A hypothetical model of the foreign antigen binding site of Class II histocompatibility molecules


*Department of Biochemistry and Molecular Biology, Harvard University, Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts 02138, USA
†Department of Medical Microbiology, Stanford University, Stanford, California 94305, USA

Class II and class I histocompatibility molecules allow T cells to recognize 'processed' polypeptide antigens1-8. The two polypeptide chains of class II molecules, α and β, are each composed of two domains (for review see ref. 6); the N-terminal domains of each, α1 and β1, are highly polymorphic9-10 and appear responsible for binding peptides at what appears to be a single site11-13 and for being recognized by MHC-restricted antigen-specific T cells14,15. Recently, the three-dimensional structure of the foreign antigen binding site of a class I histocompatibility antigen has been determined16,17. Because crystal structure of a class II molecule is not available, we have sought evidence in class II molecules for the structural features observed in the class I binding site by comparing the patterns of conserved and polymorphic residues of twenty-six class I and forty-four class II amino acid sequences. The hypothetical class II foreign-antigen binding site we present is consistent with mutation experiments18-20 and provides a structural framework for proposing peptide binding models to help understand recent peptide binding data21-23.

An overall similarity in structures of class I and class II molecules is suggested by a similarity in domain and genomic organization24, by high similarity in the C-terminal domain sequences25, and by the observation that some T cells specific for either class of molecule use the same receptor25,26. Furthermore, the close association of N-terminal residues of the α1 and β1 domains in the class II intermolecular dimer, deduced from hemi-crystalline x-ray experiments17-20, is consistent with the close association of N-terminal residues of the class I α1 and α2 domains in the intramolecular dimer observed in the crystal structure16. By imagining an attachment of the class I α1 domain to β1-microglobulin, a structural model of class II molecules can be produced from the class I structure16. Like the class I structure (Fig. 2a), such a class II molecule would have a cleft between the C-terminal α-helices of its polymorphic α1 and β1 domains with the bottom of the cleft formed by the N-terminal β-strands of each domain17.

Pairwise sequence comparisons of class II α1 domains with the polymorphic domains of class I (α1 and α2) revealed no significant sequence similarity, whereas comparisons of class II β1 domains revealed weakly significant similarity with both the α1 and α2 domains of class I17-21,25. The choice to align class II α1 to class I α1 and class II β1 to class I α2 is suggested by the glycosylation site (Ssin, X, Ser/Thr) at residue 86 (class II residues in the text have been renumbered as class II HLA-A2 numbering to facilitate comparisons to the class I three-dimensional structure, see legend to Fig. 1), which is completely conserved in all class I and class II α1 sequences, and by the disulfide-linked cysteines (101 and 164) found in all class I α1 and class II β1 sequences16 (Fig. 1). This choice has been supported both by similarities in the respective genomic structures21,25 and by similarities in the alteration of T-cell recognition by a pattern of mutation at 152, 155, and 156 in both class I α2

<table>
<thead>
<tr>
<th>Residue numbering</th>
<th>In Substitution</th>
<th>Monoclonal antibody</th>
<th>allo</th>
<th>T cells</th>
<th>ag</th>
<th>Reference</th>
<th>Location on hypothetical class II structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>79</td>
<td>I- A*</td>
<td>E - K</td>
<td>+4/5</td>
<td>-0/2</td>
<td>+2/17</td>
<td>18, 19</td>
<td>H2, solvent-accessible, pointing up</td>
</tr>
<tr>
<td>95</td>
<td>I- A*</td>
<td>Y - H</td>
<td>-0/12</td>
<td>+2/7</td>
<td>-0/3</td>
<td>20</td>
<td>SI*, pointing up into helix</td>
</tr>
<tr>
<td>95</td>
<td>I- A*</td>
<td>H - V</td>
<td>-0/8</td>
<td>ND</td>
<td>ND</td>
<td>21</td>
<td>SI*, pointing up into helix</td>
</tr>
<tr>
<td>98</td>
<td>I- A*</td>
<td>Q - K</td>
<td>-0/8</td>
<td>ND</td>
<td>ND</td>
<td>21</td>
<td>SI*, pointing down</td>
</tr>
<tr>
<td>99</td>
<td>I- A*</td>
<td>G - P</td>
<td>-0/12</td>
<td>+1/7</td>
<td>+1/3</td>
<td>20</td>
<td>SI*, pointing up in crevice</td>
</tr>
<tr>
<td>99</td>
<td>I- A*</td>
<td>P - G</td>
<td>-0/8</td>
<td>ND</td>
<td>ND</td>
<td>21</td>
<td>SI*, pointing site</td>
</tr>
<tr>
<td>100</td>
<td>I- A*</td>
<td>F - E</td>
<td>-0/8</td>
<td>ND</td>
<td>ND</td>
<td>21</td>
<td>SI*, pointing down</td>
</tr>
<tr>
<td>103</td>
<td>I- A*</td>
<td>F - Y</td>
<td>-0/8</td>
<td>ND</td>
<td>ND</td>
<td>21</td>
<td>SI*, SI*</td>
</tr>
<tr>
<td>114</td>
<td>I- A*</td>
<td>V - E</td>
<td>ND</td>
<td>+5/7</td>
<td>22</td>
<td>SZ*, pointing into site</td>
<td></td>
</tr>
<tr>
<td>114</td>
<td>I- A*</td>
<td>I - T</td>
<td>-0/8</td>
<td>ND</td>
<td>ND</td>
<td>21</td>
<td>SI*, pointing into site</td>
</tr>
<tr>
<td>126</td>
<td>I- A*</td>
<td>F - Y</td>
<td>-0/8</td>
<td>ND</td>
<td>ND</td>
<td>21</td>
<td>SI*, pointing into helix (see caption)</td>
</tr>
<tr>
<td>134</td>
<td>I- A*</td>
<td>R - H</td>
<td>+4/16</td>
<td>-0/2</td>
<td>-0/17</td>
<td>23, 24</td>
<td>SI*, solvent-accessible, pointing down towards the region of the lower domains</td>
</tr>
<tr>
<td>140</td>
<td>I- A*</td>
<td>R - C</td>
<td>+2/7</td>
<td>+1/8</td>
<td>-0/15</td>
<td>25</td>
<td>H2, solvent-accessible, pointing down</td>
</tr>
<tr>
<td>140</td>
<td>I- A*</td>
<td>E - D</td>
<td>+3/10</td>
<td>ND</td>
<td>ND</td>
<td>25</td>
<td>H2, solvent-accessible, pointing up</td>
</tr>
<tr>
<td>149</td>
<td>I- A*</td>
<td>K - S</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>21</td>
<td>H1, H2*, kink between helices</td>
</tr>
<tr>
<td>150</td>
<td>I- A*</td>
<td>Q - P</td>
<td>+7/10</td>
<td>-0/1</td>
<td>+7/15</td>
<td>25, 26</td>
<td>H1, H2*, kink between helices</td>
</tr>
<tr>
<td>150</td>
<td>I- A*</td>
<td>R - Q</td>
<td>+1/10</td>
<td>-0/1</td>
<td>+2/15</td>
<td>25, 26</td>
<td>H1, H2*, kink between helices</td>
</tr>
<tr>
<td>154</td>
<td>I- A*</td>
<td>R - K</td>
<td>+1/3</td>
<td>ND</td>
<td>ND</td>
<td>27</td>
<td>H2, solvent-accessible, pointing up</td>
</tr>
<tr>
<td>152</td>
<td>I- A*</td>
<td>F - F</td>
<td>-0/6</td>
<td>ND</td>
<td>ND</td>
<td>28</td>
<td>H2, into site</td>
</tr>
<tr>
<td>155</td>
<td>I- A*</td>
<td>R - Q</td>
<td>+4/10</td>
<td>+1/11</td>
<td>+1/15</td>
<td>25, 26</td>
<td>H2*, up into top part of the site</td>
</tr>
<tr>
<td>155</td>
<td>I- A*</td>
<td>R - Q</td>
<td>+4/6</td>
<td>ND</td>
<td>ND</td>
<td>28</td>
<td>H2*, up into top part of the site</td>
</tr>
<tr>
<td>156</td>
<td>I- A*</td>
<td>T - K</td>
<td>- (NC)</td>
<td>ND</td>
<td>ND</td>
<td>28</td>
<td>H2*, into site</td>
</tr>
<tr>
<td>160</td>
<td>I- A*</td>
<td>L - V</td>
<td>-0/7</td>
<td>ND</td>
<td>ND</td>
<td>21</td>
<td>H2, down towards strand</td>
</tr>
<tr>
<td>163</td>
<td>I- A*</td>
<td>V - A</td>
<td>-0/8</td>
<td>ND</td>
<td>ND</td>
<td>21</td>
<td>H3, into site</td>
</tr>
<tr>
<td>170</td>
<td>I- A*</td>
<td>K - G</td>
<td>-0/8</td>
<td>ND</td>
<td>ND</td>
<td>21</td>
<td>H3, up (?)</td>
</tr>
<tr>
<td>171</td>
<td>I- A*</td>
<td>T - P</td>
<td>-0/8</td>
<td>ND</td>
<td>ND</td>
<td>21</td>
<td>H3, into site (?)</td>
</tr>
<tr>
<td>174</td>
<td>I- A*</td>
<td>P - S</td>
<td>-0/8</td>
<td>ND</td>
<td>ND</td>
<td>21</td>
<td>H3, (?)</td>
</tr>
</tbody>
</table>

(x<y) Indicates a loss of reactivity with x parent in monoclonal antibodies, or x allotypic, or foreign antigen-restricted (ag) T cells out of a total of y antibodies/T-cell clones examined, as a result of the single substitution indicated. +(-) indicates the residue involved in monoclonal antibody or T-cell recognition. The mutation at 126, which does not cause loss of binding by any parent I- A*,-reactive antibodies, does cause a gain in binding by (non-parent) I- A*,-reactive antibodies. ND, no data; NC, not clear; * Gilmer et al, personal communication. The location of the altered side chains in the hypothetical class II model is consistent with these effects (see text).
Evidence for the presence of homologous α-helices in class I and class II structures is found in the pattern of conserved amino acids. In the α-helical regions of all class I sequences these occur every third or fourth residue \( \alpha \)-residues (conserved \( \alpha \)-residues 68, 72, 75, 78, 84, conserved \( \alpha \)-residues 148, 153, 154, 157, 158, 161, 164, 165, 168; solid columns in Fig. 1). These residues form one face of the α-helices in class I \( \alpha \)-domains and \( \alpha \)-domains (Fig. 2a) and their positions are also conserved in class II \( \alpha \)-domains, and \( \beta \)-domains (Fig. 1). Interspersed between the conserved residues on the α-helices of the class I structure are many of the most polymorphic residues \( \beta \)-residues. Again, a similar pattern of polymorphic residues is found in class II structures \( \alpha \)-helices (74, 79, 80, 81; \( \beta \)-residues 152, 155, 156, 159, 163 (class I numbering, see Fig. 1)). Structural similarity in the helical regions is further supported by the observation of a completely conserved i, i+4 helix-stabilizing salt bridge (Arg57–Asp61, see Fig. 2b) and the conserved Trp147 in the class I \( \alpha \)-class II \( \beta \)-domains, and the completely conserved Leu78 in class I and class II \( \beta \)-class II \( \beta \)-domains (Fig. 1).

In the β-strand region (N-terminal half of the sequences in Fig. 1), the distribution of polymorphic residues in class II sequences often match those in class I sequences. In Fig. 1, the most polymorphic positions in each strand in the class I \( \alpha \)-domain are aligned to the corresponding regions of class II sequences \( \alpha \)-helices (9 in S1, 24 in S2, 22 in S3, 45 in S4; cross-hatched columns Fig. 1, see legend for a fuller explanation of alignment, particularly in class I). A similar alignment of polymorphic residues in class I \( \alpha \)-class II \( \beta \)-domains (95, 97, 99 in S1; 114, 116 in S2; Fig. 1) is provided by matching the conserved cysteine 101. These polymorphic residues form the ‘bottom’ of the cleft in the class I structure (Fig. 2a). The striking ‘very other one’ periodicity (95, 97, 99, and 114, 116) of the polymorphic residues is that expected for β-strand residues where every other side chain faces to one side of a sheet (see legend, Fig. 1).

Structural homology in the β-strand region is further supported by the observation that the residues making up the two salt-bridges His3–Asp29 and Arg111–Asp129, which bridge across the open end of the two beta-strand ‘hairpins’ formed by S1 and S2 and by S2’ and S3’ (Fig. 2b), are conserved in all class I and class II sequences (Fig. 1b), and are the same distance from the first polymorphic residues of their β-strands in both class I and class II [in S1, I-A and DQ are aligned to other class II and class I sequences by introducing a β-bulge \( \beta \)-bulge at Val4 (Gly bulges out) which results in 1-A polymorphic position 5 (residue 10 in I-A numbering), as well as the aligned polymorphic position 9 pointing up into the proposed site]. Interestingly, the pseudo-two-fold symmetric locations His93–Asp119 and Arg21–Asp39 (dotted lines in Fig. 2b) are conserved salt links in all class I sequences, but not class II sequences.

The regions between the aligned secondary structure elements are consistent with loop structures. Class I and class II share common amino acid residues in these regions (Fig. 2b). Insertions and deletions in class II sequences relative to class I sequences occur in or at the boundary of these loops (Fig. 1). In class I, a glycosylation site at the loop containing residues 105–108 and a switch region at a loop at residues 35–38 (underneath a suggested shift in the bend of the \( \alpha \)-helix described in the legend to Fig. 3) eliminate the sequences SDWR and RFDS, respectively, previously noted as potential CD8 interaction sites in class I (see Fig. 2b).

The alignment presented in Fig. 1 allows us to propose a hypothetical structure for the foreign antigen recognition site of class II molecules based (Fig. 3a) on structural homology with the class I site\(^7\). The distribution of polymorphic residues (cross-hatched, Fig. 3a) in the site is consistent with the conclusions that some of the polymorphic residues (for example, those on the bottom of the site) probably control interaction with antigen\(^7,48\), whereas others control allele-related recognition by T cells\(^48\). It is also consistent with studies of natural site-directed mutants\(^38,52\) of class II molecules (Table 1). Residues which are monoclonal antibody epitopes (79, 134, 145, 149, 150, 154, 155, Fig. 3a) protrude into solvent and residues implicated in antigen-restricted T cell recognition face into the site (95, 99, 114, 152) or upwards (presumably toward the T cell) (79, 150, 154, 155, 156, Fig. 3a). Aspartic acid at position 143 in DQ (57 in DQ \( \beta \)-numbering) provides almost complete resistance to insulin-dependent diabetes (1DDM)\(^59,60\), is predicted to form a salt bridge with Arg84 at one end of the cleft.

The hypothetical model provides a framework for generating testable models for peptide binding. There is currently no consensus as to the conformation that peptides adopt when bound\(^13,39\), or even whether or not different peptides adopt similar backbone conformations upon binding. The size, shape, and residue composition of the site in HLA-A2 indicate that it...
Fig. 2. Similarities between the structure of class I polymorphic domains and the proposed model for class II. a. Those positions (numbered according to HLA-A2 sequence) where residues are conserved (solid) or polymorphic (hatches) in both class I and class II. Displayed are those residues forming (large symbols) or adjacent to (small symbols) the proposed processed-antigen binding site, which is located between the squa and (class I) or at (class II) the beta-strand. Positions of residues conserved in both MHC classes generally face away from the site, while those polymorphic in both classes face into the site. In both class I and class II molecules, different isotopes have different distributions of polymorphic residues.28, Nevertheless, the sequence similarity among isotopes6,7,8,9 argues that the structural features of the site are essentially the same in different isotopes and validates the construction of a single overall class II model. H1 is not shown here. The break at the kink between helices at 149 indicates a site of insertions and deletions. b. Intramolecular interactions observed in the HLA-A2 structure16 and conserved in all class I and class II sequences (Fig. 1). Three salt bridges (solid lines bounded by arrows), whose residues are conserved in all class I and class II sequences (111 is Arg in all but three sequences, 157 is Glu or Asp, and 158 is Ser or Thr) and are of polymorphic residues (see Fig. 2a and text), argue that class I and class II have homologous beta-strands and alpha-helices. Evidence for the similar positioning of the beta-strands relative to the alpha-helices includes the disulfide bridges (101 to 164) and the hydrophobic cluster (dashed lines 124, 133, 147, 153) conserved in all class I and class II sequences. Bold dots at 84, 143 indicate the position of the Tyr-Thr hydrogen bond conserved in class I, and a potential salt bridge, semi-conserved in class II and implicated in resistance to IDDM (see text)9,10,12. X show class I RFDS-like sequences (see text). Shaded loops contain similar residues in class I and class II: SI-S2 (Pro and Gly), S2-S3 (two scads), S1-S2 (class I Pro, Gly; class II Asn, Gly), S3-S4 (126, 130, 133 conserved hydrophobic), H1-H2' (extended Gly, residues in class II sequences related to class I occur in loops S1-S2, S2-S3, S3-S4, H1-H2', and at the N- and C-terminals of both alpha and beta domains. CHO (residue 86) is the common carbohydrate attachment site. The second class II CHO site (105-107) is found in the shaded loop after 103.

has space to accommodate an alpha-helical or extended conformation, or a mixture of both, and therefore leaves open many possibilities.13, In cases where the conformation of a bound peptide can be proposed from other data or where contacts between the peptide and specific residues on the histocompatibility antigens can be identified, the number of degrees of freedom may be reduced enough to allow specific models of the complex to be proposed. The hen egg-white lysozyme peptide bonding to I-Ak is such a case.14, HEL (40-61) binds to I-Ak but not to I-Aa12,13,20. In the 10-mer HEL(52-61), three MHC-contact and three T-cell contact residues segregate on the opposite sides of the bound peptide if it binds as an alpha-helix. Because Asp52 is 2.5 turns from Arg61 on an ideal helix, it is possible that one residue contacts the helix of the a domain and the other contacts the helix of the b domain.13 By sliding this 2.5 turn helix in the site, two ways are found which both stabilize the charged peptide residues by oppositely charged I-Ak helix residues and keep the T-cell contact residues per class I alpha helix, also obscuring any structural pattern. However, if the actual bound state of these peptides is as conformationally complex as these studies imply, then more detailed structural restrictions on peptide binding are a prerequisite to modelling.

Whereas the above studies locate those peptide residues that contact the MHC molecule, a recent study25 has implicated L-E residue Val114 in contacting the PCC peptide and not the T cell. This is consistent with the model in Fig. 3a, since position 114 points up to the site from the beta-sheet region and is probably inaccessible to direct interaction with the T-cell receptor. A study showing Lys103 of PCC to contact only the MHC and not the T cell indicates that PCC 103 may contact the beta-sheet region of the site.

The hypothetical model for a class II foreign antigen binding site may be useful for formulating testable hypotheses about foreign antigen and T-cell binding. We note, however, that significant shifts in the direction (up to 30°) and location (up to 7 A) of alpha-helices and beta-strands (20°, 20 Å) have been observed in the X-ray structures of the very closely related members of both the globin and the IgG families.56,68 Therefore any hypothetical model derived from even closely related sequences may differ from the true structure in important details and should be thought of only as an imperfect guide for experiments.

Boudjemaa Samraoui thanks the Algerian Ministere de l'Enseignement et de la Recherche Scientifique for a postdoctoral leave. Pamela J. Bjorkman held an American Cancer Society postdoctoral fellowship during part of the work. The research was supported by the NIH and the Howard Hughes Medical Institute.

Received 19 January; accepted 18 March 1988.

References