Structural Definition of the H-2K\textsuperscript{d} Peptide-binding Motif*

Received for publication, September 26, 2005, and in revised form, February 2, 2006. Published, JBC Papers in Press, February 10, 2006, DOI 10.1074/jbc.M510511200

Vesselin Mitaksov\textsuperscript{2} and David H. Fremont\textsuperscript{3,5}

From the \textsuperscript{1}Department of Pathology and Immunology and \textsuperscript{5}Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110

Classic major histocompatibility complex (MHC)\textsuperscript{2} proteins serve a critical role in the adaptive immune response by binding short peptide fragments intracellularly and presenting them at the cell surface for surveillance by cytotoxic T lymphocytes (1–4). Structural studies of human and murine MHC class I proteins in complex with a variety of peptides have revealed conservative structural elements that promote efficient binding and presentation of peptide epitopes (5–9). Peptides of 8–10 residues are bound in a predominantly extended conformation within a narrow groove formed by two antiparallel \(\beta\)-helices positioned above an eight-strand \(\beta\)-sheet platform. Conservative hydrogen bonding networks are established in the binding groove with peptide mainchain and terminal atoms that enable largely sequence-independent ligation.

Although low affinity, kinetically short-lived peptide-MHC complexes can be established by highly diverse epitope sequences, stable association requires the anchoring of peptide side chains into specific pockets in the MHC groove. MHC polymorphisms are clustered in these pockets (3, 10, 11), and their shape and chemistry impose constraints that are reflected by allele-specific motifs found in the sequences of naturally processed peptides (12, 13). For example, H-2\textsuperscript{d} cell lines (P815) and H-2\textsuperscript{b} cell lines (EL4) infected with the same influenza virus present different antigenic peptides for CTL recognition (14). Thus, outbred populations that express varied MHC proteins can survey diverse peptide fragments from a given pathogen despite specificity constraints imposed by each individual allele.

The sequences of a number of naturally processed peptides that are presented by K\textsuperscript{d} have been identified, including self-peptides (12, 15–21, 60) and those encoded by viruses (14, 22, 23), parasites (24, 25), and bacteria (26). The first virally encoded T-cell epitope ever described was in fact a K\textsuperscript{d}-binding peptide derived from influenza A/PR/8/34 nucleoprotein (27). Although early studies using synthetic peptides suggested that an 11-residue peptide is presented (28), sequencing of the naturally processed peptide from virally infected cells revealed that the nucleoprotein epitope is only nine residues long (residues 147–155, TYQTRALV) (14). In fact, the vast majority of K\textsuperscript{d}-binding peptides are nine residues in length, which nearly invariably contain Tyr at the second position (P2) (29–31).

To resolve the structural underpinnings of the dominant K\textsuperscript{d}-binding motif we have undertaken crystallographic studies of K\textsuperscript{d} in complex with the antigenic peptide from influenza virus nucleoprotein (Flu). The 2.6-Å resolution structure of K\textsuperscript{d}-Flu provided an excellent framework to delineate the role of polymorphic anchoring pockets in determining K\textsuperscript{d}-specific peptide binding. To extend our understanding of the overall binding motif to a broad population of K\textsuperscript{d} epitopes, we analyzed 95 naturally processed K\textsuperscript{d} peptides in conjunction with our structural data. Comparisons of K\textsuperscript{d}-Flu to other class I peptide-MHC complexes reveal that the conformation of Flu in the K\textsuperscript{d} groove is similar to that of peptides associated with D\textsuperscript{d} despite differences in anchoring strategies between the two MHC proteins. Lastly, our structural studies provide a detailed framework for understanding the role of individual peptide residues in T-cell recognition events.

MATERIALS AND METHODS

Expression and Purification—The extracellular domains of K\textsuperscript{d} (heavy chain, residues 1–283; murine \(\beta\)-2-microglobulin (m\(\beta\)-m), residues 1–99, with signal peptides omitted) were expressed separately in the bacterial strain BL21 CodonPlus\textsuperscript{\textregistered} RIL (DE3)RIL (Stratagene) as insoluble inclusion bodies. LB media (8 liters, 37 °C) was inoculated from a single colony, and protein expression was induced at A\textsubscript{600} of 0.8 with 0.5 mM isopropyl 1-thio-\(\beta\)-D-galactopyranoside. Cells were harvested and suspended in 200 ml of buffer containing 50 mM Tris, pH 8.0, 25% (w:v) sucrose, 1 mM EDTA, 10 mM dithiothreitol, and 0.01% (w:v) NaN\textsubscript{3}. Lysozyme (0.4 mg/ml), DNase I (40 \(\mu\)g/ml), and MgCl\textsubscript{2} (10 mM) were added to the suspension, and the cells were lysed by the addition of 200 ml of buffer containing 50 mM Tris, pH 8.0, 50% (v:v) sucrose, 1 mM EDTA, 10 mM dithiothreitol, and 0.01% (w:v) NaN\textsubscript{3}.


*This work was supported by National Institutes of Health Grant GM62414–04 (to D. H. F.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

\textsuperscript{1}To whom correspondence should be addressed: Dept. of Pathology and Immunology, Washington University School of Medicine, Box 8118, 660 South Euclid Ave, St. Louis, MO 63110. Tel: 314-747-6547; Fax: 314-362-8888; E-mail: fremont@immunology.wustl.edu.

\textsuperscript{2}The abbreviations used are: MHC, major histocompatibility complex; Flu, antigenic peptide derived from Influenza A/PR/8/34 nucleoprotein residues 147–155; TYQTRALV; TCR, T-cell receptor; m\(\beta\)-m, murine \(\beta\)-2-microglobulin; CTL, cytotoxic T lymphocyte; r.m.s.d., root mean square deviation; ERK, extracellular signal-regulated kinase.
Crystal Structure of Kd-Flu

The purified, detergent-free, inclusion bodies were solubilized overnight in 6 M Gdn-HCl, 10 mM Tris, pH 8.0, and 10 mM β-mercaptoethanol. To form the Kd-Flu complex, murine β1, and heavy chain were refolded under oxidative conditions in the presence of 10 molar excess of Flu (influenza A/PR8/34 nucleoprotein residues 147–155, TYQRTRALV). Refolding was performed at 4 °C using a rapid dilution method. Briefly, Flu was diluted to 15 μM in 500 ml of refolding buffer (100 mM Tris, pH 8.0, 400 mM l-Arg, 2.0 mM EDTA, 0.5 mM GSSG, 5.0 mM GSH, and protease inhibitors). Murine β1 was injected into the refolding reaction to concentration of 4.5 μM. Following 30 min of incubation the heavy chain (final concentration 15 μM) was injected in three separate batches spaced over a 24-h period. The final concentration of Gdn-HCl in the refolding reaction did not exceed 100 mM. After an overnight incubation the refolding reaction was concentrated to 4 ml using an Amicon ultrafiltration device (Millipore, Billerica, MA).

The Kd-Flu complex was purified from protein aggregates and other impurities on a Superdex75 (GE Healthcare, Piscataway, NJ) size exclusion column using a running buffer containing 20 mM HEPES, pH 7.5, 150 mM NaCl, and 0.01% NaN3. The fractions containing Kd-Flu were pooled, dilute 3-fold in buffer containing 20 mM Tris, pH 8.5, loaded on an anion exchange Mono Q column (GE Healthcare), and eluted with a NaCl gradient (0 mM to 400 mM NaCl over 30 ml). Prior to crystallization the pure Kd-Flu complex was washed with buffer containing 50 mM Tris, pH 8.0, 0.5% Triton X-100, 1% sodium deoxycholate, 100 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol, and 0.01% NaN3. After lysis, EDTA was added (12.5 mM) and insoluble protein was pelleted by centrifugation. The inclusion bodies were washed three times with buffer containing 50 mM Tris, pH 8.0, 0.5% Triton X-100, 100 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol, and 0.01% NaN3. To remove detergent the inclusion bodies were washed twice with buffer as described above but without Triton X-100. Protein purity was confirmed by SDS-PAGE.

Crystal Structure and Data Collection—Diffraction-quality crystals of Kd-Flu were obtained by hanging-drop vapor diffusion method. Protein at 6–8 mg/ml was equilibrated at 20 °C against 12% (w/v) polyethylene glycol 2000, 5% (w/v) 2-methyl-2,4-pentanediol, and 0.01% NaCl at pH 6.9. Small crystals obtained in these drops were grown overnight and grew over 3–4 days. Crystal space group P212121 with a cell of 249 frames was collected each representing a 1.5° oscillation geometry (Table 1).

Refinement—The highest signal (correlation coefficient = 33.1% and R value = 49.4% for all 15–4.0 Å data) was obtained in the P212121 space group, consistent with the systematic absences in the data. Rigid body refinement of the three domains of the search model (α, α2, α3, and murine β1) against the Kd-Flu data yielded an R value of 41.8% for the 20- to 2.6-Å resolution data. Extensive model building was performed with the molecular modeling program O (O version 6.22, Uppsala Software Factory, Sweden) using 2Fobs – Fcalc, Fcalc – Fobs, and composite omit maps (CNS, Yale University, New Haven CT (33)). Atomic refinement was done employing simulated annealing, energy minimization, and restrained B-factor refinement protocols as implemented in CNS. The final model includes a total of 383 residues (residues A1 to A275 for the heavy chain, B1 to B99 for murine β1, and P1 to P9 for Flu) and 114 water molecules. Atomic coordinates were not assigned to residues 276–283 from the heavy chain and the N-terminal methionine of β1 as no interpretable electron density was seen for these regions of Kd-Flu. Refinement of this final model against the 20- to 2.6-Å resolution data converged to an R value of 21.6% with an Rfree of 26.9% (4.3% test set) with good geometry (Table 1).

Computational Analysis—Graphical structure representations were primarily created using Ribbons (34). Molecular surfaces of the peptide-binding groove (Figs. 4A and 5) were generated using InsightII (Biosym Technologies, San Diego CA). r.m.s.d. values between the different MHC proteins were calculated using an incremental combinatorial extension algorithm (35). r.m.s.d. values between the different MHC peptides were calculated using Lsqkab (CCP4). HBPLUS (36) was used to catalogue contacting atoms and putative hydrogen bonds. Shape complementarity scores (37) were calculated using CCP4. Atomic accessible surfaces were calculated using the program NACCESS.3

### Summary of data collection and refinement statistics

<table>
<thead>
<tr>
<th>Crystal space group</th>
<th>P21,21,21</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knee set (%)</td>
<td>592 (86)</td>
</tr>
<tr>
<td>Non-hydrogen atoms</td>
<td>3,092</td>
</tr>
<tr>
<td>Solvent molecules</td>
<td>114</td>
</tr>
<tr>
<td>r.m.s. bond lengths (Å)</td>
<td>0.008</td>
</tr>
<tr>
<td>r.m.s. bond angles (°)</td>
<td>1.4</td>
</tr>
<tr>
<td>r.m.s. dihedral angles (°)</td>
<td>25.0</td>
</tr>
<tr>
<td>r.m.s. improper angles (°)</td>
<td>0.92</td>
</tr>
<tr>
<td>Average B factor for H:Kd-Flu</td>
<td>53</td>
</tr>
<tr>
<td>Average B factor for Flu peptide</td>
<td>49</td>
</tr>
<tr>
<td>Average B factor for water molecules</td>
<td>43</td>
</tr>
<tr>
<td>Est. coord. error from Luzzati plot (Å)</td>
<td>0.34</td>
</tr>
<tr>
<td>Rfree (%)</td>
<td>26.9 (36.4)</td>
</tr>
<tr>
<td>Ramachandran plot</td>
<td></td>
</tr>
<tr>
<td>Most favored/additional (%)</td>
<td>86.5/12.6</td>
</tr>
<tr>
<td>Generously allowed/disallowed (%)</td>
<td>0.9/0.0</td>
</tr>
</tbody>
</table>

3. J. Hubbard and J. M. Thornton, Dept. of Chemistry and Molecular Biology, University College London.
RESULTS

Structure Determination—The extracellular domains of Kd (heavy chain, residues 1–283 and murine β2m, residues 1–99 plus an N-terminal methionine) were expressed separately in Escherichia coli as insoluble inclusion bodies. The Kd-Flu complex was formed in vitro under oxidative refolding conditions in the presence of excess peptide and was purified using size exclusion and anion exchange chromatographies. Electrospray mass spectral analysis of the complex confirmed the presence of abundant peaks at 32,865.69 Da and 11,817.81 Da corresponding to P6–P7 peptide bond. Further inspection of the mass spectrum over lower mass ranges revealed a monoisotopic, singly charged peak at 1,106.7 Da corresponding with the predicted molecular weights of the heavy chain and murine β2m, respectively. Further inspection of the mass spectrum over lower mass ranges revealed a monoisotopic, singly charged peak at 1,106.7 Da corresponding to Flu.

The Kd-Flu complex crystallized in the orthorhombic space group P212121 with one complex per asymmetric unit. Initial phase estimates were obtained by molecular replacement. After initial refinement, easily interpretable electron density was seen for the bound peptide that improved upon further building and refinement cycles. Diffraction data to 2.6 Å resolution were used for refinement of the final atomic model, which has an R factor of 21.6% (Rfree = 26.9%) with good angle and bond geometry (Table 1). The electron density maps for the whole complex were of good quality (Fig. 1) in the electron density between the Cβ chain and the side chains of the bound peptide except for a small break. No ambiguities were seen for the main chain and the membrane-distal peptide-binding platform of Kd is depicted as a thin cyan ribbon and Flu displayed as balls and sticks, with carbon yellow, nitrogen blue, and oxygen red. B, ribbon diagram of the Kd-Flu complex. The heavy chain is colored cyan, β2m magenta, and Flu rendered as a Corey, Pauling, and Koltun model. The complex is oriented with the peptide N terminus on the left and the membrane-proximal α1 domain at the bottom. Note the protrusion of the anchor residues TyrP2, ThrP5, and ValP9 into the peptide-binding cleft toward the bottom.

Crystal Structure of Kd-Flu

Overall Structural Features of the Complex—Kd-Flu is very similar to the structures of other MHC class I proteins (Fig. 1). Minor differences were observed for the conformations of solvent-exposed loops and the terminal regions of the complex. Flu is bound in the Kd groove between the α1 and α2 helices and on top of the β-sheet platform (Fig. 1) in the canonical manner (7). Structural alignment of Kd-Flu to the structures of other murine MHC class I complexes yielded overall pairwise r.m.s.d. values of 1.14 Å (D4, 80.4% sequence identity) to 2.15 Å (D6a, 80.7% sequence identity). An alignment of the α1-α2 domains alone yielded pairwise r.m.s.d. values of 0.63 Å (Kb, 80.4% sequence identity) to 1.01 Å (Ld, 81.1% sequence identity) reflecting the high degree of general similarity among these proteins.

Backbone Conformation of Kd-presented Flu—Like most MHC-bound nonameric peptides, the backbone of Flu assumes a predominantly extended conformation with a bulge at residues P6 and P7. Surprisingly, the main-chain kink adopted by Flu results from an infrequently observed, unfavorable conformation of ArgP6 that is well supported by our experimental data (Fig. 1A). The P6–P7 bulge is associated with a hydrophobic ridge formed by polymorphic residues Tyr156 and Trp77 in the Kd groove. While favorable Ramachandran angles would be observed if the P6–P7 peptide bond were flipped, this conformation would preclude several favorable interactions and engender steric clashes with Kd (see Fig. 4A).

We compared the conformation of Flu to that of nine-residue peptides bound to Kd (38, 39), Ld (40), and D6 (41–45). The main-chain conformations of Kd-presented peptides vary significantly from that of Flu with overall r.m.s.d. values ranging from 1.43 to 1.47 Å. The greatest differences were observed in the region between P5 and P7 where the r.m.s.d. values for the Ca atoms ranged from 1.22 to 3.66 Å (Fig. 2A). The conformation of the Ld peptide resembles more closely that of Flu with r.m.s.d. 0.95 Å, but nevertheless differs significantly between P4.
Comparison of Flu with Dβ-presented peptides reveals that they adopt nearly identical main-chain conformations all the way from P1 to P9 with r.m.s.d. values ranging from 0.62 to 0.81 Å (Fig. 2A).

Interestingly, Dβ has a similar hydrophobic ridge as Kd located beneath the P6–P7 kink (43).

We also compared the dihedral angles of the aligned peptides (Fig. 2B). This analysis revealed that the backbones of the Dβ peptides adopt a P6–P7 bulge associated with unfavorable dihedral angles for their P6 residues similar to the one in Flu (Fig. 2B). This bulging at P6–P7 was absent in the Lα peptide (Fig. 2B) despite the presence of a similar hydrophobic ridge in the same region of the Lα groove (40).

Hydrogen Bonding to the Flu Backbone—The Flu main chain has 19 nitrogen and oxygen atoms, 14 of which hydrogen bond with Kd either directly or through water-mediated networks. Of the 17 MHC amino acids that participate in hydrogen bonding 8 are invariant among MHC class I proteins. These residues anchor the N- and C-terminal regions of Flu through highly conservative hydrogen bonding networks at each end of the binding groove (Fig. 3).

Eight main-chain nitrogen and oxygen atoms between GlnP3 and ValP8 mediate hydrogen bonds with polymorphic groove residues (Fig. 3). Of particular note are the hydrogen bonds to the main-chain oxygen of ArgP6 and the main-chain oxygen and nitrogen of AlaP7. The carbonyl oxygen atoms of both residues participate in a bifurcate hydrogen bonding network with the Ne1 nitrogen of Trp73, whereas on the opposite side of the Flu backbone the amide nitrogen atom of P7 hydrogen bonds with Asp152 (Fig. 3). This hydrogen bonding arrangement can only form as a result of the unfavorable turn in the Flu main chain at ArgP6.

Binding Pockets in the Kd Groove—Specificity of peptide-MHC association is imparted through a myriad of interactions with peptide anchor side chains, which are sequestered in distinct pockets of the MHC groove. To visualize these pockets in Kd we calculated a solvent-accessible surface (46) for a spherical probe with a radius of 1.4 Å for the Kd/1 domain. Five distinct pockets are clearly apparent in the Kd groove, which correspond to pockets A, B, C, D, and F according to the nomenclature of Matsumura et al. (47) (Fig. 4A). Pocket E, which is the most variable between the different MHC proteins, is absent in our structure. Instead, the polymorphic residues Trp73 and Tyr156 fill the E pocket location creating the hydrophobic ridge across the Kd groove that accommodates the P6–P7 turn in Flu (Fig. 4A). Interestingly, this ridge is absent in Kβ, which preferentially binds 8-residue peptides.

To systematically identify the MHC residues that make up the Kd pockets, we calculated solvent-accessible surfaces of the Kd/1 domain with the use of probes of different radii ranging from 1.4 to 5.0 Å. Using this method we were able to assign residues to pockets when they are accessible to a small, 1.4-Å probe but are inaccessible to a probe larger than 3.5 Å. The

![Diagram of Kd-Flu complex](image-url)
larger pockets A, B, and F are created by 8, 9, and 8 residues, respectively (Table 2), whereas the smaller C and D pockets are each formed of 5 amino acids (Table 2). Each pocket contains residues accessible only to a probe no larger than 2.5 Å (Table 2) indicating that all five Kd pockets are deep. This is in contrast, for example, to the open nature of the pockets in the Kb groove, but is similar to the discrete nature of the pockets in Db.

The adjoining pockets B and C are highly polymorphic. Seven non-conserved and two invariant residues create the B pocket, and four non-conserved residues and one invariant residue create the C pocket (Table 2). In contrast, pockets A and F are composed of mostly conserved residues, which is consistent with their roles in anchoring of the N and C termini of the bound peptide.

**Anchoring of Flu in the Kd Groove**—In the complex Flu participates in 147 van der Waals contacts, 55 hydrophobic contacts, 20 direct, and 3 water-mediated hydrogen bonds with Kd. The majority of the hydrogen bonds (19 out of 23) are directed to the main chain of the bound peptide. In contrast, most of the hydrophobic contacts associated with ligand binding (51 out of 55) are established with peptide side chains. A total of 1236 Å² (79%) of the Flu solvent-accessible surface is buried in the MHC groove, whereas 753 Å² of Kd becomes solvent-inaccessible upon complex formation.

Flu binding is associated with the complete burial of TyrP2, ThrP5, and ValP9 in pockets B, C, and F, respectively (Fig. 4A). These side chains account for 44% of the total peptide buried surface area. The interaction surface between Flu and Kd is characterized by a shape complementarity
The constituent residues of the binding pockets in the peptide-binding groove of H-2K\(^d\) and their relative accessibilities

The assignment of each MHC residue to a particular pocket or subsite is indicated. The corresponding van der Waals contacts for each MHC residue with (a) side chain(s) of the Flu peptide is(are) also listed. Conserved amino acids are underlined.

<table>
<thead>
<tr>
<th>Amino acid position</th>
<th>Residue name</th>
<th>Pocket assignment(^a)</th>
<th>Probe (Å) (^b)</th>
<th>Contacts with the Flu peptide(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Leu</td>
<td>A</td>
<td>2.5</td>
<td>Tyr P2</td>
</tr>
<tr>
<td>7</td>
<td>Tyr</td>
<td>A/B</td>
<td>3.0</td>
<td>Tyr P2</td>
</tr>
<tr>
<td>9</td>
<td>Val</td>
<td>B</td>
<td>2.5</td>
<td>Tyr P2</td>
</tr>
<tr>
<td>22</td>
<td>Phe</td>
<td>B</td>
<td>2.5</td>
<td>Tyr P2</td>
</tr>
<tr>
<td>24</td>
<td>Ala</td>
<td>B</td>
<td>2.0</td>
<td>Tyr P2</td>
</tr>
<tr>
<td>45</td>
<td>Val</td>
<td>A</td>
<td>2.0</td>
<td>Tyr P2</td>
</tr>
<tr>
<td>59</td>
<td>Tyr</td>
<td>A</td>
<td>3.5</td>
<td>Thr P1, Tyr P5</td>
</tr>
<tr>
<td>63</td>
<td>Gln</td>
<td>A</td>
<td>3.0</td>
<td>Thr P1, Tyr P5</td>
</tr>
<tr>
<td>67</td>
<td>Ala</td>
<td>B</td>
<td>2.5</td>
<td>Tyr P2</td>
</tr>
<tr>
<td>74</td>
<td>Phe</td>
<td>B/C</td>
<td>2.0</td>
<td>Val P9</td>
</tr>
<tr>
<td>77</td>
<td>Ser</td>
<td>F</td>
<td>3.0</td>
<td>Val P9</td>
</tr>
<tr>
<td>81</td>
<td>Ala</td>
<td>F</td>
<td>2.0</td>
<td>Val P9</td>
</tr>
<tr>
<td>95</td>
<td>Phe</td>
<td>F</td>
<td>2.5</td>
<td>Val P9</td>
</tr>
<tr>
<td>97</td>
<td>Arg</td>
<td>B/C/D</td>
<td>4.0</td>
<td>Tyr P2, Gln P3, Thr P5</td>
</tr>
<tr>
<td>99</td>
<td>Phe</td>
<td>B/D</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>114</td>
<td>Gln</td>
<td>D/C</td>
<td>2.0</td>
<td>Gln P3</td>
</tr>
<tr>
<td>116</td>
<td>Phe</td>
<td>C/F</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>118</td>
<td>Tyr</td>
<td>F</td>
<td>2.5</td>
<td>Val P9</td>
</tr>
<tr>
<td>123</td>
<td>Tyr</td>
<td>F</td>
<td>2.5</td>
<td>Val P9</td>
</tr>
<tr>
<td>143</td>
<td>Thr</td>
<td>F</td>
<td>3.0</td>
<td>Val P9</td>
</tr>
<tr>
<td>147</td>
<td>Trp</td>
<td>C/F</td>
<td>3.5</td>
<td>Ala P7, Val P9</td>
</tr>
<tr>
<td>159</td>
<td>Tyr</td>
<td>A/D</td>
<td>4.0</td>
<td>Thr P1, Gln P3</td>
</tr>
<tr>
<td>160</td>
<td>Leu</td>
<td>D</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>171</td>
<td>Tyr</td>
<td>A</td>
<td>2.5</td>
<td>Thr P1</td>
</tr>
</tbody>
</table>

\(^a\) Pockets A, B, C, D, and F consist of 8, 9, 5, 5, and 8 residues, respectively. Thus pockets A, B, and F are large, whereas pockets C and D are smaller. Each pocket contains at least one residue accessible only to a 1.4-Å probe indicating that all five pockets of H-2K\(^d\) are deep.

\(^b\) The radius of probe to which a residue becomes inaccessible. Although residues 62, 65, 66, 69, 70, 73, 152, 155, 156, 163, and 167 contact the peptide, they are only accessible to probes with a radius >4.0 Å and as such are defined as surface residues and omitted from this table.

\(^c\) van der Waals contacts of 4.0 Å or less between the side chains of the peptide and the MHC are included in the table.

score of 0.76 comparable to that of other peptide-MHC interfaces and slightly greater than antigen-antibody interfaces (37).

**B Pocket**—Our structural data reveal that Tyr\(^{P2}\) fits snugly in the B pocket where it makes three edge-to-face π-stacking interactions with Tyr\(^7\), Phe\(^{45}\), and Phe\(^{99}\) and mediates large hydrophobic contacts with Val\(^8\) and the aliphatic portion of Arg\(^{69}\) (Fig. 4B). The hydroxyl group of Tyr\(^{P2}\) hydrogen bonds directly to the Oα of Asp\(^{70}\) and makes favorable electrostatic contacts with the guanidinium group of Arg\(^{67}\) (Fig. 4B). A number of binding studies of P2 peptide variants (48, 49) reveals that Tyr\(^{P2}\) is the dominant anchoring residue. Conserved changes at that position result in two orders of magnitude decrease in binding affinity, whereas non-conservative mutations of the P2 Tyr completely abrogate binding.

**C Pocket**—Thr\(^{P5}\) of Flu is buried in the C pocket where it makes favorable hydrophobic contacts with the indole ring of Trp\(^{73}\) (Fig. 3B). The hydroxyl group of Thr\(^{P5}\) mediates a direct hydrogen bond to the Oα of Asp\(^{70}\) and participates in favorable van der Waals contacts with the guanidinium group of Arg\(^{67}\) (Fig. 4B). The C pocket appears larger than the small Thr residue anchored in it (Fig. 4A) indicative of a capacity to accommodate both small and medium size side chains. Indeed, binding studies indicate that only variants with large and/or charged amino acids at P5 exhibit decreased binding affinity (50).

**F Pocket**—The Flu terminal residue, Val\(^{P9}\), is fully buried in the F pocket and makes a number of favorable hydrophobic contacts with Trp\(^{73}\), Tyr\(^{44}\), Phe\(^{99}\), Tyr\(^{123}\), Thr\(^{141}\), and Trp\(^{147}\) (Fig. 4C). The Val\(^{P9}\) side chain appears smaller than the F pocket indicating that larger hydrophobic residues such as Ile and Leu would bind well in this region of the K\(^d\) groove. Mutations at the P9 position show that the F pocket tolerates broad amino acid substitutions as peptides with non-conservative P9 substitutions still retain, albeit low, binding capacity (48, 49).

**DISCUSSION**

To structurally define peptide binding and presentation by K\(^d\) we crystallized it in complex with the well characterized antigenic epitope derived from influenza nucleoprotein. Tyr\(^{P2}\), Thr\(^{P5}\), and Val\(^{P9}\) of Flu are completely buried in pockets B, C, and F, respectively. There they make a myriad of favorable hydrophobic and electrostatic contacts arguing that these polymorphic pockets are the dominant structural determinants of peptide binding specificity for K\(^d\) (Fig. 5).

**The K\(^d\) Peptide-binding Motif**—To define the peptide-binding motif of K\(^d\) in the context of relevant physiological data, we analyzed the sequences of 95 naturally processed, nonamer K\(^d\) epitopes recently identified by Unanue and colleagues (60) and further available in MHC peptide databases (51–53). Although peptides of alternate length have been shown to associate with K\(^d\), the resulting complexes are short-lived and are unlikely to be physiologically dominant (60).

As expected from the exquisite binding chemistry and geometry, tyrosine is found almost exclusively at the P2 position (96%) of the naturally processed nonamer peptides (Fig. 5). The size, shape, and chemical nature of the B pocket clearly dictate the stringent selection of tyrosine at this position. Aromatic residues are flawlessly poised for stacking interactions with the tyrosine phenyl ring; the guanidinium group of Arg\(^{67}\) neutralizes the negative electrostatic potential of the tyrosine hydroxyl moiety, whereas Asp\(^{70}\) hydrogen bonds directly to the hydroxyl oxygen.

Second only to the P2 position, the C-terminal position in the K\(^d\) peptides exhibits high levels of amino acid restriction with preference for Leu, Ile, and Val. Among these, Ile and Leu are found at nearly equal frequencies and account for >80% of the amino acids found at the P9 position in the K\(^d\) peptide pool (Fig. 5). This preference toward larger hydrophobic side chains likely results from a higher kinetic stability (lower k\(_{off}\)) of the peptide-MHC complex when a larger hydrophobic area is buried in the F pocket (54, 55). In addition, the chymotrypsin-like activity of the proteasome generates cleavage products with hydrophobic C-terminal residues further contributing to the observed hydrophobic side chain prevalence at the C termini of MHC-I-presented epitopes (56, 57).

The K\(^d\) peptide comparisons reveal that the buried P5 position tolerates a variety of amino acid residues with restriction mainly based on their size (Fig. 5). In addition to Ser and Thr, which account for 31% of the residues at P5, small and medium size hydrophobic residues are also prevalent (Fig. 5). We believe that this broader tolerance is due to the larger size of the C pocket and the ability of Asp\(^{70}\) to hydrogen bond to the P5 main chain nitrogen facilitating its burial in the complex in the absence of a hydrogen bond donor at the P5 position.

**Implication for TCR Recognition**—Our structural data and peptide analysis predict that P4 and P6 are dominantly involved in TCR recognition. These side chains extrude prominently from the K\(^d\) groove in Flu (Figs. 1B and 4A) and show the greatest amino acid variability in the K\(^d\) peptide pool. Indeed, P4 and P6 have been shown to be critical for T-cell recognition in several K\(^d\)-restricted model systems (Fig. 5) (20, 22, 23). For example, the recognition of a peptide derived from dengue type 2 NS3 protein by the 2D42 CTL clone is compromised by P6 substitutions. The recognition of NS3 peptides from all three dengue serotypes by the E10.6 clone is diminished by P4 substitutions (22). Another example is the cross-reactive CTL clone B7–B7. It recognizes HA peptides derived from two strains of influenza, A/Jap and A/PR/8, and its activity is sensitive to amino acid-substituted peptides at P4 and P6 (23). Lastly, the G9–C8 CTL clone that recog-
nizes an insulin-derived peptide is similarly dependent on the P4 and P6 peptide residues for K<sup>d</sup>-restricted cytotoxic activity (20).

The T-cell recognition data in the dengue virus and diabetes model systems also implicate residues P1, P3, P7, and P8 as additional recognition elements. The 2D42 clone distinguishes between type 2 and type 1 and 3 NS3 peptides based on a single Glu to Gly substitution at P8, whereas its activity is greatly diminished by amino acid substitutions at P7 and P8 within the type 2 NS3 peptide (22). In addition, the cytotoxic activity of the E10.6 clone is weakened by the presentation of NS3 peptides substituted at P3 (22). Likewise, alanine substitutions at P1, P3, and P8 compromise recognition of the insulin peptide by the G9–C8 clone (20). The P1 residue has also been shown to be critical for TCR recognition in the C18 T-cell rejection of transplantable fibrosarcomas. In this model, tumor elimination is dependent on the differential recognition of a variant ERK2 kinase peptide that differs from wild-type only at the P1 position (15, 17). These results are consistent with

**FIGURE 5.** The peptide-binding motif of K<sup>d</sup>-Flu and characterized TCR interactions. The solvent-accessible surface of the peptide-binding groove is displayed as a blue dotted surface. Aligned under the rendering are the sequences of Flu (boxed) and six other antigenic peptides representative of the K<sup>d</sup> peptide-binding motif as well as three antigenic peptides for which TCR contact residues (denoted by the asterisk) have been determined. Pocket B and the P2 residue are highlighted in red while the secondary anchors P5 and P9 and their respective pockets C and F are highlighted in green. Percent values in parentheses reflect the frequencies at which each amino acid is found at the indicated position; only the most prevalent amino acids are denoted. The shorthand above each aligned residue denotes the orientation of each residue of Flu in the complex. The anchor symbol denotes anchoring residues that point down toward the peptide binding platform; ↑ denotes residues pointing away from the peptide binding platform and toward solvent; ← denoted residues pointing toward the α<sub>1</sub> helix; and → denotes residues pointing toward the α<sub>2</sub> helix (see Fig. 4A).
the structure of Kd-Flu in which these auxiliary contact residues are appreciably solvent-accessible (Fig. 4A).

In Kd-Flu the P5 anchor residue is buried in the C pocket. Interestingly, Wong et al. have shown that substitutions at P5 in the insulin-derived peptide reduce or completely abolish G9–C8 cytotoxic activity, which the authors interpreted as P5 being able to directly interact with the TCR (20). It is conceivable that the insulin peptide binds to Kd with an exposed P5 residue. However, it is also feasible that reorganization events related to anchoring of different residues in the C pocket result in concerted changes in the peptide-MHC conformation. In fact, buried secondary anchor residues are known to be capable of indirectly modulating T-cell recognition events by buttressing the overall peptide-MHC surface conformation (58, 59). Taken together, the results of our study enable a detailed understanding of Kd-restricted antigen processing, presentation, and recognition in a number of murine model systems involving pathogen recognition, tumor rejection, and induction of diabetes.

Acknowledgments—We thank Drs. Ted Hansen and Emil R. Unanue for discussion and critical comments on the manuscript. We also thank Drs. Paul Allen and Eric Pamer for sharing reagents.

REFERENCES