LETTER TO THE EDITOR

Crystal structure of HLA-B*5801 with a TW10 HIV Gag epitope reveals a novel mode of peptide presentation

Xiaolong Li[1], Pedro A Lamothe[2], Bruce D Walker[2,3,4] and Jia-huai Wang[1,5]

The major histocompatibility complex (MHC) is encoded by a set of genes that play a key role in the vertebrate adaptive immune system. In humans, these genes encode proteins termed human leukocyte antigens (HLAs). The corresponding expressed cell surface glycoproteins are divided into two classes: class I (MHC I) and II (MHC II). Although MHC I and MHC II have very different chemical compositions, they adopt a similar three-dimensional structure, with two membrane-proximal Ig-like domains supporting an antigen-presenting groove. This groove is composed of a large eight-stranded β sheet that forms a platform beneath two α helices that flank the bound peptide.

Despite the overall similarity of the antigen-presenting architecture, there is a striking distinction in the preferred length of peptides loaded onto the two classes of MHC molecules. MHC I tends to present peptides of 8-10 amino acid residues in length, whereas MHC II can accommodate peptides of an almost unlimited length. Analysis of crystal structure data indicates that this difference is due to disparate specific structural features conserved in MHC I molecules and MHC II molecules.[1] One feature unique to MHC I molecules is a hydrogen bond network that binds both termini of peptides to MHC I. The second is that the conserved aromatic side-chains of MHC I block the ends of the peptide from extending outside of the groove. This fixes the distance between the Cα atoms of the terminal residues at ~ 22 Å. Consequently, for any peptide of more than 8 residues in length, the central portion must protrude outside of the class I peptide-binding groove. The carboxyl terminal group of the bound peptide on MHC I effectively points away from the groove, suggesting that it might be possible for an extension to exist at this C-end. Indeed, there is a report of such an extension at the C terminus of a peptide, where an additional residue of the peptide extends outside of the groove.[2] Since the N-terminal end is deeply buried in the groove and is not exposed like the C terminus, it appears that it would be structurally difficult for such an extension to exist at this N terminus.

Here, through structural analysis of the immunodominant HIV Gag epitope TW10 (TSTLQEQIGW, Gag residues 240-249), presented by HLA-B*5801, derived from patients bearing the potentially HIV-1-protective HLA class I allele *B*5801*, we report the surprising finding that the N-terminal residue of this antigenic peptide is not embedded in the HLA peptide-binding groove but instead unexpectedly extends out of the groove in an unconventional fashion. This observation contrasts with what occurs for another dominant Gag epitope, QW9 (QASQEVKNW, Gag residues 308-316), which binds the same HLA-B*5801 in a canonical fashion, as described previously.[3] These results reveal a previously unknown flexibility in the MHC presentation of antigenic peptides.

Figure 1a shows the crystal structure of the HLA-B*5801 molecule loaded with the TW10 peptide at a 2.0 Å resolution (Figure 1a, yellow stick/ribbon), superimposed on HLA-B*5801 loaded with the QW9 peptide (Figure 1a, orange stick/ribbon) at 2.5 Å resolution (PDB 5IM7). Figure 1b depicts the electron density fitting of the TW10 peptide. In the canonical peptide binding of QW9 on HLA-B*5801, the bulky aromatic side-chains of Trp167 and Tyr84 (Figure 1a, magenta stick) of the HLA

[1]Department of Medical Oncology and Cancer Biology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02215, USA; [2]Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology, Harvard University, Boston, MA 02139, USA; [3]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA; [4]Institute for Medical and Engineering Sciences, Massachusetts Institute of Technology, Cambridge MA 02139 USA and [5]Department of Pediatrics and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02215, USA.

Correspondence: BD Walker, Ragon Institute of MGH, MIT and Harvard, 400 Technology Square, Cambridge, MA 02139, USA or J-h Wang, Department of Medical Oncology and Cancer Biology, Dana-Farber Cancer Institute, LC-3211, 450 Brookline Ave, Boston, MA 02215, USA.
E-mail: bwalker@mgh.harvard.edu or jwang@crystal.harvard.edu
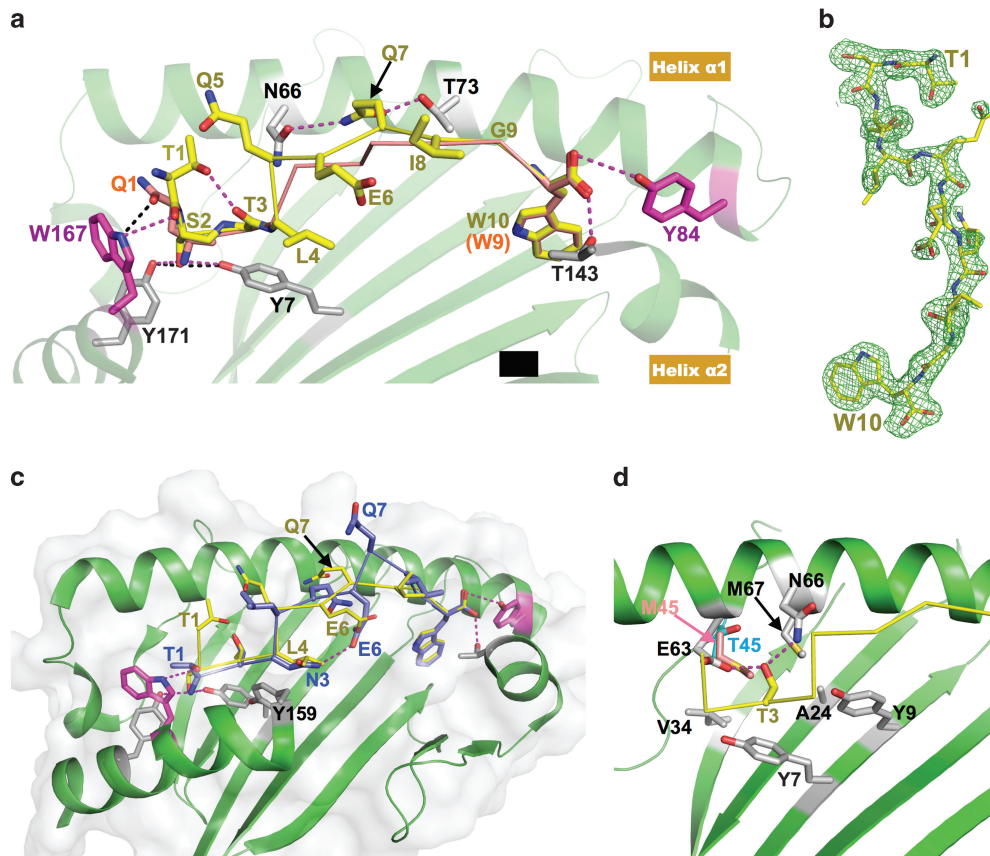
Figure 1 Structural features of immunodominant peptides bound to HLA-B*5801 and/or HLA-B*5701. (**a**) Comparison of two immunodominant antigenic peptides, TW10 and QW9, bound to the same HLA-B*5801 molecule. The green cartoon is the HLA-B*5801 structure. For clarity, the portion between Thr143 and Tyr171 of helix α2 was removed. The yellow stick/ribbon model is the TW10 peptide, and the orange stick/ribbon model is QW9. The two HLA residues that impose antigenic peptide length restriction, Trp167 and Tyr84, are highlighted with magenta sticks. The hydrogen bonds formed between TW10 and HLA-B*5801 are shown with magenta dashed lines, while the hydrogen bonds formed between Gln1 of QW9 and HLA-B*5801 are shown with black dashed lines. In TW10, the Ser2 side-chain forms two hydrogen bonds (magenta dashed lines) with Tyr7 and Tyr171 of HLA in place of the two canonical hydrogen bonds (black dashed lines) formed between the amino group of Gln1 of QW9 with Tyr7 and Tyr171. (**b**) 2Fo-Fc electron density map for TW10 bound to HLA-B*5801, contoured at 1.2 σ. The peptide TW10 (yellow stick) is depicted with its N-terminal Thr1 to the top right and its C-terminal Trp10 to the bottom. Notably, the electron density convincingly shows that the N-terminal residue Thr1 is not embedded within but instead bends away from the peptide-binding groove of the HLA molecule. (**c**) Overlay of the structure of HLA-B*5701/TW10-T3N onto that of HLA-B*5801/TW10-wt. HLA-B*5801 shown as a green ribbon drawing. For clarity, the nearly identical HLA-B*5701 molecule is not shown here. The yellow stick/ribbon model is TW10-wt, and the blue stick/ribbon model is TW10-T3N. The N-terminal amino group of TW10-T3N forms two canonical hydrogen bonds with the HLA molecule. Asn3 of TW10-T3N forms an internal hydrogen bond with Glu6 within TW10-T3N. Note that Gln7 of TW10-T3N prominently protrudes from the peptide-binding groove. (**d**) The environment of the peptide-binding pocket for the Thr3 residue of TW10. For clarity, the HLA-B*5701 molecule is not shown here. Thr3 forms hydrogen bonds with Glu63 and Asn66 located in helix α1 of HLA-B*57/5801. The peptide-binding pocket, to which the Thr3 side-chain points, is enriched in relatively hydrophobic residues. The Thr45 residue of HLA-B*5801 is shown in cyan, and Met45 of HLA-B*5701 at the same site is highlighted with a pink stick.

molecule most notably block the peptide from extension at either end. The QW9 N-terminal amino group of Gln1 forms two hydrogen bonds with the conserved Tyr7 and Tyr171 residues of HLA, whereas its C-terminal carboxyl group forms two hydrogen bonds with the side-chains of Tyr84 and Thr143 (Figure 1a). This N-terminal Gln1 uses its tip amide to form a hydrogen bond with the NE1 atom of Trp-167 of HLA and uses the

aliphatic portion of the side-chain to pack onto the bulky indole ring of Trp167.

In contrast to HLA-B*5801/QW9, for HLA-B*5801/TW10, it is the hydroxyl group of the side-chain of Ser2 of TW10 that forms the two hydrogen bonds with Tyr7 and Tyr171. The TW10 N-terminal Thr1 residue bends away from the groove, with the planar peptide group of Thr1-Ser2 packing against the HLA

Trp167 ring in a remarkably similar manner to the Gln1 side-chain of the QW9 peptide. Additionally, the carbonyl oxygen of Thr1 forms a hydrogen bond with the NE1 atom on the side-chain of Trp167, and the hydroxyl group of the Thr1 side-chain forms another hydrogen bond with the carbonyl group of Thr3 of TW10, stabilizing the side-chain conformation of Thr1 (Figure 1a). Thus, the apparently 'extra' Thr1 residue of the

TW10 peptide effectively replaces the side-chain of the 'conventional' Gln1 residue of the QW9 peptide in the interaction with HLA Trp167. In contrast, the structures of the C-terminal ends of TW10 and QW9 overlay each other extremely well (Figure 1a). Note that the overall conformation of TW10 peptide residues 2–10 matches quite well with that of QW9 peptide residues 1–9. We also solved a structure of HLA-B*5701/TW10 to a resolution of 2.9 Å. HLA-B*5701/TW10 exhibits essentially the same features as HLA-B*5801/TW10, as would be expected given the similarity of these two class I molecules (Supplementary Figures S1a and b).

Comparison of the structures of HLA-B*5801/TW10 and HLA-B*5801/QW9 revealed that neither peptide is prominently displayed on the HLA surface. Rather, the central portion of both peptides that engages the TCR appears as a flat plateau (Figure 1a). TW10 achieves this conformation by anchoring the side-chain of Ser2 into the groove, in place of the conventional amino terminus of the peptide. Since Thr1 of TW10 curves away from the groove, there are effectively only nine, rather than ten, residues in the groove. A potential advantage of this difference for antiviral immunity was suggested by an in-depth study on cytotoxic T cell recognition of influenza virus epitopes.[4] In that study, a 'feature-less' peptide-MHC (pMHC) generated a more 'public' TCR with limited diversity, whereas a peptide with more prominent features selected diverse 'private' TCRs. The authors argued that for these feature-less pMHC complexes with TCRs, the MHC molecule may play more important roles in TCR engagement. Indeed, in a previous study[3] comparing the structures of protective HLA-B*5801 with disease-progressive HLA-B*5301, we showed that an exposed surface of HLA at the N terminus of helix α1 exhibits the greatest difference between these two HLA molecules. This region mostly interacts with germline-encoded CDR1/CDR3 of TCR α. Regarding the side-chains displayed by TW10 on HLA-B*5801 (Figure 1a), the only prominent side-chain is Gln5. For Gln7 and Gly9 of TW10, the main-chain

Cα atoms are located at the highest position in the groove. Particularly notable is residue Gln7, located in the middle of the peptide, for which the direction of the side-chain from Cα to Cβ appears to point upwards. However, the rest of the side-chain bends downward, with its amide group at the tip, to form two hydrogen bonds with the side-chains of HLA residues Asn66 and Thr73 in helix α1 and is therefore not exposed for TCR recognition (Figure 1a). Thus, the TW10 peptide is indeed less prominently featured, which is even more strikingly apparent when this HLA-B*5801/TW10 structure is compared with the recently published structure of HLA-B*5701 loaded with the escape mutant TW10-T3N in complex with the natural killer receptor KIR3DL1[5] that appeared online while we were preparing this manuscript.

Clinical and immunological studies have indicated that in comparison with the other HIV Gag epitopes, the immune responses to and immune escape from TW10 appear to play a more important role in elite controller patients with *HLA-B*5801* and/or *HLA-B*5701* alleles. The TW10-T3N escape mutant arises rapidly and almost universally after acute infection in individuals expressing HLA-B*57/5801.[6] The mutant appears to dramatically impair the recognition of the naturally selected T-cell receptors.[7] Within the HLA-B*5701/TW10-T3N structure, Pymm *et al.*[5] described a register shift of the mutant peptide TW10-T3N relative to the wild-type TW10 within the binding groove, such that Thr1 no longer bends away from and, instead, anchors at the groove in the conventional peptide-binding fashion to MHC I. We superimposed this HLA-B*5701/TW10-T3N structure (PDB 5T70) onto our HLA-B*5801/TW10 structure (Figure 1c). The conformations of the wild-type peptide TW10-wt (Figure 1c, yellow stick/ribbon) and its mutant TW10-T3N (Figure 1c, blue stick/ribbon) differ substantially. In Figure 1c, it can be seen that a new internal hydrogen bond is formed between Asn3 and Glu6 within the mutant peptide TW10-T3N. The side chain of Asn3 occupies the Leu4 position of TW10-wt and packs onto the aromatic ring of HLA Tyr159 in the same way as

Leu4 of TW10-wt. We also noted that Asn3 in the TW10-T3N mutant cannot fit into the environment where Thr3 of TW10-wt is located. Figure 1d shows the structural environment of Thr3 in HLA-B*5801. The side-chain of this Thr3 forms two hydrogen bonds with HLA class I residues Glu63 and Asn66, and this binding pocket is otherwise enriched in hydrophobic HLA residues, with the exception of Thr45. However, in HLA-B*5701, this Thr45 (Figure 1d, cyan stick) is replaced by Met45 (Figure 1d, pink stick), thus becoming even more hydrophobic. It is clear that the longer and more hydrophilic side-chain of Asn3 would be difficult to accommodate in this more hydrophobic and much narrower binding pocket of HLA-B*5701. In summary, in HLA-B*5701, Asn3 of the mutant peptide TW10-T3N energetically favors moving away from where TW10-wt Thr3 sits and shifts one register to the Leu4 position of TW10-wt.

The consequences of these differences are two-fold. First, the TW10-T3N peptide loads onto HLA in a canonical fashion, with its N-terminal Thr1 anchored in the groove. Consequently, the more interesting second point is that the central portion of TW10-T3N now has to protrude outward, and the side-chain of Gln7 cannot form two hydrogen bonds with HLA, as shown in Figure 1a for Gln7 in TW10-wt. Instead, this Gln7 in TW10-T3N is forced to extend out of the groove, meaning that the TW10-T3N peptide is no longer inconspicuously featured and is instead very prominent (Figure 1c). This should give rise to a significant change in the TCR recognition landscape for the escape mutant T3N loaded onto HLA-B*5701, consistent with impaired recognition by TCRs that respond to HLA-B*5701/TW10-wt.[7]

In summary, we report an HLA-B*5801 molecule that presents an immunodominant antigenic peptide, TW10, from the HIV-1 Gag protein in a novel fashion, demonstrating that this class I molecule can present peptides in both a canonical and non-canonical fashion. We further show that the N-terminal residue of TW10 does not anchor into but instead bends away from the peptide-binding groove. Regarding the structure

of the TW10-T3N mutant, we discuss its change in register compared with TW10-wt and the possible implications for escape from expanded TCR clonotypes.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

1 Stern LJ, Wiley DC. Antigenic peptide binding by class I and class II histocompatibility proteins. *Structure* 1994; **2**: 245–251.
2 Collins EJ, Garboczi DN, Wiley DC. Three-dimensional structure of a peptide extending from one end of a class I MHC binding site. *Nature* 1994; **371**: 626–629.
3 Li X, Lamothe PA, Ng R, Xu S, Teng M, Walker BD, Wang JH. Crystal structure of HLA-B*5801, a protective HLA allele for HIV-1 infection. *Protein & Cell* 2016; **7**: 761–765.
4 Turner SJ, Kedzierska K, Komodromou H, La Gruta NL, Dunstone MA, Webb AI *et al.* Lack of prominent peptide-major histocompatibility complex features limits repertoire diversity in virus-specific CD8+ T cell populations. *Nat Immunol* 2005; **6**: 382–389.
5 Pymm P, Illing PT, Ramarathinam SH, O'Connor GM, Hughes VA, Hitchen C *et al.* MHC-I peptides get out of the groove and enable a novel mechanism of HIV-1 escape. *Nat Struct Mol Biol* 2017; **24**: 387–394.
6 Miura T, Brockman MA, Schneidewind A, Lobritz M, Pereyra F, Rathod A *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* 2009; **83**: 2743–2755.
7 Chen H, Ndhlovu ZM, Liu D, Porter LC, Fang JW, Darko S *et al.* TCR clonotypes modulate the protective effect of HLA class I molecules in HIV-1 infection. *Nat Immunol* 2012; **13**: 691–700.

Supplementary Information for this article can be found on the *Cellular & Molecular Immunology* website (http://www.nature.com/cmi)