

Breaking Confinement: Unconventional Peptide Presentation by Major Histocompatibility (MHC) Class I allele HLA-A*02:01

Soumya G. Remesh¹, Massimo Andreatta^{2,3}, Ge Ying¹, Thomas Kaever², Morten Nielsen^{3,4}, Curtis McMurtrey^{5,6}, William Hildebrand^{5,6}, Bjoern Peters², and Dirk M. Zajonc^{1,7}

From the ¹Division for Cell Biology and ²Division of Vaccine Discovery, La Jolla Institute for Allergy and Immunology, La Jolla, CA 92037, USA.

³Instituto de Investigaciones Biotecnológicas, Universidad Nacional de San Martín, CP1650, San Martín, Argentina

⁴Center for Biological Sequence Analysis, Department of Bio and Health Informatics, The Technical University of Denmark, 2800 Lyngby, Denmark

⁵University of Oklahoma Health Science Center, Department of Microbiology and Immunology, Oklahoma City, OK, USA

⁶Pure MHC LLC, Austin, TX, USA

⁷Department of Internal Medicine, Faculty of Medicine and Health Sciences, Ghent University, 9000 Ghent, Belgium

Running title: Peptide-induced opening of the F pocket of MHC class I

To whom correspondence should be addressed: Prof. Dirk M. Zajonc, Division of Cell Biology, La Jolla Institute for Allergy and Immunology, 9420 Athena Cir, La Jolla, CA 9203, Telephone: (858) 752-6605; FAX: (858) 752-6985; E-mail: dzajonc@lji.org

Keywords: major histocompatibility complex (MHC), peptide interaction, antigen presentation, protein crystallization, protein structure, T cell, Natural Killer Cell, *Toxoplasma gondii*

ABSTRACT

Peptide antigen-presentation by Major Histocompatibility Class (MHC) I proteins initiates CD8⁺ T cell mediated immunity against pathogens and cancers. MHC I molecules typically bind peptides with nine amino acids in length with both ends tucked inside the major A and F binding pocket. It has been known for a while that longer peptides can also bind by either bulging out of the groove in the middle of the peptide or by binding in a zig-zag fashion inside the groove. In a recent study, we identified an alternative binding conformation of naturally occurring peptides from *Toxoplasma gondii* bound by HLA-A*02:01. These peptides were extended at the C-terminus (PΩ) and contained charged amino acids not more than 3 residues after the anchor amino acid at PΩ, which enabled them to open the F pocket and expose their C-terminal extension into the solvent. Here, we show that the mechanism of F pocket

opening is dictated by the charge of the first charged amino acid found within the extension. While positively charged amino acid result in the Tyr84 swing, amino acids that are negatively charged induce a not previously described Lys146 lift. Further, we demonstrate that the peptides with alternative binding modes have properties that fit very poorly to the conventional MHC class I pathway, and suggest they are presented via alternative means, potentially including cross-presentation via the MHC class II pathway.

Peptide presentation by MHC class I molecules regulates which fragments of a pathogen or cancer antigen are displayed to cytotoxic T cells for immune recognition. Understanding the mechanism of antigen presentation by MHC I is crucial in an attempt to

design therapeutic strategies aimed at modulating subsequent immune responses to control disease. Toxoplasmosis is a parasitic disease caused by infection with the large intracellular protozoan *Toxoplasma gondii* (1,2). While generally asymptomatic in healthy adults, *T. gondii* infection can cause congenital toxoplasmosis during pregnancy and result in abortion or neonatal disease (1,2). T cell mediated immunity against *T. gondii* derived peptide antigens provides strong protection against *T. gondii* and involves both peptide presentation by Major Histocompatibility Class I (MHC I) and Class II (MHC II) protein (3-5). While *T. gondii* can interfere with CD4 T cell responses by downregulating MHC II expression in IFN- γ activated macrophages, immunization with *T. gondii* MHC II peptide ligands can elicit potent CD4 T cell response that can lower parasite burden in the brain (6,7). Immunocompromised individuals and patients with T-cell deficiencies are highly susceptible to *T. gondii* infections (8,9). CD8⁺ T cell responses have been studied more widely than CD4⁺ and peptide ligands for MHC I have been identified to be derived from surface proteins, or proteins of specialized secretory organelles (rhoptry proteins) that can either be secreted into the parasite cytosol or the parasitophorous vacuole (8-14).

HLA-A*02:01 has been the focus of studies aimed at identifying MHC-I restricted peptide ligands that confer protection against *T. gondii* in HLA transgenic mice (15), and as such is a suitable MHC class I allele to study the basic rules of peptide presentation. Generally, most canonical peptide ligands for MHC I are 9-10 amino acids in length. However, peptide ligands with more than 11 amino acids have been identified as ligands for MHC I in general and form the non-canonical ligand group (13,16,17). These long peptides have been shown to interact with the residues of the binding groove of HLA class I heavy (alpha) chain much like the canonical binders with some changes. The second (P2) and C-terminal (P Ω) residues of the antigen peptide anchor into the A and F pockets of the binding groove, respectively while the middle portion of these over-sized peptides either ‘bulge out’ or ‘zig-zag’ in the binding groove to be accommodated (18,19).

In contrast to these ‘bulged’ peptides, we had recently identified longer *T. gondii* peptides

eluted from HLA-A*02:01 molecules that had a conserved N-terminal start but differed in their residue composition at the C-terminus (20). We showed through crystallographic studies that in the HLA-A*02:01 complex with one 12 mer peptide residue Tyr84 of the MHC heavy chain swung out and opened the F pocket, allowing the C-terminal amino acid of the peptide to protrude into the solvent, while the nested 11mer N-terminal core peptide bound in a conventional zig-zag orientation tucked with both peptide ends inside the peptide binding groove.

To further investigate whether the opening of the binding groove could be achieved with other peptides presented on *T. gondii* infected cells, and to understand what the structural requirements are to enable such unconventional modes of binding, we crystallized complexes of HLA-A*02:01 with several pairs of core (nested) and C-terminally extended peptides. Surprisingly, we found that there are at least two distinct modes of opening the F pocket of HLA-A*02:01 involving the residues Tyr84 and Lys146. We suggest that these unconventional modes of binding will help better understand targets of MHC class I restricted epitope recognition.

RESULTS

*Crystal structures of HLA-A*02:01 in complex with conventional and extended peptides*

To identify additional peptides with likely unconventional binding motifs, we scanned the set of peptides eluted from HLA-A*02:01 for those that had poor predicted binding affinity of the full length peptide (percentile rank >10%), but contained a nested N-terminal peptide with high predicted affinity (percentile rank <2%). In our previous study we had examined one such peptide (FVLELEPEWTVK) which had a single lysine added to the C-terminus of the core peptide (FVLELEPEWTV) and induced a structural change in Tyr84 of HLA-A*02:01 (20). In contrast, in the current study we examined 3 sets of the peptides that had C-terminal amino acid additions that contained negatively charged amino acids or both negatively and positively charged residues (Figure 1). To investigate if the F pocket of HLA-A*02:01 could also be opened by these extending peptides, we refolded HLA-A*02:01 with several nested and extending peptides and determined the crystal structures of these complexes. We obtained

crystal structures for all complexes at resolutions between 1.85 Å and 2.75 Å (Table 1). Electron densities for all the peptides were well defined over the entire peptide length that is bound within the binding groove, while C-terminally extending residues that did not contact HLA-A*02:01 were disordered (Figure 1). When all the different peptides are compared, slight structural changes in HLA-A*02:01 are observed in the A pocket. Peptides with an N-terminal tyrosine (YLSPIASPL, YLSPIASPLL, and YLSPIASPLLDGKSLR) open the A pocket slightly for the bulky side chain to be accommodated, while peptides that begin with glycine (GLKEGIPAL, GLKEGIPALDN, GLLPELPAV, and GLLPELPAVGGNE) are more buried inside the A pocket, since they lack any side chain (Figure 1). In addition, subtle structural changes are observed throughout the binding groove to allow optimal binding of the different amino acid side chains. However, when structures of the core peptides are compared with their respective extended peptides, the position of Tyr84 of HLA-A*02:01 was unchanged. Surprisingly, however, Lys146 of the F pocket, which is located close to Tyr84 and forms a “lid” to bury the PΩ amino acid in the core peptides moved upwards to open the F pocket, when the extending peptides were bound (Figure 1). While Lys146 adopts slightly different positions when all the extending peptide structures are compared, in each structure the Lys146 lid was opened for the C-terminal extensions to protrude from the F pocket.

Hydrogen bond network for nested and longer peptide pairs

Next, we looked at the detailed interactions between HLA-A*02:01 and the individual peptides. In case of peptide GLKEGIPAL, an extensive hydrogen bond network is seen involving PΩ leucine residue and residues of the heavy chain that line the binding groove including Asp76, Thr80, Tyr84, Thr143, Lys1467 and Trp147 (Figure 2a. Upper panel). In contrast, for the extended peptide GLKEGIPALDN, the hydrogen bond between Lys146 and terminal carboxy group of PΩ leucine is replaced with one between Lys146 and the PΩ+1 aspartate side chain, since Lys146 adopts a

different orientation (Figure 2a. Lower panel). In case of the peptide pair GLLPELPAV and GLLPELPAVGGNE, a similar hydrogen bond network is observed for both peptides with only a minor difference in the crystal structure of the longer peptide. The hydrogen bond interaction between the terminal carboxylate of PΩ valine and Lys146 is missing in the crystal structure with the longer peptide (Figure 2b). The same is true for the peptide pair YLSPIASPL and YLSPIASPLLDGKSLR (Figure 2c). As a result, the change in the orientation of Lys146 leads to the loss of hydrogen bond formation with the carboxylate of the PΩ amino acid. However, the hydrogen bond interaction between HLA-A*02:01 residue Trp147 and the backbone oxygen of the P8 amino acid remains conserved. Depending on the amino acid following residue P8 a novel hydrogen bond can be formed with the side chain of a compatible amino acid at PΩ+1 (here Asp10). Since the C-terminally extending amino acids project away from the peptide binding groove, electron density becomes increasingly disordered as the peptide exits the F pocket (Figure 1).

*Extending peptides do not significantly destabilize HLA-A*02:01*

To determine the relative stability of the individual HLA-A*02:01-peptide complexes, we followed their thermal denaturation by differential scanning fluorimetry. The melting temperatures (T_m) obtained from the melt curves allowed us to compare the stability of the different complexes (Figure 3). We observe that complexes of HLA-A*02:01 with extended peptides had similar stability to those with their equivalent nested peptides with not more than 8°C difference between them. For example, the T_m for HLA-A*02:01 complex with GLKEGIPAL is 63°C while that with its longer peptide counterpart is 61°C. Addition of 6 extra residues to peptide YLSPIASPLL also only changes the T_m of the complex with peptide YLSPIASPLLDGKSLR by about 7°C [Figure 3 and ref. (20)]. Interestingly, some of the complexes with nested peptides are as stable as those with longer peptides (compare GLKEGIPAL with GLLPEPPVGGNE, Figure 3). It is also worth noting that there are some variations in the stability of complexes with different nested peptides. For instance, peptide

GLKEGIPAL forms a less stable complex as compared to peptide GLLPELPAV (Figure 3). Although most interactions between HLA-A*02:01 and the peptides are conserved there is a significant difference in the hydrogen bond interaction between Tyr84 with their terminal carboxylate (3.65 Å for G9L and 2.9 Å for G9V, Figure 2). The lack of an intimate hydrogen bond interaction of the terminal amino acid with Tyr84 in G9L peptide is likely a major contributor to the reduced melting temperature. Compared to our previous study, we noticed that a single positively charged amino acid addition (compare FVLELEPEWTV and FVLELEPEWTVK ref. (20)) destabilizes the protein-peptide complex more than extending peptides that follow a short (2 amino acid) negatively charged residue addition (compare GLKEGIPAL with GLKEGIPALDN, Figure 3). However, longer peptide additions (4-6 amino acids) that contain a negatively charged amino acid (YLSPIASPLL vs. YLSPIASPLLDGKSLR and GLLPELPAV vs. GLLPELPAVGGNE) reduce the protein-peptide complex to the same extent as the single “K” addition found in peptide FVLELEPEWTVK (20). This highlights that the structural change involving Tyr84 of HLA-A*02:01 is more destabilizing than that of Lys146, when the negatively charged peptide extension is very short.

Lysine 146 lift

The different orientations of Lys146 upon binding of the longer peptides open the F pocket of HLA-A*02:01 and are required for the C-terminally extending amino acids to project into the solvent, since Lys146 forms a partial “lid” above the F pocket, held in position by the hydrogen bond interaction of the carboxylate of the C-terminal amino acid (PΩ) of any nested peptide (Figures 2 and 4). Even though, the position of Lys146 is not precisely conserved between the different structures of HLA-A*02:01 bound to the extending peptide, the lift of the residue to accommodate the peptide extension seems to be consistent. In case of the crystal structure of HLA-A*02:01 with YLSPIASPL, YLSPIASPLL, and YLSPIASPLLDGKSLR there are variations in the way that the two nested peptides are accommodated in the binding groove. With YLSPIASPL, the binding of the residues is quite conventional with the P2 and PΩ anchor

residues binding to the A and F pocket. Surprisingly, however, YLSPIASPLL, the nested peptide with one extra leucine residue at the C-terminal end of the peptide undergoes certain extent of bulging to accommodate the terminal leucine residue (PΩ+1, instead of PΩ) as the anchor residue in the F pocket (Figure 4d). The difference in the binding of these two nested peptides to HLA-A*02:01 underscores the requirement for a sequence motif or particular amino acid features within the bound peptide to induce movement of Lys146 to open the F pocket. Since a mere increase in length of the peptide does not cause the change in orientation of Ly146, it is likely that the addition of charged residues within the C-terminal extension is the contributing factor to open the F pocket. Thus, in addition to the previously identified “Tyr84 swing” to accommodate the peptide FVLELEPEWTVK (UFP¹⁶⁻²⁷) (20), we observed a “Lys146 lift” in HLA-A*02:01 as a second mechanism of opening the F pocket induced by the extending peptides YLSPIASPLLDGKSLR, GLKEGIPALDN and GLLPELPAVGGNE (Figure 5).

It's a game of charge

The negatively charged amino acids in the extended peptides do not always immediately follow the nested conventional PΩ anchor residue, but can be several residues downstream (Figure 5c). The previously reported longer peptide contains a C-terminal addition of a positively charged residue (FVLELEPEWTVK) that opens the binding groove using the “Tyr84 swing” mechanism. Here, we observed that all extended peptides with a negatively charged residue only open the binding groove using the “Lys146 lift”-mechanism. This included the peptide YLSPIASPLLDGKSLR which contains a lysine (positive charge) residue following the aspartate (negative charge), but no “Tyr84 swing” was observed. This suggested that the first charged residue determined which of the two distinct structural modes of binding an extended peptide will adapt.

Extensions with negative charges are longer and more frequent than extensions with positive charges

Given the discovery of several ligands binding in an unconventional mode, we aimed to assess the generality of these findings. In the *T. gondii* peptide elution dataset, a total of 134 peptides were predicted to bind with conventional P2-P Ω anchors with predicted rank $\leq 10\%$. Of the remaining 150 peptides with poor predicted binding, 108 contained a nested strong binder (rank $\leq 2\%$) at the N-terminus, with one or more residues extending beyond P Ω . These 108 peptides are expected to be enriched for examples with a similar unconventional mode of binding as the ones in our structural studies.

We classified the ligands into having ‘negative’ or ‘positive’ extension based on the first charged residue found after the nested binding peptide. Nearly half of the extended ligands contained a negatively charged residue as the first charged amino acid within the first three residues of the extension, whereas positively charged extensions were less frequent (Figure 6a). Positive extensions were short (less than 3 residues on average), while negative extensions had an average length of 8.7 residues (Figure 6b). Notably, 66% of the positive extensions consisted of a single or two residues, compared to only 8% of negative extensions being shorter than three amino acids. Considering only the longest version of ligands with extensions of multiple sizes, the average length of negative extensions was of 11.0 residues.

Extended peptide MHC-I ligands show a putative processing motif that is more similar to MHC class II ligands than conventional MHC-I ligands

Given the unconventional length and mode of binding of the observed extended peptides, and given that *T. gondii* has an unusual compartmentalized lifecycle in the cells it infects, we wanted to examine if the unconventional ligands found had the typical motifs of peptides derived from the conventional MHC class I processing and presentation pathway. As shown in Figure 6, this was not the case. Extended ligands had significantly lower scores for TAP transport (Figure 6c) compared to canonical *T. gondii* ligands ($p=1.2 \times 10^{-5}$, Wilcoxon rank-sum test), but not significantly different from those of random

peptides ($p=0.34$). Similarly, proteasome cleavage scores (Figure 6d) for extended ligands were significantly lower compared to canonical ligands ($p=5 \times 10^{-16}$), but not significantly higher than the random natural peptides ($p=0.24$). In other words, long ligands with terminal extensions were predicted to be poor substrates both for proteasome cleavage and TAP transport, suggesting an alternative mechanism for the generation and translocation to MHC class I of these extended ligands.

Given the lifecycle of *T. gondii*, it is possible that the unconventional MHC-I peptide ligands derived from it are processed and (cross-)presented through the same pathway as MHC-II ligands. If that is the case, we would not only expect that these ligands have different amino acid motifs as those found for MHC-I ligands (as was shown above), but also that they have a pattern congruent with what is found for MHC-II ligands. Accordingly, we examined the amino acid patterns of the C-terminal residues in extended ligands. Remarkably, despite being predominantly negatively charged or uncharged in the first 3 residues of the extension, a large fraction of ligands presented either an Arginine or Lysine at the very C-terminal residue (42 out of 108). If these ligands were cross-presented, we would expect similar trimming motifs in class II ligands, and we examined published datasets of MHC-II ligands for evidence of such a motif (see Experimental Procedures). Indeed, we found that the C-terminal composition of the extended *T. gondii* class I ligands strongly correlated to the residue distribution at the C-terminus of eluted MHC class II ligands, where positively charged amino acids were also enriched (Figure 6e). Taken together, we find that the unconventional MHC-I ligands presented by *T. gondii* have sequence motifs much more consistent with cross-presentation rather than with generation through proteasomal cleavage and TAP transport.

DISCUSSION

$\alpha\beta$ T cell receptor (TCR) recognition of MHC presented microbial peptides initiates T cell mediated immunity against infection. Generally, the TCR binds with both TCR α and β chain in a diagonal orientation above the MHC molecule

(28). While the germline encoded complementary determining region (CDR) 1 and 2 loops bind to MHC, the hypervariable loops CDR3 α and 3 β specifically bind and recognize the peptide and provide antigen specificity (21). MHC I has a closed binding groove and peptides bind with both ends tucked inside the binding pocket, while MHC II has an open binding pocket and peptide ligands, typically 15-20 amino acids bind with both N- and C-termini hanging over the end of the groove. Since MHC I presents peptides in a more confined space compared to MHC II, the TCR of CD8⁺ T cells often contact and discriminate their entire peptide sequence (21). In our present study, we have focused on a panel of C-terminally extended *T. gondii* peptides that contain a negatively charged amino acid within their C-terminal extensions. Analysis of a large dataset of HLA-A*02:01 eluted *T. gondii* peptides from our previous study (20), demonstrated that C-terminally extending peptides are very common in *T. gondii* and that additions that contain negatively charged amino acids are more represented than those that contain only positively charged residues. These peptides contain a canonical HLA-A2*02:01 binding motif at their N-terminus, while addition of the C-terminal extensions render predictions about their binding to MHC I difficult. While the N-terminus of these peptides bind like canonical peptides to the MHC I allele HLA-A*02:01, the C-terminal extensions induce a structural change at the F pocket to allow their extension into the solvent. As such, algorithms aimed at predicting peptide binding to MHC I need to capture rules that allow identification of a canonical N-terminal MHC I binding motives, while adding descriptors, such as charged residues in the C-terminal extension to predict binding of C-terminally extending peptides. In principle these rules could be learned directly from peptide binding data using machine-learning techniques such as the recently described extended NNalign method (22,23). However, this is complicated by the very limited amount of quantitative data available characterized by non-canonical binding. We would envision this situation to change as more binding data becomes available and as MHC ligand data are included into the training data of MHC class I binding prediction algorithms. We showed that only charged residues that follow the P9 amino acid of a canonical peptide within 3 or

less amino acids induce the structural change that allows them to bind to MHC I and stabilize the complex. These rules can now be incorporated into existing algorithms to predict MHC I binding peptide at least for HLA-A*02:01.

Since the residues that are involved in the F pocket opening (Tyr84 and Lys146) are conserved across all HLA-A, -B, and -C alleles (and also found in many non-classical MHC I molecules, such as HLA-E, HLA-G, Qa-1, as well as viral MHC I mimics such as UL18), we postulate that the ability to open the F pocket is a universal characteristic found across many MHC I molecules. While it is not clear to what extent this structural change affects CD8⁺ T cell recognition, Killer Immunoglobulin Receptors (KIRs), a family of activating and inhibitory receptors expressed on NK cells, NKT cells and many CD4⁺ and CD8⁺ T cells, bind directly above the F pocket. As a consequence, any structural change around the F pocket would likely affect KIR binding and directly modulate host immune responses (24,25) Especially the inability of inhibitory KIRs to engage MHC I would lower the threshold of activation for many more immune cells to combat infection or cancers. Future studies will have to address the origin and potential function of these longer peptides in host immunity.

EXPERIMENTAL PROCEDURES

*HLA-A*02:01 Expression and Purification*- HLA-A*02:01 class I heavy chain ectodomain (residues 21-274) and human β -2 microglobulin (h β 2m, 1-99) were expressed as inclusion bodies and refolded as reported previously with modifications reported here. Briefly, both the heavy chain and light chain were expressed in *Escherichia coli* BL21 DE3 cells, induced at OD₆₀₀ of 0.6 with 1mM isopropyl-1-thio-D-galactopyranoside (IPTG) and cells were harvested after 4 hours by centrifugation (5000 g for 20 mins). Cells were resuspended separately in lysis buffer (100 mM Tris-HCl pH 7.0, 5mM EDTA, 5mM DTT, 0.5mM PMSF) and the cells were broken the cells were broken with 4-5 passes through a microfluidizer (20 kPa) (Microfluidics). Cell lysate was centrifuged (50,000 g for 30 min at 4°C) to collect inclusion bodies. Inclusion bodies were further resuspended in wash buffer A (100 mM Tris-HCl pH 7.0, 5 mM EDTA, 5 mM DTT, 2

M Urea, 2 % w/v Triton X-100) centrifuged again and washed in wash buffer B (100 mM Tris-HCl pH 7.0, 5 mM EDTA, 2 mM DTT). Finally the inclusion bodies were denatured in extraction buffer (50 mM Tris-HCl pH 7.0, 5 mM EDTA, 2 mM DTT, 6 M Guanidine-HCl) for subsequent refolding. 3 mg of h β 2m was added dropwise to 250 mL of refolding buffer (0.1 M Tris-HCl pH 8.0, 2 mM EDTA, 400 mM L-arginine, 5 mM oxidized glutathione, 5 mM reduced glutathione) and stirred for 1-2 hours. Between 11-15 mg of HLA-A heavy chain mixed with 2-3 mg of individual peptide (GenScript) was then added to the refolding mix and further stirred at 4°C for 72 hours. Final heavy chain:light chain:peptide ratios were in the range of 2:1:12 and 2.5:1:12 for different peptides. Following refolding, the refolding mixture was centrifuged at 50,000 g to remove any precipitated protein and the supernatant was concentrated to about 3 ml for size exclusion chromatography (SEC) using a Superdex S200 HR16/60 gel filtration column pre-equilibrated with SEC buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl). Fractions containing refolded HLA-A*02:01- β 2m-peptide complexes were pooled, concentrated to 5-12 mg/mL and used for subsequent crystallization experiments.

Crystallization and Data Collection-Initial attempts to obtain crystals for the HLA-A*02:01-peptide complexes using factorial screens were not successful. Thin needle shaped sea urchin crystals of HLA-A*02:01-F12K complex obtained in 1.2M sodium citrate were used to cross seed the other complexes. The complexes were equilibrated in 30% PEG 4000, 0.1M Tris-HCl pH 8.0, 0.2M lithium sulfate for 1-2 hours by mixing 0.15 μ l complex and 0.15 μ l of precipitant at 20°C before seeding. Thin plate-like crystals were obtained by sitting drop vapor diffusion at 20°C after 2-4 days. The crystals were flash frozen in cryoprotectant (crystallization solution: 100% glycerol - 3:1) using liquid nitrogen.

Diffraction data for HLA-A*02:01 complex with peptides G9V, G11N, G13E and Y16R were collected remotely at beamline 7.1 at the Stanford Synchrotron Radiation Light source (SSRL) and processed to 1.86 Å, 2.3 Å, 2.1 Å and 2.4 Å resolution, respectively using HKL2000. Diffraction data for HLA-A*02:01 complex with peptides G9L, Y9L and Y10L were collected remotely at beamline 12.3.1 at the Advanced Light

Source (ALS) and processed to 1.85 Å, 2.5 Å and 2.75 Å resolution, respectively using HKL2000. Phases were obtained by molecular replacement with Phaser MR (26) in ccp4i (27,28) using the protein coordinates for HLA-A*02:01 (PDB ID 3MRE) and resulted in unambiguous electron density for all the peptides. Model building was carried out using COOT (29,30). Structures were refined using Refmac (31). Data collection and structure refinement parameters are provided in Table 1.

Thermal Denaturation Assay-HLA-A*02:01- β 2m-peptide complexes with the various different peptides were analyzed for thermal denaturation by differential scanning fluorimetry using a LightCycler 480 (Roche). HLA-A*02:01- β 2m-peptide complexes with different peptides at 100 μ M in reaction buffer (20mM Tris-HCl pH 7.5, 150mM NaCl) were used as stock solutions. Each reaction mixture constituted of 1-2 μ l protein complex stock solution, 2 μ l of SYPRO Orange dye (100X, Invitrogen) made up to 20 μ l in reaction buffer in a 96-well white plate compatible with the instrument. A temperature gradient from 20°C - 85°C at steps of 0.06°C/sec and 10 acquisitions/°C was used for the experiment. Each experiment with individual protein-peptide complex was repeated thrice. A melt curve of the total fluorescence of the run was plotted against temperature. The minima of the first derivative of the melt curve from raw fluorescence data (temperature differential of absolute fluorescence versus temperature) provided the melting temperature (T_m) for individual HLA-A*02:01- β 2m peptide complexes (inflection point of the melt curve) (32).

Binding affinity predictions and analysis of extensions-Binding affinities for all 284 eluted ligands to HLA-A*02:01 in the McMurtrey data set (20) were predicted using NetMHCpan-3.0 (21). Peptides with conventional P2-P Ω anchors and predicted rank of up to 10% were considered as canonical binders. A rank score lower than 10% indicates that a peptide is among the 10% strongest binders for HLA-A*02:01 in a large pool of random natural peptides. Peptides that were not predicted to bind canonically, but containing a nested 8-11mer subsequence with predicted high affinity (rank within top 2%) at the N-terminus were classified as extended peptides. Extensions

following the P Ω of unconventional binders were categorized based on the first charged residue of the extension found within the first 3 residues of the extension. Extensions where the first charged residue was an R or K were classified as ‘positive’, while extensions where the first residues were a D or E were considered ‘negative’. As controls, we randomly picked amino acid sequences from the same set of source proteins, with the same length distribution as the observed extensions. We repeated the sampling for random extensions a thousand times to be able to calculate distributions and compared them to the observed extensions.

Derivation of a MHC-II processing motif based on published data-A large set of 16,868 unique eluted HLA class II ligands was downloaded from the IEDB (33) and inspected for amino acid enrichment at the C-terminus. We compared the frequency of the very last residue at the C-terminus in the MHCII ligands and in the *T. gondii* ligands, applying a pseudo-count correction (34) with $\beta=50$. Pseudo-counts exploit information about amino acid similarity to smooth the observed amino acid frequencies of small sequence data sets. This correction has a negligible effect on the large set of MHC II ligands, but it is important in the *T. gondii* dataset because some amino acids were never observed at the C-terminus. Enrichment scores were then calculated as $S_A = \log_2(f_A/q_A)$, where f_A is the pseudo-count corrected frequency for amino acid A, and q_A is the background frequency of A in natural proteins.

Prediction of proteasomal cleavage and TAP transport-Predictions of proteasomal cleavage and TAP transport were obtained using the MHC class I processing tools of the IEDB (35,36) both for ligands predicted to bind canonically and in the extended mode. For the processing predictions, precursors for all ligands were obtained by elongating them by one residue at the N-terminus (to allow for transport of elongated precursors by TAP) and five residues at the C-terminus (to cover the residues thought to impact cleavage by the proteasome), using the context of their source protein. As a control, we also produced proteasome and TAP scores for 100 random natural sequences extracted from Uniprot (<http://uniprot.org>).

Acknowledgements: We would like to thank the support staff at the ALS and SSRL for access to remote data collection.

Conflict of interest: The authors declare that they have no conflict of interest with the contents of this article.

Author contributions: SGM conducted most of the experiments including protein refolding, crystallization and structure determination. GY assisted in protein refolding. MA performed bioinformatic analysis of the *T. gondii* peptide data set with early contributions from TK. CM and WH identified the *T. gondii* peptides. SGM, MN, BP, and DMZ wrote the paper.

REFERENCES

1. Montoyo J. G., L. O. (2004) Toxoplasmosis. *Lancet* **363**, 1965-1976
2. Webster, J. P. (2010) Review of "Toxoplasmosis of Animals and Humans (Second Edition)" by J.P. Dubey. *Parasites & Vectors* **3**, 112
3. Cong, H., Mui, E. J., Witola, W. H., Sidney, J., Alexander, J., Sette, A., Maewal, A., El Bissati, K., Zhou, Y., Suzuki, Y., Lee, D., Woods, S., Sommerville, C., Henriquez, F. L., Roberts, C. W., and McLeod, R. (2012) Toxoplasma gondii HLA-B*0702-restricted GRA7(20-28) peptide with adjuvants and a universal helper T cell epitope elicits CD8(+) T cells producing interferon-gamma and reduces parasite burden in HLA-B*0702 mice. *Hum Immunol* **73**, 1-10
4. Blanchard, N., and Shastri, N. (2010) Topological journey of parasite-derived antigens for presentation by MHC class I molecules. *Trends Immunol* **31**, 414-421
5. Luder, C. G. K., Seeber, F. (2001) Toxoplasma gondii and MHC-restricted antigen presentation: on degradation, transport and modulation. *International Journal for Parasitology* **31**, 1355-1369
6. Leroux, L. P., Dasanayake, D., Rommereim, L. M., Fox, B. A., Bzik, D. J., Jardim, A., and Dzierszynski, F. S. (2015) Secreted Toxoplasma gondii molecules interfere with expression of MHC-II in interferon gamma-activated macrophages. *Int J Parasitol.* **45**, 319-332
7. Grover, H. S., Blanchard, N., Gonzalez, F., Chan, S., Robey, E. A., and Shastri, N. (2012) The Toxoplasma gondii peptide AS15 elicits CD4 T cells that can control parasite burden. *Infect Immun* **80**, 3279-3288
8. Israelski, D. M., Remington, J.S. (1993) Toxoplasmosis in patients with cancer. *Clin Infect Dis.* **16**, S423-435
9. Luft, B. J., Remington, J.S. (1992) Toxoplasmic encephalitis in AIDS. *Clin Infect Dis.* **15**, 211-222
10. Deckert-Schlüter M., S. D., Schmidt D., Schwendemann G., Wiestler O.D., Hof H. (1994) Toxoplasma encephalitis in congenic B10 and BALB mice: impact of genetic factors on the immune response. *Infect Immun.* **62**, 221-228
11. Sher, A., Hakim, F. T., and Gazzinelli, R. T. (1991) CD8 + T cells from mice vaccinated against Toxoplasma gondii are cytotoxic for parasite-infected or antigen-pulsed host cells *Journal of immunology* **147**, 2310-2316
12. Denkers, E. Y. (1999) T lymphocyte-dependent effector mechanisms of immunity to Toxoplasma gondii. *Microbes Infect.* **1**, 699-708
13. Hassan, C., Chabrol, E., Jahn, L., Kester, M. G., de Ru, A. H., Drijfhout, J. W., Rossjohn, J., Falkenburg, J. H., Heemskerk, M. H., Gras, S., and van Veelen, P. A. (2015) Naturally processed non-canonical HLA-A*02:01 presented peptides. *J Biol Chem* **290**, 2593-2603
14. Blanchard, N., Gonzalez, F., Schaeffer, M., Joncker, N. T., Cheng, T., Shastri, A. J., Robey, E. A., and Shastri, N. (2008) Immunodominant, protective response to the parasite Toxoplasma gondii requires antigen processing in the endoplasmic reticulum. *Nat Immunol* **9**, 937-944
15. Cong, H., Mui, E. J., Witola, W. H., Sidney, J., Alexander, J., Sette, A., Maewal, A., and McLeod, R. (2011) Towards an immunosense vaccine to prevent toxoplasmosis: protective Toxoplasma gondii epitopes restricted by HLA-A*0201. *Vaccine* **29**, 754-762
16. Schittenhelm, R. B., Sian, T.C, Wilmann, P.G., Dudek, N.L., Purcell, A.W. (2015) Revisiting the arthritogenic peptide theory: quantitative not qualitative changes in the peptide repertoire of HLA-B27 allotypes. *Arthritis Rheumatol.* **67**, 702-713
17. Burrows, S. R., Rossjohn, J., and McCluskey, J. (2006) Have we cut ourselves too short in mapping CTL epitopes? *Trends Immunol* **27**, 11-16
18. Schaible, U. E., Hagens, K., Fischer, K., Collins, H. L., and Kaufmann, S. H. (2000) Intersection of group I CD1 molecules and mycobacteria in different intracellular compartments of dendritic cells. *J Immunol* **164**, 4843-4852.
19. 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., and 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. *J Biol Chem* **280**, 23900-23909
20. McMurtrey, C., Trolle, T., Sansom, T., Remesh, S. G., Kaefer, T., Bardet, W., Jackson, K., McLeod, R., Sette, A., Nielsen, M., Zajonc, D. M., Blader, I. J., Peters, B., and Hildebrand, W. (2016) Toxoplasma gondii peptide ligands open the gate of the HLA class I binding groove. *Elife* **5**
 21. Rossjohn, J., Gras, S., Miles, J. J., Turner, S. J., Godfrey, D. I., and McCluskey, J. (2015) T cell antigen receptor recognition of antigen-presenting molecules. *Annu Rev Immunol* **33**, 169-200
 22. Andreatta, M., Nielsen, M. (2016) Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics*. **32**, 511-517
 23. Nielsen, M., and Andreatta, M. (2016) NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med* **8**, 33
 24. Saunders, P. M., Vivian, J.P., O'Connor, G.M., Sullivan, L.C., Pymm, P., Rossjohn, J., Brooks, A.G. (2015) A bird's eye view of NK cell receptor interactions with their MHC class I ligands. *Immunol Rev*. **267**, 148-166
 25. Vivier, E., and Anfossi, N. (2004) Inhibitory NK-cell receptors on T cells: witness of the past, actors of the future. *Nature Reviews Immunology* **4**, 190-198
 26. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674
 27. Collaborative Computational Project, N. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*. **50**, 760-763
 28. Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr*. **59**, 1131-1137
 29. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132
 30. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501
 31. Murshudov, G. N., Vagin, A.A., Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*. **53**, 240-255
 32. Tanford, C. (1968) Protein denaturation. *Adv Protein Chem*. **23**, 121-282
 33. Vita, R., Overton, J. A., Greenbaum, J. A., Ponomarenko, J., Clark, J. D., Cantrell, J. R., Wheeler, D. K., Gabbard, J. L., Hix, D., Sette, A., and Peters, B. (2015) The immune epitope database (IEDB) 3.0. *Nucleic Acids Res* **43**, D405-412
 34. Altschul, S. F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. **25**, 3389-3402
 35. Peters, B., Bulik, S., Tampe, R., van Endert, P. M., and Holzhutter, H. G. (2003) Identifying MHC Class I Epitopes by Predicting the TAP Transport Efficiency of Epitope Precursors. *The Journal of Immunology* **171**, 1741-1749
 36. Tenzer, S., Peters, B., Bulik, S., Schoor, O., Lemmel, C., Schatz, M. M., Kloetzel, P. M., Rammensee, H. G., Schild, H., and Holzhutter, H. G. (2005) Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell Mol Life Sci* **62**, 1025-1037

FOOTNOTES

This project has been funded in part by the National Institute of health grants AI128609 (DMZ), AI062629 (subcontract to WHH) and the Immune Epitope Database NIH contract HHSN272201200010C (BP). The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393) and the National Center for Research Resources (P41RR001209). The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

The abbreviations used are: MHC, Major Histocompatibility Complex; HLA, human leukocyte antigen; *Toxoplasma gondii*, *T. gondii*; TCR, T cell receptor; NK, Natural Killer.

TABLE 1. Data collection and refinement statistics

HLA-A*02:01 Peptide (PDB ID)	G9L (5ENW)	G11N (5F7D)	G9V (5FA3)	G13E (5EOT)	Y9L (5F9J)	Y10L (5FDW)	Y16R (5FA4)
Data collection							
Resolution range (Å) ^a	50.0-1.85 (1.89-1.85)	50.0-2.30 (2.38-2.3)	51.6-1.86 (1.89-1.85)	40-2.10 (2.18-2.10)	50.0-2.5 (2.59-2.5)	50-2.7 2.8-2.7	50.0-2.4 (2.44-2.4)
Completeness (%) ^a	93.2 (96.1)	96.6 (81.1)	98.8 (97.5)	99.6 (97.9)	100 (100)	93.4 (95.4)	97.9 (85.2)
Number of unique reflections	35,425	19,954	37,513	26,570	15,320	11,732	17,776
Redundancy	2.7	3.4	3.7	3.6	3.7	2.8	3.4
R _{sym} (%)	8.6 (53.1)	7.1 (33.5)	7.8 (31.1)	12.8 (57.0)	16.1 (71.1)	19.0 (66.4)	13.5 (66.3)
R _{pim} (%)	6.1 (38.5)	4.5 (21.7)	4.7 (18.8)	7.8 (37.7)	9.7 (42.9)	13.1(45.4)	8.4 (42.6)
I/σ ^a	21.5 (3.1)	21.2 (3.3)	20.3 (4.6)	13.3 (2.1)	7.9 (2.1)	8.3 (2.4)	11.3 (1.7)
Refinement statistics							
Number of reflections (F>0)	33,590	18,963	35,290	25,169	14,464	11,305	16,851
Maximum resolution (Å)	1.85	2.3	1.86	2.1	2.51	2.7	2.4
R _{cryst} (%)	20.8 (25.1)	20.9 (36.9)	20.9 (23.9)	20.9 (30.8)	19.9 (26.3)	21.3 (24.3)	20.4 (35.5)
R _{free} (%)	24.4 (29.4)	25.7 (34.1)	23.4 (29.4)	23.7 (31.3)	27.5 (33.3)	28.5 (32.9)	25.8 (36.3)
Number of atoms							
Protein	3371	3151	3397	3227	3170	3142	3210
Peptide	3047	3011	3060	3015	3012	2996	3032
Glycerol	62	70	59	66	67	75	77
Solvent molecules (waters)	3	2	3	0	3	1	2
Ion	241	57	280	143	71	62	83
Ramachandran statistics (%)							
Favored	98.7	97.6	98.7	98.9	97.6	96.0	97.9
Outliers	0.0	0.0	0.0	0.0	0.0	0.0	0.0
R.m.s.d. from ideal geometry							
Bond length (Å)	0.0064	0.0075	0.0061	0.0076	0.01	0.01	0.009
Bond angles (°)	1.12	1.22	1.11	1.22	1.45	1.46	1.33
Average B values (Å²)							
Protein	29.6	48.5	15.1	27.5	26.3	20.8	36.9
Peptide	17.8	48.2	13.0	26.2	25.8	26.3	28.8
Water molecules	29.4	41.7	24.8	26.7	21.0	15.8	35.2

^aNumber in parenthesis refer to highest resolution shell

FIGURE LEGENDS

FIGURE 1. HLA-A*02:01-peptide structures. a) Overall structure of HLA-A*02:01 with heavy chain in gray and light chain (β 2M) in blue. Peptide backbones are superimposed within the binding groove. b-d) Molecular surface representation of the binding groove of HLA-A*02:01 with the individual peptides bound. Peptide sequences are labeled and charged amino acids are colored in Blue (positive) and red (negative). 2FoFc Electron density for the peptide (shown in blue mesh) is contoured at 1σ .

FIGURE 2. Detailed hydrogen bond interactions around the F pocket. HLA-A*02:01 is shown in grey, peptides in color and electron density for Lys146 as a blue mesh contoured at 1σ . Hydrogen bonds between 2.5-3.65 Å are shown as blue dashed lines. Structures of nested peptides are shown in the top panel and the corresponding extending peptides below. a) Peptide pair GLKEGIPAL (green) and GLKEGIPALD (cyan); b) GLLPELPAV (pink) and GLLPELPAVGGNE (grey); c) peptide pair YLSPIASPL (violet) and YLSPIASPLLDGKSLR (orange).

FIGURE 3. Thermal Denaturation assay. First derivative of the melt curve for individual HLA-A*02:01-peptide complex. Calculated melting temperatures for each complex are provided in parenthesis.

FIGURE 4. Binding comparison between nested and extending peptides. a) Binding groove of HLA-A*02:01 with the nested peptide GLKEGIPAL (green) superimposed with extended peptide GLKEGIPALD (cyan). b) Superimposition of nested peptide GLLPELPAV (pink) with extended peptide GLLPELPAVGGNE (grey). c) Superimposition of nested peptide YLSPIASPL (violet) with extended peptide YLSPIASPLLDGKSLR (orange). d) Superimposition of nested peptide YLSPIASPL (violet) with nested peptide YLSPIASPLL (yellow).

FIGURE 5. Mechanisms of F pocket opening and peptide sequences. a) ‘Tyrosine swing’: Y84 adopts a different rotamer to open the F pocket and accommodate longer peptide FVLELEPEWTVK (UFP¹⁶⁻²⁷) (15). b) ‘Lysine lift’: Lifting of residue K146 to accommodate longer peptides including GLKEGIPALD, GLLPELPAVGGNE and YLSPIASPLLDGKSLR. c) List of nested and longer *T. gondii* peptides. Positively charged residues (blue) open the F pocket via the ‘Tyrosine swing’, while negatively charged residues (red) open the binding groove via the ‘Lysine lift’.

FIGURE 6. Properties of ligands with C-terminal extensions. a) Negatively charged extensions (D,E) in the *T. gondii* eluted set were more frequent than expected compared to a background distribution of resampled control datasets, whereas positively charged extensions (R,K) were less frequent. b) Positive extensions were on average shorter than 3 residues, negative extensions had a mean length larger than 8 residues. c) TAP transport scores for extended ligands were significantly lower than for canonical ligands, and not significantly different from random. d) Proteasome cleavage scores for extended ligands were significantly lower than for canonical ligands, and not significantly higher than random. e) The C-terminal composition of *T. gondii* ligands correlates with the C-terminal enrichment of class II ligands, with positively charged residues in the first quadrant and hydrophobic amino acids dominating the third quadrant.

Figure 1

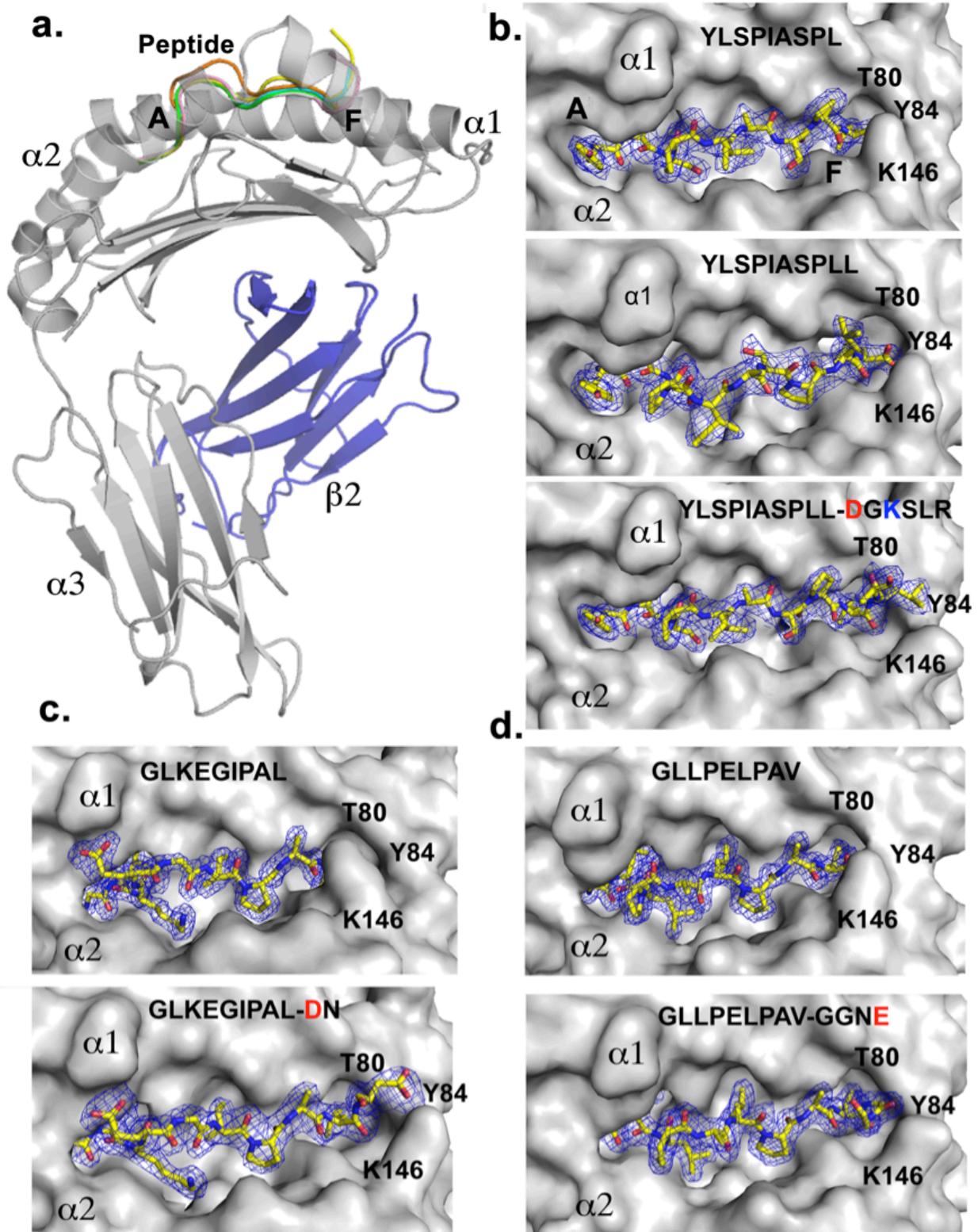


Figure 2

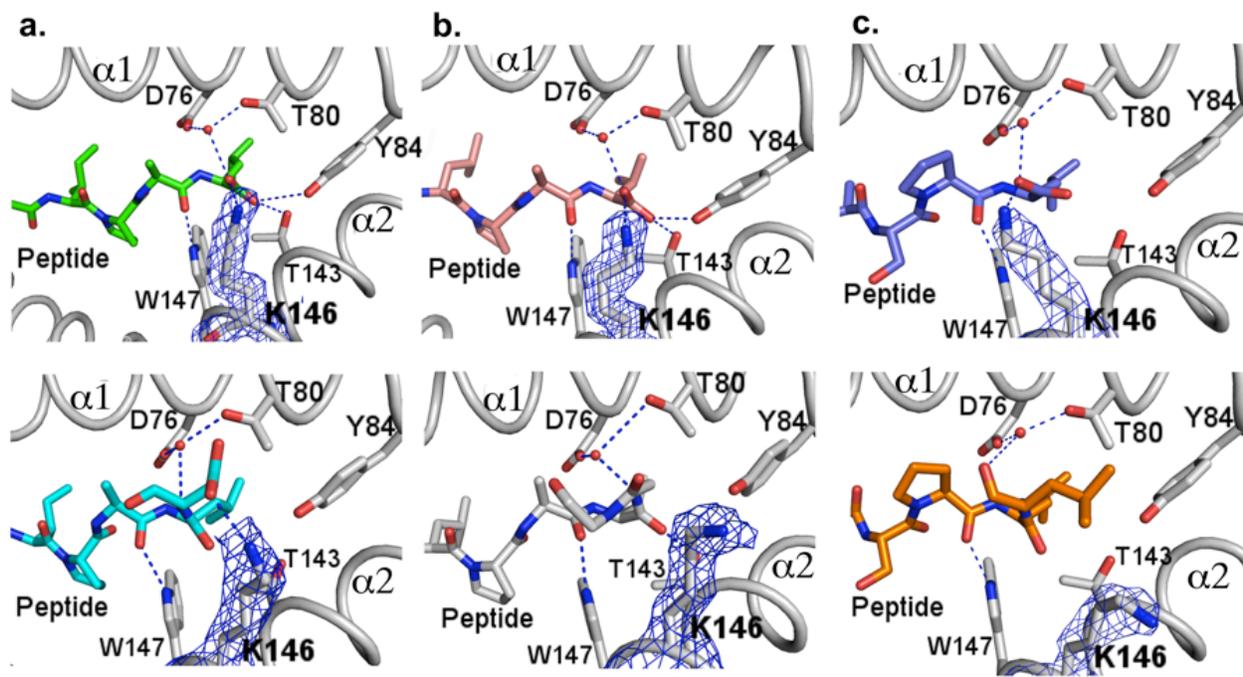


Figure 3

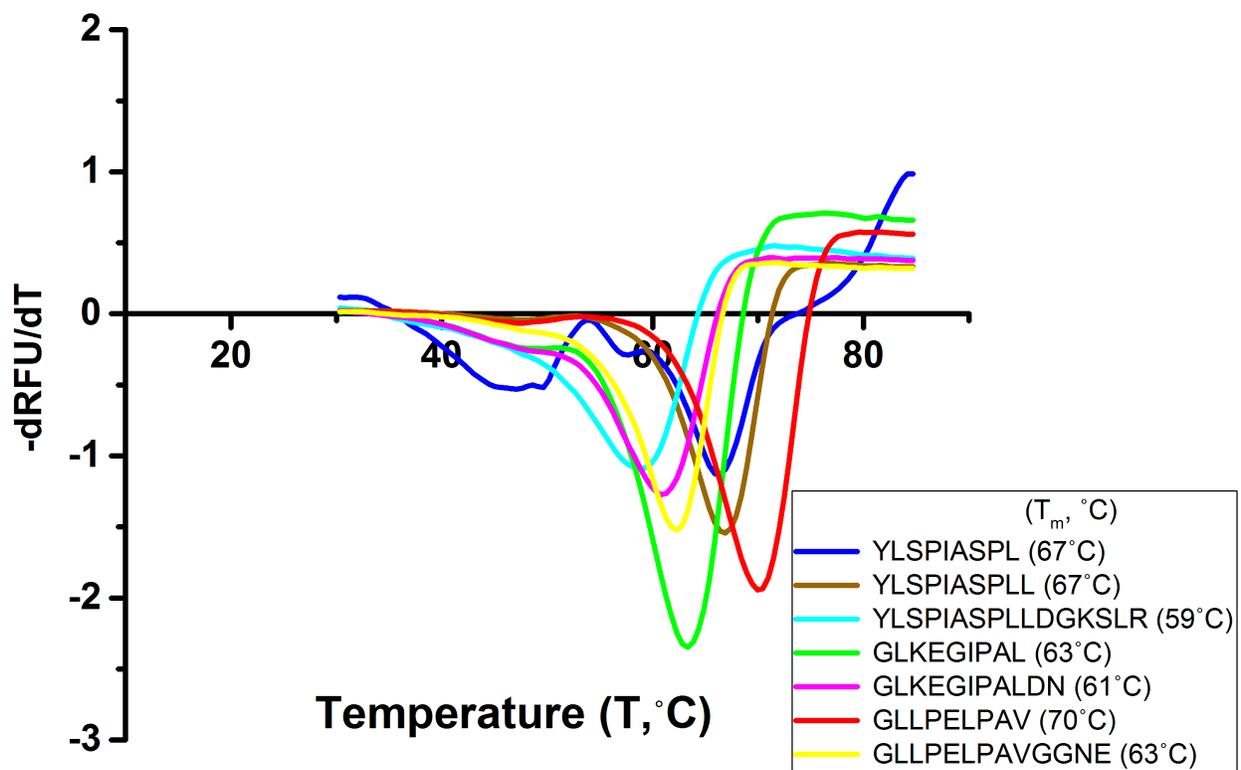


Figure 4

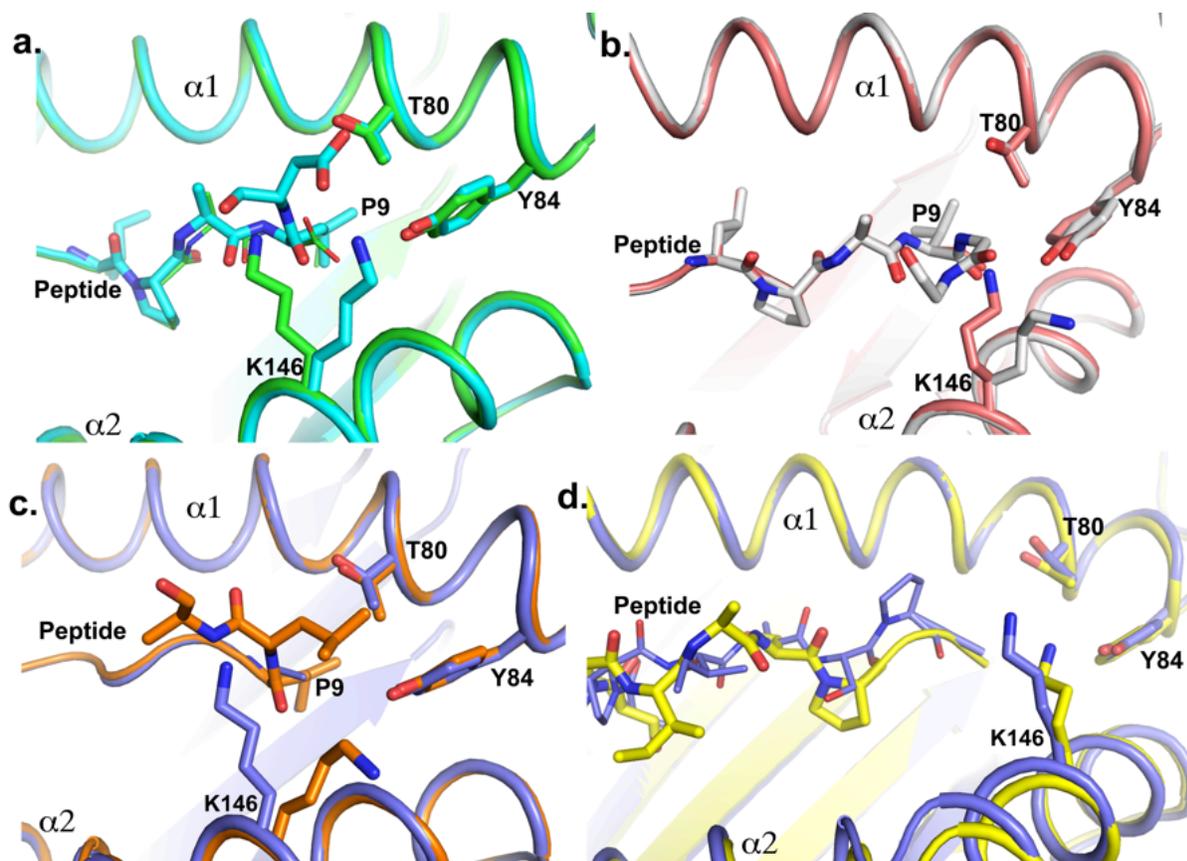


Figure 5

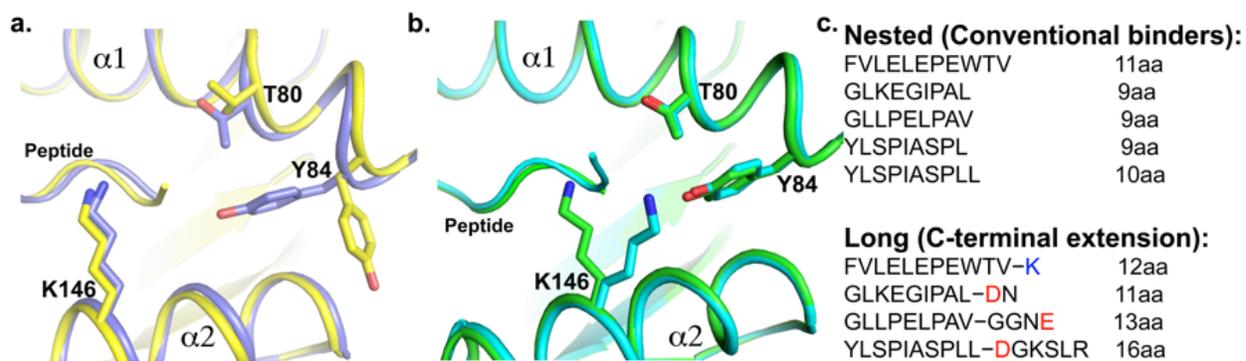
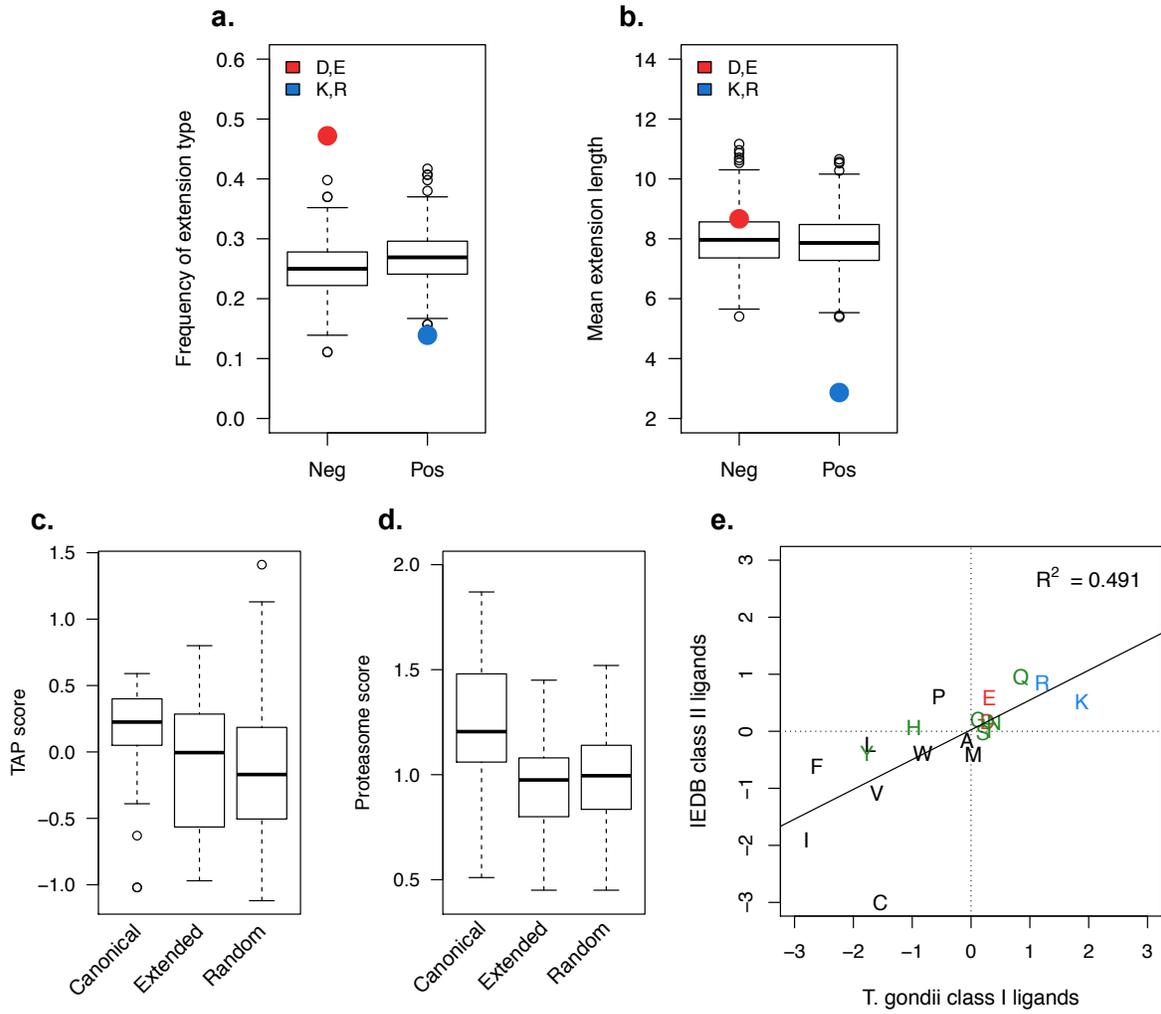


Figure 6



Breaking confinement: unconventional peptide presentation by major histocompatibility (MHC) class I allele HLA-A*02:01

Soumya G. Remesh, Massimo Andreatta, Ge Ying, Thomas Kaefer, Morten Nielsen, Curtis McMurtrey, William Hildebrand, Bjoern Peters and Dirk M. Zajonc

J. Biol. Chem. published online February 8, 2017

Access the most updated version of this article at doi: [10.1074/jbc.M117.776542](https://doi.org/10.1074/jbc.M117.776542)

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at <http://www.jbc.org/content/early/2017/02/08/jbc.M117.776542.full.html#ref-list-1>