mutations result in the loss of epitope presentation for individual T-cell populations or a population becomes exhausted[308; 309]. Notably, the patient we observed to recognise the most epitope forms and have the highest number of cross-recognising T-cell populations had the longest confirmed treatment free period at the time of sampling. The abrogation of production of these epitope forms in viral sequences containing escape mutations also suggests that reducing their production is beneficial for the virus and is selected during infection. Analysis of a larger group of patients to enable comparison of disease control among those with responses to these epitope forms and those with a single KK10 response will be needed to confirm this hypothesis.

Crystallisation of HLA B*2705 in complex with these recognised epitope forms has allowed us further insight into the accommodation of the extended epitope forms within the HLA class-1 groove and shows a shared binding mode at the termini, particularly at the N-terminus of the groove along with more permissive binding through water-mediated interactions in the centre of the groove allows this set of peptides to bind. The conserved binding at the termini, with differing structural motifs and surface exposure of residues in the central portion of the groove suggest a mechanism whereby cross-recognition and

epitope specific TCR could be produced to these epitope forms in a manner consistent with the functional data obtained.

Sequencing and characterisation of the TCR specific for KK10 epitope forms in several patients has shown some shared features which may be involved in peptide recognition as well as some differences between TCR capable of cross recognition and those that are specific for single epitope forms. These data help to show the importance of peptide structure for the development of cross-recognising TCR. Further TCR sequencing, and the generation of HLA B*2705-TCR complex crystal structures will improve our understanding of these interactions and reveal any modifications to peptide structure that occur on TCR binding, which have been reported for previous complexes[410].

The analysis of proteasomal processing of the KK10 containing region of HIV-1 p24 gag presented here shows that in the presence of the commonly seen escape mutations R132K and L136M the production of extended KK10 epitope forms is abrogated, showing that these mutants also affect proteasomal cleavage patterns. The absence of KK10 production in the L136M sequence suggests that this escape mutant, until now described variously as acting through escape of T-cell recognition in some patients and being a prerequisite compensatory mutation for R132K[58; 256], may primarily be selected for reduction in production of the KK10 epitope and the extended epitope forms. The development of new T-cells specific for the L136M KK10 form has been shown in some studies[321] possibly due to proteasomal independent production of the epitope, cross-recognition of a KL8 epitope form or continued presentation of virus lacking the L136M mutation in a small percentage of the population or at local sites. Examination of *in vivo* production of epitopes from this sequence will be required to determine the extent of abrogation that occurs at the cell surface.

In an analysis of viral sequences for the region of p24 gag containing the KK10 epitope we further identified a number of mutations, one of these was a shift in viral sequence at certain

positions from the wild-type residue of the clade B virus to that of clade C. Association of a number of these mutations, including the N120S mutation were significantly associated with patients lacking recognition of the KK10 epitope and was even more significantly associated with the absence of R132K and L136M escape mutations (P=<0.0001). This raises the possibility that substitution of residues at these positions may alter antigen processing and reduce the presentation of KK10 on the cell surface. Notably, patients without recognition of the KK10 epitope did not display significant reduction in recognition of other epitope forms (KI11 showed a trend towards reduced recognition P=0.0582), indicating that these mutations may act specifically on the KK10 epitope itself or include only certain epitope forms. Analysis of proteasomal digestions of sequences containing each of the mutations found and combinations of mutations that were linked will allow us to determine the effect these shifts have on antigen processing in the region and how this might be connected with the avoidance of KK10 recognition and development of KK10 intra-epitope escape mutations.

In addition to the effect of extended KK10 epitope forms on the immune response to this region, production of truncated KK10 epitope forms was observed in both wild-type and mutant containing sequences in molar ratios of up to 1200 times that of KK10. These epitopes consist of a maximum of 8 residues and most have been discounted by previous studies as too short for HLA class-1 binding. However, data presented here shows production of the KI4, KI5, KL6, KG7 and KL8 forms and that these epitope forms are indeed capable of binding HLA B*2705 following using *in vitro* refolding (KI4, KI5,) and determination of the crystal structures (KL6, KG7 and KL8). Data produced in collaboration with Prof Peter van Endert further shows that they are efficiently transported into the endoplasmic reticulum and I demonstrate that these epitopes do not elicit CD8+ T-cell responses from *ex vivo* patient PBMC, though TCR recognising KL8 were found in dextramer staining. My preliminary data further suggests that these epitope forms may be capable of

reducing the T-cell response to the KK10 epitope at the ratios at which they are produced by the proteasomes.

The production of these short peptide epitope forms by the proteasome, which are able to bind HLA and may block responses to cytotoxic T-cell epitopes has not, to my knowledge, previously been shown. The continued production of these epitopes in proteasomal digests in the presence of escape mutant containing sequences further suggests that these epitope forms may play a role in the modulation of the T-cell response to this region. We suggest that this may be a previously unrecognised form of viral escape. The presence of short epitope forms capable of HLA binding in other responses would help to indicate whether this is a widespread escape mechanism and the extent to which it operates in HIV-1 infection and in other viral infections.

Previous studies have indicated that the presence of KIR3DL1 in combination with the HLA*B2705 epitope may delay disease progression following infection with HIV-1[325]. Though licensing of NK cells by the presence of cognate ligand for KIR3DL1 has been suggested as a mechanism for improving NK effector function and therefore control of disease, the effects of direct interaction of KIR3DL1 containing cells with HLA B*2705^{+} HIV-1 infected cells have yet to be shown. The loss of binding observed for KK10 epitope forms in comparison to KK10 reveals a straightforward mechanism: the loss of NK inhibitory interaction with the HLA B*2705 molecule on the surface of infected cells due to the presentation of KK10 epitope forms unable to bind KIR3DL1 would result in lysis by the NK cell and so reduce overall viral replication. The analysis of lysis by NK cells of cells presenting these epitope forms would provide further information as to the validity of this suggestion. However the identification in several patient viral sequences of a mutation that I show abrogates KIR3DL1 binding of the KK10 epitope may show selection pressure to reduce recognition of infected cells and analysis of the ability of CD8+ T-cells to recognise

our proposed NK escape variant will help us to determine whether this mutation is directed by NK selection pressure alone or by a combination of CTL and NK pressures.

Previous attempts to create vaccines for HIV-1 that induce effective cellular immune responses have been largely unsuccessful. However, both CD8+ and CD4+ T-cell responses do seem to be generated when using vector-delivered DNA vaccines. These immune responses have resulted in either no effect on the rate of infection or an increase in acquisition[76; 426; 427; 428]. Current DNA vaccines rely on natural epitope and flanking sequences for commonly presented epitopes. The design of these constructs may be improved by the use of modified sequences which promote the production of viral epitopes that are readily recognised by the immune system and reducing the presence of those that are less effective or serve to obscure recognition of useful epitopes. The work here contributes towards our understanding of the control of antigen processing and presentation in this conserved region of p24 gag and with further work may be useful in aiding the design of such a vaccine.

In addition to vaccine design, this work highlights the variation in length of peptide epitopes that are capable of binding HLA B*2705 and begins to determine factors controlling the extent to which CD8+ T-cells are able to recognise and cross recognise these epitope forms. A more comprehensive understanding of these processes is essential for improving our ability to modulate their development in order to improve the immune response to viral infections.

6.1. Further studies

1. Repetition of IFNγ ELISpot assay to confirm findings that truncated KK10 epitope forms affect the response to KK10 (Figure S 18).
2. T-cell sorting and clonotyping for additional patients in order to better understand responses and the characteristics of TCR recognising these peptides.
3. Extension of viral sequencing to further patients and proteasomal digestion of sequences containing combinations of mutations of interest identified from our sequence analysis.
4. Creation of a TCR-HLA B*2705 crystal structure with the KK10 epitope forms to determine the mode of TCR binding and effects on the peptide conformation of TCR docking.
5. Extension of ICS assay to assess CD4+ T-cell cytokine production in response to KK10 epitope forms.

7. Materials and Methods

7.1. Production of HLA B*2705 Monomers

7.1.1. Expression of HLA B*2705 heavy chain and β2M

Expression of HLA B*2705 heavy chain and β2M was carried out using a slightly modified protocol to that in[399]. One Shot® BL21(DE3)pLysS Chemically Competent *E. coli* were transformed according to protocol (Invitrogen C6060-03) with the prepared vector. The *E.coli* were plated onto LB agar plates containing ampicillin at 100μg/ml concentration. A single colony was picked and transferred to 100ml LB Amp (100μg/ml ampicillin added) medium. This was then left shaking overnight at 37°C. The optical density at 600nm was measured and the volume needed to give an OD600 of 0.05 was added to 1litre LB Amp medium. This was incubated at 37°C with shaking and the OD600 checked regularly. When the OD600 reached between 0.5 and 0.8, 0.5ml 1M IPTG was added to the flask to induce expression of HLA heavy chain protein. The flask was then incubated at 37°C with shaking for a minimum of 6 hours. The *E.coli* containing growth media was transferred to 1 litre centrifuge buckets and spun in an ultra-centrifuge at 6000rpm for 15 minutes. The supernatant was then discarded and the pellet homogenised with approximately 20mls cold PBS. This was frozen at -80°C for future purification.

7.1.2. Purification of Inclusion bodies

The purification carried out was a slightly modified version of the protocol given in[399]. The pellet-PBS homogenate from the previous step was first thawed to 4°C and kept on ice. The mixture was sonicated (while still on ice) using a Sanyo Soniprep 150 sonicator at an amplitude of 8 microns for 1 minute. The sample was then left to cool for a minute before the sonication was repeated. This was done for 5 minutes total sonication time. Once

sonicated, the sample was placed in an ultracentrifuge and spun at 15000rpm for 10 minutes. The supernatant discarded and the pellet retained. The pellet was homogenised with 15ml Triton wash buffer (0.5% Triton X-100, 50mM Tris pH 8.0, 100mM NaCl, 1mM EDTA) and re-centrifuged to remove debris. This was repeated 3 times. Finally, the pellet was homogenised in 15ml resuspension buffer (50mM Tris pH 8.0, 100mM NaCl, 1mM EDTA, 1mM DTT) and centrifuged again at the same settings to remove any remaining triton detergent. The protein pellet was solubilised in approximately 5mls urea buffer (8M urea, 0.1M NaH_2PO_4, 10mM Tris pH 8.0, 0.1mM EDTA, 0.1mM DTT). The solubilised protein was returned to the centrifuge and spun for 10minutes at 15000rpm to pellet any insoluble impurities, the supernatant carefully removed and saved. The pellet was discarded. The protein concentration in the supernatant was then measured using absorbance at 280nm on a Nanodrop (Thermo-Scientific) spectrophotometer. Aliquots were stored at -80°C.

7.1.3. *In vitro* refolding by rapid dilution for HLA B*2705, β2M and peptide

Refolding of HLA B*2705 and HLA A*0201 complexes was undertaken using a rapid dilution method first given in[429] and subsequently modified by[430]. In the modified method a 500ml solution of 400mM L-Arginine monohydrochloride, 100mM Tris pH8.0 is cooled to 4°C. 5mM reduced glutathione and 0.5mM oxidised glutathione were then dissolved in this solution. Once completely dissolved 10mg β2M purified from inclusion bodies and solubilised in urea buffer (8M Urea, 0.1M NaH_2PO_4, 10mM Tris pH 8.0, 0.1mM EDTA, 0.1mM DTT) was injected into the mixture while stirring. This was followed by 1.67mg peptide solubilised in 67µl DMSO and made up to 1ml with 8M Urea (or urea buffer). After 30 minutes incubation, 10mg A*0201 or B*2705 heavy chain also in urea buffer was added dropwise to the stirring solution. The solution was then left for 24 hours stirring at 4°C. The addition of β2M, peptide and heavy chain was repeated as before and left again for 24 hours. After a final addition of

peptide and protein to bring the total added to 30mg β2M and heavy chain, and 5mg peptide. The solution was left for a minimum of 2 days stirring at 4°C before concentration.

7.1.4. Separation and purification of the HLA B*2705 heterotrimer

Size exclusion purification of HLA B*2705 and HLA A*0201 monomers was carried out following refolding using an AKTA Purifier system and a HiLoad 26/60 Superdex 75 prep grade gel filtration column (GE Healthcare) in FPLC buffer (20mM Tris pH 8.0, 100mM NaCl). The heterotrimer peak was initially checked by running eluent fractions corresponding to absorbance peaks on an SDS-page gel to identify the composition of each peak. The heterotrimer peak is identified as having roughly equal intensity of the HLA B*2705 heavy chain band and the β2M band. The heterotrimer was then given a final purification or "polish" by running the fraction through a Superdex 75 10/300 GL column (GE Healthcare) on the AKTA purifier system. Refolded heterotrimer taken from this purification was checked again on an SDS page gel before being concentrated for crystallization trials using Amicon Ultra-0.5 Centrifugal Filter Units with an Ultracel-10 membrane. Typical concentrations for trials were 13mg/ml, measured using absorbance at 280nm on a Nanodrop (Thermo-Scientific) spectrophotometer (extinction co-efficient calculated and used to determine concentration from absorbance). The optimal concentration for crystallisation was determined for each refold using a PCT™ Pre-Crystallisation Test (Hampton Research).

7.2. Production of HLA B*2705 Crystal Structures

7.2.1. Crystallisation screening

Crystallisation was carried out using the sitting drop vapor diffusion method using facilities at the division of structural biology (STRUBI) in the Wellcome Trust Centre for Human Genetics, University of Oxford. Briefly, prepared crystallisation screen (Hampton research) reagents were dispensed into a CrystalQuick 96 Well Sitting Drop Plate (Greiner) using the Hydra edrop II (Thermo-Scientific) and 10nl protein mixed with 10nl of the well buffer condition using Microsys liquid handler (Cartesian Technologies), this was dispensed as a sitting drop for each well. Crystallisation screens used for these HLA B*2705 complexes were: PEG/Ion, Index, Block 1, Block 2, Block 3 and Block 4. Crystal screens were stored using a Formulatrix storage and imaging system at 4°C.

7.2.2. Data collection and model refinement

Crystals obtained from the screens were picked from wells at 4 °C using and immersed in reservoir buffer mixed with 20% glycerol as a cryo-protectant. The crystal was then vitrified at 100K in liquid nitrogen. X-ray diffraction was carried out on beamlines at the European Synchrotron Radiation Facility (ESRF), located in Grenoble, France (www.esrf.eu) or at Diamond, Harwell, UK (www.diamond.ac.uk) as detailed in (Table 29). A model was constructed using molecular replacement based on the structure (PDB ID: 1W0V)[260] in which the original peptide was omitted from the modeling process to avoid bias. Refinement of the model and building of the peptides were then carried out using the programs COOT[431], BUSTER[432], and Phenix[433]. Validation of these models to prepare for deposition in the Protein Data Bank was carried out using MolProbity[397]. Further analysis and figures were

made using the PyMol molecular graphics program (Schrodinger). The Adaptive Poisson-Boltzmann Solver (APBS) plugin was used to calculate electrostatic potentials for the models shown through solving of the Poisson-Boltzmann equation[434].

7.3. *Ex vivo* staining of patient PBMC's with HLA B27 Dextramers

7.3.1. Donors

Cryopreserved peripheral blood mononuclear cells (PBMC) from HLA B*2705-positive HIV-1-infected individuals were provided by the Department of Infectious Diseases, Righospitalet in Copenhagen, Denmark. Ten donor patient samples were provided with four of these at more than one time-point.

7.3.2. Dextramer staining

PBMCs were resuspended in FACS buffer (PBS w. 5%FCS) and 100µl was plated per well at a density of 1×10^6/ml. PBMC were then stained with LIVE/DEAD® fixable violet dead cell stain (Invitrogen) for 20 minutes at room temperature. Cells were washed twice and then incubated with appropriate dextramers as per the manufacturer protocol[435]. The cells were then incubated with antibodies (CD3 EDC (Beckman Coulter), CD4+ APC-eFluor®E780 (eBioscience), CD8+ PerCP (BD)), for a further 30 minutes at 4 °C, washed and fixed with 1% Formaldehyde for a minimum of 1 hour prior to analysis. After fixation, the cells were analysed on a Dako CyAn ADP flow cytometer at the WIMM FACS facility using summit software. Non-identical HLA dextramers with an irrelevant epitope were used as gating controls, single color stains of cells with appropriate antibodies and unstained cells were used to calculate the compensation. Analyses were carried out using the Flowjo version 9.2

(TreeStar Inc.) software package. Graphical representation and statistical analysis was performed using Prism software (Graphpad).

7.4. Separation, extraction, cloning and sequencing of KK10 epitope form specific CD8+ T-cells

7.4.1. Staining and cell sorting

Cell sorting took place in the Henry Wellcome building for biomedical research at Cardiff University in the lab of David Price. Staining of patient PBMC with dextramer was carried out according to the manufacturers protocol435 and as described above (Chapter 7.3). Additional antibodies to those used for dextramer staining on the cyan flow cytometer for cell type and phenotypic markers were used in this assay in order to obtain improved separation and additional data on T-cell function (CD45RO, CD57, CD27, CD14, CD19, CD56). Compensation controls were set up using BD CompBeads with single colour antibody staining. Cell sorting was carried out using a customized 20-parameter FACSAria II flow cytometer (BD) and samples were sorted directly into 1.5ml tubes containing RNAlater (Invitrogen). FACSAria II Set-up and sample running were carried out with the assistance of Dr. Kristin Ladell.

7.4.2. mRNA Isolation

mRNA isolation was carried out place in the Henry Wellcome building for biomedical research at Cardiff University in the lab of David Price with the assistance of Dr James McLaren. Isolation was performed using a μMACS mRNA isolation kit according to the manufacturers protocol[436]. Briefly, cells were centrifuged in RNAlater for 7 minutes at

15000g. Supernatant was then removed from the sample, 900µl Lysis/Binding buffer added and the sample mixed. 50µl Oligo(dT) beads were added to the lysate and mixed. The lysate was then loaded onto a MACS column pre-rinsed with 100µl Lysis/Binding buffer and allowed to run through. The column was then washed with 2 x 200µl of Lysis/Binding buffer followed by 4 x 100µl Wash buffer. 27µl hot Elution buffer was then applied to the column and allowed to run through. A further 30µl hot Elution buffer was applied and the eluate collected.

7.4.3. Reverse transcription and amplification of cDNA

cDNA was made from the mRNA obtained in the previous step through the use of the SMARTer RACE cDNA amplification kit (Clontech). All reagents are those provided in this kit unless indicated otherwise.

Briefly to prepare RACE ready cDNA 1µl of the 5' CDS primer A was added to 6µl mRNA. This was heated at 72°C for 3 minutes and then held at 42°C for 2 minutes. To this was added 1µl SMARTer II anchor oligonucleotide, 2µl 5x First Strand buffer (250mM Tris-HCl (pH 8.3), 375mM KCl, 30mM $MgCl_2$), 1µl DTT (20mM), 1µl dNTPs (each 10mM), 1µl RNAse inhibitor (40U/µl) and 1µl Superscript II RNAse H^- Reverse Transcriptase (supplied at 200U/µl, Invitrogen) and the mix was heated at 42°C for 2 hours. 10µl Tricine-EDTA buffer (10mM Tricine-KOH (pH 8.5), 1.0mM EDTA) was added and the Reverse Transcriptase denatured through heating at 72°C for 10 minutes. cDNA was stored at -80°C until ready for use.

For amplification of cDNA for the TCR α and β chains the following primers were used:

MBC2 (β-chain)(25µM): 5'-tgcttctgatggctcaaacacagcgacct-3'

Hu3CAC3 (α-chain)(25µM): 5'-aataggcagacagacttgtcactgga-3'

5x Universal Primer Mix (UPM)(Clontech):

Long (0.4µM): 5'-ctaatacgactcactatagggcaagcagtggtatcaacgcagagt-3'

Short (2µM): 5'-ctaatacgactcactatagggc-3'

In a 50µl reaction, 1µl of the TCR chain specific primer was combined with 10µl UPM, 1µl dNTPs, 5µl 10x PCR buffer (400mM Tricine-KOH (pH 8.7 at 25°C), 150mM KOAc, 35mM Mg(OAc)2, 37.5µg/ml BSA, 0.05% Tween 20, 0.05% Nonidet-P40), 1µl dNTPs, 13µl cDNA from previous stage, 19µl of DEPC treated water and 1µl 50x Advantage 2 polymerase mix (TITANIUM Taq DNA polymerase, proofreading polymerase and Taqstart antibody, 1.1µg/µl in 50% glycerol, 15mM Tris-HCL (pH8.0), 75mM KCl, 0.05mM EDTA). This was mixed and underwent the following PCR program on an applied biosystems GeneAmp 2700 thermal cycler:

95°C for 30 sec

to activate Taq, then….

95°C 5 sec

72°C 2 min

for 5 cycles

95°C 5 sec

70°C 10 sec

72°C 2 min

for 5 cycles

95°C 5 sec

68°C 10 sec

72°C 2 min

for 35 cycles

Following this, the reaction mixture was run on a 1.3% Agarose gel (TAE buffered) with a 100bp DNA ladder marker and bands were visualised using SYBR Gold (Invitrogen) staining under a UV light source. Bands at approximately 500bp were cut from the gel (TCR chains run at approximately this size, with variation of around+/-100bp). cDNA was extracted from these bands using the Nucleospin extract II kit (Clontech).

7.4.4. Cloning of TCR into TOPO vector and preparation for sequencing

Purified cDNA was ligated into the pCR4-TOPO vector (Invitrogen) using the TOPO TA Cloning Kit for Sequencing (Invitrogen). 4µl of cDNA was mixed with 1µ1 salt solution (1.2M NaCl, 0.06M $MgCl_2$) and the TOPO vector (10ng/µl plasmid linearized with bound topoisomerase in 50% glycerol, 50mM Tris-HCL pH7.4, 1mM EDTA, 2mM DTT, 0.1% Triton X-100, 100µg/ml BSA 30µM phenol red). This was incubated at room temperature for 5 minutes.

The ligated construct was transformed into DH5α-$T1^R$ competent *E.coli* (Invitrogen). Briefly 2µl of the cloning reaction was added to tubes of One Shot DH5α-$T1^R$ (Invitrogen) and incubated on ice for 20 minutes. The cells were then heat shocked at 42°C for 30 seconds and returned to ice for 2 minutes. 250µl S.O.C medium (2% tryptone, 0.5% yeast extract, 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$, 20mM glucose) was added and the cells incubated at 37°C for 45 minutes. Cells were then plated onto LB AIX plates (LB agar with 0.1mM IPTG and 40µg/ml X-gal) and incubated overnight at 37°C.

Following incubation, colony PCR was performed for amplification of single TCR sequences. Platinum High Fidelity Taq DNA polymerase (Invitrogen) was used for this PCR. 96 well PCR plates (Bio-Rad) were prepared with 2.5µl 10x HiFi buffer (600mM Tris-SO4 (pH8.9),

180mM Ammonium Sulfate), 1µl $MgSO_4$ (50mM), 0.5µl dNTPs, 0.14µl HiFi Taq, 18.86µl PCR grade H_20 and 1µl each of the following primers:

M13F (5µM): 5'-ttttcccagtcacgac-3'

M13R (5µM): 5'-caggaaacagctatgac-3'

Individual white colonies were transferred from the AIX plates into each well. One 96 well plate was completed for each KK10 epitope specific T-cell population. Two wells were reserved for negative controls, one without a colony and one with a blue colony added. PCR used an applied biosystems GeneAmp 2700 thermal cycler, the program was as follows:

95°C 30sec

57°C 30sec

68°C 3min

Repeated for 35 cycles

Prior to sequencing, amplification of cDNA was confirmed by screening of 5µl of 12 colonies from across the plate and the negative controls on a 1.3% Agarose gel for appropriate sized bands.

7.4.5. Sequencing of TCR

Sequencing was carried out using the MBC2 or HU3CA3 primers as appropriate at the Weatherall Institute of Molecular Medicine core sequencing facility using a 48 capillary ABI-3730 DNA analyser. Sequences were analysed by import into Sequencher (Gene Codes Corporation, Ann Arbor, MI, USA) aligned and manually edited.

7.5. Intracellular cytokine staining of patient PBMC

Patient PBMC was thawed and resuspended in RPMI-1640 medium (Sigma) containing 10% Fetal Calf Serum, 100U penicillin/ml, 100µg streptomycin/ml, 2mM L-glutamine (R10 medium) along with 10µl DNase (2mg/ml) and incubated at 37°C for 1 hour at 1×10^6 cells/ml. Cells were then washed in R10 and at this stage approximately 1.5×10^5 cells were taken for later gag sequencing and stored in white blood cell lysis buffer (Qiagen). Cells were then plated out at 100µl per well and stimulated with 1µl CD28/CD49d Co-stimulatory agent (BD Biosciences), 500ng/ml ionomycin and 50ng/ml phorbol 12-myristate 13acetate (PMA). At this stage 4µl CD107a-FITC was also added to samples along with peptides (dissolved in DMSO and diluted in R10) at the required concentration. Unstimulated control samples received only the co-stimulatory agent. Following incubation at 37°C for 1 hour 1µl of Golgiplug (containing Brefeldin A) and 0.2µl of Monensin (final conc 2µM) (ebioscience) were added per well. Samples were then incubated at 37°C for 5.5 hours.

Samples were then washed twice in FACS buffer and resuspended in FACS buffer with LIVE/DEAD® fixable violet dead cell stain (Invitrogen) for 20 minutes at room temperature. Cells were washed twice, fixed and permeabilised using Cytofix/Cytoperm solution and wash (BD biosciences), they were then stained with the following antibody panel, CD3 EDC (Beckman Coulter), CD4+ APC-eFluor®E780 (eBioscience), CD8+ PerCP (BD), IL-2 PE, IFN-g Pe-Cy7 and TNF-a APC (all ebioscience), for 30 minutes at 4°C. Cells were washed and resuspended in 1% formaldehyde for analysis.

7.6. Sequencing of HIV-1 p17 and p24 gag from patient PBMC

Cells obtained from patient samples either during dextramer staining or ICS analysis were stored in Cell Lysis Solution (Qiagen) and DNA was extracted using the Qiagen Gentra Puregene Blood Core Kit A according to manufacturers protocol. Briefly lysed cells were treated with RNase and protein was precipitated. DNA was the precipitated using isopropanol, washed in ethanol and resuspended in PCR grade water.

An established nested PCR technique was then used to amplify the gag sequences from the proviral DNA. We used the Advantage 2 polymerase kit (Clontech) for both the1st and 2nd round PCR reactions. Buffers and enzyme specifications can be found in section 7.4.3

1st Round reaction

These were used in a 25µl reaction mix in combination with 2µl extracted DNA and 0.25µl of each of the following primers:

Forward primer: *Gag* positions 728-751: 5'-cgaggggcggcgactggtgagtac-3'

Reverse primer: *Gag* positions 1941-1916: 5'-cctaaaattgcctctctgcatcatta-3'

This underwent the following PCR program on an applied biosystems GeneAmp 2700 thermal cycler:

95°C 1min

Then

95°C 30sec

50°C 15sec

68°C 3min

-For 5 cycles

95°C 30sec

55°C 15sec

68°C 3min

-For 5 cycles

95°C 30sec

65°C 3sec

68°C 3min

-For 25 cycles

2nd Round reaction

Following the amplification in the 1st reaction, 3µl of the reaction mix was used in a 50µl mix with 0.5µl of the following primers:

Forward primer: *Gag* positions 763-788: 5'-tgactagcggaggctagaaggagaga-3'

Reverse primer: *Gag* positions 1911-1884: 5'-agctgaatttgttacttggctcattgcc-3'

This underwent the following PCR program on an applied biosystems GeneAmp 2700 thermal cycler:

95°C 1min

Then

95°C 30sec

58°C 15sec

68°C 3min

-For 10 cycles

95°C 30sec

65°C 3sec

68°C 3min

-For 25 cycles

Following this, amplification was confirmed and the DNA purified by running on a 1% Agarose gel with a 1kb DNA ladder marker. A single band at approximately 1kb was extracted and purified using the Nucleospin extract II kit (Clontech). Purified DNA samples were cloned into the PCR4 TOPO vector for sequencing for further amplification using the method described in 7.4.4. 10 colonies were then picked, placed in 5µl LB medium with ampicillin (100µg/ml) and incubated overnight. Plasmid DNA was extracted from these cultures using the Qiaprep spin miniprep kit (Qiagen) according to manufacturers protocol. Prior to sequencing, ligation and transformation were checked by performing an ECoR1 restriction digest (NEB) on 2µl of the purified DNA and running the resulting reaction mix on a 1% Agarose gel. Samples containing a band at 1kb were sent for sequencing at the Weatherall Institute of Molecular Medicine core sequencing facility using a 48 capillary ABI-3730 DNA analyser. Sequences were initially analysed by import into Sequencher (Gene Codes Corporation, Ann Arbor, MI, USA) aligned and manually edited. Further analysis was performed through manual alignment and editing in MacClade.

7.7. T2 cell line KK10 epitope form binding assay

T2 cells transfected with HLA B*2705 were thawed and expanded for several days in R10 medium. Cells were washed and resuspended in RPMI-1640 medium (Sigma) containing 100U penicillin/ml, 100µg streptomycin/ml, 2mM L-glutamine (R0 medium), 1.5×10^5 cells were plated into wells of a 96 well tissue culture plate and incubated with KK10 epitope form peptides KL8, KK10, KI11, KR13 and KM14 at concentrations of 0-100µM. TW10 and KL8 KKM were used as non-HLA B*2705 binding control peptides. Incubation was at 28°C for 16 hours. Cells were then washed and resuspended in PBS and stained with LIVE/DEAD® fixable violet dead cell stain (Invitrogen) for 20 minutes at room temperature. Cells were washed again and stained with HLA B*2705 specific ME-1 antibody (produced by HIU lab management) and then washed and a secondary anti mouse IgG FITC conjugated antibody added, each incubation was for 30 minutes at 4°C. Finally cells were washed and fixed in 1% formaldehyde prior to analysis.

7.8. T2 cell IFNγ ELISpot competition assay

7.8.1. Expansion of T-cell clones

Fresh blood from 3 healthy donors was collected and lymphocytes were isolated through Ficoll-Paque (GE Healthcare) density gradient separation. Donor lymphocytes were mixed and resuspended to a density of 2×10^6 cells/ml in RPMI-1640 medium (Sigma) containing 10% Human AB Serum, 100U penicillin/ml, 100µg streptomycin/ml and 2mM L-glutamine (H10 medium). The mixed lymphocytes were then irradiated at 3000 rads for 15 minutes.

In a T25 flask, thawed vials of 1×10^6cells of the KK10 specific T-cell clone G12C was combined with 20mls H10 medium combined with 25ng/ml IL15, 1µg/ml phytohemagglutinin (PHA), 20iU/ml IL2 and 1×10^7 cells irradiated donor PBMC. Flasks are then incubated undisturbed at 37°C for 5 days. At day 5, 10mls of media are removed from the flask without disturbing the cells and replaced with H10 containing 25ng/ml IL15 and 20iU/ml IL2. This is incubated for a further 2 days. Finally, cells are counted and plated in H10 containing 25ng/ml IL15 and 200iU/ml IL2 at 3×10^6 cells/well in a 24 well plate.

7.8.2. IFNγ ELISpot assay

An ELISpot assay was used to determine the effects of truncated peptides on the IFNγ response to KK10 by CD8+ T-cells (Millipore IFNγ ELISpot kit). A KK10 specific clone, G12C (kindly provided by Prof Victor Appay) was used with a T2 tap deficient, HLA class-1 deficient murine B-cell line transfected with HLA B*2705 (kindly provided by Dr Simon Kollnberger) as target cells. These cells were incubated with an IC50 concentration of KK10 determined through titration of the KK10 response in an initial ELISpot assay. Truncated peptides were incubated with samples in increasing concentrations against the IC50 KK10 peptide to determine the presence of any effect on the KK10 response. Cells incubated with either IC50 KK10 alone or each of the truncated peptides alone at their highest concentration were used as controls. Truncated peptide alone wells provided baseline readings for analysis. Samples were tested in duplicate.

Briefly, ELISpot 96 well plates (Millipore) were coated with capture antibody at 15µg/ml. Plates were incubated for 2 hours at 37°C before washing 6x with RPMI-1640 (Sigma). Before use plates were blocked with R10 medium for 1 hour at room temperature. In a round bottomed 96 well tissue culture plate 5×10^3 T2-B27 cells were plated in each well

(50µl) and appropriate concentrations of the truncated peptides KL6, KG7 and KL8 or a pool of these were added and incubated at 37°C for 30 minutes. KK10 at the IC50 concentration was then added to samples and incubated for a further 30 minutes. The G12C clone was then added to the wells at 200 cells/well. Following this, blocking medium was removed from the ELISpot plate and samples transferred. The plate was incubated overnight at 37°C.

Cells were then removed from the plate and the plate washed 6x with PBS containing 0.05% Tween-20. Detector antibody (50µl per well) was then incubated on the plate for 3 hours at room temperature before repeating the wash step. This process was repeated for the conjugate antibody, (50µl per well, 1.5 hours incubation) before developing the plate and leaving to dry for two days with ventilation. Spots were counted using an ELISpot plate reader and converted to spot forming units (SFU).

7.9. KIR3DL1 binding Assay

BAF/3 IL3 dependent murine pro-B lymphoid cell lines stably transfected with KIR3DL1, KIR3DL2 and without transfection (Kind gifts from Dr Simon Kollnberger) were expanded in R10 medium with 0.5ng/ml mouse IL3. Cells were washed in FACS buffer (PBS with 5% FCS) and plated at 100µl/well at a density of 2×10^6 cells/ml. Cells were then stained with LIVE/DEAD® fixable violet dead cell stain (Invitrogen) for 20 minutes at room temperature. Cells were washed twice and then incubated with appropriate dextramers as per the manufacturer protocol[435]. Following this, cells were washed and fixed in 1% formaldehyde for a minimum of 1 hour prior to analysis. Samples were prepared in duplicate for each test (up to 3 tests for some dextramer). Controls included a negative unconjugated PE/APC only stain and untransfected BAF cells with identical dextramer stains. Expression of KIR3DL1 was confirmed using a KIR3DL1 specific DX9 antibody (FITC conjugated, BD biosciences).

8. References

1. World Health Organisation, D. o. H. A. (2011). Global HIV/AIDS response: epidemic update and health sector progress towards universal access: progress report 2011. In *Global HIV/AIDS Response: Progress reports*. World Health Organisation Geneva.
2. Temin HM, M. S. (1970). Viral RNA-dependent DNA Polymerase-RNA-dependent DNA Polymerase in Virions of Rous Sarcoma Virus. *Nature* **226**, 1211-1213.
3. Franchini G, B. M. (1989). Genetic Relatedness of the Human Immunodeficiency Viruses Type 1 and 2 (HIV-1, HIV-2) and the Simian Immunodeficiency Virus (SIV). *Annals of the New York Academy of Sciences* **554**, 81-87.
4. McDougal JS, M. A., Cort SP, Nicholson JK, Cross GD, Scheppler-Campbell JA, Hicks D, Sligh J. (1985). Cellular tropism of the human retrovirus HTLV Ill:LAy. I. role of T cell activation and expression of the T4 antigen. *Journal of Immunology* **135**, 3151-3162.
5. David Klatzmann, E. C., Sophie Chamaret, Jacqueline Gruest, Denise Guetard, Thierry Hercend, Jean-Claude Gluckman & Luc Montagnier (1984). T-lymphocyte T4 molecule behaves as the receptor for human retroviral LAy. *Nature* **312**, 767.
6. Groenink M, F. R., Broersen S, Baker CH, Koot M, van't Wout AB, Huisman HG, Miedema F, Tersmette M, Schuitemaker H. (1993). Relation of phenotype evolution of HIV-1 to envelope V2 configuration. *Science* **260**, 1513-1516.
7. Bushman FD, F. T., Craigie R. (1990). Retroviral DNA Integration Directed by HIV Integration Protein in Vitro. *Science* **249**, 1555-1558.
8. Haseltine, C. F. a. W. (1991). Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. *Journal of Virology* **65**, 1910-1915.
9. Farnet CM, H. W. (1990). Integration of human immunodeficiency virus type 1 DNA in vitro. *Proceedings of the National Academy of Sciences* **87**, 4164-4168.
10. Brigati, C., Giacca, M., Noonan, D. M. & Albini, A. (2003). HIV Tat, its TARgets and the control of viral gene expression. *FEMS Microbiology Letters* **220**, 57-65.
11. Lee JH, L. Y., Doerre S, Sista P, Ballard DW, Greene WC, Franza BR Jr. (1991). A member of the set of kappa B binding proteins, HIVEN86A, is a product of the human c-rel proto-oncogene. *Oncogene* **6**, 665-667.
12. Böhnlein E, L. J., Siekevitz M, Ballard DW, Franza BR, Greene WC. (1988). The same inducible nuclear proteins regulates mitogen activation of both the interleukin-2 receptor-alpha gene and type 1 HIV. *Cell* **53**, 827-836.
13. H Mitsuya, K. J. W., P A Furman, M H St Clair, S N Lehrman, R C Gallo, D Bolognesi, D W Barry, and S Broder. (1985). 3'-Azido-3'-deoxythymidine (BW A509U)- an antiviral agent that inhibits the infectivity and cytopathic effect of human T-lymphotropic virus type III:lymphadenopathy-associated virus in vitro. *Proceedings of the National Academy of Sciences* **82**, 7096-7100.
14. Broder, H. M. a. S. (1986). Inhibition of the in vitro infectivity and cytopathic effect of human T-lymphotrophic virus type III:lymphadenopathy-associated virus (HTLV-III:LAV) by 2',3'-dideoxynucleosides. *Proceedings of the National Academy of Sciences* **83**, 1911-1915.
15. P A Furman, J. A. F., M H St Clair, K Weinhold, J L Rideout, G A Freeman, S N Lehrman, D P Bolognesi, S Broder, H Mitsuya. (1986). Phosphorylation of 3'-azido-3'-deoxythymidine and selective interaction of the 5'-triphosphate with human immunodeficiency virus reverse transcriptase. *Proceedings of the National Academy of Sciences* **83**, 8333-8337.

16. Abram, M. E., Ferris, A. L., Shao, W., Alvord, W. G. & Hughes, S. H. (2010). Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *Journal of Virology* **84**, 9864-78.
17. Leslie, A. J., Pfafferott, K. J., Chetty, P., Draenert, R., Addo, M. M., Feeney, M., Tang, Y., Holmes, E. C., Allen, T., Prado, J. G., Altfeld, M., Brander, C., Dixon, C., Ramduth, D., Jeena, P., Thomas, S. A., St John, A., Roach, T. A., Kupfer, B., Luzzi, G., Edwards, A., Taylor, G., Lyall, H., Tudor-Williams, G., Novelli, V., Martinez-Picado, J., Kiepiela, P., Walker, B. D. & Goulder, P. J. (2004). HIV evolution: CTL escape mutation and reversion after transmission. *Nature medicine* **10**, 282-9.
18. Karlsson, A. C., Iversen, A. K., Chapman, J. M., de Oliviera, T., Spotts, G., McMichael, A. J., Davenport, M. P., Hecht, F. M. & Nixon, D. F. (2007). Sequential broadening of CTL responses in early HIV-1 infection is associated with viral escape. *PloS one* **2**, e225.
19. Boutwell, C. L., Rolland, M. M., Herbeck, J. T., Mullins, J. I. & Allen, T. M. (2010). Viral evolution and escape during acute HIV-1 infection. *The Journal of infectious diseases* **202 Suppl 2**, S309-14.
20. Justice, A., Modur, S., Tate, J., Althoff, K., Jacobson, L., Gebo, K., Kitahata, M., Horberg, M., Brooks, J., Buchacz, K., Rourke, S., Rachlis, A., Napravnik, S., Eron, J., Willig, J., Moore, R., Kirk, G., Bosch, R., Rodriguez, B., Hogg, R., Thorne, J., Goedert, J., Klein, M., Gill, M., Deeks, S., Sterling, T., Anastos, K. & Gange, S. (2012). Predictive Accuracy of the Veterans Aging Cohort Study (VACS) Index for Mortality with HIV Infection: A North American Cross Cohort Analysis. *Journal of acquired immune deficiency syndromes*.
21. Pai NP, T. J., Lawrence J, Colford JM Jr, Reingold AL. (2005). Structured treatment interruptions (STI) in chronic suppressed HIV infection in adults. *Cochrane Database of Systematic Reviews* **4**, CD005482.
22. Sedaghat, A. R., Siliciano, R. F. & Wilke, C. O. (2008). Low-level HIV-1 replication and the dynamics of the resting CD4+ T cell reservoir for HIV-1 in the setting of HAART. *BMC infectious diseases* **8**, 2.
23. Chun TW, S. L., Mizell SB, Ehler LA, Mican JA, Baseler M, Lloyd AL, Nowak MA, Fauci AS. (1997). Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proceedings of the National Academy of Sciences* **94**, 13193-13197.
24. (2012). UNAIDS Report on the Global AIDS EpidemIc 2012. UNAIDS.
25. Salomon JA, H. D. (2008). Evaluating the impact of antiretroviral therapy on HIV transmission. *AIDS* **22 (Suppl 1)**, S149-59.
26. Hollingsworth, T. D., Anderson, R. M. & Fraser, C. (2008). HIV-1 transmission, by stage of infection. *The Journal of infectious diseases* **198**, 687-93.
27. Roof, P., Ricci, M., Genin, P., Montano, M. A., Essex, M., Wainberg, M. A., Gatignol, A. & Hiscott, J. (2002). Differential regulation of HIV-1 clade-specific B, C, and E long terminal repeats by NF-kappaB and the Tat transactivator. *Virology* **296**, 77-83.
28. Gavin Morrow, L. V., Panagiotis Vagenas and Melissa Robbiani. (2007). Current Concepts of HIV Transmission. *Current HIV/AIDS Reports* **4**, 29-35.
29. Cameron PU, F. P., Barker JM, Gezelter S, Inaba K, Steinman RM. (1992). Dendritic cells exposed to human immunodeficiency virus type-1 transmit a vigorous cytopathic infection to CD4+ T cells. *Science* **257**, 383-387.
30. Turville, S. G., Cameron, P. U., Handley, A., Lin, G., Pohlmann, S., Doms, R. W. & Cunningham, A. L. (2002). Diversity of receptors binding HIV on dendritic cell subsets. *Nature immunology* **3**, 975-83.
31. Geijtenbeek TB, K. D., Torensma R, van Vliet SJ, van Duijnhoven GC, Middel J, Cornelissen IL, Nottet HS, KewalRamani VN, Littman DR, Figdor CG, van Kooyk Y.

(2000). DC-SIGN, a Dendritic Cell–Specific HIV-1-Binding Protein that Enhances trans-Infection of T Cells. *Cell* **100**, 587-597.

32. Salazar-Gonzalez, J. F., Salazar, M. G., Keele, B. F., Learn, G. H., Giorgi, E. E., Li, H., Decker, J. M., Wang, S., Baalwa, J., Kraus, M. H., Parrish, N. F., Shaw, K. S., Guffey, M. B., Bar, K. J., Davis, K. L., Ochsenbauer-Jambor, C., Kappes, J. C., Saag, M. S., Cohen, M. S., Mulenga, J., Derdeyn, C. A., Allen, S., Hunter, E., Markowitz, M., Hraber, P., Perelson, A. S., Bhattacharya, T., Haynes, B. F., Korber, B. T., Hahn, B. H. & Shaw, G. M. (2009). Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *The Journal of experimental medicine* **206**, 1273-89.
33. Fischer W, G. V., Giorgi EE, Hraber PT, Keele BF, Leitner T, Han CS, Gleasner CD, Green L, Lo CC, Nag A, Wallstrom TC, Wang S, McMichael AJ, Haynes BF, Hahn BH, Perelson AS, Borrow P, Shaw GM, Bhattacharya T, Korber BT. (2010). Transmission of Single HIV-1 Genomes and Dynamics of Early Immune Escape Revealed by Ultra-Deep Sequencing. *PloS one* **5**, e12303.
34. Bar, K. J., Li, H., Chamberland, A., Tremblay, C., Routy, J. P., Grayson, T., Sun, C., Wang, S., Learn, G. H., Morgan, C. J., Schumacher, J. E., Haynes, B. F., Keele, B. F., Hahn, B. H. & Shaw, G. M. (2010). Wide variation in the multiplicity of HIV-1 infection among injection drug users. *Journal of Virology* **84**, 6241-7.
35. Keele, B. F., Li, H., Learn, G. H., Hraber, P., Giorgi, E. E., Grayson, T., Sun, C., Chen, Y., Yeh, W. W., Letvin, N. L., Mascola, J. R., Nabel, G. J., Haynes, B. F., Bhattacharya, T., Perelson, A. S., Korber, B. T., Hahn, B. H. & Shaw, G. M. (2009). Low-dose rectal inoculation of rhesus macaques by SIVsmE660 or SIVmac251 recapitulates human mucosal infection by HIV-1. *The Journal of experimental medicine* **206**, 1117-34.
36. Lee, H. Y., Giorgi, E. E., Keele, B. F., Gaschen, B., Athreya, G. S., Salazar-Gonzalez, J. F., Pham, K. T., Goepfert, P. A., Kilby, J. M., Saag, M. S., Delwart, E. L., Busch, M. P., Hahn, B. H., Shaw, G. M., Korber, B. T., Bhattacharya, T. & Perelson, A. S. (2009). Modeling sequence evolution in acute HIV-1 infection. *Journal of theoretical biology* **261**, 341-60.
37. Keele, B. F., Giorgi, E. E., Salazar-Gonzalez, J. F., Decker, J. M., Pham, K. T., Salazar, M. G., Sun, C., Grayson, T., Wang, S., Li, H., Wei, X., Jiang, C., Kirchherr, J. L., Gao, F., Anderson, J. A., Ping, L. H., Swanstrom, R., Tomaras, G. D., Blattner, W. A., Goepfert, P. A., Kilby, J. M., Saag, M. S., Delwart, E. L., Busch, M. P., Cohen, M. S., Montefiori, D. C., Haynes, B. F., Gaschen, B., Athreya, G. S., Lee, H. Y., Wood, N., Seoighe, C., Perelson, A. S., Bhattacharya, T., Korber, B. T., Hahn, B. H. & Shaw, G. M. (2008). Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 7552-7.
38. Emau, P., Jiang, Y., Agy, M. B., Tian, B., Bekele, G. & Tsai, C. C. (2006). Post-exposure prophylaxis for SIV revisited: animal model for HIV prevention. *AIDS research and therapy* **3**, 29.
39. Chun TW, E. D., Berrey MM, Shea T, Corey L, Fauci AS. (1998). Early establishment of a pool of latently infected, resting CD4+ T cells during primary HIV-1 infection. *Proceedings of the National Academy of Sciences* **95**, 8869-8873.
40. Cicala, C., Martinelli, E., McNally, J. P., Goode, D. J., Gopaul, R., Hiatt, J., Jelicic, K., Kottilil, S., Macleod, K., O'Shea, A., Patel, N., Van Ryk, D., Wei, D., Pascuccio, M., Yi, L., McKinnon, L., Izulla, P., Kimani, J., Kaul, R., Fauci, A. S. & Arthos, J. (2009). The integrin alpha4beta7 forms a complex with cell-surface CD4+ and defines a T-cell subset that is highly susceptible to infection by HIV-1. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 20877-82.

41. Mehandru, S., Poles, M. A., Tenner-Racz, K., Horowitz, A., Hurley, A., Hogan, C., Boden, D., Racz, P. & Markowitz, M. (2004). Primary HIV-1 infection is associated with preferential depletion of CD4+ T lymphocytes from effector sites in the gastrointestinal tract. *The Journal of experimental medicine* **200**, 761-70.
42. Brenchley, J. M., Schacker, T. W., Ruff, L. E., Price, D. A., Taylor, J. H., Beilman, G. J., Nguyen, P. L., Khoruts, A., Larson, M., Haase, A. T. & Douek, D. C. (2004). CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. *The Journal of experimental medicine* **200**, 749-59.
43. Gasper-Smith, N., Crossman, D. M., Whitesides, J. F., Mensali, N., Ottinger, J. S., Plonk, S. G., Moody, M. A., Ferrari, G., Weinhold, K. J., Miller, S. E., Reich, C. F., 3rd, Qin, L., Self, S. G., Shaw, G. M., Denny, T. N., Jones, L. E., Pisetsky, D. S. & Haynes, B. F. (2008). Induction of plasma (TRAIL), TNFR-2, Fas ligand, and plasma microparticles after human immunodeficiency virus type 1 (HIV-1) transmission: implications for HIV-1 vaccine design. *Journal of Virology* **82**, 7700-10.
44. Levesque, M. C., Moody, M. A., Hwang, K. K., Marshall, D. J., Whitesides, J. F., Amos, J. D., Gurley, T. C., Allgood, S., Haynes, B. B., Vandergrift, N. A., Plonk, S., Parker, D. C., Cohen, M. S., Tomaras, G. D., Goepfert, P. A., Shaw, G. M., Schmitz, J. E., Eron, J. J., Shaheen, N. J., Hicks, C. B., Liao, H. X., Markowitz, M., Kelsoe, G., Margolis, D. M. & Haynes, B. F. (2009). Polyclonal B cell differentiation and loss of gastrointestinal tract germinal centers in the earliest stages of HIV-1 infection. *PLoS medicine* **6**, e1000107.
45. Schacker TW, H. J., Shea T, Coombs RW, Corey L. (1998). Biological and Virologic Characteristics of Primary HIV Infection. *Annals of Internal Medicine* **128**, 613-620.
46. Stacey, A. R., Norris, P. J., Qin, L., Haygreen, E. A., Taylor, E., Heitman, J., Lebedeva, M., DeCamp, A., Li, D., Grove, D., Self, S. G. & Borrow, P. (2009). Induction of a striking systemic cytokine cascade prior to peak viremia in acute human immunodeficiency virus type 1 infection, in contrast to more modest and delayed responses in acute hepatitis B and C virus infections. *Journal of Virology* **83**, 3719-33.
47. Nilsson J, K.-d.-L. S., Granath A, Sönnerborg A, Goh LE, Andersson J. (2007). Early immune activation in gut-associated and peripheral lymphoid tissue during acute HIV infection. *AIDS* **21**, 565-574.
48. Centers for Disease Control and Prevention. (1999). CDC Guidelines for National Human Immunodeficiency Virus Case Surveillance, Including Monitoring for Human Immunodeficiency Virus Infection and Acquired Immunodeficiency Syndrome. Prevention", C. f. D. C. a.
49. Myron S. Cohen, G. M. S., Andrew J. McMichael, and Barton F. Haynes. (2011). Acute HIV-1 Infection. *New England Journal of Medicine* **364**, 1943-1954.
50. R A Koup, J. T. S., Y Cao, C A Andrews, G McLeod, W Borkowsky, C Farthing and D D Ho. (1994). Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *Journal of Virology* **68**, 4650-4655.
51. Goonetilleke, N., Liu, M. K., Salazar-Gonzalez, J. F., Ferrari, G., Giorgi, E., Ganusov, V. V., Keele, B. F., Learn, G. H., Turnbull, E. L., Salazar, M. G., Weinhold, K. J., Moore, S., Letvin, N., Haynes, B. F., Cohen, M. S., Hraber, P., Bhattacharya, T., Borrow, P., Perelson, A. S., Hahn, B. H., Shaw, G. M., Korber, B. T. & McMichael, A. J. (2009). The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *The Journal of experimental medicine* **206**, 1253-72.
52. Ferrari G, K. B., Goonetilleke N, Liu MK, Turnbull EL, Salazar-Gonzalez JF, Hawkins N, Self S, Watson S, Betts MR, Gay C, McGhee K, Pellegrino P, Williams I, Tomaras GD, Haynes BF, Gray CM, Borrow P, Roederer M, McMichael AJ, Weinhold KJ. (2011).

Relationship between Functional Profile of HIV-1 Specific CD8+ T Cells and Epitope Variability with the Selection of Escape Mutants in Acute HIV-1 Infection. *PLoS pathogens* **7**, e1001273.

53. Lichterfeld, M., Yu, X. G., Le Gall, S. & Altfeld, M. (2005). Immunodominance of HIV-1-specific CD8(+) T-cell responses in acute HIV-1 infection: at the crossroads of viral and host genetics. *Trends in immunology* **26**, 166-71.
54. Lichterfeld, M., Yu, X. G., Cohen, D., Addo, M. M., Malenfant, J., Perkins, B., Pae, E., Johnston, M. N., Strick, D., Allen, T. M., Rosenberg, E. S., Korber, B., Walker, B. D. & Altfeld, M. (2004). HIV-1 Nef is preferentially recognized by CD8+ T cells in primary HIV-1 infection despite a relatively high degree of genetic diversity. *AIDS* **18**, 1383-1392.
55. Bernardin, F., Kong, D., Peddada, L., Baxter-Lowe, L. A. & Delwart, E. (2005). Human immunodeficiency virus mutations during the first month of infection are preferentially found in known cytotoxic T-lymphocyte epitopes. *Journal of Virology* **79**, 11523-8.
56. Pantaleo G, D. J., Soudeyns H, Graziosi C, Denis F, Adelsberger JW, Borrow P, Saag MS, Shaw GM, Sekaly RP, et al. (1994). Major expansion of CD8+ T cells with a predominant Vβ usage during the primary immune response to HIV. *Nature* **370**, 463-467.
57. Altfeld, M., Kalife, E. T., Qi, Y., Streeck, H., Lichterfeld, M., Johnston, M. N., Burgett, N., Swartz, M. E., Yang, A., Alter, G., Yu, X. G., Meier, A., Rockstroh, J. K., Allen, T. M., Jessen, H., Rosenberg, E. S., Carrington, M. & Walker, B. D. (2006). HLA Alleles Associated with Delayed Progression to AIDS Contribute Strongly to the Initial CD8(+) T Cell Response against HIV-1. *PLoS medicine* **3**, e403.
58. Kelleher AD, L. C., Holmes EC, Allen RL, Wilson J, Conlon C, Workman C, Shaunak S, Olson K, Goulder P, Brander C, Ogg G, Sullivan JS, Dyer W, Jones I, McMichael AJ, Rowland-Jones S, Phillips RE. (2001). Clustered Mutations in HIV-1 gag Are Consistently Required for Escape from HLA-B27–restricted Cytotoxic T Lymphocyte Responses. *Journal of Experimental Medicine* **193**, 375-385.
59. Crawford, H., Prado, J. G., Leslie, A., Hue, S., Honeyborne, I., Reddy, S., van der Stok, M., Mncube, Z., Brander, C., Rousseau, C., Mullins, J. I., Kaslow, R., Goepfert, P., Allen, S., Hunter, E., Mulenga, J., Kiepiela, P., Walker, B. D. & Goulder, P. J. (2007). Compensatory mutation partially restores fitness and delays reversion of escape mutation within the immunodominant HLA-B*5703-restricted Gag epitope in chronic human immunodeficiency virus type 1 infection. *Journal of Virology* **81**, 8346-51.
60. Tomaras, G. D., Yates, N. L., Liu, P., Qin, L., Fouda, G. G., Chavez, L. L., Decamp, A. C., Parks, R. J., Ashley, V. C., Lucas, J. T., Cohen, M., Eron, J., Hicks, C. B., Liao, H. X., Self, S. G., Landucci, G., Forthal, D. N., Weinhold, K. J., Keele, B. F., Hahn, B. H., Greenberg, M. L., Morris, L., Karim, S. S., Blattner, W. A., Montefiori, D. C., Shaw, G. M., Perelson, A. S. & Haynes, B. F. (2008). Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *Journal of Virology* **82**, 12449-63.
61. Li, B., Gladden, A. D., Altfeld, M., Kaldor, J. M., Cooper, D. A., Kelleher, A. D. & Allen, T. M. (2007). Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. *Journal of Virology* **81**, 193-201.
62. Allen, T. M., Altfeld, M., Yu, X. G., O'Sullivan, K. M., Lichterfeld, M., Le Gall, S., John, M., Mothe, B. R., Lee, P. K., Kalife, E. T., Cohen, D. E., Freedberg, K. A., Strick, D. A., Johnston, M. N., Sette, A., Rosenberg, E. S., Mallal, S. A., Goulder, P. J., Brander, C. &

Walker, B. D. (2004). Selection, transmission, and reversion of an antigen-processing cytotoxic T-lymphocyte escape mutation in human immunodeficiency virus type 1 infection. *Journal of Virology* **78**, 7069-78.

63. Alter, G., Rihn, S., Walter, K., Nolting, A., Martin, M., Rosenberg, E. S., Miller, J. S., Carrington, M. & Altfeld, M. (2009). HLA class I subtype-dependent expansion of KIR3DS1+ and KIR3DL1+ NK cells during acute human immunodeficiency virus type 1 infection. *Journal of Virology* **83**, 6798-805.
64. Isabel Luque, R. S., M. Dolores Galiani, Rafael González, Fernando García, José A. López de Castro, José Peña. (1996). Threonine 80 on HLA-B27 confers protection against lysis by a group of natural killer clones. *European journal of immunology* **26**, 1974-1977.
65. Matano T, S. R., Siemon C, Connors M, Lane HC, Martin MA. (1998). Administration of an anti-CD8+ monoclonal antibody interferes with the clearance of chimeric simian:human immunodeficiency virus during primary infections of rhesus macaques. *Journal of Virology* **72**, 164-169.
66. Fellay, J., Ge, D., Shianna, K. V., Colombo, S., Ledergerber, B., Cirulli, E. T., Urban, T. J., Zhang, K., Gumbs, C. E., Smith, J. P., Castagna, A., Cozzi-Lepri, A., De Luca, A., Easterbrook, P., Gunthard, H. F., Mallal, S., Mussini, C., Dalmau, J., Martinez-Picado, J., Miro, J. M., Obel, N., Wolinsky, S. M., Martinson, J. J., Detels, R., Margolick, J. B., Jacobson, L. P., Descombes, P., Antonarakis, S. E., Beckmann, J. S., O'Brien, S. J., Letvin, N. L., McMichael, A. J., Haynes, B. F., Carrington, M., Feng, S., Telenti, A. & Goldstein, D. B. (2009). Common genetic variation and the control of HIV-1 in humans. *PLoS genetics* **5**, e1000791.
67. Pereyra, F., Jia, X., McLaren, P. J., Telenti, A., de Bakker, P. I., Walker, B. D., Ripke, S., Brumme, C. J., Pulit, S. L., Carrington, M., Kadie, C. M., Carlson, J. M., Heckerman, D., Graham, R. R., Plenge, R. M., Deeks, S. G., Gianniny, L., Crawford, G., Sullivan, J., Gonzalez, E., Davies, L., Camargo, A., Moore, J. M., Beattie, N., Gupta, S., Crenshaw, A., Burtt, N. P., Guiducci, C., Gupta, N., Gao, X., Qi, Y., Yuki, Y., Piechocka-Trocha, A., Cutrell, E., Rosenberg, R., Moss, K. L., Lemay, P., O'Leary, J., Schaefer, T., Verma, P., Toth, I., Block, B., Baker, B., Rothchild, A., Lian, J., Proudfoot, J., Alvino, D. M., Vine, S., Addo, M. M., Allen, T. M., Altfeld, M., Henn, M. R., Le Gall, S., Streeck, H., Haas, D. W., Kuritzkes, D. R., Robbins, G. K., Shafer, R. W., Gulick, R. M., Shikuma, C. M., Haubrich, R., Riddler, S., Sax, P. E., Daar, E. S., Ribaudo, H. J., Agan, B., Agarwal, S., Ahern, R. L., Allen, B. L., Altidor, S., Altschuler, E. L., Ambardar, S., Anastos, K., Anderson, B., Anderson, V., Andrady, U., Antoniskis, D., Bangsberg, D., Barbaro, D., Barrie, W., Bartczak, J., Barton, S., Basden, P., Basgoz, N., Bazner, S., Bellos, N. C., Benson, A. M., Berger, J., Bernard, N. F., Bernard, A. M., Birch, C., Bodner, S. J., Bolan, R. K., Boudreaux, E. T., Bradley, M., Braun, J. F., Brndjar, J. E., Brown, S. J., Brown, K., Brown, S. T., et al. (2010). The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* **330**, 1551-7.
68. Limou, S., Le Clerc, S., Coulonges, C., Carpentier, W., Dina, C., Delaneau, O., Labib, T., Taing, L., Sladek, R., Deveau, C., Ratsimandresey, R., Montes, M., Spadoni, J. L., Lelievre, J. D., Levy, Y., Therwath, A., Schachter, F., Matsuda, F., Gut, I., Froguel, P., Delfraissy, J. F., Hercberg, S. & Zagury, J. F. (2009). Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). *The Journal of infectious diseases* **199**, 419-26.
69. Richman, D. D., Wrin, T., Little, S. J. & Petropoulos, C. J. (2003). Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 4144-9.

70. Wei X, D. J., Wang S, Hui H, Kappes JC, Wu X, Salazar-Gonzalez JF, Salazar MG, Kilby JM, Saag MS, Komarova NL, Nowak MA, Hahn BH, Kwong PD, Shaw GM. (2003). Antibody neutralization and escape by HIV-1. *Nature* **422**, 307-312.
71. Simek, M. D., Rida, W., Priddy, F. H., Pung, P., Carrow, E., Laufer, D. S., Lehrman, J. K., Boaz, M., Tarragona-Fiol, T., Miiro, G., Birungi, J., Pozniak, A., McPhee, D. A., Manigart, O., Karita, E., Inwoley, A., Jaoko, W., Dehovitz, J., Bekker, L. G., Pitisuttithum, P., Paris, R., Walker, L. M., Poignard, P., Wrin, T., Fast, P. E., Burton, D. R. & Koff, W. C. (2009). Human immunodeficiency virus type 1 elite neutralizers: individuals with broad and potent neutralizing activity identified by using a high-throughput neutralization assay together with an analytical selection algorithm. *Journal of Virology* **83**, 7337-48.
72. Euler, Z., van Gils, M. J., Bunnik, E. M., Phung, P., Schweighardt, B., Wrin, T. & Schuitemaker, H. (2010). Cross-reactive neutralizing humoral immunity does not protect from HIV type 1 disease progression. *The Journal of infectious diseases* **201**, 1045-53.
73. van Gils, M. J., Euler, Z., Schweighardt, B., Wrin, T. & Schuitemaker, H. (2009). Prevalence of cross-reactive HIV-1-neutralizing activity in HIV-1-infected patients with rapid or slow disease progression. *AIDS* **23**, 2405-14.
74. Trkola, A., Kuster, H., Rusert, P., Joos, B., Fischer, M., Leemann, C., Manrique, A., Huber, M., Rehr, M., Oxenius, A., Weber, R., Stiegler, G., Vcelar, B., Katinger, H., Aceto, L. & Gunthard, H. F. (2005). Delay of HIV-1 rebound after cessation of antiretroviral therapy through passive transfer of human neutralizing antibodies. *Nature medicine* **11**, 615-22.
75. Balazs, A. B., Chen, J., Hong, C. M., Rao, D. S., Yang, L. & Baltimore, D. (2012). Antibody-based protection against HIV infection by vectored immunoprophylaxis. *Nature* **481**, 81-4.
76. Buchbinder SP, M. D., Duerr A, Fitzgerald DW, Mogg R, Li D, Gilbert PB, Lama JR, Marmor M, Del Rio C, McElrath MJ, Casimiro DR, Gottesdiener KM, Chodakewitz JA, Corey L, Robertson MN; Step Study Protocol Team. (2008). Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial. *The Lancet* **372**, 1881-1893.
77. Rerks-Ngarm S, P. P., Nitayaphan S, Kaewkungwal J, Chiu J, Paris R, Premsri N, Namwat C, de Souza M, Adams E, Benenson M, Gurunathan S, Tartaglia J, McNeil JG, Francis DP, Stablein D, Birx DL, Chunsuttiwat S, Khamboonruang C, Thongcharoen P, Robb ML, Michael NL, Kunasol P, Kim JH; MOPH-TAVEG Investigators. (2009). Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *New England Journal of Medicine* **361**, 2209-2220.
78. Magdalena Magierowska, I. T., Patrice Debré, Françoise Sanson, Brigitte Autran, Yves Rivière, Dominique Charron, French ALT, IMMUNOCO Study Groups and Dominique Costagliola. (1999). Combined genotypes of CCR5, CCR2 SDF1, and HLA genes that can predict the long-term nonprogressor status in human immunodeficiency virus–1 infected individuals. *Blood* **93**, 936-941.
79. R.A. Kaslow, M. C., R. Apple, L. Park, A. Muñoz, A.J. Saah, J.J. Goedert, C. Winkler, S.J. O'Brien, C. Rinaldo, R. Detels, W. Blattner, J. Phair, H. Erlich & D.L. Mann. (1996). Influence of combinations of human major histocompatibility complex genes on the course of HIV−1 infection. *Nature Medicine* **2**, 405-411.
80. Sambandam, A., Bell, J. J., Schwarz, B. A., Zediak, V. P., Chi, A. W., Zlotoff, D. A., Krishnamoorthy, S. L., Burg, J. M. & Bhandoola, A. (2008). Progenitor migration to the thymus and T cell lineage commitment. *Immunologic research* **42**, 65-74.

81. Carpenter, A. C. & Bosselut, R. (2010). Decision checkpoints in the thymus. *Nature immunology* **11**, 666-73.
82. Schwarz BA, B. A. (2006). Trafficking from the bone marrow to the thymus- a prerequisite for thymopoiesis. *Immunological Reviews* **209**, 47-57.
83. Wilkinson B, O. J., Jenkinson EJ. (1999). Factors regulating stem cell recruitment to the fetal thymus. *Journal of Immunology* **162**, 3873-3881.
84. Maillard, I., Fang, T. & Pear, W. S. (2005). Regulation of lymphoid development, differentiation, and function by the Notch pathway. *Annual review of immunology* **23**, 945-74.
85. Pui, J. C., Allman, D., Xu, L., DeRocco, S., Karnell, F. G., Bakkour, S., Lee, J. Y., Kadesch, T., Hardy, R. R., Aster, J. C. & Pear, W. S. (1999). Notch1 Expression in Early Lymphopoiesis Influences B versus T Lineage Determination. *Immunity* **11**, 299-308.
86. Radtke F, W. A., Stark G, Bauer M, van Meerwijk J, MacDonald HR, Aguet M. (1999). Deficient T Cell Fate Specification in Mice with an Induced Inactivation of Notch1. *Immunity* **10**, 547-558.
87. Rothenberg, E. V., Moore, J. E. & Yui, M. A. (2008). Launching the T-cell-lineage developmental programme. *Nature reviews. Immunology* **8**, 9-21.
88. Pai SY, T. M., Ting CN, Leiden JM, Glimcher LH, Ho IC. (2003). Critical roles for transcription factor GATA-3 in thymocyte development. *Immunity* **19**, 863-875.
89. Growney, J. D., Shigematsu, H., Li, Z., Lee, B. H., Adelsperger, J., Rowan, R., Curley, D. P., Kutok, J. L., Akashi, K., Williams, I. R., Speck, N. A. & Gilliland, D. G. (2005). Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. *Blood* **106**, 494-504.
90. Taghon, T., Yui, M. A. & Rothenberg, E. V. (2007). Mast cell lineage diversion of T lineage precursors by the essential T cell transcription factor GATA-3. *Nature immunology* **8**, 845-55.
91. Garcia KC, D. M., Stanfield RL, Brunmark A, Jackson MR, Peterson PA, Teyton L, Wilson IA. (1996). An αβ T cell receptor structure at 2.5 A and its orienta- tion in the TCR-MHC complex. *Science* **274**, 209-219.
92. Garcia KC, T. L., Wilson IA. (1999). Structural basis of T cell recognition. *Annual review of immunology* **17**, 369-397.
93. Poupot M, F. J. (2004). Non-peptide antigens activating human Vgamma9/Vdelta2 T lymphocytes. *Immunology Letters* **95**, 129-138.
94. Bonneville M, F. J. (2005). Sensing cell stress and transformation through Vgamma9V- delta2 T cell-mediated recognition of the isoprenoid pathway metabolites. *Microbes and Infection* **7**, 503-509.
95. MS, K. (2009). Mechanics of T cell receptor gene rearrangement. *Current Opinion in Immunology* **21**, 133-139.
96. Kreslavsky, T., Garbe, A.I., Krueger, A. & von Boehmer, H. (2008). T cell receptor-instructed αβ versus γδ lineage commitment revealed by single-cell analysis. *Journal of Experimental Medicine* **205**, 1173-1186.
97. Ciofani M, K. G., Wiest DL, von Boehmer H, Zuniga-Pflucker JC. (2006). Stage-specific and differential notch dependency at the αβ and γδ T lineage bifurcation. *Immunity* **25**, 105-116.
98. Melichar, H. J. e. a. (2007). Regulation of αβ versus γδ T lymphocyte differentiation by the transcription factor SOX13. *Science* **315**, 230-233.
99. Bassing CH, S. W., Alt FW. (2002). The mechanism and regulation of chromosomal V(D)J recombination. *Cell* **109**, s45-55.
100. Van Gent DC, R. D., Gellert M (1996). The RAG1 and RAG2 Proteins Establish the 12:23 Rule in V(D)J Recombination. *Cell* **85**, 107-113.

101. Gauss, G. H. & Lieber, M. R. (1992). The basis for the mechanistic bias for deletional over inversional V(D)J recombination. *Genes & Development* **6**, 1553-1561.
102. Pan PY, L. M., Teale JM. (1997). The role of recombination signal sequences in the preferential joining by deletion in DH–JH recombination and in the ordered rearrangement of the IgH locus. *International Immunology* **9**, 515-522.
103. Stephen V. Desiderio, G. D. Y., Michael Paskind, Elise Thomas, Michael A. Boss, Nathaniel Landau, Frederick W. Alt & David Baltimore. (1984). Insertion of N regions into heavy chain genes is correlated with expression of terminal deoxytransferase in B cells. *Nature* **311**, 752-755.
104. Susan Gilfillan, A. D., Marianne Lemeur, Christophe Benoist and Diane Mathis. (1993). Mice lacking Tdt: mature animals with an immature lymphocyte repertoire. *Science* **261**, 1175-1178.
105. Toshihisa Komori, A. O., Valerie Stewart and Frederick W. Alt. (1993). Lack of N regions in antigen receptor variable region genes of Tdt-deficient lymphocytes. *Science* **261**, 1171-1175.
106. Borg, N. A., Ely, L. K., Beddoe, T., Macdonald, W. A., Reid, H. H., Clements, C. S., Purcell, A. W., Kjer-Nielsen, L., Miles, J. J., Burrows, S. R., McCluskey, J. & Rossjohn, J. (2005). The CDR3 regions of an immunodominant T cell receptor dictate the 'energetic landscape' of peptide-MHC recognition. *Nature immunology* **6**, 171-80.
107. Scott, D. R., Borbulevych, O. Y., Piepenbrink, K. H., Corcelli, S. A. & Baker, B. M. (2011). Disparate degrees of hypervariable loop flexibility control T-cell receptor cross-reactivity, specificity, and binding mechanism. *Journal of molecular biology* **414**, 385-400.
108. Maynard, J., Petersson, K., Wilson, D. H., Adams, E. J., Blondelle, S. E., Boulanger, M. J., Wilson, D. B. & Garcia, K. C. (2005). Structure of an autoimmune T cell receptor complexed with class II peptide-MHC: insights into MHC bias and antigen specificity. *Immunity* **22**, 81-92.
109. al., Y. S. e. (2006). Mechanistic basis of pre-T cell receptor-mediated autonomous signaling critical for thymocyte development. *Nature Immunology* **7**, 67-75.
110. Stanhope-Baker P, H. K., Shaffer AL, Constantinescu A, Schlissel MS. (1996). Cell type-specific chromatin structure determines the targeting of V(D)J recombinase activity in vitro. *Cell* **85**, 887-897.
111. Janas, M. L., Varano, G., Gudmundsson, K., Noda, M., Nagasawa, T. & Turner, M. (2010). Thymic development beyond beta-selection requires phosphatidylinositol 3-kinase activation by CXCR4. *The Journal of experimental medicine* **207**, 247-61.
112. Yamasaki, S., Ishikawa, E., Sakuma, M., Ogata, K., Sakata-Sogawa, K., Hiroshima, M., Wiest, D. L., Tokunaga, M. & Saito, T. (2006). Mechanistic basis of pre-T cell receptor-mediated autonomous signaling critical for thymocyte development. *Nature immunology* **7**, 67-75.
113. Schatz, D. G. & Swanson, P. C. (2011). V(D)J recombination: mechanisms of initiation. *Annual review of genetics* **45**, 167-202.
114. Starr, T. K., Jameson, S. C. & Hogquist, K. A. (2003). Positive and negative selection of T cells. *Annual review of immunology* **21**, 139-76.
115. Sant'Angelo, D. B. & Janeway, C. A., Jr. (2002). Negative selection of thymocytes expressing the D10 TCR. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 6931-6.
116. Mathis, D. & Benoist, C. (2009). Aire. *Annual review of immunology* **27**, 287-312.
117. Murata, S., Sasaki, K., Kishimoto, T., Niwa, S., Hayashi, H., Takahama, Y. & Tanaka, K. (2007). Regulation of CD8+ T cell development by thymus-specific proteasomes. *Science* **316**, 1349-53.

118. Bautista, J. L., Lio, C. W., Lathrop, S. K., Forbush, K., Liang, Y., Luo, J., Rudensky, A. Y. & Hsieh, C. S. (2009). Intraclonal competition limits the fate determination of regulatory T cells in the thymus. *Nature immunology* **10**, 610-7.
119. Josefowicz, S. Z. & Rudensky, A. (2009). Control of regulatory T cell lineage commitment and maintenance. *Immunity* **30**, 616-25.
120. Liu, X. & Bosselut, R. (2004). Duration of TCR signaling controls CD4-CD8+ lineage differentiation in vivo. *Nature immunology* **5**, 280-8.
121. Brugnera E, B. A., Cibotti R, Yu Q, Guinter TI, Yamashita Y, Sharrow SO, Singer A. (2000). Coreceptor Reversal in the Thymus- Signaled CD4+8+ Thymocytes Initially Terminate CD8+ Transcription Even When Differentiating into CD8+ T Cells. *Immunity* **13**, 59-71.
122. Germain, R. N. (2002). T-cell development and the CD4-CD8+ lineage decision. *Nature reviews. Immunology* **2**, 309-22.
123. Rudolph, M. G., Stanfield, R. L. & Wilson, I. A. (2006). How TCRs bind MHCs, peptides, and coreceptors. *Annual review of immunology* **24**, 419-66.
124. Jia-huai Wang, E. L. R. (2012). The structural basis of ab T-lineage immune recognition- TCR docking topologies, mechanotransduction, and co-receptor function. *Immunological Reviews* **250**, 102-119.
125. Yin, L., Huseby, E., Scott-Browne, J., Rubtsova, K., Pinilla, C., Crawford, F., Marrack, P., Dai, S. & Kappler, J. W. (2011). A single T cell receptor bound to major histocompatibility complex class I and class II glycoproteins reveals switchable TCR conformers. *Immunity* **35**, 23-33.
126. Matechak, E. O., Killeen, N., Hedrick, S.M. & Fowlkes, B.J. (1996). MHC class II-specific T cells can develop in the CD8+ lineage when CD4+ is absent. *Immunity* **4**, 337-347.
127. Jennifer Buslepp, H. W., William E. Biddison, Ettore Appella, and Edward J. Collins. (2003). A Correlation between TCR Vα Docking on MHC and CD8+ Dependence: Implications for T Cell Selection. *Immunity* **19**, 595-606.
128. Obata-Onai A, H. S., Onai N, Kurachi M, Nagai S, Shizuno K, Nagahata T, Matsushima K. (2002). Comprehensive gene expression analysis of human NK cells and CD8(+) T lymphocytes. *International Immunology* **14**, 1085-1098.
129. Abdalla AO, K. S., Hansson L, Rossmann ED, Jeddi-Tehrani M, Shokri F, Osterborg A, Mellstedt H, Rabbani H. (2003). Kinetics of cytokine gene expression in human CD4+ and CD8+ T-lymphocyte subsets using quantitative real-time PCR. *Scandinavian Journal of Immunology* **58**, 601-606.
130. Sandberg JK, F. N., Nixon DF. (2001). Functional Heterogeneity of Cytokines and C+ytolytic Effector Molecules in Human CD8+ T Lymphocytes. *Journal of Immunology* **167**, 181-187.
131. Zimmerli, S. C., Harari, A., Cellerai, C., Vallelian, F., Bart, P. A. & Pantaleo, G. (2005). HIV-1-specific IFN-gamma/IL-2-secreting CD8+ T cells support CD4-independent proliferation of HIV-1-specific CD8+ T cells. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 7239-44.
132. Techakriengkrai, N., Tansiri, Y. & Hansasuta, P. (2013). Poor HIV control in HLA-B*27 and B*57/58 noncontrollers is associated with limited number of polyfunctional Gag p24-specific CD8+ T cells. *AIDS* **27**, 17-27.
133. Zehn D, K. C., Bevan MJ, Palmer E. (2012). TCR signaling requirements for activating T cells and for generating memory. *Cellular and molecular life sciences : CMLS* **69**, 1565-1575.
134. McGill, J., Van Rooijen, N. & Legge, K. L. (2008). Protective influenza-specific CD8+ T cell responses require interactions with dendritic cells in the lungs. *The Journal of experimental medicine* **205**, 1635-46.

135. Williams, M. A., Tyznik, A. J. & Bevan, M. J. (2006). Interleukin-2 signals during priming are required for secondary expansion of CD8+ memory T cells. *Nature* **441**, 890-3.
136. Mescher MF, C. J., Agarwal P, Casey KA, Gerner M, Hammerbeck CD, Popescu F, Xiao Z. (2006). Signals required for programming effector and memory development by CD8+ T cells. *Immunological Reviews* **211**, 81-92.
137. Wiesel, M. & Oxenius, A. (2012). From crucial to negligible: functional CD8(+) T-cell responses and their dependence on CD4(+) T-cell help. *European journal of immunology* **42**, 1080-8.
138. Bai A, H. H., Yeung M, Chen J. (2007). Kruppel-like factor 2 controls T cell trafficking by activating L-selectin (CD62L) and sphingosine-1-phosphate receptor 1 transcription. *Journal of Immunology* **178**, 7632-7639.
139. Matloubian M, L. C., Cinamon G, Lesneski MJ, Xu Y, Brinkmann V, Allende ML, Proia RL, Cyster JG. (2004). Lymphocyte egress from thymus and peripheral lymphoid organs is dependent on S1P receptor 1. *Nature* **427**, 355-360.
140. Lichtenheld MG, O. K., Lu P, Lowrey DM, Hameed A, Hengartner H, Podack ER. (1988). Structure and function of human perforin. *Nature* **335**, 448-451.
141. Voskoboinik I, D. M., Baran K, Whisstock JC, Trapani JA. (2010). Perforin- structure, function, and role in human immunopathology. *Immunological Reviews* **235**, 35-54.
142. Cullen, S. P. & Martin, S. J. (2008). Mechanisms of granule-dependent killing. *Cell death and differentiation* **15**, 251-62.
143. Esser MT, D. R., Krishnamurthy B, Gullo CA, Graham MB, Braciale VL. (1997). IL-2 induces Fas ligand:Fas (CD95L:CD95) cytotoxicity in CD8+ and CD4+ +T lymphocyte clones. *journal of Immunology* **158**, 5612-5618.
144. Jellison ER, K. S., Welsh RM. (2005). Cutting Edge- MHC Class II-Restricted Killing In Vivo during Viral Infection. *Journal of Immunology* **174**, 614-618.
145. Norris, P. J., Moffett, H. F., Yang, O. O., Kaufmann, D. E., Clark, M. J., Addo, M. M. & Rosenberg, E. S. (2004). Beyond help: direct effector functions of human immunodeficiency virus type 1-specific CD4(+) T cells. *Journal of Virology* **78**, 8844-51.
146. Zaunders JJ, D. W., Wang B, Munier ML, Miranda-Saksena M, Newton R, Moore J, Mackay CR, Cooper DA, Saksena NK, Kelleher AD. (2004). Identification of circulating antigen-specific CD4+ T lymphocytes with a CCR5+, cytotoxic phenotype in an HIV-1 long-term nonprogressor and in CMV infection. *Blood* **103**, 2238-2247.
147. Martorelli, D., Muraro, E., Merlo, A., Turrini, R., Rosato, A. & Dolcetti, R. (2010). Role of CD4+ cytotoxic T lymphocytes in the control of viral diseases and cancer. *International reviews of immunology* **29**, 371-402.
148. Zhu, J., Yamane, H. & Paul, W. E. (2010). Differentiation of effector CD4+ T cell populations. *Annual review of immunology* **28**, 445-89.
149. Dardalhon, V., Awasthi, A., Kwon, H., Galileos, G., Gao, W., Sobel, R. A., Mitsdoerffer, M., Strom, T. B., Elyaman, W., Ho, I. C., Khoury, S., Oukka, M. & Kuchroo, V. K. (2008). IL-4 inhibits TGF-beta-induced Foxp3+ T cells and, together with TGF-beta, generates IL-9+ IL-10+ Foxp3(-) effector T cells. *Nature immunology* **9**, 1347-55.
150. Veldhoen, M., Uyttenhove, C., van Snick, J., Helmby, H., Westendorf, A., Buer, J., Martin, B., Wilhelm, C. & Stockinger, B. (2008). Transforming growth factor-beta 'reprograms' the differentiation of T helper 2 cells and promotes an interleukin 9-producing subset. *Nature immunology* **9**, 1341-6.
151. Groux H, O. G. A., Bigler M, Rouleau M, Antonenko S, de Vries JE, Roncarolo MG. (1997). A CD4+ T-cell subset inhibits antigen-specific T-cell responses and prevents colitis. *Nature* **389**, 737-742.

152. Chen Y, K. V., Inobe J, Hafler DA, Weiner HL. (1994). Regulatory T Cell Clones Induced by Oral Tolerance- Suppression of Autoimmune Encephalomyelitis. *Science* **265**, 1237-1240.
153. Zhu, J., Yamane, H. & Paul, W. E. (2010). Differentiation of effector CD4+ T cell populations (*). *Annual review of immunology* **28**, 445-89.
154. Ansel, K. M., Djuretic, I., Tanasa, B. & Rao, A. (2006). Regulation of Th2 differentiation and Il4 locus accessibility. *Annual review of immunology* **24**, 607-56.
155. Mullen, A. C., Hutchins, A. S., High, F. A., Lee, H. W., Sykes, K. J., Chodosh, L. A. & Reiner, S. L. (2002). Hlx is induced by and genetically interacts with T-bet to promote heritable T(H)1 gene induction. *Nature immunology* **3**, 652-8.
156. Szabo SJ, K. S., Costa GL, Zhang X, Fathman CG, Glimcher LH. (2000). A novel transcription factor, T-bet, directs Th1 lineage commitment. *Cell* **100**, 655-669.
157. Seder RA, G. R., Sher A, Paul WE. (1993). Interleukin 12 acts directly on CD4+ T cells to enhance priming for interferon gamma production and diminishes interleukin 4 inhibition of such priming. *Proceedings of the National Academy of Sciences* **90**, 10188-10192.
158. Hsieh CS, M. S., Tripp CS, Wolf SF, O'Garra A, Murphy KM. (1993). Development of TH1 CD4+ T Cells Through IL-12 Produced by Listeria Induced Macrophages. *Science* **260**, 547-549.
159. Wan, Y. Y. (2010). Multi-tasking of helper T cells. *Immunology* **130**, 166-71.
160. Schroder, K., Hertzog, P. J., Ravasi, T. & Hume, D. A. (2004). Interferon-gamma: an overview of signals, mechanisms and functions. *Journal of leukocyte biology* **75**, 163-89.
161. Mori, M. & Gotoh, T. (2000). Regulation of nitric oxide production by arginine metabolic enzymes. *Biochemical and biophysical research communications* **275**, 715-9.
162. Loke P, N. M., Parkinson J, Guiliano D, Blaxter M, Allen JE. (2002). IL-4 dependent alternatively-activated macrophages have a distinctive in vivo gene expression phenotype. *BMC Immunology* **3**, 7.
163. D F Fiorentino, M. W. B., T R Mosmann. (1989). Two types of mouse T helper cell. IV. Th2 clones secrete a factor that inhibits cytokine production by Th1 clones. *Journal of Experimental Medicine* **170**, 2081-2095.
164. Zhu, Z., Zheng, T., Homer, R. J., Kim, Y. K., Chen, N. Y., Cohn, L., Hamid, Q. & Elias, J. A. (2004). Acidic mammalian chitinase in asthmatic Th2 inflammation and IL-13 pathway activation. *Science* **304**, 1678-82.
165. Gounni AS, G. B., Nutku E, Aris F, Latifa K, Minshall E, North J, Tavernier J, Levit R, Nicolaides N, Robinson D, Hamid Q. (2000). Interleukin-9 enhances interleukin-5 receptor expression, differentiation, and survival of human eosinophils. *Blood* **96**, 2163-2171.
166. Ye P, R. F., Kanaly S, Stocking KL, Schurr J, Schwarzenberger P, Oliver P, Huang W, Zhang P, Zhang J, Shellito JE, Bagby GJ, Nelson S, Charrier K, Peschon JJ, Kolls JK. (2001). Requirement of Interleukin 17 Receptor Signaling for Lung CXC Chemokine and Granulocyte Colony-stimulating Factor Expression, Neutrophil Recruitment, and Host Defense. *Journal of Experimental Medicine* **194**, 519-527.
167. Kolls, J. K. & Linden, A. (2004). Interleukin-17 family members and inflammation. *Immunity* **21**, 467-76.
168. Kovacsovics-Bankowski M, R. K. (1995). A phagosome-to-cytosol pathway for exogenous antigens presented on MHC class I molecules. *Science* **267**, 243-246.
169. Joffre, O. P., Segura, E., Savina, A. & Amigorena, S. (2012). Cross-presentation by dendritic cells. *Nature reviews. Immunology* **12**, 557-69.

170. Jonathan W. Yewdell, J. R. B. a. Y. H. (1988). Cells process exogenous pro-teins for recognition by cytotoxic T lymphocytes. *Science* **239**, 637-640.
171. Moore MW, C. F., Bevan MJ. (1986). Introduction of soluble protein into the class I pathway of antigen processing and presentation. *Cell* **54**, 777-785.
172. Pinet V, M. M., Long EO. (1994). Two processing pathways for the MHC class II-restricted presentation of exogenous influenza virus antigen. *Journal of Immunology* **152**, 4852-4860.
173. Reith, W., LeibundGut-Landmann, S. & Waldburger, J. M. (2005). Regulation of MHC class II gene expression by the class II transactivator. *Nature reviews. Immunology* **5**, 793-806.
174. Watts, C. (1997). Capture and processing of exogenous antigens for presentation on MHC molecules. *Annual review of immunology* **15**, 821-850.
175. Manoury B, H. E., Morrice N, Dando PM, Barrett AJ, Watts C. (1998). An asparaginyl endopeptidase processes a microbial antigen for class II MHC presentation. *Nature* **396**, 695-699.
176. Nakagawa TY, R. A. (1999). The role of lysosomal proteinases in MHC class Il-mediated antigen processing and presentation. *Immunological Reviews* **172**, 121-129.
177. Harold A. Chapman, R. J. R., and Guo-Ping Shi. (1997). Cathepsins and compartmentalization in antigen presentation. *Annual Review of Physiology* **59**.
178. Jones PP, M. D., Hewgill D, McDevitt HO. (1979). Detection of a common polypeptide chain in I--A and I--E sub-region immunoprecipitates. *Molecular immunology* **16**, 51-60.
179. Günter J. Hämmerling, J. M. (1990). The function of the invariant chain in antigen presentation by MHC class II molecules. *Immunology Today* **11**, 337-340.
180. R.J. Riese, P. R. W., D. Bromme, L.R. Natkin, J.A. Villadangos, H.A. Ploegh, H.A. Chapman. (1996). Essential Role for Cathepsin S in MHC Class II–Associated Invariant Chain Processing and Peptide Loading. *Immunity* **4**, 357-366.
181. van Ham SM, T. E., Lillemeier BF, Grüneberg U, van Meijgaarden KE, Pastoors L, Verwoerd D, Tulp A, Canas B, Rahman D, Ottenhoff TH, Pappin DJ, Trowsdale J, Neefjes J. (1997). HLA-DO is a negative modulator of HLA-DM-mediated MHC class II peptide loading. *Current Biology* **7**, 950-957.
182. Denzin, L. K. (1997). Negative Regulation by HLA-DO of MHC Class II-Restricted Antigen Processing. *Science* **278**, 106-109.
183. Morris P, S. J., Attaya M, Amaya M, Goodman S, Bergman C, Monaco JJ, Mellins E. (1994). An essential role for HLA-DM in antigen presentation by class II major histocompatibility molecules. *Nature* **368**, 551-554.
184. Fling SP, A. B., Pious D. (1994). HLA-DMA and -DMB genes are both required for MHC class II:peptide complex formation in antigen- presenting cells. *Nature* **368**, 554-558.
185. Denzin LK, C. P. (1995). HLA-DM induces CLIP dissociation from MHC class II alpha beta dimers and facilitates peptide loading. *Cell* **82**, 155-165.
186. Michalek MT, G. E., Rock KL. (1996). Chemical denaturation and modification of ovalbumin alters its dependence on ubiquitin conjugation for class I antigen presentation. *Journal of Immunology* **157**, 617-624.
187. Michalek MT, G. E., Gramm C, Goldberg AL, Rock KL. (1993). A role for the ubiquitin-dependent proteolytic pathway in MHC class I-restricted antigen presentation. *Nature* **363**, 552-554.
188. Schubert U, A. L., Gibbs J, Norbury CC, Yewdell JW, Bennink JR. (2000). Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature* **404**, 770-774.

189. Jan Lowe, D. S., Bing Jap, Peter Zwickl, Wolfgang Baumeister, Robert Hubert. (1995). Crystal Structure of the 20S Proteasome from the Archaeon T. acidophilum at 3.4 Å Resolution. *Science* **268**, 533-539.
190. Dubiel W, F. K., Rechsteiner M. (1995). Subunits of the regulatory complex of the 26S protease. *Molecular Biology Reports* **21**, 27-34.
191. Glickman MH, R. D., Fu H, Larsen CN, Coux O, Wefes I, Pfeifer G, Cjeka Z, Vierstra R, Baumeister W, Fried V, Finley D. (1999). Functional analysis of the proteasome regulatory particle. *Molecular Biology Reports* **26**, 21-28.
192. Glickman MH, R. D., Coux O, Wefes I, Pfeifer G, Cjeka Z, Baumeister W, Fried VA, Finley D. (1998). A Subcomplex of the Proteasome Regulatory Particle Required for Ubiquitin-Conjugate Degradation and Related to the COP9-Signalosome and eIF3. *Cell* **94**, 615-623.
193. Kisselev, A. F., Kaganovich, D. & Goldberg, A. L. (2002). Binding of hydrophobic peptides to several non-catalytic sites promotes peptide hydrolysis by all active sites of 20 S proteasomes. Evidence for peptide-induced channel opening in the alpha-rings. *The Journal of biological chemistry* **277**, 22260-70.
194. Beck F, U. P., Bohn S, Schweitzer A, Pfeifer G, Sakata E, Nickell S, Plitzko JM, Villa E, Baumeister W, Förster F. (2012). Near-atomic resolution structural model of the yeast 26S proteasome. *Proceedings of the National Academy of Sciences* **109**, 14870-14875.
195. Unno M, M. T., Morimoto Y, Tomisugi Y, Tanaka K, Yasuoka N, Tsukihara T. (2002). The Structure of the Mammalian 20S Proteasome at 2.75 Å Resolution. *Structure* **10**, 609-618.
196. Kisselev, A. F., Garcia-Calvo, M., Overkleeft, H. S., Peterson, E., Pennington, M. W., Ploegh, H. L., Thornberry, N. A. & Goldberg, A. L. (2003). The caspase-like sites of proteasomes, their substrate specificity, new inhibitors and substrates, and allosteric interactions with the trypsin-like sites. *The Journal of biological chemistry* **278**, 35869-77.
197. Eleuteri AM, K. R., Cardozo C, Orlowski M. (1997). Bovine spleen multicatalytic proteinase complex (proteasome)- replacement of X, Y, and Z subunits by LMP7, LMP2 and MECL1 and changes in properties and specificity. *Journal of Biological Chemistry* **272**, 11824-11831.
198. Orlowski M, C. C., Michaud C. (1993). Evidence for the presence of five distinct proteolytic components in the pituitary multicatalytic proteinase complex. Properties of two components cleaving bonds on the carboxyl side of branched chain and small neutral amino acids. *Biochemistry* **32**, 1563-1572.
199. Alexei F. Kisselev, T. N. A., Kee Min Woo and Alfred L. Goldberg. (1999). The sizes of peptides generated from protein by mammalian 26 and 20S proteasomes. Implications for understanding the degradative mechanism and antigen presentation. *Journal of Biological Chemistry* **274**, 3363-3371.
200. Toes RE, N. A., Degermann S, Schirle M, Emmerich NP, Kraft M, Laplace C, Zwinderman A, Dick TP, Müller J, Schönfisch B, Schmid C, Fehling HJ, Stevanovic S, Rammensee HG, Schild H. (2001). Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. *Journal of Experimental Medicine* **194**, 1-12.
201. Gaczynska M, R. K., Goldberg AL. (1993). Gamma interferon and expression of MHC genes regulate peptide hydrolysis. *Nature* **365**, 264-267.
202. Früh K, G. M., Wang K, Bujard H, Peterson PA, Yang Y. (1994). Displacement of housekeeping proteasome sub-units by MHC-encoded LMPs- a newly discovered

mechanism for modulating the multicatalytic proteinase complex. *The EMBO Journal* **13**, 3236-3244.
203. Gaczynska M, R. K., Spies T, Goldberg AL. (1994). Peptidase activities of proteasomes are differentially regulated by the major histocompatibility complex-encoded genes for LMP2 and LMP7. *Proceedings of the National Academy of Sciences* **91**, 9213-9217.
204. Aki M, S. N., Takashina M, Akiyama K, Kagawa S, Tamura T, Tanahashi N, Yoshimura T, Tanaka K, Ichihara A. (1994). Interferon-gamma induces different subunit organizations and functional diversity of proteasomes. *Journal of Biochemistry* **115**, 257-269.
205. Driscoll J, B. M., Finley D, Monaco JJ. (1993). MHC-linked LMP gene products specifically alter peptidase activities of the proteasome. *Nature* **365**, 262-264.
206. Van den Eynde BJ, M. S. (2001). Differential processing of class-I-restricted epitopes by the standard proteasome and the immunoproteasome. *Current Opinion in Immunology* **13**, 147-153.
207. Rammensee, H.-G. F., Thomas; Stevanovic,Stefan (1995). MHC ligands and peptide motifs- first listing. *Immunogenetics* **41**, 178-228.
208. Teoh, C. Y. & Davies, K. J. (2004). Potential roles of protein oxidation and the immunoproteasome in MHC class I antigen presentation: the 'PrOxI' hypothesis. *Archives of biochemistry and biophysics* **423**, 88-96.
209. Rock KL, G. C., Rothstein L, Clark K, Stein R, Dick L, Hwang D, Goldberg AL. (1994). Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. *Cell* **78**, 761-771.
210. Craiu A, A. T., Goldberg A, Rock KL. (1997). Two distinct proteolytic processes in the generation of a major histocompatibility complex class I-presented peptide. *Proceedings of the National Academy of Sciences* **94**, 10850-10855.
211. Geier E, P. G., Wilm M, Lucchiari-Hartz M, Baumeister W, Eichmann K, Niedermann G. (1999). A giant protease with potential to substitute for some functions of the proteasome. *Science* **283**, 978-981.
212. Tanioka, T., Hattori, A., Masuda, S., Nomura, Y., Nakayama, H., Mizutani, S. & Tsujimoto, M. (2003). Human leukocyte-derived arginine aminopeptidase. The third member of the oxytocinase subfamily of aminopeptidases. *The Journal of biological chemistry* **278**, 32275-83.
213. Seifert, U., Maranon, C., Shmueli, A., Desoutter, J. F., Wesoloski, L., Janek, K., Henklein, P., Diescher, S., Andrieu, M., de la Salle, H., Weinschenk, T., Schild, H., Laderach, D., Galy, A., Haas, G., Kloetzel, P. M., Reiss, Y. & Hosmalin, A. (2003). An essential role for tripeptidyl peptidase in the generation of an MHC class I epitope. *Nature immunology* **4**, 375-9.
214. Saric, T., Beninga, J., Graef, C. I., Akopian, T. N., Rock, K. L. & Goldberg, A. L. (2001). Major histocompatibility complex class I-presented antigenic peptides are degraded in cytosolic extracts primarily by thimet oligopeptidase. *The Journal of biological chemistry* **276**, 36474-81.
215. York IA, M. A., Lemerise K, Zeng W, Shen Y, Abraham CR, Saric T, Goldberg AL, Rock KL. (2003). The cytosolic endopeptidase, thimet oligopeptidase, destroys antigenic peptides and limits the extent of MHC class I antigen presentation. *Immunity* **18**, 429-440.
216. Stoltze L, S. M., Schwarz G, Schröter C, Thompson MW, Hersh LB, Kalbacher H, Stevanovic S, Rammensee HG, Schild H. (2000). Two new proteases in the MHC class I processing pathway. *Nature immunology* **1**, 413-418.

217. Beninga J, R. K., Goldberg AL. (1998). Interferon-γ can stimulate post-proteaso- mal trimming of the N terminus of an antigenic peptide by inducing leucine aminopeptidase. *Journal of Biological Chemistry* **273**, 18734-18742.
218. Saveanu, L., Carroll, O., Lindo, V., Del Val, M., Lopez, D., Lepelletier, Y., Greer, F., Schomburg, L., Fruci, D., Niedermann, G. & van Endert, P. M. (2005). Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. *Nature immunology* **6**, 689-97.
219. Saric, T., Chang, S. C., Hattori, A., York, I. A., Markant, S., Rock, K. L., Tsujimoto, M. & Goldberg, A. L. (2002). An IFN-gamma-induced aminopeptidase in the ER, ERAP1, trims precursors to MHC class I-presented peptides. *Nature immunology* **3**, 1169-76.
220. Serwold T, G. F., Kim J, Jacob R, Shastri N. (2002). ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. *Nature* **419**, 480-483.
221. Kleijmeer MJ, K. A., Geuze HJ, Slot JW, Townsend A, Trowsdale J. (1992). Location of MHC-encoded transporters in the endoplasmic reticulum and cis-Golgi. *Nature* **357**, 342-344.
222. Androlewicz MJ, A. K., Cresswell P. (1993). Evidence that transporters associated with antigen processing translocate a major histocompatibility complex class I-binding peptide into the endoplasmic reticulum in an ATP-dependent manner. *Proceedings of the National Academy of Sciences* **90**, 9130-9134.
223. Procko, E., Ferrin-O'Connell, I., Ng, S. L. & Gaudet, R. (2006). Distinct structural and functional properties of the ATPase sites in an asymmetric ABC transporter. *Molecular Cell* **24**, 51-62.
224. Smith PC, K. N., Millen L, Moody JE, Rosen J, Thomas PJ, Hunt JF. (2002). ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Molecular Cell* **10**, 139-149.
225. Gubler B, D. S., Armandola EA, Hammer J, Caillat-Zucman S, van Endert PM. (1998). Substrate selection by transporters associated with antigen processing occurs during peptide binding to TAP. *Molecular immunology* **35**, 427-433.
226. Procko E, G. R. (2009). Antigen processing and presentation- TAPping into ABC transporters. *Current opinion in immunology* **21**, 84-91.
227. Gaudet R, W. D. (2001). Structure of the ABC ATPase domain of human TAP1, the transporter associated with antigen processing. *The EMBO Journal* **20**, 4964-4972.
228. Deverson EV, L. L., Seelig A, Coadwell WJ, Tredgett EM, Butcher GW, Howard JC. (1998). Functional analysis by site-directed mutagenesis of the complex polymorphism in rat transporter associated with antigen processing. *Journal of Immunology* **160**, 2767-2779.
229. Dawson, R. J. & Locher, K. P. (2006). Structure of a bacterial multidrug ABC transporter. *Nature* **443**, 180-5.
230. Grégoire Lauvau, K. K., Gabriele Niedermann, Marina Ostankovitch, PatriciaYotnda, Hüseyin Firat, Francis V. Chisari, and Peter M. van Endert. (1999). Human Transporters Associated with Antigen Processing (Taps) Select Epitope Precursor Peptides for Processing in the Endoplasmic Reticulum and Presentation to T Cells. *Journal of Experimental Medicine* **190**, 1227-1239.
231. Schmitt L, T. R. (2000). Affinity, specificity, diversity- a challenge for the ABC transporter TAP in cellular immunity. *Chembiochem* **1**, 16-35.
232. Uebel S, K. W., Kienle S, Wiesmüller KH, Jung G, Tampé R. (1997). Recognition principle of the TAP transporter disclosed by combinatorial peptide libraries. *Proceedings of the National Academy of Sciences* **94**, 8976-8981.

233. van Endert PM, R. D., Greco G, Fleischhauer K, Sidney J, Sette A, Bach JF. (1995). The peptide-binding motif for the human transporter associated with antigen processing. *Journal of Experimental Medicine* **182**, 1883-1895.
234. Rufer E, L. R., Knittler MR. (2007). Molecular Architecture of the TAP-Associated MHC Class I Peptide-Loading Complex. *Journal of Immunology* **179**, 5717-5727.
235. Wearsch, P. A. & Cresswell, P. (2008). The quality control of MHC class I peptide loading. *Current opinion in cell biology* **20**, 624-31.
236. Bangia N, L. P., Hughes EA, Surman M, Cresswell P. (1999). The N-terminal region of tapasin is required to stabilize the MHC class I loading complex. *European journal of immunology* **29**, 1858-1870.
237. Lehner PJ, S. M., Cresswell P. (1998). Soluble Tapasin Restores MHC Class I Expression and Function in the Tapasin-Negative Cell Line .220. *Immunity* **8**, 221-231.
238. Tenzer, S., Wee, E., Burgevin, A., Stewart-Jones, G., Friis, L., Lamberth, K., Chang, C. H., Harndahl, M., Weimershaus, M., Gerstoft, J., Akkad, N., Klenerman, P., Fugger, L., Jones, E. Y., McMichael, A. J., Buus, S., Schild, H., van Endert, P. & Iversen, A. K. (2009). Antigen processing influences HIV-specific cytotoxic T lymphocyte immunodominance. *Nature immunology* **10**, 636-46.
239. Thorsby, E. (2009). A short history of HLA. *Tissue antigens* **74**, 101-16.
240. Horton, R., Wilming, L., Rand, V., Lovering, R. C., Bruford, E. A., Khodiyar, V. K., Lush, M. J., Povey, S., Talbot, C. C., Jr., Wright, M. W., Wain, H. M., Trowsdale, J., Ziegler, A. & Beck, S. (2004). Gene map of the extended human MHC. *Nature reviews. Genetics* **5**, 889-99.
241. Sullivan, L. C., Hoare, H. L., McCluskey, J., Rossjohn, J. & Brooks, A. G. (2006). A structural perspective on MHC class Ib molecules in adaptive immunity. *Trends in immunology* **27**, 413-20.
242. Vance RE, K. J., Altman JD, Jensen PE, Raulet DH. (1998). Mouse CD94:NKG2A Is a Natural Killer Cell Receptor for the Nonclassical Major Histocompatibility Complex (MHC) Class I Molecule Qa-1. *Journal of Experimental Medicine* **188**, 1841-1848.
243. Rajagopalan, S., Bryceson, Y. T., Kuppusamy, S. P., Geraghty, D. E., van der Meer, A., Joosten, I. & Long, E. O. (2006). Activation of NK cells by an endocytosed receptor for soluble HLA-G. *PLoS biology* **4**, e9.
244. Brown JH, J. T., Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **364**, 33-39.
245. Carrington, M. (1999). HLA and HIV-1: Heterozygote Advantage and B*35-Cw*04 Disadvantage. *Science* **283**, 1748-1752.
246. Gilbert, S. C. (1998). Association of Malaria Parasite Population Structure, HLA, and Immunological Antagonism. *Science* **279**, 1173-1177.
247. Hill AV, E. J., Willis AC, Aidoo M, Allsopp CE, Gotch FM, Gao XM, Takiguchi M, Greenwood BM, Townsend AR, et al. (1992). Molecular analysis of the association of HLA-B53 and resistance to severe malaria. *Nature* **360**, 434-439.
248. Wang, J. H., Zheng, X., Ke, X., Dorak, M. T., Shen, J., Boodram, B., O'Gorman, M., Beaman, K., Cotler, S. J., Hershow, R. & Rong, L. (2009). Ethnic and geographical differences in HLA associations with the outcome of hepatitis C virus infection. *Virology journal* **6**, 46.
249. A.V. Hill, C. E. A., D. Kwiatkowski, N.M. Anstey, P. Twumasi, P.A. Rowe, S. Bennett, D. Brewster, A.J. McMichael, B.M. Greenwood. (1991). Common west African HLA antigens are associated with protection from severe malaria. *Nature* **352**, 595-600.

250. Lambotte O, B. F., Madec Y, Nguyen A, Goujard C, Meyer L, Rouzioux C, Venet A, Delfraissy JF. (2005). HIV controllers- a homogeneous group of HIV-1-infected patients with spontaneous control of viral replication. *Clinical Infectious diseases* **41**, 1053-1056.
251. Bailey, J. R., Williams, T. M., Siliciano, R. F. & Blankson, J. N. (2006). Maintenance of viral suppression in HIV-1-infected HLA-B*57+ elite suppressors despite CTL escape mutations. *The Journal of experimental medicine* **203**, 1357-69.
252. Migueles, S. A., Sabbaghian, M. S., Shupert, W. L., Bettinotti, M. P., Marincola, F. M., Martino, L., Hallahan, C. W., Selig, S. M., Schwartz, D., Sullivan, J. & Connors, M. (2000). HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 2709-14.
253. Tomiyama H, M. K., Shiga H, Moore YI, Oka S, Iwamoto A, Kaneko Y, Takiguchi M. (1997). Evidence of presentation of multiple HIV-1 cytotoxic T lymphocyte epitopes by HLA-B*3501 molecules that are associated with the accelerated progression of AIDS. *Journal of Immunology* **158**, 5026-5034.
254. Goulder, P. J. & Watkins, D. I. (2008). Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nature reviews. Immunology* **8**, 619-30.
255. Guillaume B. E. Stewart-Jones, G. G., Ian M. Overton, Rupert Kaul, Philippe Roche, Andrew J. McMichael, Sarah Rowland-Jones and E. Yvonne Jones. (2005). Structures of Three HIV-1 HLA-B*5703-Peptide Complexes and Identification of Related HLAs Potentially Associated with Long-Term Nonprogression. *The Journal of Immunology* **175**, 2459-2468.
256. Goulder PJ, P. R., Colbert RA, McAdam S, Ogg G, Nowak MA, Giangrande P, Luzzi G, Morgan B, Edwards A, McMichael AJ, Rowland-Jones S. (1997). Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nature Medicine*, 212-217.
257. Martinez-Picado, J., Prado, J. G., Fry, E. E., Pfafferott, K., Leslie, A., Chetty, S., Thobakgale, C., Honeyborne, I., Crawford, H., Matthews, P., Pillay, T., Rousseau, C., Mullins, J. I., Brander, C., Walker, B. D., Stuart, D. I., Kiepiela, P. & Goulder, P. (2006). Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. *Journal of Virology* **80**, 3617-23.
258. H.M. Berman, K. H., H. Nakamura. (2003). Announcing the worldwide Protein Data Bank. *Nature Structural Biology* **10**.
259. Guillame B. E. Stewart-Jones, K. d. G., Simon Kollnberger, Andrew J. McMichael, E. Yvonne Jones and Paul Bowness. (2005). Crystal structures and KIR3DL1 recognition of three immunodominant viral peptides complexed to HLA-B*2705. *European Journal of Immunology* **35**, 341-351.
260. Hulsmeyer, M., Welfle, K., Pohlmann, T., Misselwitz, R., Alexiev, U., Welfle, H., Saenger, W., Uchanska-Ziegler, B. & Ziegler, A. (2005). Thermodynamic and structural equivalence of two HLA-B27 subtypes complexed with a self-peptide. *Journal of molecular biology* **346**, 1367-79.
261. Hulsmeyer, M., Hillig, R. C., Volz, A., Ruhl, M., Schroder, W., Saenger, W., Ziegler, A. & Uchanska-Ziegler, B. (2002). HLA-B27 subtypes differentially associated with disease exhibit subtle structural alterations. *The Journal of biological chemistry* **277**, 47844-53.
262. M. A. Sapert, P. J. B., D. C. Wiley. (1991). Refined Structure of the Human Histocompatibility Antigen HLA-A2 at 2.6 A Resolution. *Journal of Molecular Biology* **219**, 277-319.

263. McLaren, P. J., Ripke, S., Pelak, K., Weintrob, A. C., Patsopoulos, N. A., Jia, X., Erlich, R. L., Lennon, N. J., Kadie, C. M., Heckerman, D., Gupta, N., Haas, D. W., Deeks, S. G., Pereyra, F., Walker, B. D. & de Bakker, P. I. (2012). Fine-mapping classical HLA variation associated with durable host control of HIV-1 infection in African Americans. *Human molecular genetics* **21**, 4334-47.
264. McMichael, A. J. & Yvonne Jones, E. (2010). First-Class Control of HIV-1. *Science* **330**, 1488-1490.
265. David N. Garboczi, P. G., Ursula Utz, Qing R. Fan, William E. Biddison & Don C. Wiley. (1996). Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* **384**, 93-192.
266. Borg, N. A., Wun, K. S., Kjer-Nielsen, L., Wilce, M. C., Pellicci, D. G., Koh, R., Besra, G. S., Bharadwaj, M., Godfrey, D. I., McCluskey, J. & Rossjohn, J. (2007). CD1d-lipid-antigen recognition by the semi-invariant NKT T-cell receptor. *Nature* **448**, 44-9.
267. Hahn, M., Nicholson, M. J., Pyrdol, J. & Wucherpfennig, K. W. (2005). Unconventional topology of self peptide-major histocompatibility complex binding by a human autoimmune T cell receptor. *Nature immunology* **6**, 490-6.
268. Reinherz, E. L. (1999). The Crystal Structure of a T Cell Receptor in Complex with Peptide and MHC Class II. *Science* **286**, 1913-1921.
269. Gakamsky, D. M., Lewitzki, E., Grell, E., Saulquin, X., Malissen, B., Montero-Julian, F., Bonneville, M. & Pecht, I. (2007). Kinetic evidence for a ligand-binding-induced conformational transition in the T cell receptor. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 16639-44.
270. Wu LC, T. D., Lyons DS, Garcia KC, Davis MM. (2002). Two-step binding mechanism for T-cell receptor recognition of peptide–MHC. *Nature* **418**, 552-556.
271. Davis-Harrison RL, I. F., Baker BM. (2007). T Cell Receptor Binding Transition States and Recognition of Peptide/MHC. *Biochemistry* **46**, 1840-1850.
272. M-K. Teng, A. S., A.G.D. Tse, J-H. Liu, J. Liu, R.E. Hussey, S.G. Nathenson, H-C. Chang, E.L. Reinherz and J-H. Wang. (1998). Identification of a common docking topology with substantial variation among different TCR–peptide–MHC complexes. *Current Biology* **8**, 409-412.
273. Markus G Rudolph, I. A. W. (2002). The specificity of TCR:pMHC interaction. *Current Opinion in Immunology* **14**, 52-64.
274. Arstila, T. P. (1999). A Direct Estimate of the Human T Cell Receptor Diversity. *Science* **286**, 958-961.
275. Mason, D. (1998). A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunology Today* **19**, 395-404.
276. Bjorkman, M. M. D. P. J. (1988). T-cell antigen receptor genes and T-cell recognition. *Nature* **334**, 395-402.
277. John J. Miles, N. A. B., Rebekah M. Brennan, Fleur E. Tynan, Lars Kjer-Nielsen, Sharon L. Silins, Melissa J. Bell, Jacqueline M. Burrows, James McCluskey, Jamie Rossjohn and Scott R. Burrows. (2006). TCRα Genes Direct MHC Restriction in the Potent Human T Cell Response to a Class I-Bound Viral Epitope. *The Journal of Immunology* **177**, 6804-6814.
278. Adams, J. J., Narayanan, S., Liu, B., Birnbaum, M. E., Kruse, A. C., Bowerman, N. A., Chen, W., Levin, A. M., Connolly, J. M., Zhu, C., Kranz, D. M. & Garcia, K. C. (2011). T cell receptor signaling is limited by docking geometry to peptide-major histocompatibility complex. *Immunity* **35**, 681-93.
279. Takamasa Ueno, H. T. a. M. T. (2002). Single T Cell Receptor-Mediated Recognition of an Identical HIV-Derived Peptide Presented by Multiple HLA Class I Molecules. *The Journal of Immunology* **169**, 4961-4969.

280. Hiroko Tomiyama, N. Y., Hiroki Komatsu, Kazuo Hirayama and Masafumi Takiguchi. (2000). A single CTL clone can recognize a naturally processed HIV-1 epitope presented by two different HLA class I molecules. *European journal of immunology* **30**, 2521-2530.
281. Liu, Y. C., Chen, Z., Burrows, S. R., Purcell, A. W., McCluskey, J., Rossjohn, J. & Gras, S. (2012). The energetic basis underpinning T-cell receptor recognition of a super-bulged peptide bound to a major histocompatibility complex class I molecule. *The Journal of biological chemistry* **287**, 12267-76.
282. Borbulevych, O. Y., Piepenbrink, K. H., Gloor, B. E., Scott, D. R., Sommese, R. F., Cole, D. K., Sewell, A. K. & Baker, B. M. (2009). T cell receptor cross-reactivity directed by antigen-dependent tuning of peptide-MHC molecular flexibility. *Immunity* **31**, 885-96.
283. Christopher J. Holland, P. J. R., Sabrina Vollers, J. Mauricio Calvo-Calle, Florian Madura, Anna Fuller, Andrew K. Sewell, Lawrence J. Stern, Andrew Godkin and David K. Cole. (2012). Minimal conformational plasticity enables TCR cross-reactivity to different MHC class II heterodimers. *Scientific Reports* **2:629**, 1-8.
284. Ely, L. K., Beddoe, T., Clements, C. S., Matthews, J. M., Purcell, A. W., Kjer-Nielsen, L., McCluskey, J. & Rossjohn, J. (2006). Disparate thermodynamics governing T cell receptor-MHC-I interactions implicate extrinsic factors in guiding MHC restriction. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 6641-6.
285. Rubtsova K, S.-B. J., Crawford F, Dai S, Marrack P, Kappler JW. (2009). Different Vβ CDR3s Can Reveal the Inherent MHC Reactivity of Germline-Encoded TCR V Regions. *Proceedings of the National Academy of Sciences* **106**, 7951-7956.
286. Bee-Cheng Sim, L. Z., Mark I. Greene, Nicholas. R. J. Gascoigne. (1996). Control of MHC restriction of TCR Vα CDR1 and CDR2. *Science* **273**, 963-966.
287. Feng, D., Bond, C. J., Ely, L. K., Maynard, J. & Garcia, K. C. (2007). Structural evidence for a germline-encoded T cell receptor-major histocompatibility complex interaction 'codon'. *Nature immunology* **8**, 975-83.
288. Shaodong Dai, E. S. H., Kira Rubtsova, James Scott-Browne, Frances Crawford, Whitney A. Macdonald, John W. Kappler, Philippa Marrack. (2008). Crossreactive T cells spotlight the germline rules for αβ T cell receptor interactions with MHC molecules. *Immunity* **28**, 324-334.
289. Scott R. Burrows, Z. C., Julia K. Archbold, Fleur E. Tynan, Travis Beddoe, Lars Kjer-Nielsen, John J. Miles, Rajiv Khanna, Denis J. Moss, Yu Chih Liu, Stephanie Gras, Lyudmila Kostenko, Rebekah M. Brennan, Craig S. Clements, Andrew G. Brooks, Anthony W. Purcell, James McCluskey, and Jamie Rossjohn. (2010). Hard wiring of T cell receptor specificity for the major histocompatibility complex is underpinned by TCR adaptability. *Proceedings of the National Academy of Sciences* **107**, 10608-10613.
290. M. J. Wilson, M. T., J. Trowsdale. (1997). Genomic organization of a human killer cell inhibitory receptor gene. *Tissue antigens*, 574-579.
291. A. Selvakumar, U. S., N.Palanisamy, R. S.K. Chaganti, B. Dupont. (1997). Genomic organization and allelic polymorphism of the human killer cell inhibitory receptor gene KIR 103. *Tissue antigens* **49**, 564-573.
292. B.Dupont, A. S., U.Steffens. (1997). The killer cell inhibitory receptor genomic region on human chromosome 19q13.4. *Tissue antigens* **49**, 557-563.
293. E. Baker, A. D. A., J. H. Phillips, G. R. Sutherland and L. L. Lanier. (1995). Natural killer cell receptor for HLA-B allotypes, NKB1- map position 19q13.4. *Chromosome Research* **3**, 511.

294. Roberto Biassoni, C. C., Michela Falco, Simonetta Verdiani, Cristina Bottino, Massimo Vitale, Romana Conte, Alessandro Poggi, Alessandro Moretta and Lorenzo Moretta. (1996). The human leukocyte antigen (HLA)-C-specific "activatory" or "inhibitory" natural killer cell receptors display highly homologous extracellular domains but differ in their transmembrane and intracytoplasmic portions. *Journal of Experimental Medicine* **183**, 645-650.
295. Katsumi Maenaka, T. J., David I Stuart and E Yvonne Jones. (1999). Crystal structure of the human p58 killer cell inhibitory receptor (KIR2DL3) specific for HLA-Cw3-related MHC class I. *Structure* **7**, 391-398.
296. Fan QR, L. E., Wiley DC. (2001). Crystal structure of the human natural killer cell inhibitory receptor KIR2DL1–HLA-Cw4 complex. *Nature immunology* **2**, 452-460.
297. Greg A. Snyder, A. G. B., and Peter D. Sun. (1999). Crystal structure of the HLA-Cw3 allotype-specific killer cell inhibitory receptor KIR2DL2. *Proceedings of the National Academy of Sciences* **96**, 3864-3869.
298. Jeffrey C. Boyington, S. A. M., Peter Schuck, Andrew G. Brooks & Peter D. Sun. (2000). Crystal structure of an NK cell immunoglobulin-like receptor in complex with its class I MHC ligand. *Nature* **405**, 537-543.
299. Graef, T., Moesta, A. K., Norman, P. J., Abi-Rached, L., Vago, L., Older Aguilar, A. M., Gleimer, M., Hammond, J. A., Guethlein, L. A., Bushnell, D. A., Robinson, P. J. & Parham, P. (2009). KIR2DS4 is a product of gene conversion with KIR3DL2 that introduced specificity for HLA-A*11 while diminishing avidity for HLA-C. *The Journal of experimental medicine* **206**, 2557-72.
300. Vivian, J. P., Duncan, R. C., Berry, R., O'Connor, G. M., Reid, H. H., Beddoe, T., Gras, S., Saunders, P. M., Olshina, M. A., Widjaja, J. M. L., Harpur, C. M., Lin, J., Maloveste, S. M., Price, D. A., Lafont, B. A. P., McVicar, D. W., Clements, C. S., Brooks, A. G. & Rossjohn, J. (2011). Killer cell immunoglobulin-like receptor 3DL1-mediated recognition of human leukocyte antigen B. *Nature* **479**, 401-405.
301. Saulquin, X., Gastinel, L. N. & Vivier, E. (2003). Crystal structure of the human natural killer cell activating receptor KIR2DS2 (CD158j). *The Journal of experimental medicine* **197**, 933-8.
302. Fu, L., Hazes, B. & Burshtyn, D. N. (2011). The first Ig domain of KIR3DL1 contacts MHC class I at a secondary site. *Journal of immunology* **187**, 1816-25.
303. Douek, D. C., Colantonio, A. D., Bimber, B. N., Neidermyer, W. J., Reeves, R. K., Alter, G., Altfeld, M., Johnson, R. P., Carrington, M., O'Connor, D. H. & Evans, D. T. (2011). KIR Polymorphisms Modulate Peptide-Dependent Binding to an MHC Class I Ligand with a Bw6 Motif. *PLoS pathogens* **7**, e1001316.
304. Campbell, K. S. & Purdy, A. K. (2011). Structure/function of human killer cell immunoglobulin-like receptors: lessons from polymorphisms, evolution, crystal structures and mutations. *Immunology* **132**, 315-25.
305. Jeffrey C. Boyington, A. G. B., Peter D. Sun. (2001). Structure of killer cell immunoglobulin-like receptors and their recognition of the class I MHC molecules. *Immunological Reviews* **181**, 66-78.
306. Jeffrey C. Boyington, P. D. S. (2001). A structural perspective on MHC class I recognition by killer cell immunoglobulin-like receptors. *Molecular immunology* **38**, 1007-1021.
307. Yu, B., Fonseca, D. P. A. J., O'Rourke, S. M. & Berman, P. W. (2009). Protease Cleavage Sites in HIV-1 gp120 Recognized by Antigen Processing Enzymes Are Conserved and Located at Receptor Binding Sites. *Journal of Virology* **84**, 1513-1526.
308. Chen, H., Ndhlovu, Z. M., Liu, D., Porter, L. C., Fang, J. W., Darko, S., Brockman, M. A., Miura, T., Brumme, Z. L., Schneidewind, A., Piechocka-Trocha, A., Cesa, K. T., Sela, J.,

Cung, T. D., Toth, I., Pereyra, F., Yu, X. G., Douek, D. C., Kaufmann, D. E., Allen, T. M. & Walker, B. D. (2012). TCR clonotypes modulate the protective effect of HLA class I molecules in HIV-1 infection. *Nature immunology* **13**, 691-700.

309. Almeida, J. R., Price, D. A., Papagno, L., Arkoub, Z. A., Sauce, D., Bornstein, E., Asher, T. E., Samri, A., Schnuriger, A., Theodorou, I., Costagliola, D., Rouzioux, C., Agut, H., Marcelin, A. G., Douek, D., Autran, B. & Appay, V. (2007). Superior control of HIV-1 replication by CD8+ T cells is reflected by their avidity, polyfunctionality, and clonal turnover. *The Journal of experimental medicine* **204**, 2473-85.
310. Turnbull, E. L., Wong, M., Wang, S., Wei, X., Jones, N. A., Conrod, K. E., Aldam, D., Turner, J., Pellegrino, P., Keele, B. F., Williams, I., Shaw, G. M. & Borrow, P. (2009). Kinetics of expansion of epitope-specific T cell responses during primary HIV-1 infection. *Journal of immunology* **182**, 7131-45.
311. Mathias Lichterfeld, D. M., Thai Duong Hong Cung, Katie L. Williams, Michael T. Waring, Jinghe Huang, Florencia Pereyra, Alicja Trocha, Gordon J. Freeman, Eric S. Rosenberg, Bruce D. Walker and Xu G. Yu. (2008). Telomerase activity of HIV-1–specific CD8+ T cells- constitutive up-regulation in controllers and selective increase by blockade of PD ligand 1 in progressors. *Blood* **112**, 3679-3687.
312. Migueles, S. A., Laborico, A. C., Shupert, W. L., Sabbaghian, M. S., Rabin, R., Hallahan, C. W., Van Baarle, D., Kostense, S., Miedema, F., McLaughlin, M., Ehler, L., Metcalf, J., Liu, S. & Connors, M. (2002). HIV-specific CD8+ T cell proliferation is coupled to perforin expression and is maintained in nonprogressors. *Nature immunology* **3**, 1061-8.
313. Betts, M. R., Nason, M. C., West, S. M., De Rosa, S. C., Migueles, S. A., Abraham, J., Lederman, M. M., Benito, J. M., Goepfert, P. A., Connors, M., Roederer, M. & Koup, R. A. (2006). HIV nonprogressors preferentially maintain highly functional HIV-specific CD8+ T cells. *Blood* **107**, 4781-9.
314. Douglas F. Nixon, A. R. M. T., John G. Elvin, Charles R. Rizza, John Gallwey & Andrew J. Mcmichael. (1988). HIV-1 gag-specific cytotoxic T lymphocytes defined with recombinant vaccinia virus and synthetic peptides. *Nature* **336**, 484-486.
315. Streeck, H., Lichterfeld, M., Alter, G., Meier, A., Teigen, N., Yassine-Diab, B., Sidhu, H. K., Little, S., Kelleher, A., Routy, J. P., Rosenberg, E. S., Sekaly, R. P., Walker, B. D. & Altfeld, M. (2007). Recognition of a defined region within p24 gag by CD8+ T cells during primary human immunodeficiency virus type 1 infection in individuals expressing protective HLA class I alleles. *Journal of Virology* **81**, 7725-31.
316. Gao, X., Bashirova, A., Iversen, A. K., Phair, J., Goedert, J. J., Buchbinder, S., Hoots, K., Vlahov, D., Altfeld, M., O'Brien, S. J. & Carrington, M. (2005). AIDS restriction HLA allotypes target distinct intervals of HIV-1 pathogenesis. *Nature medicine* **11**, 1290-2.
317. Schneidewind, A., Brockman, M. A., Yang, R., Adam, R. I., Li, B., Le Gall, S., Rinaldo, C. R., Craggs, S. L., Allgaier, R. L., Power, K. A., Kuntzen, T., Tung, C. S., LaBute, M. X., Mueller, S. M., Harrer, T., McMichael, A. J., Goulder, P. J., Aiken, C., Brander, C., Kelleher, A. D. & Allen, T. M. (2007). Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication. *Journal of Virology* **81**, 12382-93.
318. Feeney, M. E., Tang, Y., Roosevelt, K. A., Leslie, A. J., McIntosh, K., Karthas, N., Walker, B. D. & Goulder, P. J. R. (2004). Immune Escape Precedes Breakthrough Human Immunodeficiency Virus Type 1 Viremia and Broadening of the Cytotoxic T-Lymphocyte Response in an HLA-B27-Positive Long-Term-Nonprogressing Child. *Journal of Virology* **78**, 8927-8930.
319. Phillips RE, R.-J. S., Nixon DF, Gotch FM, Edwards JP, Ogunlesi AO, Elvin JG, Rothbard JA, Bangham CR, Rizza CR, et al. (1991). Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature* **354**, 453-459.

320. Goulder PJ, B. C., Tang Y, Tremblay C, Colbert RA, Addo MM, Rosenberg ES, Nguyen T, Allen R, Trocha A, Altfeld M, He S, Bunce M, Funkhouser R, Pelton SI, Burchett SK, McIntosh K, Korber BT, Walker BD. (2001). Evolution and transmission of stable CTL escape mutations in HIV infection. *Nature* **412**, 334-338.
321. Lichterfeld, M., Kavanagh, D. G., Williams, K. L., Moza, B., Mui, S. K., Miura, T., Sivamurthy, R., Allgaier, R., Pereyra, F., Trocha, A., Feeney, M., Gandhi, R. T., Rosenberg, E. S., Altfeld, M., Allen, T. M., Allen, R., Walker, B. D., Sundberg, E. J. & Yu, X. G. (2007). A viral CTL escape mutation leading to immunoglobulin-like transcript 4-mediated functional inhibition of myelomonocytic cells. *The Journal of experimental medicine* **204**, 2813-24.
322. Dean R. Madden, J. C. G., Jack L. Strominger, Don C. Wiley. (1992). The Three-Dimensional Structure of HLA-B27 at 2.1 A Resolution Suggests a General Mechanism for Tight Peptide Binding to MHC. *Cell* **70**, 1035-1048.
323. T. S. Jardetzky, W. S. L., R. A. Robinson, D. R. Madden & D. C. Wiley. (1991). Identification of self peptides bound to purified HLA-B27. *Nature*, 326-329.
324. Payne, R. P., Kloverpris, H., Sacha, J. B., Brumme, Z., Brumme, C., Buus, S., Sims, S., Hickling, S., Riddell, L., Chen, F., Luzzi, G., Edwards, A., Phillips, R., Prado, J. G. & Goulder, P. J. R. (2010). Efficacious Early Antiviral Activity of HIV Gag- and Pol-Specific HLA-B*2705-Restricted CD8+ T Cells. *Journal of Virology* **84**, 10543-10557.
325. Martin, M. P., Qi, Y., Gao, X., Yamada, E., Martin, J. N., Pereyra, F., Colombo, S., Brown, E. E., Shupert, W. L., Phair, J., Goedert, J. J., Buchbinder, S., Kirk, G. D., Telenti, A., Connors, M., O'Brien, S. J., Walker, B. D., Parham, P., Deeks, S. G., McVicar, D. W. & Carrington, M. (2007). Innate partnership of HLA-B and KIR3DL1 subtypes against HIV-1. *Nature genetics* **39**, 733-40.
326. Ying Qi, M. P. M., Xiaojiang Gao, Lisa Jacobson, James J. Goedert, Susan Buchbinder, Gregory D. Kirk, Stephen J. O'Brien, John Trowsdale, Mary Carrington. (2006). KIR/HLA Pleiotropism: Protection against Both HIV and Opportunistic Infections. *PLoS pathogens* **2**, e79.
327. Martin, M. P., Gao, X., Lee, J. H., Nelson, G. W., Detels, R., Goedert, J. J., Buchbinder, S., Hoots, K., Vlahov, D., Trowsdale, J., Wilson, M., O'Brien, S. J. & Carrington, M. (2002). Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nature genetics* **31**, 429-34.
328. Alter, G., Martin, M. P., Teigen, N., Carr, W. H., Suscovich, T. J., Schneidewind, A., Streeck, H., Waring, M., Meier, A., Brander, C., Lifson, J. D., Allen, T. M., Carrington, M. & Altfeld, M. (2007). Differential natural killer cell mediated inhibition of HIV-1 replication based on distinct KIR/HLA subtypes. *Journal of Experimental Medicine* **204**, 3027-3036.
329. Kim, S., Poursine-Laurent, J., Truscott, S. M., Lybarger, L., Song, Y. J., Yang, L., French, A. R., Sunwoo, J. B., Lemieux, S., Hansen, T. H. & Yokoyama, W. M. (2005). Licensing of natural killer cells by host major histocompatibility complex class I molecules. *Nature* **436**, 709-13.
330. Anfossi, N., Andre, P., Guia, S., Falk, C. S., Roetynck, S., Stewart, C. A., Breso, V., Frassati, C., Reviron, D., Middleton, D., Romagne, F., Ugolini, S. & Vivier, E. (2006). Human NK cell education by inhibitory receptors for MHC class I. *Immunity* **25**, 331-42.
331. Fernandez, N. C., Treiner, E., Vance, R. E., Jamieson, A. M., Lemieux, S. & Raulet, D. H. (2005). A subset of natural killer cells achieves self-tolerance without expressing inhibitory receptors specific for self-MHC molecules. *Blood* **105**, 4416-23.
332. Cooley, S., Xiao, F., Pitt, M., Gleason, M., McCullar, V., Bergemann, T. L., McQueen, K. L., Guethlein, L. A., Parham, P. & Miller, J. S. (2007). A subpopulation of human

peripheral blood NK cells that lacks inhibitory receptors for self-MHC is developmentally immature. *Blood* **110**, 578-86.

333. Kim, S., Sunwoo, J. B., Yang, L., Choi, T., Song, Y. J., French, A. R., Vlahiotis, A., Piccirillo, J. F., Cella, M., Colonna, M., Mohanakumar, T., Hsu, K. C., Dupont, B. & Yokoyama, W. M. (2008). HLA alleles determine differences in human natural killer cell responsiveness and potency. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 3053-8.
334. O'Connor, G. M., Yamada, E., Rampersaud, A., Thomas, R., Carrington, M. & McVicar, D. W. (2011). Analysis of binding of KIR3DS1*014 to HLA suggests distinct evolutionary history of KIR3DS1. *Journal of immunology* **187**, 2162-71.
335. Barbour JD, S. U., Caillier SJ, Levy JA, Hecht FM, Oksenberg JR. (2007). Synergy or independence? Deciphering the interaction of HLA Class I and NK cell KIR alleles in early HIV-1 disease progression. *PLoS pathogens* **3**, e43.
336. Iversen, A. K., Stewart-Jones, G., Learn, G. H., Christie, N., Sylvester-Hviid, C., Armitage, A. E., Kaul, R., Beattie, T., Lee, J. K., Li, Y., Chotiyarnwong, P., Dong, T., Xu, X., Luscher, M. A., MacDonald, K., Ullum, H., Klarlund-Pedersen, B., Skinhoj, P., Fugger, L., Buus, S., Mullins, J. I., Jones, E. Y., van der Merwe, P. A. & McMichael, A. J. (2006). Conflicting selective forces affect T cell receptor contacts in an immunodominant human immunodeficiency virus epitope. *Nature immunology* **7**, 179-89.
337. Koenig S, C. A., Brewah YA, Jones GM, Leath S, Boots LJ, Davey V, Pantaleo G, Demarest JF, Carter C, et al. (1995). Transfer of HIV-1-specific cytotoxic T lymphocytes to an AIDS patient leads to selection for mutant HIV variants and subsequent disease progression. *Nature Medicine* **1**, 330-336.
338. Wolinsky SM, K. B., Neumann AU, Daniels M, Kunstman KJ, Whetsell AJ, Furtado MR, Cao Y, Ho DD, Safrit JT. (1996). Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. *Science* **272**, 537-542.
339. Borrow, P., Lewicki, H., Wei, X., Horwitz, M. S., Peffer, N., Meyers, H., Nelson, J. A., Gairin, J. E., Hahn, B. H., Oldstone, M. B. A. & Shaw, G. M. (1997). Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nature Medicine* **3**, 205-211.
340. Price DA, G. P., Klenerman P, Sewell AK, Easterbrook PJ, Troop M, Bangham CR, Phillips RE. (1997). Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. *Proceedings of the National Academy of Sciences* **94**, 1890-1895.
341. Allen TM, O. C. D., Jing P, Dzuris JL, Mothé BR, Vogel TU, Dunphy E, Liebl ME, Emerson C, Wilson N, Kunstman KJ, Wang X, Allison DB, Hughes AL, Desrosiers RC, Altman JD, Wolinsky SM, Sette A, Watkins DI. (2000). Tat-specific cytotoxic T lymphocytes select for SIV escape variants during resolution of primary viraemia. *Nature* **407**, 386-390.
342. Iglesias, M. C., Almeida, J. R., Fastenackels, S., van Bockel, D. J., Hashimoto, M., Venturi, V., Gostick, E., Urrutia, A., Wooldridge, L., Clement, M., Gras, S., Wilmann, P. G., Autran, B., Moris, A., Rossjohn, J., Davenport, M. P., Takiguchi, M., Brander, C., Douek, D. C., Kelleher, A. D., Price, D. A. & Appay, V. (2011). Escape from highly effective public CD8+ T-cell clonotypes by HIV. *Blood* **118**, 2138-49.
343. Eggers M, B.-F. B., Ruppert T, Kloetzel PM, Koszinowski UH. (1995). The cleavage preference of the proteasome governs the yield of antigenic peptides. *Journal of Experimental Medicine* **182**, 1865-1870.
344. Tatos N. Akopian, A. F. K., and Alfred L. Goldberg. (1997). Processive Degradation of Proteins and Other Catalytic Properties of the Proteasome from Thermoplasma acidophilum. *The Journal of Biological Chemistry* **272**, 1791-1798.

345. Lawrence R. Dick, C. R. M., George N. DeMartino, and Clive A. Slaughter. (1991). Degradation of Oxidized Insulin B Chain by the Multiproteinase Complex Macropain (Proteasome). *Biochemistry* **30**, 2725-2734.
346. Iztok Dolenc, E. S. l., Wolfgang Baumeister. (1998). Decelerated degradation of short peptides by the 20S proteasome. *FEBS Letters* **434**, 357-361.
347. Lawrence R. Dick, C. R. M., George N. DeMartino, Clive A. Slaughter. (1991). Degradation of Oxidized Insulin B Chain by the Multiproteinase Complex Macropain (Proteasome). *Biochemistry* **30**, 2725-2734.
348. Zimbwa, P., Milicic, A., Frater, J., Scriba, T. J., Willis, A., Goulder, P. J., Pillay, T., Gunthard, H., Weber, J. N., Zhang, H. T. & Phillips, R. E. (2007). Precise identification of a human immunodeficiency virus type 1 antigen processing mutant. *Journal of Virology* **81**, 2031-8.
349. Draenert, R., Le Gall, S., Pfafferott, K. J., Leslie, A. J., Chetty, P., Brander, C., Holmes, E. C., Chang, S. C., Feeney, M. E., Addo, M. M., Ruiz, L., Ramduth, D., Jeena, P., Altfeld, M., Thomas, S., Tang, Y., Verrill, C. L., Dixon, C., Prado, J. G., Kiepiela, P., Martinez-Picado, J., Walker, B. D. & Goulder, P. J. (2004). Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection. *The Journal of experimental medicine* **199**, 905-15.
350. Anita Milicic, D. A. P., Peter Zimbwa, Bruce L. Booth, Helen L. Brown, Philippa J. Easterbrook, Kara Olsen, Nicola Robinson, Uzi Gileadi, Andrew K. Sewell, Vincenzo Cerundolo and Rodney E. Phillips. (2005). CD8+ T cell epitope-flanking mutations disrupt proteasomal processing of HIV-1 Nef. *The Journal of Immunology* **175**, 4618-4626.
351. Yokomaku, Y., Miura, H., Tomiyama, H., Kawana-Tachikawa, A., Takiguchi, M., Kojima, A., Nagai, Y., Iwamoto, A., Matsuda, Z. & Ariyoshi, K. (2004). Impaired Processing and Presentation of Cytotoxic-T-Lymphocyte (CTL) Epitopes Are Major Escape Mechanisms from CTL Immune Pressure in Human Immunodeficiency Virus Type 1 Infection. *Journal of Virology* **78**, 1324-1332.
352. Zhang, S. C., Martin, E., Shimada, M., Godfrey, S. B., Fricke, J., Locastro, S., Lai, N. Y., Liebesny, P., Carlson, J. M., Brumme, C. J., Ogbechie, O. A., Chen, H., Walker, B. D., Brumme, Z. L., Kavanagh, D. G. & Le Gall, S. (2012). Aminopeptidase substrate preference affects HIV epitope presentation and predicts immune escape patterns in HIV-infected individuals. *Journal of immunology* **188**, 5924-34.
353. Clive M. Gray, J. L., Jonathan M. Schapiro, John D. Altman, Mark A. Winters, Meg Crompton, Muoi Loi, Smriti K. Kundu, Mark M. Davis and Thomas C. Merigan. (1999). Frequency of Class I HLA-Restricted Anti-HIV CD8+ + T Cells in Individuals Receiving Highly Active Antiretroviral Therapy (HAART). *The Journal of Immunology* **162**, 1780-1788.
354. Song H, P. J., Cai F, Bhattacharya T, Li H, Iyer SS, Bar KJ, Decker JM, Goonetilleke N, Liu MK, Berg A, Hora B, Drinker MS, Eudailey J, Pickeral J, Moody MA, Ferrari G, McMichael A, Perelson AS, Shaw GM, Hahn BH, Haynes BF, Gao F. (2012). Impact of immune escape mutations on HIV-1 fitness in the context of the cognate transmitted: founder genome. *Retrovirology* **9**, 89.
355. Edwards, C. T., Pfafferott, K. J., Goulder, P. J., Phillips, R. E. & Holmes, E. C. (2005). Intrapatient escape in the A*0201-restricted epitope SLYNTVATL drives evolution of human immunodeficiency virus type 1 at the population level. *Journal of Virology* **79**, 9363-6.
356. Leslie, A., Kavanagh, D., Honeyborne, I., Pfafferott, K., Edwards, C., Pillay, T., Hilton, L., Thobakgale, C., Ramduth, D., Draenert, R., Le Gall, S., Luzzi, G., Edwards, A., Brander, C., Sewell, A. K., Moore, S., Mullins, J., Moore, C., Mallal, S., Bhardwaj, N., Yusim, K.,

Phillips, R., Klenerman, P., Korber, B., Kiepiela, P., Walker, B. & Goulder, P. (2005). Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA. *The Journal of experimental medicine* **201**, 891-902.

357. Kawashima, Y., Pfafferott, K., Frater, J., Matthews, P., Payne, R., Addo, M., Gatanaga, H., Fujiwara, M., Hachiya, A., Koizumi, H., Kuse, N., Oka, S., Duda, A., Prendergast, A., Crawford, H., Leslie, A., Brumme, Z., Brumme, C., Allen, T., Brander, C., Kaslow, R., Tang, J., Hunter, E., Allen, S., Mulenga, J., Branch, S., Roach, T., John, M., Mallal, S., Ogwu, A., Shapiro, R., Prado, J. G., Fidler, S., Weber, J., Pybus, O. G., Klenerman, P., Ndung'u, T., Phillips, R., Heckerman, D., Harrigan, P. R., Walker, B. D., Takiguchi, M. & Goulder, P. (2009). Adaptation of HIV-1 to human leukocyte antigen class I. *Nature* **458**, 641-5.
358. Kessler, J. H., Khan, S., Seifert, U., Le Gall, S., Chow, K. M., Paschen, A., Bres-Vloemans, S. A., de Ru, A., van Montfoort, N., Franken, K. L., Benckhuijsen, W. E., Brooks, J. M., van Hall, T., Ray, K., Mulder, A., Doxiadis, II, van Swieten, P. F., Overkleeft, H. S., Prat, A., Tomkinson, B., Neefjes, J., Kloetzel, P. M., Rodgers, D. W., Hersh, L. B., Drijfhout, J. W., van Veelen, P. A., Ossendorp, F. & Melief, C. J. (2011). Antigen processing by nardilysin and thimet oligopeptidase generates cytotoxic T cell epitopes. *Nature immunology* **12**, 45-53.
359. Cascio P, H. C., Kisselev AF, Rock KL, Goldberg AL. (2001). 26S proteasomes and immunoproteasomes produce mainly N-extended versions of an antigenic peptide. *The EMBO Journal* **20**, 2357-2366.
360. Goldberg AL, C. P., Saric T, Rock KL. (2002). The importance of the proteasome and subsequent proteolytic steps in the generation of antigenic peptides. *Molecular Immunology* **39**, 147-164.
361. Uma Malhotra, S. H., Sujay Dutta, M. Michelle Berrey, Elizabeth Delpit, David M. Koelle, Alessandro Sette, Lawrence Corey, and M. Juliana McElrath. (2001). Role for HLA class II molecules in HIV-1 suppression and cellular immunity following antiretroviral treatment. *The Journal of Clinical Investigation* **107**, 505-517.
362. Betts, M. R., Exley, B., Price, D. A., Bansal, A., Camacho, Z. T., Teaberry, V., West, S. M., Ambrozak, D. R., Tomaras, G., Roederer, M., Kilby, J. M., Tartaglia, J., Belshe, R., Gao, F., Douek, D. C., Weinhold, K. J., Koup, R. A., Goepfert, P. & Ferrari, G. (2005). Characterization of functional and phenotypic changes in anti-Gag vaccine-induced T cell responses and their role in protection after HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 4512-7.
363. Soghoian, D. Z., Jessen, H., Flanders, M., Sierra-Davidson, K., Cutler, S., Pertel, T., Ranasinghe, S., Lindqvist, M., Davis, I., Lane, K., Rychert, J., Rosenberg, E. S., Piechocka-Trocha, A., Brass, A. L., Brenchley, J. M., Walker, B. D. & Streeck, H. (2012). HIV-specific cytolytic CD4+ T cell responses during acute HIV infection predict disease outcome. *Science translational medicine* **4**, 123ra25.
364. Ranasinghe, S., Flanders, M., Cutler, S., Soghoian, D. Z., Ghebremichael, M., Davis, I., Lindqvist, M., Pereyra, F., Walker, B. D., Heckerman, D. & Streeck, H. (2012). HIV-specific CD4+ T cell responses to different viral proteins have discordant associations with viral load and clinical outcome. *Journal of Virology* **86**, 277-83.
365. Kaufmann, D. E., Bailey, P. M., Sidney, J., Wagner, B., Norris, P. J., Johnston, M. N., Cosimi, L. A., Addo, M. M., Lichterfeld, M., Altfeld, M., Frahm, N., Brander, C., Sette, A., Walker, B. D. & Rosenberg, E. S. (2004). Comprehensive Analysis of Human Immunodeficiency Virus Type 1-Specific CD4+ Responses Reveals Marked Immunodominance of gag and nef and the Presence of Broadly Recognized Peptides. *Journal of Virology* **78**, 4463-4477.

366. Boaz MJ, W. A., Murad S, Easterbrook PJ, D'Sousa E, van Wheeley C, Vyakarnam A. (2003). CD4+ responses to conserved HIV-1 T helper epitopes show both negative and positive associations with virus load in chronically infected subjects. *Clinical & experimentall immunology* **134**, 454-463.
367. Wilson, C. C., Palmer, B., Southwood, S., Sidney, J., Higashimoto, Y., Appella, E., Chesnut, R., Sette, A. & Livingston, B. D. (2001). Identification and antigenicity of broadly cross-reactive and conserved human immunodeficiency virus type 1-derived helper T-lymphocyte epitopes. *Journal of Virology* **75**, 4195-207.
368. Altman JD, M. P., Goulder PJ, Barouch DH, McHeyzer-Williams MG, Bell JI, McMichael AJ, Davis MM. (1996). Phenotypic analysis of antigen-specific T lymphocytes. *Science* **274**, 94-96.
369. Vivian, J. P., Duncan, R. C., Berry, R., O'Connor, G. M., Reid, H. H., Beddoe, T., Gras, S., Saunders, P. M., Olshina, M. A., Widjaja, J. M., Harpur, C. M., Lin, J., Maloveste, S. M., Price, D. A., Lafont, B. A., McVicar, D. W., Clements, C. S., Brooks, A. G. & Rossjohn, J. (2011). Killer cell immunoglobulin-like receptor 3DL1-mediated recognition of human leukocyte antigen B. *Nature* **479**, 401-5.
370. Blondelle, S. E., Moya-Castro, R., Osawa, K., Schroder, K. & Wilson, D. B. (2008). Immunogenically optimized peptides derived from natural mutants of HIV CTL epitopes and peptide combinatorial libraries. *Biopolymers* **90**, 683-94.
371. Linda Wooldridge, J. E.-M., Hugo A. van den Berg, Anna Skowera, John J. Miles, Mai Ping Tan, Garry Dolton, Mathew Clement, Sian Llewellyn-Lacey, David A. Price, Mark Peakman, and Andrew K. Sewell. (2011). A Single Autoimmune T Cell Receptor Recognizes More Than a Million Different Peptides. *The Journal of Biological Chemistry* **287**, 1168-1177.
372. Harkiolaki, M., Holmes, S. L., Svendsen, P., Gregersen, J. W., Jensen, L. T., McMahon, R., Friese, M. A., van Boxel, G., Etzensperger, R., Tzartos, J. S., Kranc, K., Sainsbury, S., Harlos, K., Mellins, E. D., Palace, J., Esiri, M. M., van der Merwe, P. A., Jones, E. Y. & Fugger, L. (2009). T cell-mediated autoimmune disease due to low-affinity crossreactivity to common microbial peptides. *Immunity* **30**, 348-57.
373. Chastain EM, M. S. (2012). Molecular mimicry as an inducing trigger for CNS autoimmune demyelinating disease. *Immunological Reviews* **245**, 227-238.
374. Stephan Kissler, S. M. A. a. D. C. W. (2002). Cross-reactivity and T-cell Receptor Antagonism of Myelin Basic Protein-reactive T cells is Modulated by the Activation State of the Antigen Presenting Cell. *Journal of Autoimmunity* **19**, 183-193.
375. Henrik Pedersen, J. S., Liselotte Brix. Immudex.
376. Yin, Y., Wang, X. X. & Mariuzza, R. A. (2012). Crystal structure of a complete ternary complex of T-cell receptor, peptide-MHC, and CD4. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 5405-10.
377. Gangadharan, D. & Cheroutre, H. (2004). The CD8+ isoform CD8alphaalpha is not a functional homologue of the TCR co-receptor CD8alphabeta. *Current opinion in immunology* **16**, 264-70.
378. Alexander Yu Rudensky, P. P.-H., Soon-Cheol Hong, Avlin Barlow & Charles A. Janeway Jr. (1991). Sequence analysis of peptides bound to MHC class II molecules. *Nature* **353**, 622-627.
379. Chicz RM, U. R., Lane WS, Gorga JC, Stern LJ, Vignali DA, Strominger JL. (1992). Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size. *Nature* **358**, 764-768.
380. Kirsten Falk, O. R. t., Stefan Stevanovic, Günther Jung and Hans-Georg Rammensee. (1991). Allele-Specific Motifs Revealed by Sequencing of Self-Peptides Eluted from MHC Molecules. *Nature* **351**, 290-296.

381. D. R. Madden, J. C. G., J. L. Strominger and D. C. Wiley. (1991). The structure of HLA-B27 reveals nonamer self-peptides bound in an extended conformation. *Nature* **353**, 321-325.
382. Rötzschke O, F. K. (1991). Naturally-occurring peptide antigens derived from the MHC class-I-restricted processing pathway. *Immunology Today* **12**, 447-455.
383. Tynan, F. E., Burrows, S. R., Buckle, A. M., Clements, C. S., Borg, N. A., Miles, J. J., Beddoe, T., Whisstock, J. C., Wilce, M. C., Silins, S. L., Burrows, J. M., Kjer-Nielsen, L., Kostenko, L., Purcell, A. W., McCluskey, J. & Rossjohn, J. (2005). T cell receptor recognition of a 'super-bulged' major histocompatibility complex class I-bound peptide. *Nature immunology* **6**, 1114-22.
384. Woods, T. P. H. a. K. R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proceedings of the National Academy of Sciences* **78**, 3824-3828.
385. George F.Gao, J. T., Ulrich C.Gerth, Jessica R. Wyer, Andrew J. McMichael, David I. Stuart, John I. Bell, E. Yvonne Jones & Bent K. Jakobsen. (1997). Crystal structure of the complex between human CD8+ alpha-alpha and HLA-A2. *Nature* **387**, 630-634.
386. Shore, D. A., Issafras, H., Landais, E., Teyton, L. & Wilson, I. A. (2008). The crystal structure of CD8+ in complex with YTS156.7.7 Fab and interaction with other CD8+ antibodies define the binding mode of CD8+ alphabeta to MHC class I. *Journal of molecular biology* **384**, 1190-202.
387. Tara L. Chapman, A. P. H., Anthony P. West, Jr. and Pamela J. Bjorkman. (2000). Crystal Structure and Ligand Binding Properties of the D1D2 Region of the Inhibitory Receptor LIR-1 (ILT2). *Immunity* **13**, 727-736.
388. Shiroishi, M., Tsumoto, K., Amano, K., Shirakihara, Y., Colonna, M., Braud, V. M., Allan, D. S., Makadzange, A., Rowland-Jones, S., Willcox, B., Jones, E. Y., van der Merwe, P. A., Kumagai, I. & Maenaka, K. (2003). Human inhibitory receptors Ig-like transcript 2 (ILT2) and ILT4 compete with CD8+ for MHC class I binding and bind preferentially to HLA-G. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 8856-61.
389. Wooldridge, L., Clement, M., Lissina, A., Edwards, E. S., Ladell, K., Ekeruche, J., Hewitt, R. E., Laugel, B., Gostick, E., Cole, D. K., Debets, R., Berrevoets, C., Miles, J. J., Burrows, S. R., Price, D. A. & Sewell, A. K. (2010). MHC class I molecules with Superenhanced CD8+ binding properties bypass the requirement for cognate TCR recognition and nonspecifically activate CTLs. *Journal of immunology* **184**, 3357-66.
390. Tynan, F. E., Borg, N. A., Miles, J. J., Beddoe, T., El-Hassen, D., Silins, S. L., van Zuylen, W. J., Purcell, A. W., Kjer-Nielsen, L., McCluskey, J., Burrows, S. R. & Rossjohn, J. (2005). High resolution structures of highly bulged viral epitopes bound to major histocompatibility complex class I. Implications for T-cell receptor engagement and T-cell immunodominance. *The Journal of biological chemistry* **280**, 23900-9.
391. Kumar, P., Vahedi-Faridi, A., Saenger, W., Merino, E., Lopez de Castro, J. A., Uchanska-Ziegler, B. & Ziegler, A. (2009). Structural basis for T cell alloreactivity among three HLA-B14 and HLA-B27 antigens. *The Journal of biological chemistry* **284**, 29784-97.
392. Rammensee, H.-G. (1995). Chemistry of peptides associated with MHC class I and class II molecules. *Current Opinion in Immunology* **7**, 85-96.
393. Wiley, M. n. B. a. D. C. (1994). Importance of Peptide Amino and Carboxyl Termini to the Stability of MHC Class I Molecules. *Science* **265**, 398-402.
394. Burrows, S. R., Rossjohn, J. & McCluskey, J. (2006). Have we cut ourselves too short in mapping CTL epitopes? *Trends in immunology* **27**, 11-6.
395. Urban RG, C. R., Lane WS, Strominger JL, Rehm A, Kenter MJ, UytdeHaag FG, Ploegh H, Uchanska-Ziegler B, Ziegler A. (1994). A subset of HLA-B27 molecules contains

peptides much longer than nonamers. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 1534-1538.
396. Michael Probst-Kepper, H.-J. r. H., Hanne Herrmann, Viktoria Janke, Frank Ocklenburg, Jürgen Klempnauer, Benoit J. van den Eynde and Siegfried Weiss. (2004). Conformational restraints and flexibility in 14-meric peptides in complex with HLA B*3501. *The Journal of Immunology* **173**, 5610-5616.
397. Chen, e. a. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica* **D66**, 12-21.
398. Garstka, M. A., Fritzsche, S., Lenart, I., Hein, Z., Jankevicius, G., Boyle, L. H., Elliott, T., Trowsdale, J., Antoniou, A. N., Zacharias, M. & Springer, S. (2011). Tapasin dependence of major histocompatibility complex class I molecules correlates with their conformational flexibility. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **25**, 3989-98.
399. Burgess, R. R. (2009). Chapter 17 Refolding Solubilized Inclusion Body Proteins. **463**, 259-282.
400. Glithero, A., Tormo, J., Doering, K., Kojima, M., Jones, E. Y. & Elliott, T. (2006). The crystal structure of H-2D(b) complexed with a partial peptide epitope suggests a major histocompatibility complex class I assembly intermediate. *The Journal of biological chemistry* **281**, 12699-704.
401. Kiley R. Prilliman, K. W. J., Mark Lindsey, Jihua Wang, David Crawford and William H. Hildebrand. (1999). HLA-B15 Peptide Ligands Are Preferentially Anchored at Their C Termini. *The Journal of Immunology* **162**, 7277-7284.
402. Hillig, R. C., Hulsmeyer, M., Saenger, W., Welfle, K., Misselwitz, R., Welfle, H., Kozerski, C., Volz, A., Uchanska-Ziegler, B. & Ziegler, A. (2004). Thermodynamic and structural analysis of peptide- and allele-dependent properties of two HLA-B27 subtypes exhibiting differential disease association. *The Journal of biological chemistry* **279**, 652-63.
403. Bade-Doding, C., Theodossis, A., Gras, S., Kjer-Nielsen, L., Eiz-Vesper, B., Seltsam, A., Huyton, T., Rossjohn, J., McCluskey, J. & Blasczyk, R. (2011). The impact of human leukocyte antigen (HLA) micropolymorphism on ligand specificity within the HLA-B*41 allotypic family. *Haematologica* **96**, 110-8.
404. Hulsmeyer, M., Fiorillo, M. T., Bettosini, F., Sorrentino, R., Saenger, W., Ziegler, A. & Uchanska-Ziegler, B. (2004). Dual, HLA-B27 subtype-dependent conformation of a self-peptide. *The Journal of experimental medicine* **199**, 271-81.
405. Fiorillo, M. T., Ruckert, C., Hulsmeyer, M., Sorrentino, R., Saenger, W., Ziegler, A. & Uchanska-Ziegler, B. (2005). Allele-dependent similarity between viral and self-peptide presentation by HLA-B27 subtypes. *The Journal of biological chemistry* **280**, 2962-71.
406. Nurzia E, N. D., Cauli A, Mathieu A, Tedeschi V, Caristi S, Sorrentino R, Böckmann RA, Fiorillo MT. (2012). Interaction Pattern of Arg 62 in the A-Pocket of Differentially Disease-Associated HLA-B27 Subtypes Suggests Distinct TCR Binding Modes. *PLoS One* **7**, e32865.
407. Wiley, L. S. A. D. C. (1994). Antigenic peptide binding by class I and class IIhistocompatibility proteins. *Structure* **2**, 245-251.
408. Kim, S. T., Takeuchi, K., Sun, Z. Y., Touma, M., Castro, C. E., Fahmy, A., Lang, M. J., Wagner, G. & Reinherz, E. L. (2009). The alphabeta T cell receptor is an anisotropic mechanosensor. *The Journal of biological chemistry* **284**, 31028-37.
409. Jenny E. Gumperz, L. D. B., Nicholas M. Valiante, Lucy Percival, Joseph H. Phillips, Lewis L. Lanier, and Peter Parham. (1997). Conserved and Variable Residues Within

the Bw4 Motif of HLA-B Make Separable Contributions to Recognition by the NKBl Killer Cell-Inhibitory Receptor. *The Journal of Immunology* **158**, 5237-5241.
410. Tynan, F. E., Reid, H. H., Kjer-Nielsen, L., Miles, J. J., Wilce, M. C., Kostenko, L., Borg, N. A., Williamson, N. A., Beddoe, T., Purcell, A. W., Burrows, S. R., McCluskey, J. & Rossjohn, J. (2007). A T cell receptor flattens a bulged antigenic peptide presented by a major histocompatibility complex class I molecule. *Nature immunology* **8**, 268-76.
411. Krissinel, E. & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology* **372**, 774-97.
412. Jones, N. A., Wei, X., Flower, D. R., Wong, M., Michor, F., Saag, M. S., Hahn, B. H., Nowak, M. A., Shaw, G. M. & Borrow, P. (2004). Determinants of human immunodeficiency virus type 1 escape from the primary CD8+ cytotoxic T lymphocyte response. *The Journal of experimental medicine* **200**, 1243-56.
413. Margarita Del Val, H.-J. S., Thomas Ruppert, Matthias J. Reddehase, Ulrich H. Koszinowski. (1991). Efficient processing of an antigenic sequence for presentation by MHC class I molecules depends on its neighboring residues in the protein. *Cell* **66**, 1145-1153.
414. Le Gall, S., Stamegna, P. & Walker, B. D. (2007). Portable flanking sequences modulate CTL epitope processing. *The Journal of clinical investigation* **117**, 3563-75.
415. Brackenridge, S., Evans, E. J., Toebes, M., Goonetilleke, N., Liu, M. K. P., di Gleria, K., Schumacher, T. N., Davis, S. J., McMichael, A. J. & Gillespie, G. M. (2011). An Early HIV Mutation within an HLA-B*57-Restricted T Cell Epitope Abrogates Binding to the Killer Inhibitory Receptor 3DL1. *Journal of Virology* **85**, 5415-5422.
416. Carrington, M., Martin, M. P. & van Bergen, J. (2008). KIR-HLA intercourse in HIV disease. *Trends in microbiology* **16**, 620-7.
417. Salix Boulet, S. S., Nancy Simic, Julie Bruneau, Jean-Pierre Routy, Christos M. Tsoukas and Nicole F. Bernard. (2008). Increased proportion of KIR3DS1 homozygotes in HIV-exposed uninfected individuals. *AIDS* **22**, 595-599.
418. Hansasuta, P., Dong, T., Thananchai, H., Weekes, M., Willberg, C., Aldemir, H., Rowland-Jones, S. & Braud, V. M. (2004). Recognition of HLA-A3 and HLA-A11 by KIR3DL2 is peptide-specific. *European journal of immunology* **34**, 1673-9.
419. Kollnberger, S., Chan, A., Sun, M.-Y., Ye Chen, L., Wright, C., di Gleria, K., McMichael, A. & Bowness, P. (2007). Interaction of HLA-B27 homodimers with KIR3DL1 and KIR3DL2, unlike HLA-B27 heterotrimers, is independent of the sequence of bound peptide. *European journal of immunology* **37**, 1313-1322.
420. Tynan, F. E., Burrows, S. R., Buckle, A. M., Clements, C. S., Borg, N. A., Miles, J. J., Beddoe, T., Whisstock, J. C., Wilce, M. C., Silins, S. L., Burrows, J. M., Kjer-Nielsen, L., Kostenko, L., Purcell, A. W., McCluskey, J. & Rossjohn, J. (2005). T cell receptor recognition of a 'super-bulged' major histocompatibility complex class I–bound peptide. *Nature immunology* **6**, 1114-1122.
421. Brodin, P., Karre, K. & Hoglund, P. (2009). NK cell education: not an on-off switch but a tunable rheostat. *Trends in immunology* **30**, 143-9.
422. Karina Yusim, B. T. M. K., Christian Brander, Barton F. Haynes, Richard Koup, John P. Moore, Bruce D. Walker, and David I. Watkins. (2009). HIV Molecular Immunology, pp. LA-UR 09-05941. Los Alamos National Laboratory, Theoretical Biology and Biophysics.
423. Marta Peruzzi, K. C. P., Eric O. Long, and Mauro S. Mahati. (1996). Peptide Sequence Requirements for the Recognition of HLA-B*2705 by Specific Natural Killer Cells. *The Journal of Immunology* **157**, 3350-3356.
424. Andrews, D. M., Estcourt, M. J., Andoniou, C. E., Wikstrom, M. E., Khong, A., Voigt, V., Fleming, P., Tabarias, H., Hill, G. R., van der Most, R. G., Scalzo, A. A., Smyth, M. J. &

Degli-Esposti, M. A. (2010). Innate immunity defines the capacity of antiviral T cells to limit persistent infection. *The Journal of experimental medicine* **207**, 1333-43.
425. Helen C. Su, K. B. N., Thais P. Salazar-Mather, Melanie C. Ruzek, Marc Y. Dalod and Christine A. Biron. (2001). NK cell functions restrain T cell responses during viral infections. *European journal of immunology* **31**, 3048-3055.
426. McElrath, M. J. & Haynes, B. F. (2010). Induction of immunity to human immunodeficiency virus type-1 by vaccination. *Immunity* **33**, 542-54.
427. Papagno, L., Alter, G., Assoumou, L., Murphy, R. L., Garcia, F., Clotet, B., Larsen, M., Braibant, M., Marcelin, A. G., Costagliola, D., Altfeld, M., Katlama, C. & Autran, B. (2011). Comprehensive analysis of virus-specific T-cells provides clues for the failure of therapeutic immunization with ALVAC-HIV vaccine. *AIDS* **25**, 27-36.
428. Tenbusch, M., Ignatius, R., Temchura, V., Nabi, G., Tippler, B., Stewart-Jones, G., Salazar, A. M., Sauermann, U., Stahl-Hennig, C. & Uberla, K. (2012). Risk of immunodeficiency virus infection may increase with vaccine-induced immune response. *Journal of Virology* **86**, 10533-9.
429. Garboczi DN, H. D., Wiley DC. (1992). HLA-A2-peptide complexes: refolding and crystallization of molecules expressed in Escherichia coli and complexed with single antigenic peptides. *Proceedings of the National Academy of Sciences* **89**, 3429-3433.
430. Tsumoto, K., Ejima, D., Kumagai, I. & Arakawa, T. (2003). Practical considerations in refolding proteins from inclusion bodies. *Protein Expression and Purification* **28**, 1-8.
431. Paul Emsley, B. L., William G. Scott and Kevin Cowtan. (2010). Features and Development of Coot. *Acta Crystallographica Section D - Biological Crystallography* **66**, 486-501.
432. Bricogne G., B. E., Brandl M., Flensburg C., Keller P., Paciorek W., & Roversi P, S. A., Smart O.S., Vonrhein C., Womack T.O. (2011). BUSTER 2.11.2 edit. Global Phasing Ltd., Cambridge.
433. P. D. Adams, P. V. A., G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echoo ls, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica* **D66**, 213-221.
434. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 10037-41.
435. Immudex. (2010). General staining procedure MHC Dextramer - PBMC's. In *Protocols*, Copenhagen.
436. Miltenyi-Biotec. Micro-MACS mRNA Isolation Kit - Short Protocol, Auburn, CA.
437. Appel H, K. W., Kuhne M, Hülsmeyer M, Kollnberger S, Kuhlmann S, Weiss E, Zeitz M, Wucherpfennig K, Bowness P, Sieper J. (2004). The Solvent-Inaccessible Cys67 Residue of HLA-B27 Contributes to T Cell Recognition of HLA-B27/Peptide Complexes. *Journal of Immunology* **173**, 6564-6573.
438. Alvarez, I., Marti, M., Vazquez, J., Camafeita, E., Ogueta, S. & Lopez de Castro, J. A. (2001). The Cys-67 residue of HLA-B27 influences cell surface stability, peptide specificity, and T-cell antigen presentation. *The Journal of biological chemistry* **276**, 48740-7.

9. Appendices

9.1. HLA B*2705 and A*0201 Tetramer Production

9.1.1. Production of B27 heavy chain constructs for protein expression

Production of B27 heavy chain constructs for protein expression was undertaken to allow for refolding and production of the HLA B2705 molecule bound to peptides of interest identified in the above and previous antigen processing studies. These B2705 complexes will be used for tetramer production and structural studies. The sequence used is seen in figure 5 and was obtained from the Anthony Nolan HLA database (www.ebi.ac.uk/imgt/hla/).

ATGCGGGTCACGGCGCCCCGAACCCTCCTCCTGCTGCTCTGGGGGGCAGTGGCCCTGACCGAGACCTGGGCTATGGGCTC
CCACTCCATGAGGTATTTCCACACCTCCGTGTCCCGGCCCGGCCGCGGGGAGCCCCGCTTCATCACCGTGGGCTACGTGG
ACGACACGCTGTTCGTGAGGTTCGACAGCGACGCCGCGAGTCCGAGAGAGGAGCCGCGGGCGCCGTGGATAGAGCAGGAG
GGGCCGGAGTATTGGGACCGGGAGAC*ACAGATCTGCAAGGCCAAGGC*ACAGACTGACCGAGAGGACCTGCGGACCCTGCT
CCGCTACTACAACCAGAGCGAGGCCGGGTCTCACACCCTCCAGAATATGTATGGCTGCGACGTGGGGCCGGACGGGCGCC
TCCTCCGCGGGTACCACCAGGACGCCTACGACGGCAAGGATTACATCGCCCTGAACGAGGACCTGAGCTCCTGGACCGCC
GCGGACACGGCGGCTCAGATCACCCAGCGCAAGTGGGAGGCGGCCCGTGTGGCGGAGCAGCTGAGAGCCTACCTGGAGGG
CGAGTGCGTGGAGTGGCTCCGCAGATACCTGGAGAACGGGAAGGAGACGCTGCAGCGCGCGGACCCCCCAAAGACACACG
TGACCCACCACCCCATCTCTGACCATGAGGCCACCCTGAGGTGCTGGGCCCTGGGCTTCTACCCTGCGGAGATCACACTG
ACCTGGCAGCGGGATGGCGAGGACCAAACTCAGGACACTGAGCTTGTGGAGACCAGACCAGCAGGAGATAGAACCTTCCA
GAAGTGGGCAGCTGTGGTGGTGCCTTCTGGAGAAGAGCAGAGATACACATGCCATGTACAGCATGAGGGGCTGCCGAAGC
CCCTCACCCTGAGATGGGAGCCGTCTTCCCAGTCCACCGTCCCCATCGTGGGCATTGTTGCTGGCCTGGCTGTCCTAGCA
GTTGTGGTCATCGGAGCTGTGGTCGCTGCTGTGATGTGTAGGAGGAAGAGCTCAGGTGGAAAAGGAGGGAGCTACTCTCA
GGCTGCGTGCAGCGACAGTGCCCAGGGCTCTGATGTGTCTCTCACAGCTTGA

Figure S 1: HLA B*2705 cDNA sequence taken from the Anthony Nolan HLA sequence database. Reference number HLA00225. Highlighted area shows protein-coding sequence cloned into plasmid vector. Underline indicates sequence used to create primers to amplify sequence from human cDNA. Italicized area shows sequence primers designed for mutation of free cysteine residue to serine.

This sequence was used to design the following primers (fig 6) and then to amplify the sequence from human cDNA extracted from a B cell line (ref). The sequence was originally to be produced both with and without the cysteine residue at position 67, as this is a free cysteine, which may have reduced refolding efficiency. Therefore, primers were also designed for this section of the sequence with the single letter mutation to encode a serine instead of a cysteine at this position. The two halves of the sequence were then digested by restriction endonucleases at a cleavage site encoded into the primers and re-ligated. However, since other work suggested that this mutation affected epitope binding to the complex[437; 438] this construct was not continued.

A. Forward Crystallisation Primer

5' CAGACCTCATATG**ATG**GGCTCCCAC '3

C. Forward Tetramerisation Primer

5'CAGACCTCC**ATGG**GCTCCCACTCC '3

B. Reverse Crystallisation Primer

5' CACGTCCTCTCGAGCTCCCATCTCA '3

D. Reverse Tetramerisation Primer

5'TGCAGATGGATCCTCCCATCTCA'3

Figure S 2: Primer sequences for amplification of B2705 Heavy chain and ligation into both pET 22b (A. and B.) and pET 23d (C. and D.) vectors. Light grey shows the section of the primer corresponding to a stretch of B2705 sequence. Dark grey indicates the restriction endonuclease recognition site. The ATG start codon is highlighted in bold type. The restriction endonucleases used are A. NDE1 B. XHO1 C. NCO1 AND D. BAMH1. All primers were tested for compatibility and melting temperature using Netprimer (http://www.premierbiosoft.com/netprimer/index.html).

A. Forward Site Directed Mutagenesis Primer

5' AGGAGGGGCCGGAGTATTGG 3'

B. Reverse Site Directed Mutagenesis Primer

5' GTCTCCCGGTCCCAATACTCC 3'

Figure S 3: Primer Sequences for Site directed mutagenesis of Cytosine 178 to Thymine. The mutated base is shown in dark grey. Primers were creating using guidance from the QuikChange® Primer Design

Program. All primers were tested for compatibility and melting temperature using Netprimer (http://www.premierbiosoft.com/netprimer/index.html).

It was also necessary to mutate the sequence as after sequencing the resulting constructs it was discovered that the B cell line was in fact subtype B*2703 and not subtype B*2705 as required. The single nucleotide modification of the construct was done using site directed mutagenesis with specific primers to the region to be mutated (Fig.7).

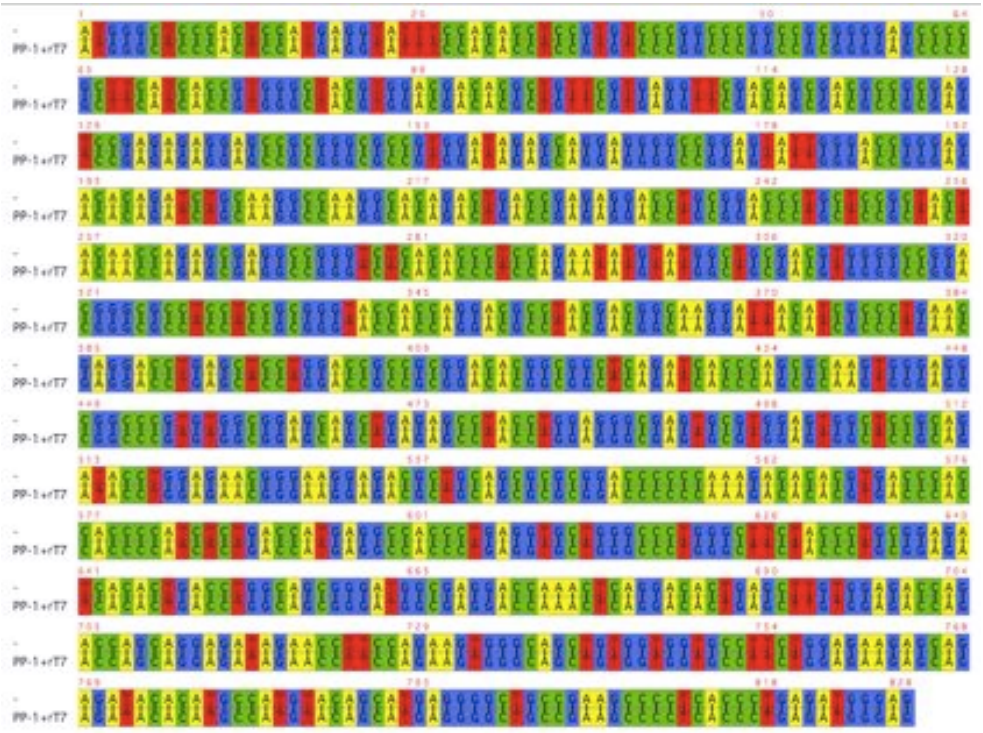


Figure S 4: Sequence alignment of HLA B*2705 sequence from Anthony Nolan trust (top sequence) and sequence obtained from sequencing insert from pet22b vector with B2705 amplified from human cDNA (PP-1+rT7) using primers for the T7 and reverse T7 promoters.

The final construct was then sequenced and compared to the reference sequence for validation. The sequence alignment is shown in figure 8 for the pet 22b vector construct. The vector used for the tetramerisation construct was a pET23d plasmid containing a

biotinylation tag and a His tag. The vector used for the crystallisation construct was a pET22b plasmid without a biotinylation tag.

Despite sequencing the insert and surrounding portions of both vectors showing these were identical to the reference and vector sequences, good expression of the B27 heavy chain protein was not obtained for either construct. Expression conditions were varied and several bacterial strains were used but none increased expression significantly. Fortunately, two HLA B2705 heavy chain constructs donated by Paul Bowness for comparison studies could be used. Both of these were the PLM-1 vector, with the tetramerisation vector again having a His tag and a biotinylation tag. The crystallisation vector has a His tag only.

Expression from these vectors using a standard protocol (Method in Chapter 7.1.1) was robust and the B*27 band at 36kDa can be seen in lane A of the gel in figure 9. β2M, seen at 11kDa in band B, was expressed from a pET23d vector construct that had been previously generated in the lab.

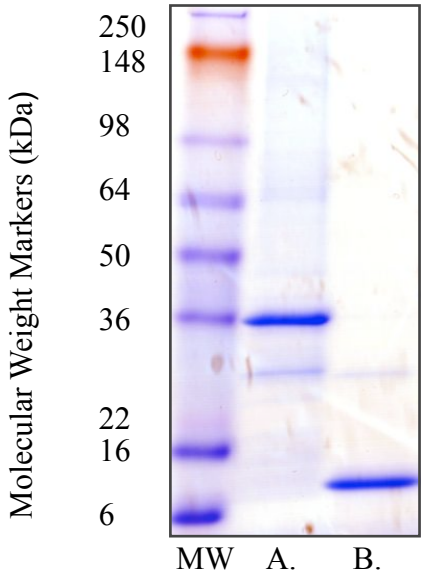


Figure S 5: 4-20% Tris-Glycine SDS Novex PAGE gel showing A. B*2705 band at 36kDa and B. β2m band at 11kDa, both purified from BL21 *E.coli* inclusion bodies.

The B27 heavy chains for both tetramerisation and crystallisation have so far been refolded with β2m and the peptides shown in Table 1 to produce HLA B27 molecules containing the KK10 processing variants identified in antigen processing studies. Figure 10B shows an absorption graph for the size exclusion purification of a refold of B27 heavy chain with his tag, β2m and a KI11 peptide, figure 10C shows the heterotrimer during further size exclusion purification. Figure 11 shows the protein from each peak run on an SDS page gel. This shows the presence of both B27 heavy chain and/or β2m in the fractions. In peak C, both B27 and β2m are present in equal amounts, suggesting it contains correctly folded protein and is therefore the peak of interest. Refolding without the presence of a peptide can also be used for comparison and is shown in figure 10A. Table 1 shows the HLA A*0201 SL9, SY10 and SL9/SY10 escape mutant tetramers that have been created for comparison of CD8+ T cell responses with those to the KK10 variants complexed with B*2705.

MHC Class 1 Allele	HLA A201		HLA B2705
	peptide	comments	peptide
Peptides Refolded	SLYNTVATL SLYNTVAVLY SLFNTIAVL SLFNTIAVLY SLFNTVAVL SLFNTIATLY		KRWIILGL KRWIIMGL KKWIIMGL KRWIILGLNK KRWIILGLNKI KRWIILGLNKIVR VGEIYKRWIILGLNK GEIYKRWIILGLNKI GEIYKRWIILGLNKIVR

Table S 1: Showing the MHC class 1 monomers refolded for tetramerisation to date during my DPhil. Not validated indicates it has not yet been tested for correct folding but has undergone all stages up to this point.

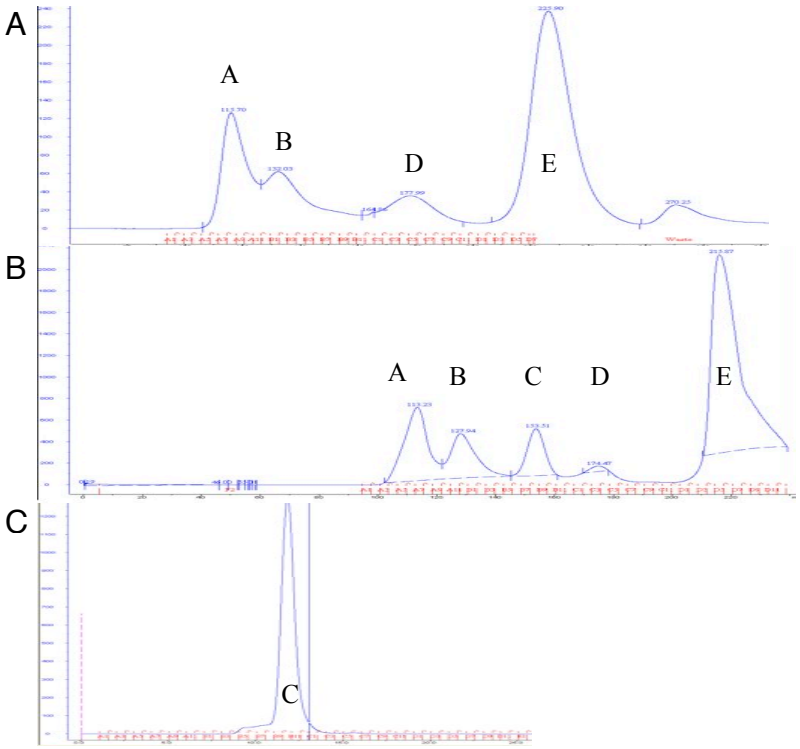


Figure S 6: UV280nm absorbance measurements from size exclusion chromatography of B27 heavy chain and β2M subjected to a rapid dilution refolding procedure A. in the absence of peptide. B. With peptide KRWIILGLNKI showing peaks A-E. Peak A contains heavy aggregates, peak B is B*2705 homodimer, peak C corresponds to the heterotrimer peak D contains the β2M homodimer and peak E light aggregates. C. Showing the B27 heterotrimer peak with KRWIILGLNKI following a second size exclusion.

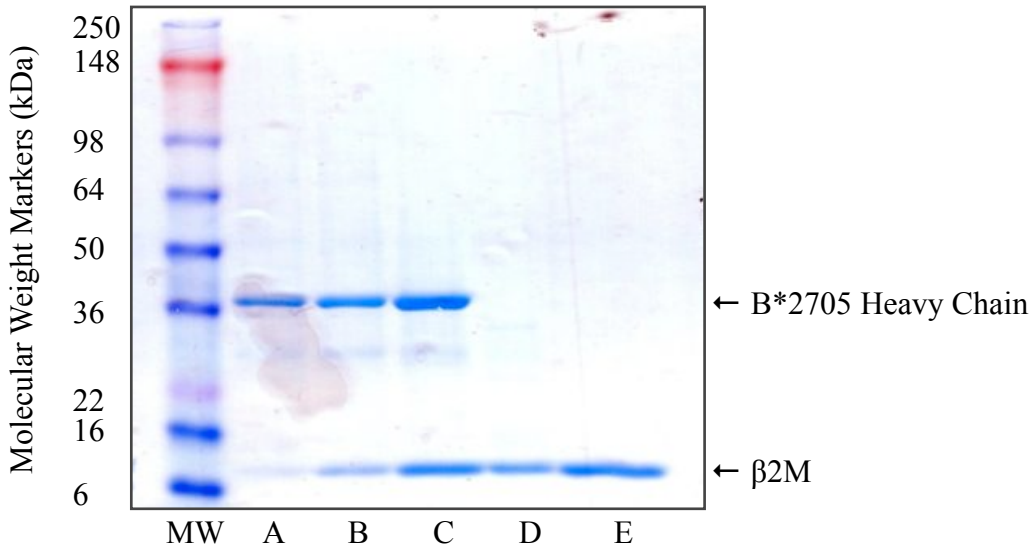


Figure S 7: This shows eluted fractions corresponding to peaks A-E in figure 10B. This shows the presence of proteins at the correct molecular weight for B27 and β2M in the various fractions. Peak C is the peak of interest as it contains both B27 and β2M and they are present in similar amounts. It is therefore likely that this peak contains correctly refolded HLA B27 molecules

9.1.2. Production and validation of HLA B2705 tetramers complexed with various natural KK10 epitope peptide forms

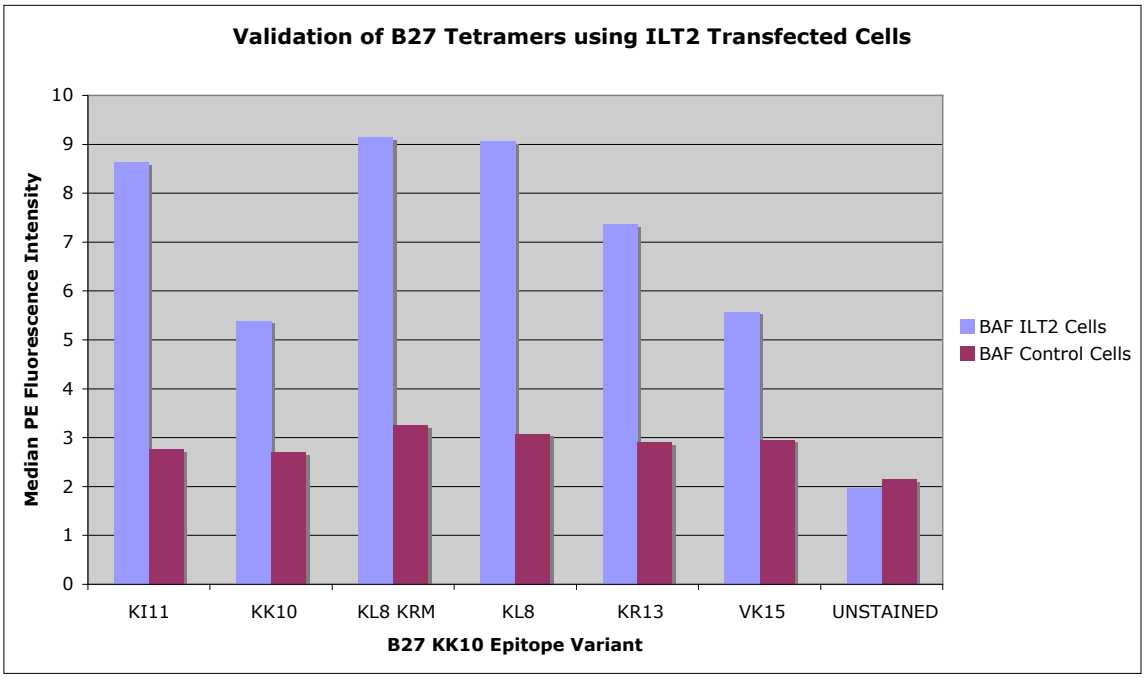


Figure S 8: Median fluorescence intensity of BAF ILT2 transfected cells and untransfected BAF cells after incubation with B27 KK10 variant tetramers. Increases in fluorescence seen in the ILT2 transfected cells over the control cells show binding of the tetramer above background levels.

Monomers of biotinylated HLA B2705 bound to KK10 processing variants were produced as shown in Table 1. Aliquots of the refolded HLA B27 molecules were then conjugated with either streptavidin PE (Sigma Aldrich) or streptavidin APC (Invitrogen) to form fluorescent-labeled tetramers of HLA B27. To validate the tetramers, both the tetramerisation process and the correct refolding of the HLA must be assessed. Validation of the tetramerisation process was shown by performing a dot blot of the streptavidin - HLA mixture as tetramerisation progressed (data not shown). The validation of refolding was performed using BAF cells transfected with the human ILT2 receptor (kind gift from Paul Bowness). This receptor binds all correctly folded MHC class 1 and so binding of the tetramer will validate its correct folding. The binding of each tetramer was assessed for both transfected

and untransfected BAF cells and the differences compared (Fig. 16). All of the tetramers shown here have clear differences between binding to the ILT2 transfected and un-transfected cells, validating their folding.

9.1.3. Creation of the HLA B*2705 expression construct

Extraction of RNA from a human B cell line was performed using an RNeasy Mini Kit (Qiagen) following the "Purification of Total RNA from Animal Cells using Spin Technology" protocol. Briefly, the cells were counted and an appropriate number pelleted, lysed and homogenised, the lysate was passed through a Qiashredder column and centrifuged. The lysate then re-homogenised and mixed with 70% ethanol, transferred to an RNeasy spin column, centrifuged and washed. The column was transferred to a new collection tube and the RNA eluted in RNase free water. The concentration and quality of the RNA was then checked using absorbance and 260/280-ratio absorbance on a nanodrop spectrophotometer (Thermo Scientific).

First Strand cDNA Synthesis using MLV-RT was carried out as described in (http://tools.invitrogen.com/content/sfs/manuals/mmlv_rt_man.pdf) and the resulting cDNA was purified using the QIAquick Gel Extraction Kit (www.qiagen.com) following separation by electrophoresis on 1% Agarose. PCR amplification of HLA B*27 sequence from human cDNA was carried out using the Roche Expand High FidelityPLUS PCR System (http://www.roche-applied-science.com/pack-insert/3300242a.pdf) using specific primers (fig. 6) with the following PCR program:

Start: 94°C for 10 minutes

Cycle 1-10: 94°C for 30 seconds,

65°C for 1 minute (decrease by 0.5°C per cycle),

72°C for 2 minutes

Cycle 11-30: 94°C for 30 seconds,

60°C for 1 minute,

72°C for 2 minutes

End: 72°C for 10 minutes, 4°C ∞

The resulting HLA B*27 DNA was separated by electrophoresis and purified as above. It was then cloned into a TOPO TA plasmid using the Invitrogen TOPO TA Cloning Kit for Sequencing (http://tools.invitrogen.com/content/sfs/manuals/topotaseq_man.pdf) with the pCR 4 TOPO plasmid vector. One Shot® TOP10 chemically competent *E. coli* were transformed with two µl of the TOPO cloning reaction and plated on LB Agar (100µg/ml ampicillin) at 37°C overnight.

Colonies obtained from this were grown in 5ml LB media overnight at 37°C and the DNA extracted and purified using the QIAprep Spin Miniprep Kit (www.qiagen.com). The TOPO vectors were sent for sequencing using M13 forward and reverse primers (Fig. 20)(Messing 1983). Sequencing was performed at the WIMM MRC core sequencing facility. Sequences were aligned with the HLA B*2705 DNA sequence (http://www.ebi.ac.uk/imgt/hla/allele.html; IMGT/HLA Acc No: HLA00225) using eBioX (http://www.ebioinformatics.org/ebiox/) to confirm that they were correct.

Site directed mutagenesis was then carried out to obtain the desired B*2705 subtype. This was done using the QuikChange XL Site-Directed Mutagenesis Kit (Stratagene) using specific primers (Fig. 7) designed using the QuikChange® Primer Design Program and XL-Gold Ultracompetent cells. The procedure was carried out as per protocol (www.genomics.agilent.com Cat. no. 200517) including all control reactions.

Correct sequences were digested from the TOPO plasmid using the NEB restriction endonucleases NCO1 and BAM H1. Buffer conditions and reaction times were calculated using the NEB double digest finder (www.neb.com/nebecomm/DoubleDigestCalculator.asp?). Digestion of the receiving vector (pET22b or pET23d) was performed with the same enzymes and conditions but 0.5 unit/µg vector DNA Calf Intestinal Phosphatase was added in the final 60 minutes of the incubation to prevent re-circularisation of the vector. Both digests were separated by electrophoresis and purified as previously described.

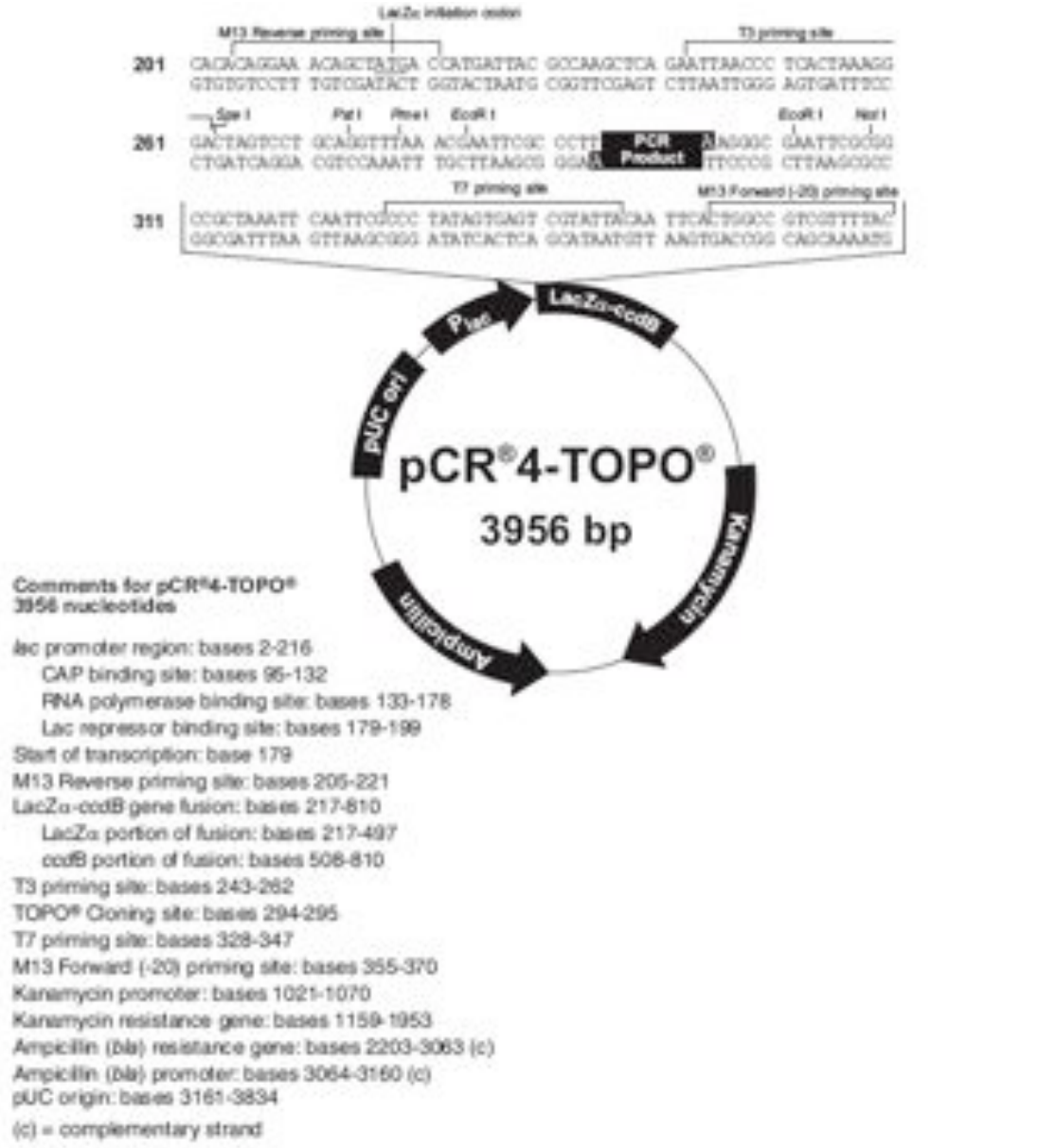


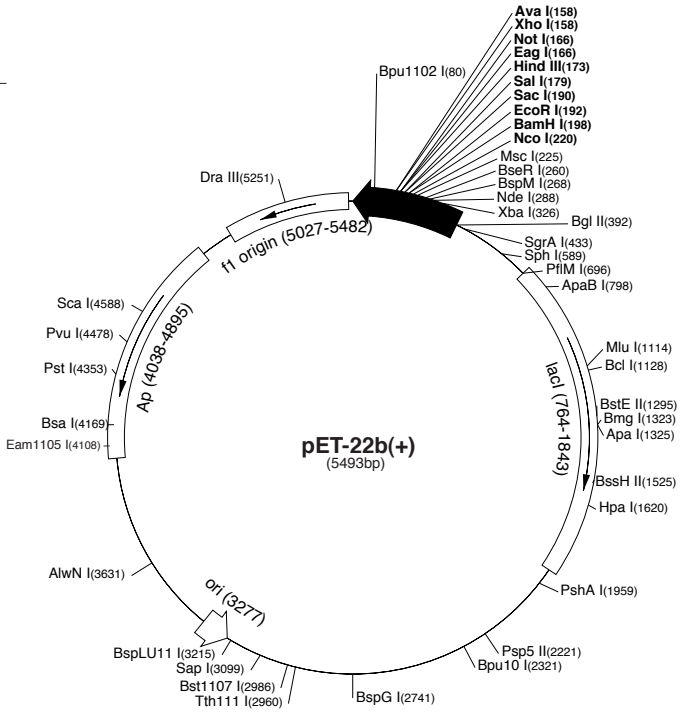
Figure S 9: Map of the pCR 4 TOPO plasmid with positions of primer sites and resistance markers.

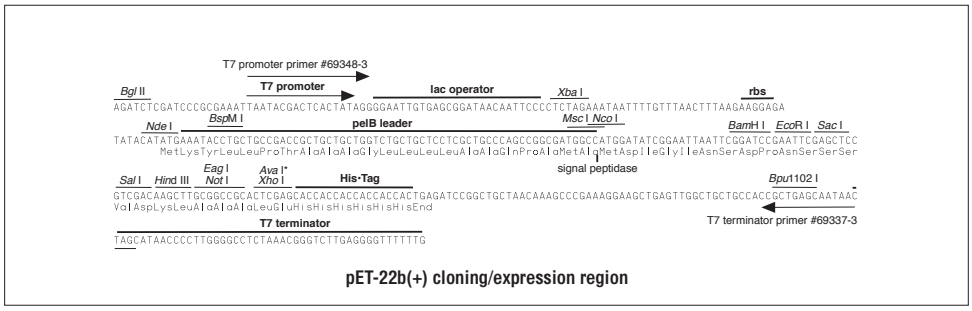
Ligation of HLA B27 sequence with pET 22b and pET23d plasmids were carried out using the Roche Rapid DNA Ligation Kit (www.roche-applied-science.com Cat. no.11635379001). Two µl of the resulting DNA was transformed into One Shot® TOP10 chemically competent *E. coli* following the standard protocol (www.invitrogen.com C4040-03). The transformed *E. coli* were incubated on LB Agar (100µg/ml ampicillin) at 37°C overnight. Colonies from these plates were then picked and individually grown in 5ml LB media (100µg/ml

ampicillin) at 37°C overnight. DNA was extracted and purified for each as previously described (http://www.qiagen.com/products/plasmid/qiaprepminiprepsystem/qiaprepspinminiprepkit.aspx#Tabs=t2).

Sequencing of pET plasmid vectors containing HLA B27 insert was carried out with the T7 forward and reverse primers (Wallis RB 1981)(Fig. 21) at the WIMM MRC core sequencing facility. Sequences were then aligned with the HLA B*2705 DNA sequence (http://www.ebi.ac.uk/imgt/hla/allele.html; IMGT/HLA Acc No: HLA00225) as well as with two regions of vector sequence flanking the insert site (Figs. 21 and 22) using eBioX (http://www.ebioinformatics.org/ebiox/).

pET-22b(+) sequence landmarks	
T7 promoter	361-377
T7 transcription start	360
pelB coding sequence	224-289
Multiple cloning sites	
(*Nco* I - *Xho* I)	158-225
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	764-1843
pBR322 origin	3277
bla coding sequence	4038-4895
f1 origin	5027-5482





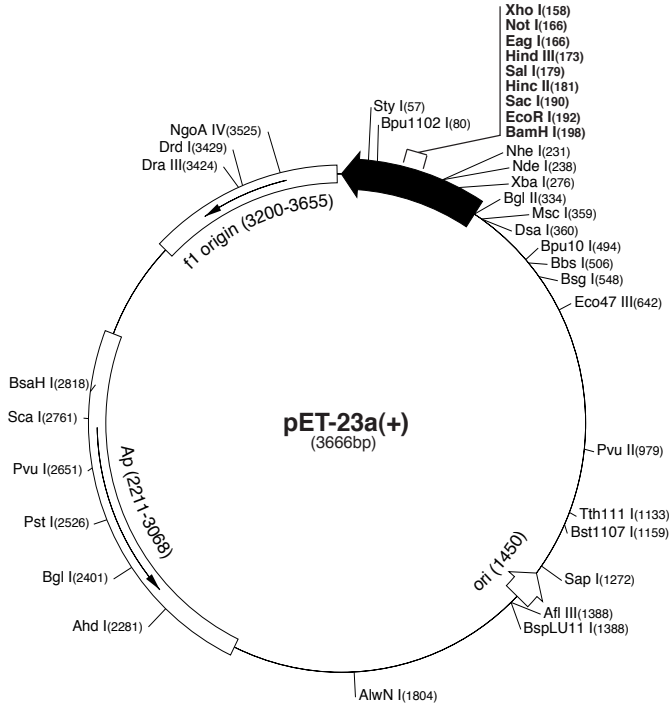
pET-22b(+) cloning/expression region

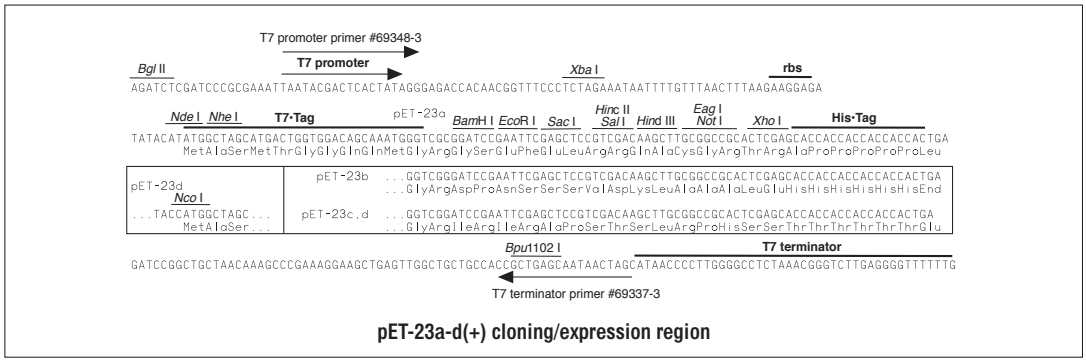
Figure S 10: Map of the pET 22b (+) plasmid vector.

pET-23a(+) sequence landmarks

T7 promoter	303-319
T7 transcription start	302
T7•Tag coding sequence	207-239
Multiple cloning sites	
(*Bam*H I - *Xho* I)	158-203
His•Tag coding sequence	140-157
T7 terminator	26-72
pBR322 origin	1450
bla coding sequence	2211-3068
f1 origin	3200-3655

The maps for pET-23b(+), pET-23c(+) and pET-23d(+) are the same as pET-23a(+) (shown) with the following exceptions: pET-23b(+) is a 3665bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-23c(+) is a 3664bp plasmid; subtract 2bp from each site beyond *Bam*H I at 198. pET-23d(+) is a 3663bp plasmid; the *Bam*H I site is in the same reading frame as in pET-23c(+). An *Nco* I site is substituted for the *Nde* I site with a net 1bp deletion at position 238 of pET-23c(+). As a result, *Nco* I cuts pET-23d(+) at 234, and *Nhe* I cuts at 229. For the rest of the sites, subtract 3bp from each site beyond position 239 in pET-23a(+). *Nde* I does not cut pET-23d(+). Note also that *Sty* I is not unique in pET-23d(+).





pET-23a-d(+) cloning/expression region

Figure S 11: Map of the pET 23a-d (+) plasmid vectors.

9.1.4. Biotinylation of HLA B*2705 and HLA A*0201 Monomers and Buffer Exchange

Biotinylation of refolded MHC with a biotinylation tag on the heavy chain was performed prior to purification of the refold. The concentrated refolding solution was filtered and desalted using PD-10 columns. This gave 7mls final desalted solution, which was incubated overnight at room temperature with 5µl BirA enzyme, 875µl Biomix A (10X concentration: 0.5 M bicine buffer, pH 8.3), 875µl Biomix B (10X concentration: 100 mM ATP, 100 mM MgOAc, 500 µM d-biotin) and 100µl d-biotin (500µM).

Size exclusion purification of HLA B*2705 and HLA A*0201 monomers was carried out following refolding (and biotinylation if necessary) using an Akta Purifier system and a HiLoad 26/60 Superdex 75 prep grade gel filtration column (GE Healthcare) in FPLC buffer (20mM Tris pH 8.0, 100mM NaCl). Refolded heterotrimers taken from this purification were concentrated to 1ml and the protein concentration measured using absorbance at 280nm on a Nanodrop (Thermo-Scientific) spectrophotometer. The purified HLA complex was then aliquoted and stored at -80°C for future tetramerisation.

9.1.5. Creation and validation of HLA B*2705 and HLA A*0201 Tetramers

Extravidin Phycoerythrin (Sigma Aldrich) or Streptavidin conjugated Allophycocyanin (Invitrogen) was added to Monomeric HLA B*2705/HLA A*0201 at a molar ratio of 4:1 (HLA: Streptavidin conjugate). Streptavidin conjugate was added in 1/10th volumes with at least 15 minutes between each addition. The progression of tetramerisation was checked with a dot-blot nitrocellulose enzyme immunoassay. 1µl of the HLA- Streptavidin conjugate mix was blotted onto a PVDF nitrocellulose membrane and left to dry. The membrane was blocked with a 4% dried milk powder, PBS solution and washed with PBS-T (PBS with 1% Tween). A 1:10000 dilution of Streptavidin HRP in ddH_20 was incubated with the membrane

for 30 minutes, and then washed as before. 1ml of each of the chemiluminescent detection reagents (Amersham ECL™ Western Blotting Detection Reagents) were added to the membrane and Hyperfilm™ ECL was exposed to the membrane. The film was then developed, the intensity of the dots on the film correlating to the amount of free biotin left in the sample.

Once tetramerisation was complete, the tetramers were incubated with BAF cells (a murine pro B-cell clone), which were either transfected with a vector producing the ILT2 receptor or remained untransfected (kind gifts from Paul Bowness, L. Lanier, UCSF, San Francisco, CA, USA). 4µl of each tetramer was incubated with both the transfected and untransfected cells (separately) in PBS with 5%FCS at room temperature for 30 minutes. The unbound tetramer was washed off and the results analysed on a Dako CyAn ADP flow cytometer at the WIMM FACS facility.

9.2. Supplementary Figures

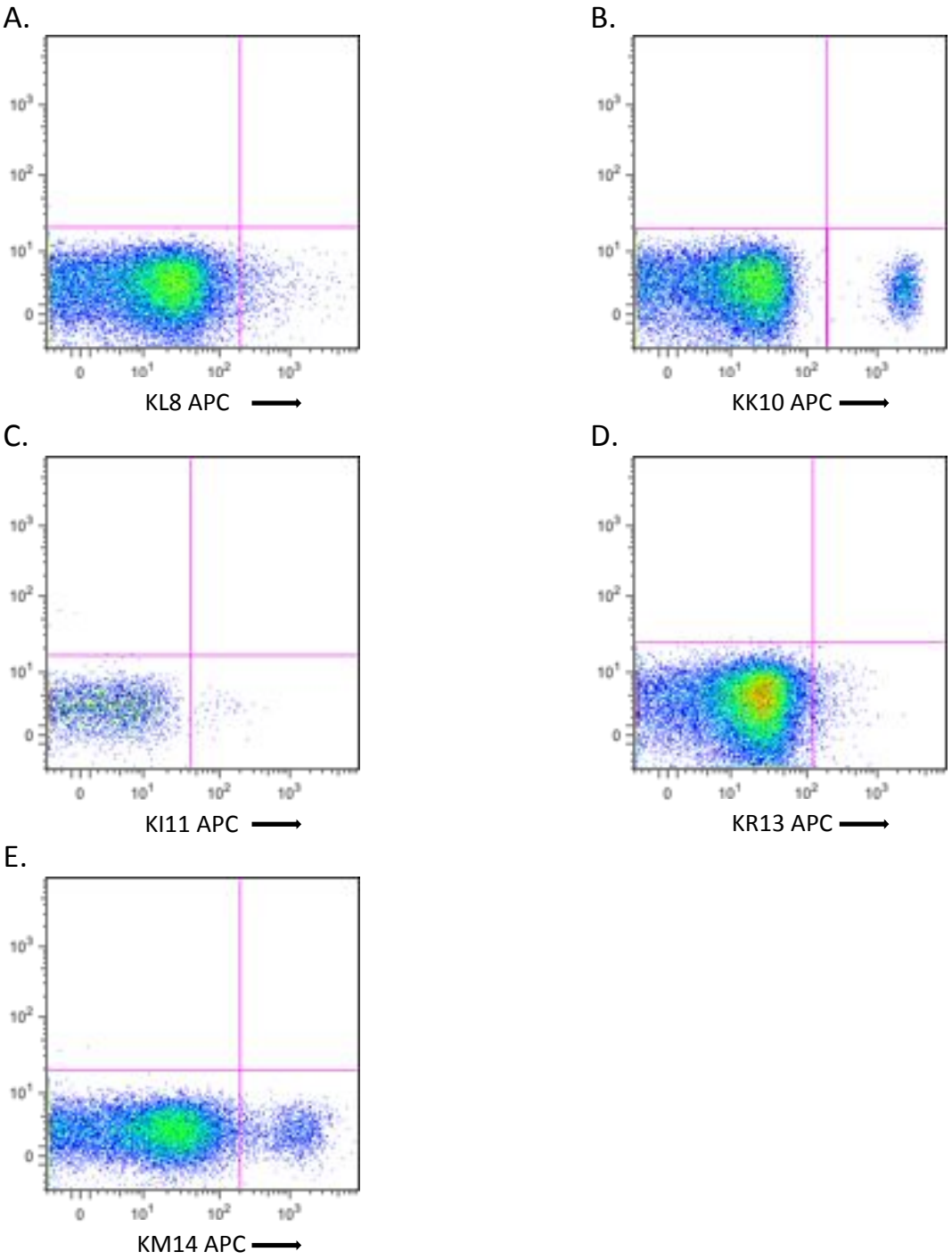


Figure S 12: Fluorescence of CD8+ T-cells from patient PBMC stained with HLA B*2705 APC-conjugated dextramer with KK10 epitope forms A. KL8 B. KK10 C. KI11 D. KR13 and E. KM14. Patients are A, B and E. 060473, C. 050875 and D. 101074. Plots are representative of positive staining for each KK10 epitope form.

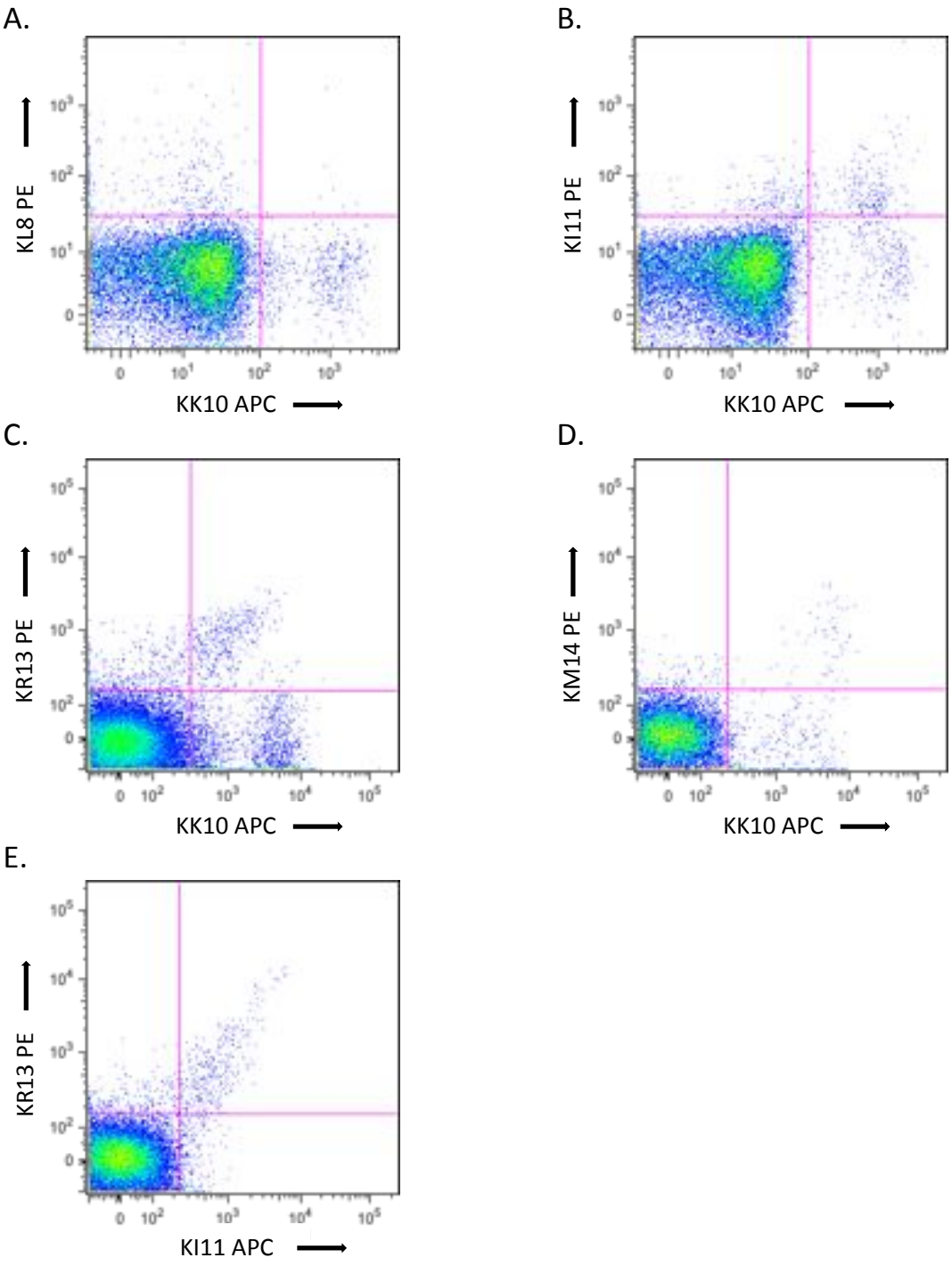


Figure S 13: APC and PE fluorescence of CD8+ T-cells from patient PBMC stained with HLA B*2705 KK10 epitope form APC-conjugated and HLA B*2705 PE conjugated dextramers A. KL8 and KK10 (030869) showing little cross-recognition. B. KK10 and KI11 (030869) with single stained and cross-recognising populations. C. KK10 and KR13 (241242) showing KK10 single and cross-recognising populations. D. KK10 and KM14 (241242) with single and cross-recognising populations and E. KI11 and KR13 (241242) showing high levels of cross-recognition. Plots represent commonly seen staining patterns for dual dextramer stains.

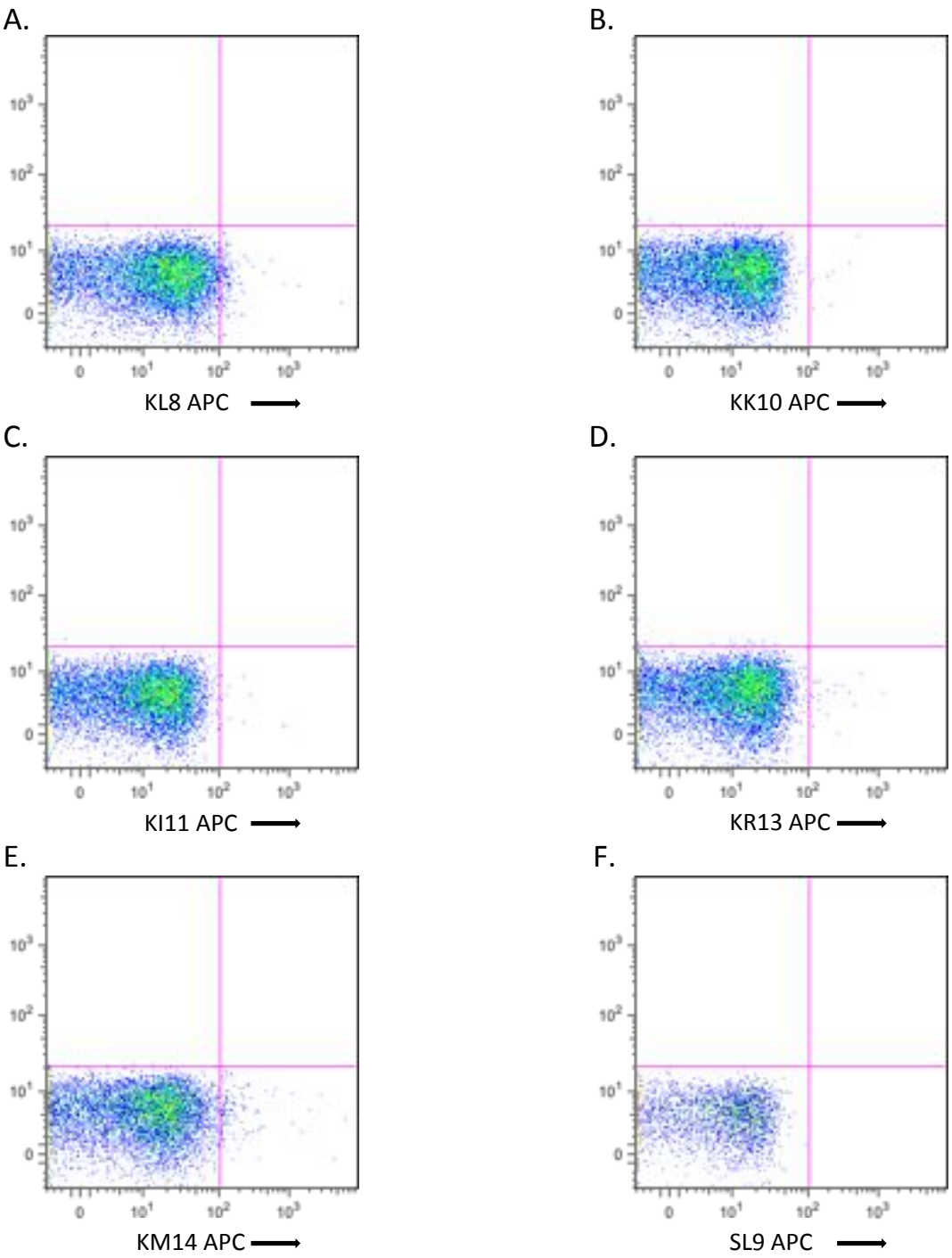


Figure S 14: Fluorescence of CD4+ T-cells from patient 111156 PBMC stained with HLA B*2705 APC-conjugated dextramer with KK10 epitope forms A. KL8 B. KK10 C. KI11 D. KR13 E. KM14 and F. HLA A*0201-SL9 as a control.

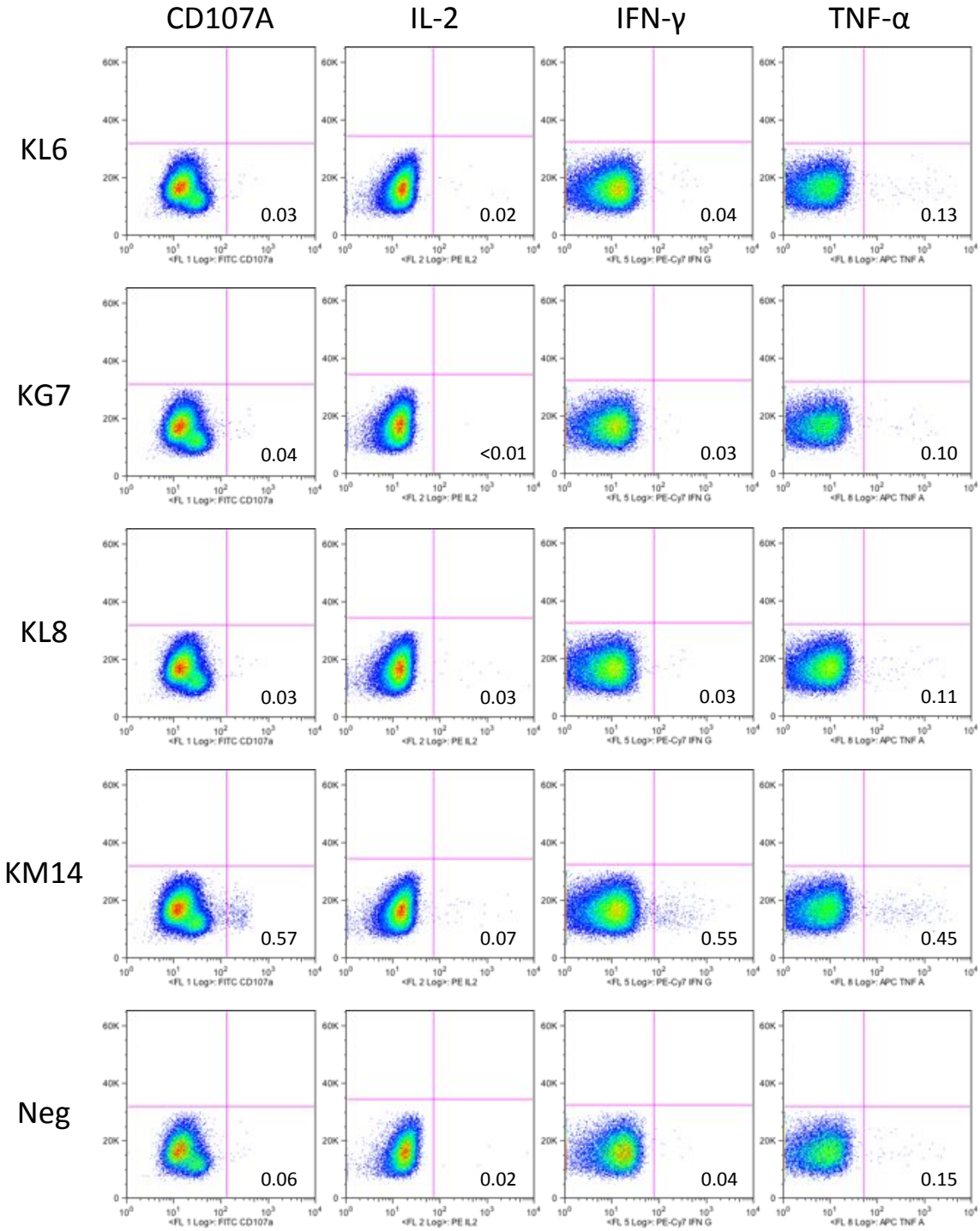


Figure S 15: Cytokine positive CD8+ T-cell populations following stimulation of patient 081077 PBMC with KK10 epitope form truncated and elongated peptides in an intracellular cytokine-staining assay. Plots show representative responses of T-cells to each peptide.

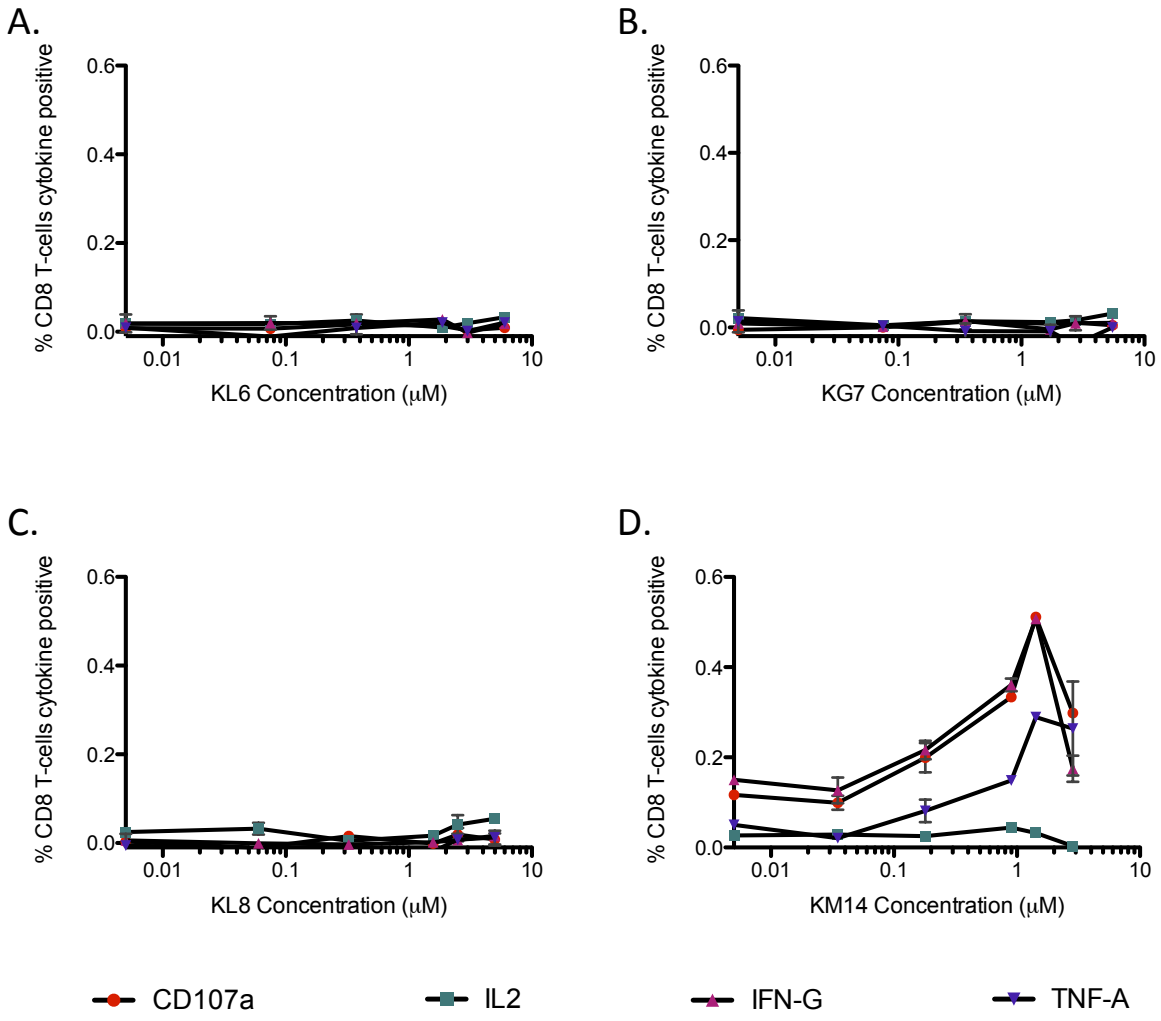


Figure S 16: Patient 081077 CD8+ T-cell cytokine responses following stimulation with increasing concentrations of KK10 epitope form peptide A. KL6 B. KG7 C. KL8 and D. KM14.

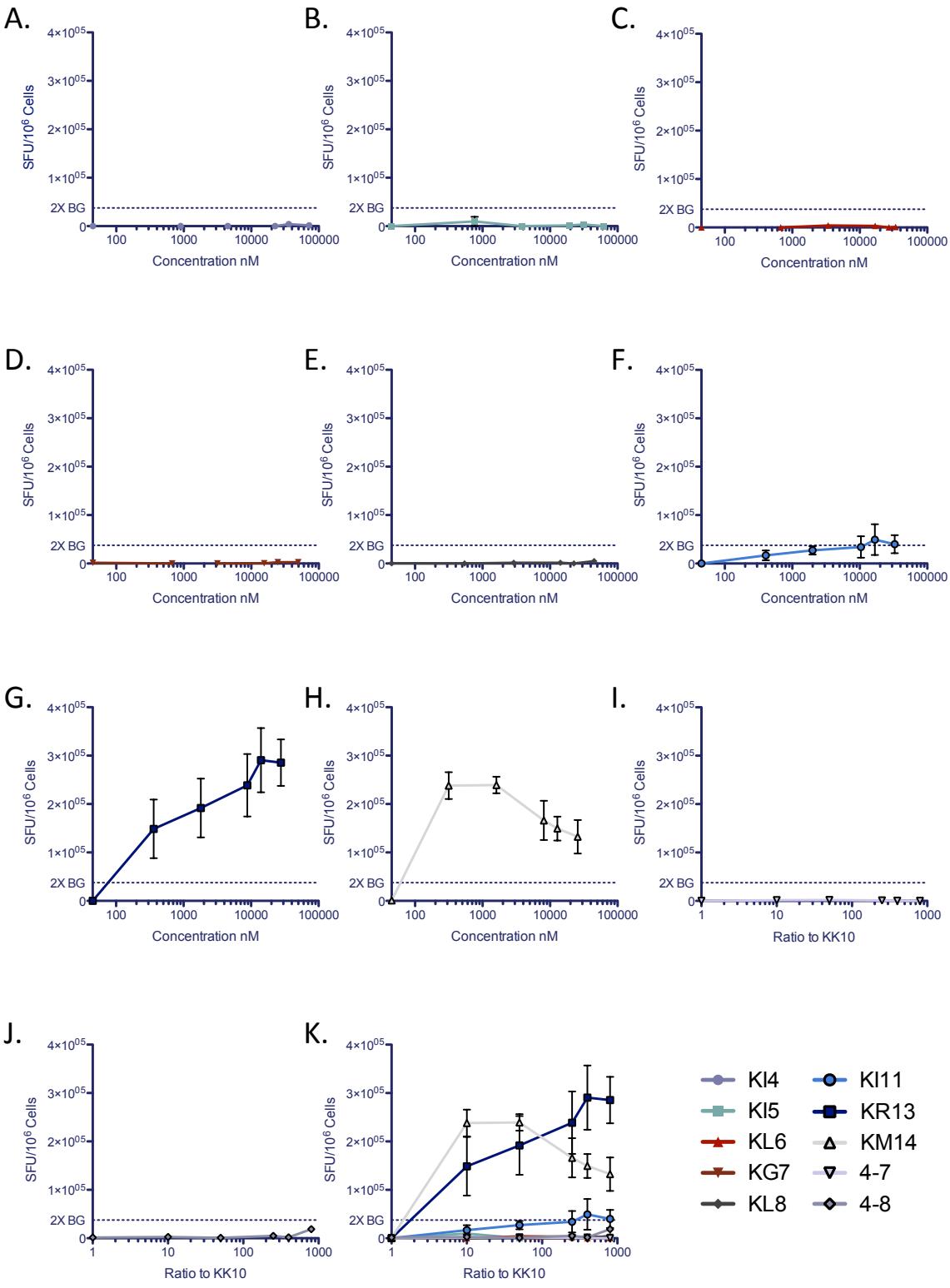


Figure S 17: IFNγ ELISpot assay for responses to KK10 epitope forms A. KI4 B. KI5 C. KL6 D.KG7 E. KL8 F. KI11 G. KR13 H. KM14 I. KI4, KI5, KL6 and KG7 J. KI4, KI5, KL6, KG7 and KL8 K. graph for all forms.

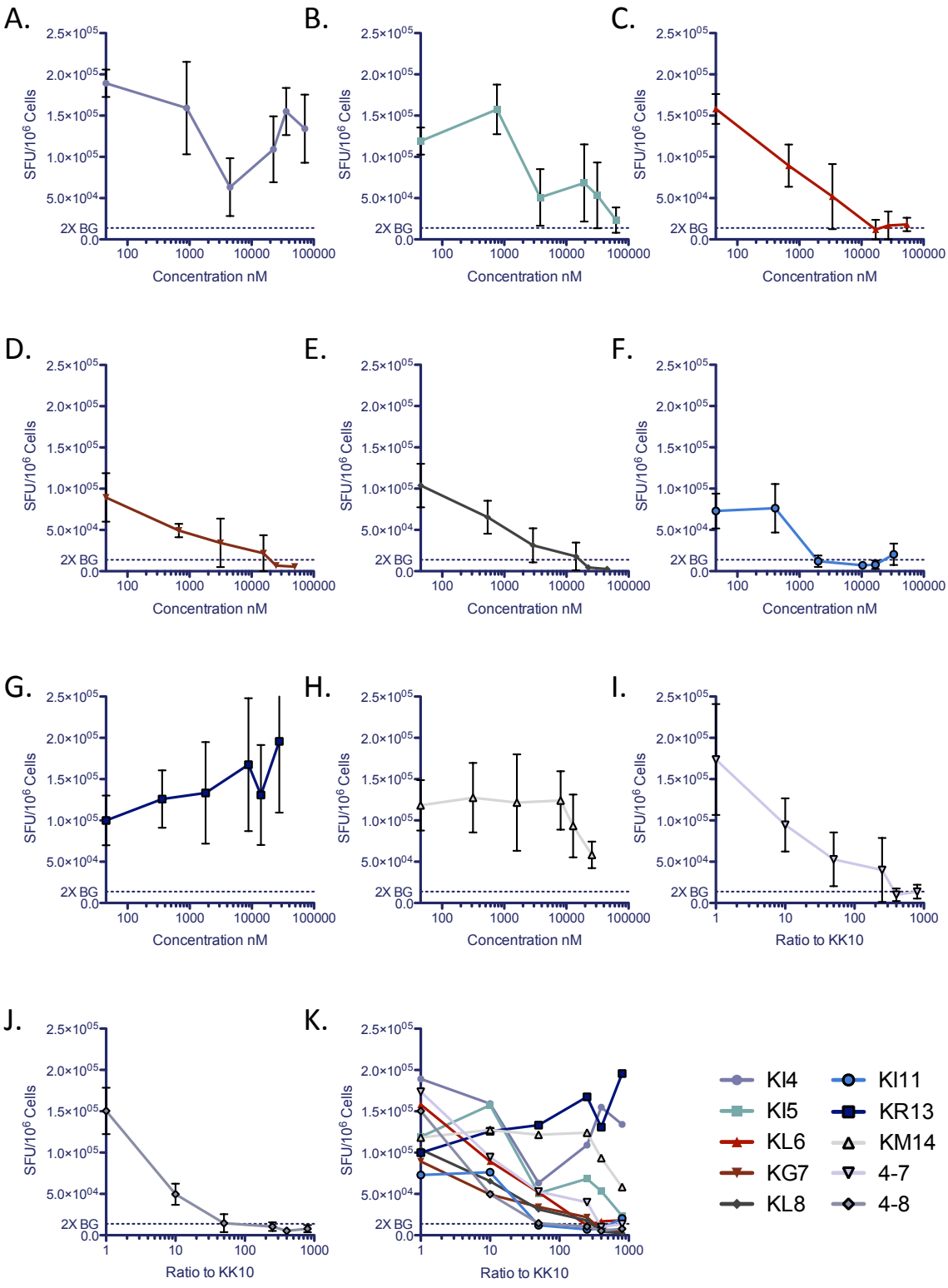


Figure S 18: IFNγ ELISpot competition assay for KK10 epitope forms versus KK10 at IC50 concentration A. KI4 B. KI5 C. KL6 D.KG7 E. KL8 F. KI11 G. KR13 H. KM14 I. KI4, KI5, KL6 and KG7 J. KI4, KI5, KL6, KG7 and KL8 K. graph for all forms.