Unconventional Peptide Presentation by Major Histocompatibility Complex (MHC) Class I Allele HLA-A*02:01

BREAKING CONFINEMENT

Soumya G. Remesh*, Massimo Andreatta§, Ge Ying‡, Thomas Kaever¶, Morten Nielsen∥§, Curtis McMurtrey++‡‡, William Hildebrand∗∗††, Bjoern Peters§, and Dirk M. Zajonc§§¶¶†††

From the ∗Division for Cell Biology and §Division of Vaccine Discovery, La Jolla Institute for Allergy and Immunology, La Jolla, California 92037, §§Instituto de Investigaciones Biotecnológicas, Universidad Nacional de San Martín, CP1650 San Martín, Argentina, †Center for Biological Sequence Analysis, Department of Bio and Health Informatics, The Technical University of Denmark, 2800 Lyngby, Denmark, ¤¤Department of Microbiology and Immunology, University of Oklahoma Health Science Center, Oklahoma City, Oklahoma 73104, ‡‡Pure MHC LLC, Austin, Texas 78229, and §§¶¶†††Department of Internal Medicine, Faculty of Medicine and Health Sciences, Ghent University, 9000 Ghent, Belgium

Edited by Luke O’Neill

Peptide antigen presentation by major histocompatibility complex (MHC) class I proteins initiates CD8+ T cell-mediated immunity against pathogens and cancers. MHC I molecules typically bind peptides with 9 amino acids in length with both ends tucked inside the major A and F binding pockets. 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 zigzag 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 (Pφ1) and contained charged amino acids not more than 3 residues after the anchor amino acid at Pφ2, 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. Although positively charged amino acids result in the Tyr-84 swing, amino acids that are negatively charged induce a not previously described Lys-146 lift. Furthermore, 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). Although 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 complex class I (MHC I) and class II (MHC II) proteins (3–5). Although T. gondii can interfere with CD4 T cell responses by down-regulating MHC II expression in IFN-γ-activated macrophages, immunization with T. gondii MHC II peptide ligands can elicit a 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 be secreted into either 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 (α) chain much like the canonical binders with some changes. The second (P2) and C-terminal (Pφ3) residues of the antigen peptide anchor into the A and F pockets of the binding
Peptide-induced Opening of the F Pocket of MHC Class I

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 Tyr-84 of HLA-A*02:01 (20). In contrast, in the current study, we examined three sets of peptides that had C-terminal amino acid additions that contained negatively charged amino acids or both negatively and positively charged residues (Fig. 1). To investigate whether 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 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, whereas C-terminally extending residues that did not contact HLA-A*02:01 were disordered (Fig. 1). When all the different peptides are compared, slight structural changes are observed throughout the binding groove to allow optimal binding of the different amino acid side chains.

In contrast to these “bulged” peptides, we 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 Tyr-84 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, whereas the nested 11-mer N-terminal core peptide bound in a conventional zigzag 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 nested and extending 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 Tyr-84 and Lys-146. We suggest that these unconventional modes of binding will help better understand targets of MHC class I-restricted epitope recognition.
### TABLE 1
Data collection and refinement statistics

*Numbers in parentheses refer to highest resolution shell.*

<table>
<thead>
<tr>
<th>Data collection</th>
<th>HLA-A*02:01-peptide (Protein Data Bank code)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G9L (5ENW)</td>
</tr>
<tr>
<td>Resolution range (Å)*</td>
<td>50.0–1.85 (1.89–1.85)</td>
</tr>
<tr>
<td>Completeness (%)*</td>
<td>93.2 (96.1)</td>
</tr>
<tr>
<td>Number of unique reflections</td>
<td>35,425</td>
</tr>
<tr>
<td>Redundancy</td>
<td>2.7</td>
</tr>
<tr>
<td>$R_{sym}$ (%)</td>
<td>8.6 (53.1)</td>
</tr>
<tr>
<td>$R_{sym}$ (%)</td>
<td>6.1 (38.5)</td>
</tr>
<tr>
<td>$I/\sigma$*</td>
<td>21.5 (3.1)</td>
</tr>
<tr>
<td>Refinement statistics</td>
<td></td>
</tr>
<tr>
<td>Number of reflections ($F &gt; 0$)</td>
<td>33,590</td>
</tr>
<tr>
<td>Maximum resolution (Å)</td>
<td>1.85</td>
</tr>
<tr>
<td>$R_{cryst}$ (%)</td>
<td>20.8 (25.1)</td>
</tr>
<tr>
<td>$R_{free}$ (%)</td>
<td>24.4 (29.4)</td>
</tr>
<tr>
<td>Number of atoms</td>
<td>3.37</td>
</tr>
<tr>
<td>Protein</td>
<td>3.047</td>
</tr>
<tr>
<td>Peptide</td>
<td>62</td>
</tr>
<tr>
<td>Glycerol</td>
<td>3</td>
</tr>
<tr>
<td>Solvent molecules (waters)</td>
<td>241</td>
</tr>
<tr>
<td>Ramachandran statistics</td>
<td></td>
</tr>
<tr>
<td>Favored</td>
<td>98.7</td>
</tr>
<tr>
<td>Outliers</td>
<td>0.0</td>
</tr>
<tr>
<td>r.m.s.d. from ideal geometry</td>
<td></td>
</tr>
<tr>
<td>Bond length (Å)</td>
<td>0.0064</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.12</td>
</tr>
<tr>
<td>Average B values (Å²)</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>29.6</td>
</tr>
<tr>
<td>Peptide</td>
<td>17.8</td>
</tr>
<tr>
<td>Water molecules</td>
<td>29.4</td>
</tr>
</tbody>
</table>
Peptide-induced Opening of the F Pocket of MHC Class I

![Image of hydrogen bond interactions around the F pocket.](Image)

**FIGURE 2. Detailed hydrogen bond interactions around the F pocket.** HLA-A*02:01 is shown in gray, peptides are shown in color, and electron density for Lys-146 is shown as a blue mesh contoured at 1σ. Hydrogen bonds between 2.5 and 3.65 Å are shown as blue dashed lines. Structures of nested peptides are shown in the top panel, and the corresponding extending peptides are shown below. a, peptide pair GLKEGIPAL (green) and GLKEGIPALD (cyan). b, GLLPELPAV(pink) and GLLPELPAVGGNE (gray). c, peptide pair YLSPIASPL (violet) and YLSPIASPLDLLGKSLR (orange).

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 P1 valine and Lys-146 is missing in the crystal structure with the longer peptide (Fig. 2b). The same is true for the peptide pair YLSPIASPL and YLSPIASPLDLLGKSLR (Fig. 2c). As a result, the change in the orientation of Lys-146 leads to the loss of hydrogen bond formation with the carboxylate of the P1 amino acid. However, the hydrogen bond interaction between HLA-A*02:01 residue Trp-147 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 P1+1 (here Asp-10). Because 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 (Fig. 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 (Fig. 3). We observed 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, whereas with its longer peptide counterpart is 61 °C. Addition of 6 extra residues to peptide YLSPIASPL also only changes the T_m of the complex with peptide YLSPIASPLDLLGKSLR by about 7 °C (Fig. 3 and Ref. 20). Interestingly, some of the complexes with nested peptides are as stable as those with longer peptides (compare GLKEGIPAL with GLLPELPAVGGNE; Fig. 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 with peptide GLLPELPAV (Fig. 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 Tyr-84 and their terminal carboxylate (3.65 Å for G9L and 2.9 Å for G9V; Fig. 2). The lack of an intimate hydrogen bond interaction of the terminal amino acid with Tyr-84 in G9L peptide is likely a major contributor to the reduced melting temperature. Compared with 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 GLKEGIPALD; Fig. 3). However, longer peptide additions (4–6 amino acids) that contain a negatively charged amino acid (YLSPIASPL versus YLSPIASPLDLLGKSLR and GLLPELPAV versus GLLPELPAVGGNE) reduce the protein-peptide complex to the same extend as the single “Lys” addition found in peptide FVLELEPEWTVK (20). This highlights that the structural change involving Tyr-84 of HLA-A*02:01 is more destabilizing than that of Lys-146 when the negatively charged peptide extension is very short.

**Lysine 146 Lift**—The different orientations of Lys-146 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 because Lys-146 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 (P1) of any nested peptide (Figs. 2 and 4). Although the position of Lys-146 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 the case of the crystal structure of HLA-A*02:01 with YLSPIASPL, YLSPIASPL, and YLSPIASPLDLLGKSLR, 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 P1 anchor residues binding to the A and F pockets. Surprisingly, however, YLSPIASPL, the nested peptide with one extra leucine residue at the C-terminal end of the peptide, undergoes a certain extent of bulging to accommodate the terminal leucine residue (P1+1 instead of P1) as the anchor residue in the F pocket (Fig. 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 Lys-146 to open the F pocket. Because a mere increase
These 108 peptides are expected to be enriched for examples of several ligands binding in an unconventional mode, we induce in HLA-A*02:01. Thus, in addition to the previously identified "Tyr-84 swing" to accommodate the peptide FVLELEPEWTVK (UFP(16–27)) (20), we observed a “Lys-146 lift” in HLA-A*02:01 as a second mechanism of opening the F pocket induced by the extending peptides YLSPIASPL (violet) with extended peptide YLSPIASPLDGKSLR (orange). d, superimposition of nested peptide YLSPIASPL (violet) with extended peptide YLSPIASPLDGKSLR (yellow).

in length of the peptide does not cause the change in orientation of Lys-146, it is likely that the addition of charged residues within the C-terminal extension is the contributing factor to open the F pocket. Thus, 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 3 residues of the extension, whereas positively charged extensions were less frequent (Fig. 6a). Positive extensions were short (less than 3 residues on average), whereas negative extensions had an average length of 8.7 residues (Fig. 6b). Notably, 66% of the positive extensions consisted of a single or 2 residues compared with only 8% of negative extensions being shorter than 3 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 life cycle in the cells it infects, we wanted to examine whether the unconventional ligands found had the typical motifs of peptides derived from the conventional MHC class I processing and presentation pathway. As shown in Fig. 6, this was not the case. Extended ligands had significantly lower scores for TAP2 transport (Fig. 6c) compared with canonical T. gondii ligands (p = 1.2 × 10^{-5}, Wilcoxon rank sum test) but were not significantly different from those of random peptides (p = 0.34). Similarly, proteasome cleavage scores (Fig. 6d) for extended ligands were significantly lower compared with canonical ligands (p = 5 × 10^{-16}) but were 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 life cycle 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 of 108). If these ligands were cross-presented, we would expect similar trimming motifs in class II ligands, and we examined published data sets 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

Extended ligands had significantly lower scores for TAP2 transport (Fig. 6c) compared with canonical T. gondii ligands (p = 1.2 × 10^{-5}, Wilcoxon rank sum test) but were not significantly different from those of random peptides (p = 0.34). Similarly, proteasome cleavage scores (Fig. 6d) for extended ligands were significantly lower compared with canonical ligands (p = 5 × 10^{-16}) but were 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 life cycle 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 of 108). If these ligands were cross-presented, we would expect similar trimming motifs in class II ligands, and we examined published data sets 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

The abbreviations used are: TAP, transporter associated with antigen processing; TCR, T cell receptor; KIR, killer immunoglobulin receptor; β₂m, human β₂-microglobulin.
terminus of eluted MHC class II ligands where positively charged amino acids were also enriched (Fig. 6e). Taken together, we found that the unconventional MHC I ligands presented by *T. gondii* have sequence motifs much more consistent with cross-presentation than with generation through proteasomal cleavage and TAP transport.

**Discussion**

αβ 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 chains in a diagonal orientation above the MHC molecule (28). Although the germ line-encoded complementarity-determining region 1 and 2 loops bind to MHC, the hypervariable loops complementarity-determining regions 3α and 3β 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, whereas 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. Because MHC I presents peptides in a more confined space compared with MHC II, the TCR of CD8+ T cells often contacts and discriminates their entire peptide sequence (21). In our present study, we have focused on a
Peptide-induced Opening of the F Pocket of MHC Class I

Panel of C-terminally extended T. gondii peptides that contain a negatively charged amino acid within their C-terminal extensions. Analysis of a large data set 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 termini, but addition of the C-terminal extensions renders predictions about their binding to MHC I difficult. Although the N termini 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 motif, whereas adding descriptors, such as charged residues in the C-terminal extension, is necessary 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 characterized by non-canonical binding available. We would envision this situation to change as more binding data become available and as MHC ligand data are included in 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 fewer 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 peptides at least for HLA-A*02:01.

Because the residues that are involved in the F pocket opening (Tyr-84 and Lys-146) are conserved across all HLA-A, -B, and -C alleles (and found in many non-classical MHC I molecules, such as HLA-E, HLA-G, and Qa-1 as well as viral MHC I and -C alleles (and found in many non-classical MHC I molecules), such as HLA-E, HLA-G, and Qa-1) as well as viral MHC I and -C alleles), we postulate that the residue can form a universal characteristic found across many MHC I molecules. Although 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 natural killer cells, natural killer T 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). In particular, 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 β2-microglobulin (β2m) (residues 1–99) were expressed as inclusion bodies and refolded as reported previously (20) with modifications reported here. Briefly, both the heavy chain and light chain were expressed in Escherichia coli BL21 DE3 cells and induced at A600 of 0.6 with 1 mM isopropyl-1-thio-β-d-galactopyranoside, and cells were harvested after 4 h by centrifugation (5000 × g for 20 min). Cells were resuspended separately in lysis buffer (100 mM Tris-HCl, pH 7.0, 5 mM EDTA, 5 mM DTT, 0.5 mM PMSE), and the cells were broken with four to five passes through a microfluidizer (20 kilopascals) (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 mM 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 mM guanidine HCl) for subsequent refolding. 3 mg of β2m 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 h. Between 11 and 15 mg of HLA-A heavy chain mixed with 2–3 mg of individual peptide (GenScript) was then added to the refolding mixture and further stirred at 4 °C for 72 h. Final heavy chain:light chain:peptide ratios were in the range of 2:1:12 and 2.5:1:12 for the 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 using a Superdex S200 HR16/60 gel filtration column pre-equilibrated with size exclusion chromatography buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl). Fractions containing refolded HLA-A*02:01-β2m-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-UFP(16–27) complex obtained in 1.2 M sodium citrate were used to cross-seed the other complexes. The complexes were equilibrated in 30% PEG 4000, 0.1 M Tris-HCl, pH 8.0, 0.2 M lithium sulfate for 1–2 h by mixing 0.15 μl of complex and 0.15 μl 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 Lightsource and processed to 1.86, 2.3, 2.1, 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 and processed to 1.85, 2.5, 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 (Protein Data Bank code 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-β2m-peptide complexes with the various peptides were analyzed for thermal denaturation by differential scanning fluorimetry using a LightCycler 480 (Roche Applied Science). HLA-A*02:01-β2m-peptide complexes with different peptides at 100 μM in reaction buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl) were used as stock solutions. Each reaction mixture constituted 1–2 μl of protein complex stock solution and 2 μl of SYPRO Orange dye (100×; Invitrogen) made up to 20 μl in reaction buffer in a 96-well white plate compatible with the instrument. A temperature gradient from 20 to 85 °C at steps of 0.06 °C/s and 10 acquisitions/°C was used for the experiment. Each experiment with an 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 $T_m$ for individual HLA-A*02:01-β2m-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 data set of McMurtrey et al. (20) were predicted using NetMHCpan-3.0 (21). Peptides with conventional P2-P9 anchors and predicted rank of up to 10% were classified 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 contained a nested 8–11-mer subsequence with predicted high affinity (rank within top 2%) at the N terminus were classified as extended peptides. Extensions following the P9 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 Arg or Lys were classified as positive, whereas extensions where the first residue was a Asp or Glu 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 with 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 Immune Epitope Database (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 MHC II ligands and in the T. gondii ligands, applying a pseudocount correction (34) with $\beta = 50$. Pseudocounts 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 data set 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 pseudocount corrected frequency for amino acid $A$ and $q_A$ is the background frequency of $A$ in natural proteins.

**Prediction of Proteosomal Cleavage and TAP Transport**—Predictions of proteosomal cleavage and TAP transport were obtained using the MHC class I processing tools of the Immune Epitope Database (35, 36) for ligands predicted to bind both canonically and in the extended mode. For the processing predictions, precursors for all ligands were obtained by elongating them by 1 residue at the N terminus (to allow for transport of elongated precursors by TAP) and 5 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.

**Author Contributions**—S. G. M. conducted most of the experiments including protein refolding, crystallization, and structure determination. G. Y. assisted in protein refolding. M. A. performed bioinformatics analysis of the T. gondii peptide data set with early contributions from T. K. C. M. and W. H. identified the T. gondii peptides. S. G. M., M. N., B. P., and D. M. Z. wrote the paper.

**Acknowledgments**—We thank the support staff at the Advanced Light Source and Stanford Synchrotron Radiation Lightsource for access to remote data collection. The Stanford Synchrotron Radiation Lightsource Structural Molecular Biology Program is supported by the United States Department of Energy Office of Biological and Environmental Research and by the National Institute of General Medical Sciences (including Grant P41GM103393) and National Center for Research Resources (Grant P41RR001209), National Institutes of Health. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the United States Department of Energy under contract DE-AC02-05CH11231.

**References**

Peptide-induced Opening of the F Pocket of MHC Class I

11. Hakim, F. T., Gazzinelli, R. T., Denkers, E., Hieny, S., Shearer, G. M., and Sher, A. (1991) CD8+ T cells from mice vaccinated against Toxoplasma gondii are cytotoxic for parasite-infected or antigen-pulsed host cells. J. Immunol. 147, 2310–2316