Citation for final published version:

Pymm, Phillip, Illing, Patricia T., Ramarathinam, Sri H., O'Connor, Geraldine M., Hughes, Victoria A., Hitchen, Corinne, Price, David A. , Ho, Bosco K., McVicar, Daniel W., Brooks, Andrew G., Purcell, Anthony W., Rossjohn, Jamie and Vivian, Julian P. 2017. MHC-I peptides get out of the groove and enable a novel mechanism of HIV-1 escape. Nature Structural and Molecular Biology 24 , pp. 387-394. 10.1038/nsmb.3381

Publishers page: http://dx.doi.org/10.1038/nsmb.3381

# MHC-I peptides get out of the groove and enable a novel mechanism of HIV-1 escape

Phillip Pymm[1,2*], Patricia T. Illing[1*], Sri H. Ramarathinam[1*], Geraldine O'Connor[3], Victoria A. Hughes[1,2], Corinne Hitchen[1], David A. Price[4,5], Bosco K. Ho[1], Daniel W. McVicar[6], Andrew G. Brooks[7], Anthony W. Purcell[#1], Jamie Rossjohn[#1,2,4] & Julian P. Vivian[#1,2]

[1]Infection and Immunity Program & Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria 3800, Australia

[2]Australian Research Council Centre of Excellence in Advanced Molecular Imaging, Monash University, Clayton, Victoria 3800, Australia

[3]Department of Biological Sciences, University of Chester, Chester, CH1 4BJ, UK

[4]Institute of Infection and Immunity, Cardiff University School of Medicine, Heath Park, Cardiff, CF14 4XN, UK

[5]Human Immunology Section, Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA

[6]Cancer and Inflammation Program, National Cancer Institute-Frederick, Frederick, MD 21702, USA

[7]Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, The University of Melbourne, Parkville, Victoria 3010, Australia

* Joint first authors

[#]Joint senior and corresponding authors: Anthony Purcell (anthony.purcell@monash.edu), Jamie Rossjohn (jamie.rossjohn@monash.edu) and Julian Vivian ( julian.vivian@monash.edu)

## Abstract

**MHC-I molecules play a crucial role in immunity by capturing peptides for presentation to T-cells and natural killer cells, The peptide termini are tethered within the MHC-I, although whether other presentation modes occur is unknown. We show that 20% of the HLA-B*57:01-peptide repertoire comprises N-terminally extended sets characterised by a common motif at position (P)-1 to P2. Structures of HLA-B*57:01 presenting N-terminally extended peptides, including the immunodominant HIV-1 Gag epitope TW10 (TSTLQEQIGW), revealed the N-terminus protruded from the peptide-binding groove. The common escape mutant TSNLQEQIGW bound HLA-B*57:01 canonically, adopting a dramatically different conformation than the TW10 peptide. This impacted recognition by the killer immunoglobulin-like receptor (KIR)3DL1. We define an uncharacterised feature of the HLA-I immunopeptidome with implications for viral immune escape. We suggest a "molecular tension" model between the adaptive and innate immune systems in recognising this HIV epitope.**

## Introduction

Human leukocyte antigen class I (HLA-I) molecules are a vital component of cellular immunity, presenting endogenous peptide epitopes on the cell surface for scrutiny by $CD8^+$ T-cells and natural killer cells (NK) cells. These peptides are derived from intracellular proteins that have been proteolytically degraded and during the course of a viral infection will include epitopes originating from the invading pathogen. HLA-I molecules bind compatible peptides via a specialised groove capped by a conserved network of residues that form hydrogen (H) bonds with the N- (P1) and C-termini (PΩ) of the peptide[1]. The peptide is further anchored at specific pockets spaced along the groove, typically at the P2 and PΩ positions for many HLA allotypes [2-5]. The nature of these 'anchoring pockets' differs between HLA-I molecules thus defining their unique peptide repertoires [5].

The closed-end nature of the HLA-I antigen (Ag)-binding cleft generally constrains the repertoire of peptides to 9−11 amino acids in length. Some longer HLA-I-restricted peptides have been described, but on account of the conserved polar interaction network with the N- and C-termini in the P1 and PΩ □□□□□□□, longer peptides are canonically accommodated in the groove by adopting a 'bulged' conformation, where the central portion of the peptide epitope projects from the groove [6,7]. The T-cell receptor (TCR) binds these bulged regions of the peptide whilst simultaneously contacting the HLA-I molecule. The Killer cell immunoglobulin-like receptors (KIR), while displaying broader peptide recognition than TCR, also bind HLA in a peptide dependent manner [8-10]. Particularly important are residues at positions PΩ-1 and PΩ-2 that often directly contact the KIR [11,12]. Isolated reports have described peptides that protrude at the C-terminus from the peptide-binding groove with the PΩ-1 residue acting as an alternate C-terminal anchor, although the extent to which this represents a common occurrence remains unclear [13-15]. Such non-canonical HLA-I-peptide landscapes are important because they potentially form unique contact surfaces for TCR and killer immunoglobulin-like receptor (KIR) recognition with attendant implications for disease-relevant immune responses [14,16,17]. In contrast, it is not known whether peptides can protrude from the N-terminus of the HLA-I Ag-binding cleft.

The *HLA-B*57:01* allele is associated with delayed disease progression in HIV-1 infected indiviudals [18-21] and hypersensitivity to the anti-viral drug abacavir [22-24]. In HLA-B*57:01$^+$ HIV-1-infected individuals, robust CD8$^+$ T-cell responses specific for four epitopes within the p24 Gag protein (TW10, IW9, QW9 and KF11) are associated with such protective immunity [19,25-27]. Cellular immune responses against these epitopes drive the selection of amino acid mutations that allow viral escape from the immune system [28-30]. A common escape mutation in the p24 Gag epitope TW10 (TSTLQEQIGW; residues 240–249) is Thr242Asn, which lies at position (P)3 of the epitope (T3N) [31-33]. The emergence of this variant coincides with loss of the original CD8$^+$ T-cell response in many patients [27,32] and impaired viremic control, especially after the acquisition of compensatory mutations [30,34]. In addition to acting as a restriction element for CD8+ T cell responses, HLA-B*57:01 is also a potent ligand for the KIR3DL1 receptor that regulates NK cell function, Furthermore genetic association studies have implicated the interaction of HLA-B*57 with certain allelic variants of KIR3DL1 with delayed progression to AIDS [35-37].

Here, we show that the repertoire of endogenous and viral peptides presented by HLA-B*57:01 contains sets of both N- and C-terminally extended epitopes. A characteristic motif was identified as a signature of N-terminally extended peptides and structural studies revealed a non-canonical binding mode whereby the N-terminal residue overhangs the peptide-binding groove of HLA-B*57:01. This novel mode of peptide binding enables HIV-1 to escape immune recognition via a register shift within the TW10 epitope that alters the presented antigenic structure upon mutation to T3N.  In addition to impacting on the adaptive immune response, this mutation also impacted on KIR3DL1 recognition, thereby highlighting differences and “molecular tension” in the adaptive and innate immune systems’ recognition of this common HIV-1 epitope.

## Results

*Structure of HLA-B*57:01-TW10 in complex with KIR3DL1*

To extend our previous analysis of KIR3DL1 bound to HLA-B*57:01 presenting a self-peptide, we determined the crystal structure of KIR3DL1*001-HLA-B*57:01 in complex with the HIV-1 p24 Gag epitope TW10 (TSTLQEQIGW; residues 240–249). Crystals of KIR3DL1-HLA-B*57:01-TW10 diffracted to a resolution of 2.0 Å and the final model was refined to $R_{\text{factor}}$ and $R_{\text{free}}$ values of 20.3 % and 23.9 %, respectively (**Table 1**).

The mode of KIR3DL1 recognition of the viral epitope was essentially identical to that of the self-peptide bound to HLA-B*57:01 [12] (**Figure 1a**). The three KIR3DL1 domains engaged HLA-B*57:01 in a conserved orientation with the D0 domain docking at the side and the D1 and D2 domains docking at the C-terminal end of the peptide binding cleft, above the $\alpha1$ and $\alpha2$ helices respectively (**Figure 1a**). The D1 domain interacted with the Bw4 motif on the $\alpha1$ helix. In the course of our analysis, we noted that the TW10 epitope adopted a unique conformation not previously observed for HLA-I-peptide complexes. Namely, the N-terminal region of the TW10 epitope protruded from the Ag-binding cleft and thus did not occupy the A pocket (**Figure 1b**). Instead, the second and third residues of the TW10 epitope were located in the A and B pockets, respectively (**Figure 1b**). Comparison with available HLA-B*57:01-peptide structures [12,23,38] showed that, with the exception of a 0.7 Å shift of Trp167 (not shown), there were minimal structural distortions to the peptide-binding cleft that accompanied this previously unobserved peptide conformation (r.m.s.d. < 0.25 Å over C$\alpha$ positions 1–180). Due to the closed nature of the peptide-binding groove, the P-1 residue of the peptide projected at a right-angle from the plane of the floor of the groove, with the side-chain pointing back along the length of the peptide (**Figure 1c**). Contacts between P-1-Thr and HLA-B*57:01 were limited to water-mediated H-bonds between Trp167 and the P-1-Thr peptide backbone and van der Waals interactions between Glu63 and Leu163 and the P-1-Thr side-chain (**Figure 1c).** Contacts between the P-1Thr and the rest of the peptide included water-mediated H-bonds with P2-Thr and P4-Gln (**Figure 1c**).

The A pocket is characterised by a highly conserved set of residues including Tyr7 and Tyr171 which contribute a network of H-bonds that stabilises the N-terminus of the presented peptide. In the HLA-B*57:01-TW10 complex, the P1-Ser residue of the TW10 epitope occupied the A pocket of HLA-B*57:01 and, as such, maintained a network of H-bonds with Tyr7 and Tyr171 (**Figure 1d**). In so doing, the P1-Ser sidechain was rotated approximately 180° from the canonical positioning of a P1 anchored residue. Accordingly, the slippage of TW10 from the groove of HLA-B*57:01 was accompanied by distortion of the peptide at the P1 residue and not by structural distortion of the peptide-binding groove. As peptide repertoire data for HLA-B*57:01 had previously demonstrated the P2 anchor preference for Ser or Thr [23] it was conceivable that either P2-Ser or P3-Thr could act as N-terminal anchors for TW10. However, the high-resolution structure of HLA-B*57:01-TW10 displayed unambiguous electron density across the peptide, thereby indicating that only one conformation of TW10 was present in the crystal lattice. Further, it was confirmed by the use of hydrogen-deuterium exchange with mass spectrometry that, compared to the peptide free in solution, TW10 adopts a single conformation whilst in complex with B*57:01 (**Supplementary Figure 1**). Thus, the P-1 to P2 residues enable the peptide to extend from the peptide-binding cleft of HLA-B*57:01.

### *N-terminally extended peptides within the HLA-B*57:01 immunopeptidome*

To determine whether N-terminally extended peptides are a general feature of the HLA-B*57:01 restricted immunopeptidome, we sought to define the range of self-peptides and HIV-Gag epitopes presented by HLA-B*57:01. Peptides were isolated from C1R cells transfected with expression constructs for HLA-B*57:01 and HIV-1-Gag and sequenced using tandem mass spectrometry. This data set was: (i) filtered for peptides known to bind the endogenous HLA-I and HLA-II of the parental cell line, or bearing the peptide binding motifs of HLA-Cw4 and HLA-B*35:03 (endogenous HLA-I of C1R cells)[17]; and (ii) further augmented with previously defined peptides from the immunopeptidome of HLA-B*57:01[23], which collectively amounted to 11954 peptides (**Supplementary Table 1**). As described previously, HLA-B*57:01 ligands were predominantly 9-11 amino acids in length and showed enrichment of S/T/A at the P2 anchor, and aromatic residues at the C-terminus (**Figure 2**) [23]. In addition, a thorough bioinformatic interrogation of this expanded HLA-B57 peptidome revealed a number of peptides that contained these shorter 9–11 amino acid peptides with an N-terminal extension, and were isolated from both HIV-1-Gag$^+$ and HIV-1-Gag$^-$ cells.

Extended sets of peptides were defined based on the following criteria: (i) minimal core sequence of 7-11 amino acids and (ii) maximal sequence > 9 amino acids. A total of 1275 sets of peptides met these criteria, of which 17%, 18% and 12% of 9, 10 and 11 residue peptides fell within these criteria and possessed only N-terminal extensions, whilst far fewer were purely C-terminally extended (3%, 3% and 2% respectively) (**Figure 2a**). To define the sequence features that were predisposed to N-terminal extensions, peptides exhibiting N-terminal heterogeneity were aligned based on the minimal sequence P1 (**Supplementary Table 2**). The motif was visualised as a sequence logo encompassing P-1 (1 residue N-terminal of P1) to P3 using iceLogo [39]. A similar motif was also generated from the

P1 to P3 of all 9–11 residue peptides within the HLA-B*57:01 peptide data set and aligned at P1 for comparison (**Figures 2b and 2c**). Whilst the global HLA-B*57:01 9-11 amino acid peptide motif shows enrichment of Lys/Ile/Arg at P1, the N-terminally extended peptides were enriched for Ser/Thr/Ala at P1 and P-1. Given that Ser/Thr/Ala are preferred P2 anchors for HLA-B*57:01, these residues could potentially be acting as alternate anchor sites in longer peptides, resulting in bulged conformations within the peptide-binding groove. However given the appearance of numerous sets where a minimal core sequence was found with several extension lengths (e.g. Set 233: STTSVASSW, TSTTSVASSW, DTSTTSVASSW, HDTSTTSVASSW, SHDTSTTSVASSW, SSHDTSTTSVASSW and TASSHDTSTTSVASSW) not all of which contained Ser/Thr/Ala at the second residue, it seemed likely that the excess residues might overhang the cleft. Thus, we defined the N-terminal protrusion motif as Ser/Thr/Ala at P-1 to P2.

### *The TW10 epitope is part of a N-terminally extended set of peptides restricted by HLA-B*57:01*

Of the 11954 peptides used in this analysis 9 were derived from the Gag polypeptide. Amongst the Gag derived epitopes, two N-terminally extended epitope sets were observed, QW9 and TW10 (**Figure 3a**, Set 33 and Set 500 **Supplementary Table 2**). The TW10 epitope was the minimal peptide observed within its N-terminally extended set that contained multiple extension lengths (**Figure 3a**). To determine the relative abundance of each peptide in the set, we measured the extracted ion chromatogram specific for each of the peptides (**Figure 3b**). The N-terminally extended AGTTSTLQEQIGW peptide was the most intense ion and by inference the most abundant, followed by TTSTLQEQIGW, GTTSTLQEQIGW and TW10 (**Figure 3b**). Two longer N-terminally extended peptides were also identified (DIAGTTSTLQEQIGW and SDIAGTTSTLQEQIGW) at significantly lower levels (**Figure 3b**). Taken together, our data suggest that a common motif characterises N-terminally protruding HLA-B*57:01-restricted peptides regardless of self or viral origin.

### *Structure of HLA-B*57:01 in complex with N-terminally extended self peptides*

To further probe the structural nature of the N-terminal extension motif we investigated three epitopes from the self-repertoire of HLA-B*57:01. The selected peptides were TSTTSVASSW (UNP Q14679), TSTFEDVKILAF (UNP Q6YHU6) and SSTRGISQLW (UNP A8CG34) each of which formed part of a nested set (Set 233, Set 926 and Set 736, respectively; **Supplementary Table 2**). Crystals of the refolded HLA-B*57:01-peptide complexes formed in the space group $P2_12_12_1$ with unit cell dimensions of ~$a$ = 50, $b$ = 82, and $c$ = 110 Å, and diffracted to high resolution (**Table 1**).

All these HLA-B*57:01-peptide complexes adopted the non-canonical conformation observed for HLA-B*57:01-TW10. Namely, the Ser residue in the second position of the TSTFEDVKILAF

(Figure 4a), TSTTSVASSW, and SSTRGISQLW peptides occupied the A pocket, which normally accommodates the N-terminal amide. As these self-peptide complexes were highly similar we restricted analysis to the HLA-B*57:01-TSTFEDVKILAF structure. The water mediated H-bonding network to the Thr at position P-1 to residues Asn66 and Trp167 of HLA-B*57:01 was highly conserved (**Figures 1c and 4b**). Similarly, the H-bonding network at the N-terminus of the peptide-binding groove was similarly conserved, with the hydroxyl of the P1-Ser maintaining the contacting Tyr7, Tyr59, Glu63 and Tyr171 (**Figures 1d and 4c**). Accordingly, the P-1 to P2 motif identified in the repertoire analysis is a common motif that enables the peptide to protrude from the N-terminus of HLA-B*57:01.

### *The structural basis of viral escape via the T3N mutation in TW10*

The TW10 epitope undergoes a common and rapid substitution at the N-terminus, namely T3N (TSNLQEQIGW), to facilitate HIV-1 immune escape in HLA-B*57:01$^+$ individuals. To address this phenomenon at the molecular level, we determined the crystal structure of HLA-B*57:01 in complex with the T3N peptide and KIR3DL1 (**Table 1**). Overall, the contacts between KIR3DL1 and HLA-B*57:01 were conserved between the TW10 and T3N structures. However, contacts between KIR3DL1 and the HLA-bound peptide were markedly different for TW10 and T3N. KIR3DL1 formed two contacts with the T3N peptide. The D1 Gly116 bound P7-Gln, whilst Tyr200 at the D1/D2 hinge-loop bound P8-Ile (**Figure 5a**). In contrast, KIR3DL1 did not directly contact the TW10 peptide. Instead, water-mediated contacts were formed between Tyr200 and the backbones of P8-Gly and P9-Trp (**Figure 5b**). KIR3DL1 therefore interacted differently with TW10 and T3N as a result of changes in the bound peptide conformation (detailed below).

The Thr3Asn substitution caused the peptide to adopt a conventional anchoring mode bound to HLA-B*57:01 with the A and B pockets occupied by the Thr and Ser in the first two peptide positions respectively (**Figure 5c**). Notably, the HLA-B*57:01-TW10 complex was more stable than the HLA-B*57:01-T3N complex, with a melting temperature of 61.8°C compared with 55.0°C in circular dichroism assays (**Supplementary Figure 2**). The inability of Asn3 to occupy the B pocket is consistent with previous HLA-B*57:01 peptide elution data in which Asn was not observed as an anchor residue at position 2 [23]. The change in anchoring at P2 shifted the peptide register across all positions except the C-terminal anchor and P9-Gly (**Figure 5d**). Accordingly, the secondary anchor positions also differed between TW10 and T3N. Specifically, the TW10 secondary anchors at P3-Leu, P6-Gln and P7-Ile were reassigned to P3-Asn, P5-Gln and P6-Glu, respectively, in T3N. Consequently, the P-1-Thr, P4-Gln and P5-Glu side-chains were considerably solvent exposed in TW10, whereas P4-Leu, P5-Gln and P8-Ile were considerably solvent exposed in T3N (**Figure 5d**).

Thus, the Thr3Asn mutation shifts the peptide register within the peptide-binding groove, resulting in the presentation of an altered peptide conformation to immune cell receptors.

*Escape mutants in the TW10 epitope attenuate KIR3DL1 recognition*

To determine whether the T3N escape mutation affects binding to KIR3DL1, we used surface plasmon resonance to measure the affinity of KIR3DL1*001 for HLA-B*57:01-TW10 and HLA-B*57:01-T3N. The TW10 complex bound more strongly to KIR3DL1 ($K_D \sim 60$ μM) compared with the T3N complex ($K_D \sim 100$ μM) (**Figure 6a**). Fluorochrome-labelled tetrameric forms of HLA-B*5701-TW10 and HLA-B*5701-T3N were then used to stain HEK293T cells transfected with prevalent KIR3DL1 allotypes (*001, *005 and *015) (parts of the data for TW10 have been published previously[40]). The T3N mutation almost entirely abrogated HLA-B*57:01 tetramer binding to all three surface-expressed allotypic variants of KIR3DL1 (**Figure 6b**). However, some differences were observed between allotypes, with KIR3DL1*015 being the most sensitive to changes in the epitope.

The relative contribution of specific KIR3DL1*001 residues to the HLA-B*57:01-TW10 and HLA-B*57:01-T3N interfaces was further probed by mutational analysis. Overall, there were distinct differences in the recognition of these HLA-B*57:01-peptide complexes (**Figure 6c, d**). Starting with the D0 domain, the T3N complex was comparatively less sensitive than the TW10 complex to substitution at Phe9, but more sensitive to substitution at Phe34 (**Figure 6c, d**). In the D1 domain, substitutions at Gly138 and Ser140 dramatically altered recognition of the TW10 complex, yet minimally affected recognition of the T3N complex, whilst the Leu166Ala mutation was sufficient to restore T3N binding to wild-type (TW10) levels (**Figure 6c, d**). Substitutions across the D2 domain indicated similar contributions from Tyr200, Glu201, Ser227, Ser228, Phe276, His278 and Ser279. In contrast, the Arg277Ala substitution abrogated binding to the T3N complex whilst improving recognition of the TW10 complex.

Taken together, these analyses demonstrate that the viral escape mutant T3N attenuates KIR3DL1 recognition in an allotype-dependent manner. Furthermore, the KIR3DL1 residue Leu166, which sat proximal to the interface with the peptide in the ternary structure, plays a central role in this effect. The altered peptide conformation of T3N therefore limits the interaction despite the presence of additional direct contacts with KIR3DL1. Accordingly, whilst the HIV escape mutants within the TW10 epitope evade T-cell recognition, these same substitutions can impair the interaction with the prototypical inhibitory receptor KIR3DL1.

## Discussion

HLA-II molecules possess open-ended Ag-binding clefts that allow peptides to extend from the groove at both the N- and C-termini. In contrast, the N- and C-terminal pockets of HLA-I molecules are closed, thereby restricting the bound peptides to a typical length of 9-11 amino acids. Nevertheless, longer MHC-I-restricted epitopes have been described that adopt a centrally bulged conformation whilst maintaining the N- and C-terminal anchor residues. In addition, there have been isolated reports of peptides with residues overhanging the C-terminal pocket of HLA-A*02:01 [14,41] and H2-M3 [42]. It has been previously proposed that N-terminally extended peptides could be presented[43,44], yet whether such peptides could overhang the more buried N-terminal pocket or whether they were centrally bulged was unclear.

Our comprehensive analysis of the repertoire of peptides naturally presented by HLA B*57:01, shows that approximately 27% are clustered into extended-sets, most of which are solely N-terminally extended (20 %) extended, whilst the others are either C-terminally extended or both N- and C-terminally extended. We also demonstrate that N-terminally extended peptides are accommodated via protrusion from the groove. From the amino acid enrichment data for N-terminally extended HLA-B*57:01 epitopes and our structure-based insights, we can devise the following rules that predict the P-1 overhang. Namely: (i) the peptide should be a minimum of 1 residue longer than the minimal epitope length that can be presented; (ii) the P2 residue should be a residue that favours binding in the B pocket and (iii) the P1 residue should be smaller than leucine to enable a 180° rotation within the A pocket. Notably, none of these criteria are specific to HLA-B*57:01 and thus we speculate that N-terminally protruded peptides will be a common feature of immunopeptidomes. Accordingly, we define a new mechanism that allows MHC-I molecules to present lengthy epitopes.

Viruses have developed numerous strategies to evade HLA-mediated immune detection, including mechanisms that directly interfere with the presentation of viral epitopes and HLA molecules on the cell surface [45,46]. For example, HIV-1 can increase cellular endocytosis of HLA molecules via *nef* [47], limit HLA transcription and peptide processing via *tat* [48] and suppress TAP-mediated peptide transport into the endoplasmic reticulum [49]. Under selection pressure from the immune system, HIV-1 also mutates rapidly to evade cytotoxic T lymphocyte responses. These epitope-centric escape mutations typically alter the conformation of exposed residues that interact with the T-cell receptor [50-52] or abrogate peptide presentation via effects on antigen processing or HLA-I binding [53,54]. Here, we report a novel mode of escape whereby the common T3N mutant exploits the P-1 overhang of TW10 and the anchor residue preferences in the A and B pockets to shift the register of the peptide within HLA-B*57:01. This structural rearrangement alters the conformation of exposed peptide residues to enable immune escape [31-33,55].

There has been significant interest in the idea that KIR-dependent recognition of HLA-I can shape the nature of viral escape mutations, with some studies suggesting that inhibitory receptors expressed by NK cells may exert some selective pressure during HIV infection [56,57]. While immune recognition and the subsequent immune escape observed for the TW10 epitope is likely a predominantly T-cell mediated phenomenon, it is plausible that this epitope also impacts upon NK cell recognition. Such effects on NK cell recognition would be accentuated by the viral modulation of the peptide repertoire through the inclusion of viral peptides, which have been observed to comprise up to 50% of the immunopeptidome in the early stages of vaccinia infection[58], and alterations to self-peptide presentation. However, it is unclear whether the T3N escape variant actually impacts NK cell inhibition via engagement of KIR3DL1. A previous study using surface plasmon resonance reported a reduced interaction with T3N presented by HLA-B*57:03 [33]. In contrast, another study based on tetramer staining of KIR3DL1-transfected cell lines reported no interaction [59]. Although this discrepancy likely reflects the different experimental approaches, it is clear from our structural, surface plasmon resonance and tetramer staining data that KIR3DL1 recognises T3N albeit with a lower affinity relative to TW10 in the context of HLA-B*57:01. There are also well-described C-terminal escape mutations within the TW10 epitope, such as G9D, that abrogate KIR3DL1 recognition [40]. Notably, these effects are KIR3DL1 allotype-dependent, consistent with the findings of a recent study [60]. This is in-line with recent reports demonstrating allotypic variation in the capacity of KIR3DL1 to bind defined HLA-I/peptide complexes[40,60]. Accordingly, there appears to be a "molecular tension" between the inhibitory germline-encoded and activating rearranging receptors of the innate and adaptive arms of the immune system as they converge to target a common HIV-1 determinant and its escape variants presented by HLA-B*57:01. In summary, our structural and proteomics analyses reveal a novel mechanism of viral immune escape, whereby HIV-1 mutates to exploit a previously undescribed mode of peptide presentation by MHC-I.

## Methods

### *Isolation of MHC-I peptide ligands*

C1R HLA B*57:01 transfectants as generated for[38] were grown to high density in RPMI 1640 (Thermo Fisher Scientific, Waltham, MA) supplemented with 10 % fetal calf serum (FCS; Bovogen Biologicals Pty. Ltd., Melbourne, Australia), 7.5 mM HEPES (MP Biomedicals, Eschwege, Germany), 150 µg/mL streptomycin (Sigma-Aldrich, St. Louis, MO), 150 U/mL benzylpenicillin (CSL, Parkville, Australia), 2 mM L-glutamine (MP Biomedicals), 76 µM β-mercaptoethylamine (Sigma-Aldrich) and 150 µM non-essential amino acids (Life Technologies, Carlsbad, CA). The Gag plasmid was a kind gift from Johnson Mak (Deakin University, Melbourne Australia) and introduced into cell lines by electroporation. Maintenance of Gag expression in these transfectants was assayed by western blot using an anti-gag antibody generated by Johnson Mak (Deakin University, Melbourne Australia). Transfectants were maintained under hygromycin (0.3 mg/mL, HLA-B*57:01) and G418 (0.5 mg/mL,Gag) selection. Cells were tested for mycoplasma contamination at regular intervals in house and for continued expression of HLA B*57:01 using the 3E12 antibody[61]. Cells were harvested by centrifugation (1200 g, 20 min, 4 °C) and snap frozen in liquid nitrogen. Clarified lysates were generated from 5 x $10^9$ cells using a combination of cryogenic milling and detergent-based lysis. HLA-peptide complexes were immunoaffinity-purified from cell lysates using the W6/32 monoclonal antibody in solid phase as described previously [62,63]. Bound complexes were eluted by acidification with 10 % acetic acid and fractionated using a 4.6 mm internal diameter x 100 mm long monolithic reversed-phase (RP) C18 high performance liquid chromatography (HPLC) column (Chromolith SpeedROD; Merck Millipore, Darmstadt, Germany) with the ÄKTAmicro™ HPLC system (GE Healthcare, Little Chalfont, UK). The mobile phase consisted of buffer A (0.1 % trifluoroacetic acid; Thermo Fisher Scientific) and buffer B (80 % acetonitrile, 0.1 % trifluoroacetic acid; Thermo Fisher Scientific). HLA-peptide mixtures were loaded onto the column at a flow rate of 1 mL/min with separation based on a B gradient of 2−40% for 4 min, 40−45% for another 4 min and a final rapid 2 min increase to 100%. Fractions (500 µL) were collected, vacuum-concentrated to 10 µL and diluted in 0.1% formic acid to reduce the acetonitrile concentration.

### *Liquid chromatography-tandem mass spectrometry (LC-MS/MS) sequencing of MHC-I-bound peptides*

For LC-MS/MS acquisition, peptide-containing fractions were loaded onto a microfluidic trap column packed with ChromXP C18-CL 3 µm particles (300 Å nominal pore size; equilibrated in 0.1% formic acid/ 2% acetonitrile) at 5 µL/min using a NanoUltra cHiPLC system (Eksigent). An analytical (75 µm x 15cm ChromXP C18-CL 3µm, 120A, Eksigent) microfluidic column was switched in line and peptides separated using linear gradient elution 0-30% buffer B (80% acetonitrile, 0.1% formic acid) over 50 min and 30-80% over 5 min flowing at 300 nL/min. Separated peptides were analysed using an AB SCIEX 5600+ TripleTOF mass spectrometer equipped with a Nanospray III ion source and accumulating up to 20 MS/MS spectra per second. The following instrument parameters were used: ion spray voltage (ISVF) was set at 2400 V, curtain gas (CUR) at 25 L/min, ion source gas (GS1) at 10 L/min and an interface heater temperature (IHT) setting of 150 ºC. MS/MS switch criteria included ions of m/z >200 amu, charge state +2 to +5, intensity >40 cps and the top 20 ions meeting this criterion were selected for MS/MS per cycle. The instrument was calibrated every four LC runs using [Glu1]-Fibrinopeptide B standard as per manufacturer’s instructions.

LC-MS/MS data was searched against the human proteome (UniProt/SwissProt v2014_10) and the HIV NL4-3-gag and AD8-env sequence using ProteinPilot™ software (version 4.5, SCIEX). A 5% FDR cut-off was applied. Peptides known to bind the endogenous MHCI of C1R cells (HLA-C*04:01 and HLA-B*35:03) [17] and those identified as binders of MHCII in similar MHCII isolations in the laboratory and commonly observed contaminants were removed from the data set prior to subsequent analysis of HLA-B*57:01 peptide ligands.

*Analysis of the HLA peptidome for extended sets*

The final combined data set used to estimate the prevalence of N-terminally extended HLA-B*57:01 ligands and characterise the extension motif (**Supplementary Table 1**) contained 11954 unique peptides. 2746 were identified as constitutive HLA-B*57:01 ligands in studies that examined the impact of abacavir on the HLA-B*57:01 peptidome[23]. In this study, 8432 peptides were eluted from same parental CIR cell line in the absence of the HIV Gag antigen, and 8233 from HIV Gag transfected cells. Extended sets were identified through an iterative process of examining longer peptides for encompassment of shorter peptides. Sets containing minimal core sequences of 7-11 amino acids consistent with canonical class I ligands and maximal sequences greater than 9 amino acids were defined as extended sets. Motifs showing the enrichment (compared to the human proteome) of specific amino acids at defined locations within the peptide ligands were generated using iceLogo stand alone version using the static reference method[39]. Letter height corresponds to the difference in frequency of the amino acid compared to the human proteome. Only significantly regulated amino acids, those for which the z-score falls outside the confidence interval for a p value of 0.05, are shown.

*Protein expression and purification*

The HLA-B*57:01 and β2-microglobulin genes were sub-cloned into the pET-30 expression vector and expressed separately into inclusion bodies in *E. coli* prior to refolding and purification as described previously [64]. Briefly, HLA-B*57:01 was refolded by rapid dilution in a solution containing 100 mM Tris-HCl pH 8.0, 400 mM L-arginine-HCl, 5 mM reduced glutathione, and 0.5 mM oxidized glutathione in the presence of β2-microglobulin the appropriate peptide. The refolded HLA-B*57:01 complex was purified by anion exchange and size exclusion chromatography.

KIR3DL1*001 (residues 1 – 299) was sub-cloned into the pHLSec mammalian expression vector with N-terminal 6xHis and secretion tags. KIR3DL1 was expressed and secreted from transiently transfected HEK 293S cells and harvested from the culture media after three by nickel affinity and gel filtration chromatography using an S200 16/60 column (GE Healthcare) in 10 mM Tris pH 8.0, 300 mM NaCl. Purified KIR3DL1 was then concentrated to 15 mg/mL and deglycosylated with endoglycosidase H (New England Biolabs, Ipswich, MA). The extent of deglycosylation was monitored by SDS-PAGE prior to crystallization trials. For surface plasmon resonance studies, a similar construct of KIR3DL1*001 was prepared in the pFastBac vector and expressed from Hi-5 insect cells (Invitrogen, Carlsbad, CA). KIR3DL1 was purified as above excluding the endoglycosidase H deglycosylation step.

*Hydrogen/deuterium exchange and analysis by LC-MS.*

A sample containing 5 µg (0.1 nmol) of peptide or protein complex containing peptide was diluted 24-fold with 50 mM Tris and 50 mM NaCl dissolved in $D_2O$ (Cambridge Isotope Laboratories) at 25 °C to label the sample. The deuteration reaction was quenched at 10 seconds by adding an equal volume of 100 mM $NaH_2PO_4$ (pH 2.4) and quickly frozen in a dry ice−ethanol bath. The frozen sample was quickly thawed and immediately injected onto a micropeptide trap column connected to a C18 HPLC column coupled to a Bruker Micro quadrupole time of flight mass spectrometer. The HLA-bound peptide was separated using a 12 min gradient of 10−45% acetonitrile at a flow rate of 50 µL/min. The micropeptide trap and C18 HPLC column were immersed in ice to minimize back exchange. Because the mass of a peptide increases by one for every amide hydrogen atom exchanged with deuterium, the amount of deuterium can be determined by comparing the mass of the labelled peptide with the mass of the same peptide without the label. The centroid mass of each peptide was determined using the software package MagTran [65]. Data are representative of three independent experiments.

*Crystallisation, data collection, structure determination and refinement*

The peptide sequences crystallised in complex with HLA-B*57:01 are summarized in Table 2. HLA-B*57:01 binary and ternary complexes with KIR3DL1*001 were concentrated to ~ 10 mg/mL and crystallized at 294 K by the hanging-drop vapour-diffusion method. Binary complexes (TSTTSVASSW, TSTFEDVKILAF and SSTRGISQLW) crystallised from a solution comprising 12 -20 % PEG 4000, 0.2 M ammonium acetate and 0.1 M tri-sodium citrate pH 5.4 – 5.6. Ternary complexes (KIR3DL1-TW10 and KIR3DL1-T3N) crystallised from a solution comprising 16% PEG 3350, 2% tacsimate pH 5.0 and 0.1M tri-sodium citrate pH 5.6. Prior to data collection, crystals were equilibrated in reservoir solution with 10% glycerol added as a cryoprotectant and then flash-cooled in a stream of liquid nitrogen at 100 K. Data sets were collected at the MX2 beamline (Australian Synchrotron, Victoria). The data were recorded on a Quantum-315 CCD detector and were integrated and scaled using MOSFLM and SCALA from the CCP4 program suite [66-68]. Details of the data processing statistics are summarised in Table 1. The crystal structures were solved by molecular replacement, as implemented in PHASER [69] with HLA-B*57:01-LF9 used as the search model (Protein Data Bank accession number: 2RFX). Refinement of the models proceeded with iterative rounds of manual building in COOT [70] and refinement in PHENIX [71]. The structures were validated with MOLPROBITY [72]. Refinement statistics are summarised in Table 1. Coordinates and structure factors were deposited in the Protein Data Bank under accession numbers (TSTTSVASSW 5T6X, TSTFEDVKILAF 5T6Y, SSTRGISQLW 5T6W, KIR3DL1-TW10 5T6Z and KIR3DL1-T3N 5T70).

*Surface plasmon resonance*

Surface plasmon resonance experiments were conducted at 298 K on a Biacore 3000 instrument using HBS buffer (10 mM HEPES-HCl pH 7.4, 150 mM NaCl and 0.005% surfactant P20). The W6/32 antibody was immobilised on a CM5 chip via amine coupling to capture HLA-peptide complexes, creating a surface density of approximately 700 response units (RU). KIR3DL1*001 (2.37–300 μM) was injected over the chip at a flow rate of 5 μL/min. The response to W6/32 alone was subtracted from the response to KIR3DL1*001-HLA-B*5701-peptide. Equilibrium data were analysed using GraphPad Prism. All data are representative of two independent experiments. Error bars represent standard error.

*Transfection studies*

FLAG-tagged KIR3DL1*001 was cloned into the pEF6 vector. Specific nucleotide residues were mutated using a QuikChange II Site Directed Mutagenesis Kit (Stratagene), and constructs were introduced into HEK293T cells using FuGene® 6 Transfection Reagent (Roche). After 48 hr, the cells were harvested and stained with anti-FLAG antibody (clone M2, Sigma-Aldrich) or tetramer for 30 min at 4 °C. The cells were then washed and analysed on a Fortessa flow cytometer (BD Biosciences). Staining with the anti-FLAG antibody showed that none of the introduced mutations substantially affected cell surface expression of KIR3DL1*001 (data not shown). KIR3DL1 mutant data are representative of three independent experiments. KIR3DL1 allotype data are representative of four independent experiments, data were not analysed if FLAG expression was present on >5% of cells, giving a minimum of two replicates in this assay. Error bars represent standard error of the mean. HEK293T cells were regularly tested and maintained mycoplasma free for the duration of these studies.

## ACKNOWLEDGEMENTS

This work was supported by project grants from the National Health and Medical Research Council of Australia (NH&MRC, APP1063829) and Australian Research Council (ARC, DP150104503).

A.W.P. is a NH&MRC Senior Research Fellow. P.T.I. is a NH&MRC Early Career Fellow. S.H.R. is the recipient of an Australian Postgraduate Award. D.A.P. is supported by a Wellcome Trust Senior Investigator Award. J.R. is supported by an ARC Laureate Fellowship. This research was undertaken in part on the MX2 beamline at the Australian Synchrotron, Victoria, Australia.

## REFERENCES

1. Saper MA, B.P., Wiley DC. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 A resolution. *Journal of Molecular Biology* **219**, 277-319 (1991).
2. Rammensee HG, F.T., Stevanoviíc S. MHC ligands and peptide motifs-first listing. *Immunogenetics* **41**, 178-228 (1995).
3. Deres K, B.W., Faath S, Jung G, Rammensee HG. MHC:peptide binding studies indicate hierarchy of anchor residues.pdf. *Cellular Immunology* **151**, 158-167 (1993).
4. Wilson IA, F.D. Structural analysis of MHC class I molecules with bound peptide antigens. *Seminars in Immunology* **5**, 75-80 (1993).
5. Garrett TP, S.M., Bjorkman PJ, Strominger JL, Wiley DC. Specificity pockets for the side chains of peptide antigens in HLA-Aw68. *Nature* **342**, 692-696 (1989).
6. Speir JA, S.J., Joly E, Butcher GW, Wilson IA. Two Different, Highly Exposed, Bulged Structures for an Unusually Long Peptide Bound to Rat MHC Class I RT1-A. *Immunity* **14**, 81-92 (2001).
7. Tynan, F.E. et al. 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. *J Biol Chem* **280**, 23900-9 (2005).
8. Malnati MS, e.a. Peptide specificity in the recognition of MHC class I by natural killer cell clones. *Science* **267**, 1016-1018 (1995).
9. Peruzzi M, P.K., Long EO, Malnati MS. Peptide Sequence Requirements for the Recognition of HLA-B*2705 by Specific Natural Killer Cells. *J Immunol* **102**, 13224-13229 (1996).
10. Stewart-Jones GB, d.G.K., Kollnberger S, McMichael AJ, Jones EY, Bowness P. Crystal structures and KIR3DL1 recognition of three immunodominant viral peptides complexed to HLA-B*2705. *Eur J Immunol* **35**, 341-351 (2005).
11. Fan QR, L.E., Wiley DC. Crystal structure of the human natural killer cell inhibitory receptor KIR2DL1–HLA-Cw4 complex. *Nature Immunology* **2**, 452-460 (2001).
12. Vivian, J.P. et al. Killer cell immunoglobulin-like receptor 3DL1-mediated recognition of human leukocyte antigen B. *Nature* **479**, 401-5 (2011).
13. Collins EJ, G.D., Wiley DC. Three-dimensional structure of a peptide extending from one end of a class I MHC binding site. *Nature* **371**, 626-629 (1994).
14. Tenzer, S. et al. Antigen processing influences HIV-specific cytotoxic T lymphocyte immunodominance. *Nat Immunol* **10**, 636-46 (2009).
15. McMurtrey, C. et al. Toxoplasma gondii peptide ligands open the gate of the HLA class I binding groove. *Elife* **5**(2016).
16. Anette Stryhn, L.Ø.P., Arne Holm and Søren Buus. Longer peptide can be accommodated in the MHC class I binding site by a protrusion mechanism. *Eur J Immunol* **30**, 3089-3099 (2000).
17. Schittenhelm, R.B., Dudek, N.L., Croft, N.P., Ramarathinam, S.H. & Purcell, A.W. A comprehensive analysis of constitutive naturally processed and presented HLA-C*04:01 (Cw4)-specific peptides. *Tissue Antigens* **83**, 174-9 (2014).

18. Carrington, M. & O'Brien, S.J. The influence of HLA genotype on AIDS. *Annu Rev Med* **54**, 535-51 (2003).
19. Migueles, S.A. et al. HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proc Natl Acad Sci U S A* **97**, 2709-14 (2000).
20. Gao, X. et al. AIDS restriction HLA allotypes target distinct intervals of HIV-1 pathogenesis. *Nat Med* **11**, 1290-2 (2005).
21. Kaslow RA, C.M., Apple R, Park L, Muñoz A, Saah AJ, Goedert JJ, Winkler C, O'Brien SJ, Rinaldo C, Detels R, Blattner W, Phair J, Erlich H, Mann DL. Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nature Medicine* **2**, 405-411 (1996).
22. Mallal, S. et al. Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *The Lancet* **359**, 727-732 (2002).
23. Illing, P.T. et al. Immune self-reactivity triggered by drug-modified HLA-peptide repertoire. *Nature* **486**, 554-8 (2012).
24. Hetherington, S. et al. Genetic variations in HLA-B region and hypersensitivity reactions to abacavir. *The Lancet* **359**, 1121-1122 (2002).
25. Goulder PJ, B.M., Krausa P, McIntyre K, Crowley S, Morgan B, Edwards A, Giangrande P, Phillips RE, McMichael AJ. Novel, Cross-Restricted, Conserved, and Immunodominant Cytotoxic T Lymphocyte Epitopes in Slow Progressors in HIV Type 1 Infection. *AIDS Res Hum Retroviruses* **12**, 1691-1698 (1996).
26. Klein MR, v.d.B.S., Hovenkamp E, Holwerda AM, Drijfhout JW, Melief CJ, Miedema F. Characterization of HLA-B57-restricted human immunodeficiency virus type 1 Gag- and RT-specific cytotoxic T lymphocyte responses. *The Journal of General Virology* **79**, 2191-2201 (1998).
27. Bailey, J.R., Williams, T.M., Siliciano, R.F. & Blankson, J.N. Maintenance of viral suppression in HIV-1-infected HLA-B*57+ elite suppressors despite CTL escape mutations. *J Exp Med* **203**, 1357-69 (2006).
28. Bernardin, F., Kong, D., Peddada, L., Baxter-Lowe, L.A. & Delwart, E. Human immunodeficiency virus mutations during the first month of infection are preferentially found in known cytotoxic T-lymphocyte epitopes. *J Virol* **79**, 11523-8 (2005).
29. Ganusov, V.V. et al. Fitness costs and diversity of the cytotoxic T lymphocyte (CTL) response determine the rate of CTL escape during acute and chronic phases of HIV infection. *J Virol* **85**, 10518-28 (2011).
30. Martinez-Picado, J. et al. Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. *J Virol* **80**, 3617-23 (2006).
31. Novitsky, V. et al. Dynamics and timing of in vivo mutations at Gag residue 242 during primary HIV-1 subtype C infection. *Virology* **403**, 37-46 (2010).
32. Miura, T. et al. HLA-B57/B*5801 human immunodeficiency virus type 1 elite controllers select for rare gag variants associated with reduced viral replication capacity and strong cytotoxic T-lymphocyte [corrected] recognition. *J Virol* **83**, 2743-55 (2009).
33. Brackenridge, S. et al. An early HIV mutation within an HLA-B*57-restricted T cell epitope abrogates binding to the killer inhibitory receptor 3DL1. *J Virol* **85**, 5415-22 (2011).
34. Crawford, H. et al. 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. *J Virol* **81**, 8346-51 (2007).

35. Alter, G. et al. Differential natural killer cell-mediated inhibition of HIV-1 replication based on distinct KIR/HLA subtypes. *J Exp Med* **204**, 3027-36 (2007).
36. Martin, M.P. et al. Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nat Genet* **31**, 429-34 (2002).
37. Qi, Y. et al. KIR/HLA pleiotropism: protection against both HIV and opportunistic infections. *PLoS Pathog* **2**, e79 (2006).
38. Chessman, D. et al. Human leukocyte antigen class I-restricted activation of CD8+ T cells provides the immunogenetic basis of a systemic drug hypersensitivity. *Immunity* **28**, 822-32 (2008).
39. Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J. & Gevaert, K. Improved visualization of protein consensus sequences by iceLogo. *Nat Methods* **6**, 786-7 (2009).
40. O'Connor, G.M. et al. Mutational and structural analysis of KIR3DL1 reveals a lineage-defining allotypic dimorphism that impacts both HLA and peptide sensitivity. *J Immunol* **192**, 2875-84 (2014).
41. Collins, E.J., Garboczi, D.N. & Wiley, D.C. Three-dimensional structure of a peptide extending from one end of a class I MHC binding site. *Nature* **371**, 626-9 (1994).
42. Wang, C.R. et al. Nonclassical binding of formylated peptide in crystal structure of the MHC class Ib molecule H2-M3. *Cell* **82**, 655-64 (1995).
43. Escobar, H. et al. Large scale mass spectrometric profiling of peptides eluted from HLA molecules reveals N-terminal-extended peptide motifs. *J Immunol* **181**, 4874-82 (2008).
44. Samino, Y. et al. A long N-terminal-extended nested set of abundant and antigenic major histocompatibility complex class I natural ligands from HIV envelope protein. *J Biol Chem* **281**, 6358-65 (2006).
45. Petersen, J.L., Morris, C.R. & Solheim, J.C. Virus evasion of MHC class I molecule presentation. *J Immunol* **171**, 4473-8 (2003).
46. Jost, S. & Altfeld, M. Evasion from NK cell-mediated immune responses by HIV-1. *Microbes Infect* **14**, 904-15 (2012).
47. Schwartz, O., Marechal, V., Le Gall, S., Lemonnier, F. & Heard, J.M. Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein. *Nat Med* **2**, 338-42 (1996).
48. Seeger, M., Ferrell, K., Frank, R. & Dubiel, W. HIV-1 tat inhibits the 20 S proteasome and its 11 S regulator-mediated activation. *J Biol Chem* **272**, 8145-8 (1997).
49. Kutsch, O., Vey, T., Kerkau, T., Hunig, T. & Schimpl, A. HIV type 1 abrogates TAP-mediated transport of antigenic peptides presented by MHC class I. Transporter associated with antigen presentation. *AIDS Res Hum Retroviruses* **18**, 1319-25 (2002).
50. Iglesias, M.C. et al. Escape from highly effective public CD8+ T-cell clonotypes by HIV. *Blood* **118**, 2138-49 (2011).
51. Liu, Y.C. et al. A molecular basis for the interplay between T cells, viral mutants and human leukocyte antigen micropolymorphism. *J Biol Chem* (2014).
52. Ladell, K. et al. A molecular basis for the control of preimmune escape variants by HIV-specific CD8+ T cells. *Immunity* **38**, 425-36 (2013).
53. Goulder, P.J. et al. Evolution and transmission of stable CTL escape mutations in HIV infection. *Nature* **412**, 334-8 (2001).
54. Schneidewind, A. et al. 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. *J Virol* **81**, 12382-93 (2007).
55. Crawford, H. et al. Evolution of HLA-B*5703 HIV-1 escape mutations in HLA-B*5703-positive individuals and their transmission recipients. *J Exp Med* **206**, 909-21 (2009).
56. Alter, G. & Altfeld, M. NK cells in HIV-1 infection: evidence for their role in the control of HIV-1 infection. *J Intern Med* **265**, 29-42 (2009).

57. Lichterfeld, M. et al. A viral CTL escape mutation leading to immunoglobulin-like transcript 4-mediated functional inhibition of myelomonocytic cells. *J Exp Med* **204**, 2813-24 (2007).
58. Croft, N.P. et al. Kinetics of antigen expression and epitope presentation during virus infection. *PLoS Pathog* **9**, e1003129 (2013).
59. Fadda, L. et al. Common HIV-1 peptide variants mediate differential binding of KIR3DL1 to HLA-Bw4 molecules. *J Virol* **85**, 5970-4 (2011).
60. Saunders, P.M. et al. Killer cell immunoglobulin-like receptor 3DL1 polymorphism defines distinct hierarchies of HLA class I recognition. *J Exp Med* **213**, 791-807 (2016).
61. Kostenko, L. et al. Rapid screening for the detection of HLA-B57 and HLA-B58 in prevention of drug hypersensitivity. *Tissue Antigens* **78**, 11-20 (2011).
62. Dudek, N.L. et al. Constitutive and inflammatory immunopeptidome of pancreatic beta-cells. *Diabetes* **61**, 3018-25 (2012).
63. Purcell, A.W. & Gorman, J.J. The use of post-source decay in matrix-assisted laser desorption/ionisation mass spectrometry to delineate T cell determinants. *J Immunol Methods* **249**, 17-31 (2001).
64. Clements, C.S. et al. The production, purification and crystallization of a soluble heterodimeric form of a highly selected T-cell receptor in its unliganded and liganded state. *Acta Crystallogr D Biol Crystallogr* **58**, 2131-4 (2002).
65. Zhang, Z. & Marshall, A.G. A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra. *J Am Soc Mass Spectrom* **9**, 225-33 (1998).
66. Collaborative. The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D* **50**, 760-763 (1994).
67. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**, 72-82 (2006).
68. Leslie, A.G.W. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* **26**(1992).
69. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
70. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).
71. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).
72. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).

## FIGURE LEGENDS:

**Figure 1:** HLA-B*57:01 in complex with the TW10 peptide. (**a**) The overall structure of KIR3DL1 in complex with HLA-B*57:01-TW10. The structure of KIR3DL1 as bound to HLA B*57:01-LF9 is overlayed (grey) as a comparison for the binding mode. (**b**) Cartoon representation of the crystal structure of HLA B*57:01 (light grey) complexed with the TW10 peptide-TSTLQEQIGW (orange) shown against the α1 helix of the HLA and oriented N-C terminal from left. Anchor pockets of the HLA binding groove are indicated at position P1 (A) P2 (B) and PΩ (F). (**c**). Orientation of the protruding Thr residue at P-1 and **(d)** The conserved hydrogen bonding network at the N-terminal end of the HLA B*57:01 is maintained to Ser at P1. Hydrogen bonds are displayed as blue dashed lines.

**Figure 2:** **(a)** Length distribution of HLA-B*57:01 ligands, showing those classified within nested sets. Peptides within the nested sets are then further broken down into the following categories: (i) peptides with versions extended at the N-terminus, but not the C-terminus (blue, N-terminally extended); (ii) peptides with versions extended at the C-terminus, but not the N-terminus (red, C-terminally extended); (iii) peptides with versions extended at the N-terminus and versions extended at the C-terminus (green, N- or C-terminally extended); (iv) peptides with versions extended at both the N- and C-terminus but not at either terminus alone (purple, N- and C-terminally extended); and (v) peptides that are the maximal sequence of an extended set (turquoise). Peptides that were not classified as part of an extended set are shown in orange. However, it should be noted that extended sets were defined as having a minimal sequence of < 12 amino acids to minimise ambiguity during motif analysis. Thus, peptides of 12 amino acids or greater that are not part of the extended sets as defined but have extended versions are present in this category. Percentage values show the percentage of 9, 10 and 11 amino acid peptides that are part of extended sets and are extended at the N-terminus alone. (**b**) and (**c**) Sequence logos showing the percentage difference in abundance of amino acids at each location in the N-terminal portion of HLA-B*57:01 ligands in comparison to their abundance in the human proteome. Logos were generated from the N-terminal portion of all 9–11 residue peptides in the HLA-B*57:01 data set (n=8268), aligned based on first residue assignment of P1 **(b),** and from the N-terminal portion of the maximal sequences of nested sets containing purely N-terminal extensions (n=972), aligned based on the minimal sequences of the nested set and possessing C-terminal aromatic anchors **(c)**. Sequence logos were generated using the iceLogo stand alone version [39].

**Figure 3:** The HIV-1 Gag repertoire of HLA-B*57:01. **(a)** HIV-1 Gag epitopes presented by HLA-B*57:01. The QW9 and TW10 epitopes form part of N-terminally extended sets. **(b)** Relative levels

of N-terminally extended variants compared to the HIV-Gag TW10 epitope. Error bars represent mean with SD; n =3.

**Figure 4:** Cartoon representation of the crystal structure of HLA-B*57:01 (light grey) complexed with the TSTFEDVKILAF peptide (blue). **(a)** HLA-B*57:01 complexed with the TSTFEDVKILAF peptide shown against the α1 helix of the HLA and oriented N–C terminal from left. **(b)** The network of direct and water-mediated hydrogen bonds (dark blue dashed lines) around the protruding P-1 residue showing the interaction of P-1-Thr with Trp167 and Asn66 of the HLA. **(c)** The network of conserved hydrogen bonds at the N-terminus of the HLA-B*57:01 peptide-binding groove showing the P1-Ser1 side-chain replacing contacts normally mediated by the N-terminus of the peptide.

**Figure 5:** Comparison of the HLA-B*57:01-TW10 and HLA-B*57:01-T3N ternary complex structures with KIR3DL1*001. **(a)** Cartoon representation of the interactions between KIR3DL1 (teal) and the T3N peptide (blue) presented by HLA-B*57:01 (light grey). **(b)** The interactions between KIR3DL1 (teal) and the TW10 peptide (orange). (**c**) Cartoon representation of the crystal structure of HLA-B*57:01 (light grey) in complex with the T3N peptide-TSNLQEQIGW (dark blue) shown against the α1 helix of the HLA and oriented N-C terminal from left. Hydrogen bonds are displayed as blue dashed lines. (**d**) Overlay of the TW10 (orange) and T3N (blue) peptide conformations.

**Figure 6:** HLA B*57:01 TW10 and T3N binding to KIR3DL1 **(a)** (i) SPR injection series for KIR3DL1 binding to the TW10 (top) and T3N (bottom) HLA B*57:01 complexes. (ii) SPR based affinity measurements of the interaction between KIR3DL1*001 and B57:01 T3N and (iii) B*57:01 TW10 complexes **(b)** Staining of KIR3DL1 allotypes with HLA B*57:01 TW10 (orange) and T3N (blue) tetramers (0.2 μg each), normalised to TW10 binding **(c)** HLA B*57:01 TW10 (orange) and T3N (blue) tetramer staining of HEK293 cells transfected with KIR3DL1*001 and a panel of KIR3DL1*001 interface residue mutants, normalised to TW10 tetramer binding to KIR3DL1*001 or **(d)** to the respective tetramer binding of KIR3DL1*001 transfectants.

**Supplementary Figure 1:** Hydrogen/deuterium exchange spectra for the TW10 peptide in complex with HLA-B*57:01 and free in solution. (**a**) Peptide in complex with HLA-B*57:01 at 0 seconds showing the normal isotopic distribution for a singly charged peptide (**b**) The peptide in complex with HLA-B*57:01 after 10 seconds incubation in $D_2O$ showing a single Gaussian distribution indicative of a single bound conformation and (**c**) the peptide free in a solution of $D_2O$ for 10 seconds showing a bimodal distribution suggesting multiple conformations in solution. Spectra displayed are from a single experiment and represent data from three independent experiments.

**Supplementary Figure 2:** Circular dichroism readings taken at 222nm over a temperature range of 20–90 °C for the HLA-B*57:01-TW10 complex **(a)** and the HLA-B*57:01-T3N complex **(b)**. $T_m$ was calculated by fitting a sigmoidal dose-response curve and taking the IC50 value of the curve.

**Supplementary Table 1:** Master list of peptides isolated from HLA-B*57:01 in this study (8432 from HIV Gag$^{neg}$ cells, and 8233 from HIV Gag$^{pos}$ transfected cells) and in previous work (Illing et al Nature 2012, n=2746)[23] used in nested set analysis (combined data set, n=11954). Peptides contained in nested sets are indicated. Peptides considered unlikely to be true HLA-B*57:01 ligands due to lack of conformation to the HLA-B*57:01 consensus motif are noted.

**Supplementary Table 2:** 1000 purely N-terminally extended nested sets of peptides identified in isolates from HLA-B*57:01. Peptides considered unlikely to be true HLA-B*57:01 ligands due to lack of conformation to consensus motif are noted. Alignment of sets is based on P1 of the minimal peptide except where the minimal peptide is 7 or 8 amino acids in length (in which case residue 1 is P3 or P2 respectively). Likely misaligns due to non-favourable amino acid at P2, and likely contaminants are noted. Maximal peptides used to describe the N-terminal extension motif are indicated. Peptides for which structures have been resolved and the register of the overhang observed are noted. The detection of these peptides in samples from HLA-B*57:01 of HIV Gag$^{neg}$ and HIV Gag transfected cells is noted, showing that N-terminal extended sets are not an artefact of Gag expression.

**Supplementary Table 3:** Extended sets containing C-terminal or N and C-terminal extensions. Peptides considered unlikely to be true HLA-B*57:01 ligands due to lack of conformation to consensus motif are noted.