Use of Complete Eluted Peptide Sequence Data from HLA-DR and -DQ Molecules to Predict T Cell Epitopes, and the Influence of the Nonbinding Terminal Regions of Ligands in Epitope Selection

Andrew J. Godkin, Miles P. Davenport, Anthony Willis, Derek P. Jewell and Adrian V. S. Hill

*J Immunol* 1998; 161:850-858; 
http://www.jimmunol.org/content/161/2/850

---

**Why The JI?**

- **Rapid Reviews! 30 days** from submission to initial decision
- **No Triage!** Every submission reviewed by practicing scientists
- **Speedy Publication!** 4 weeks from acceptance to publication

*average

---

**References**

This article cites 35 articles, 9 of which you can access for free at:
http://www.jimmunol.org/content/161/2/850.full#ref-list-1

**Subscription**

Information about subscribing to *The Journal of Immunology* is online at:
http://jimmunol.org/subscription

**Permissions**

Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

**Email Alerts**

Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts
Use of Complete Eluted Peptide Sequence Data from HLA-DR and -DQ Molecules to Predict T Cell Epitopes, and the Influence of the Nonbinding Terminal Regions of Ligands in Epitope Selection

Andrew J. Godkin, Miles P. Davenport, Anthony Willis, Derek P. Jewell, and Adrian V. S. Hill*

In diseases with a strong association with an HLA haplotype, identification of relevant T cell epitopes may allow alteration of the pathologic process. In this report we use a reverse immunogenetic approach to predict possible HLA class II-restricted T cell epitopes by using complete pool sequencing data. Data from HLA-DR2(B1*1501), -DR3(B1*0301), -DQ2(A1*0501, B1*0201), and -DQ8(A1*0301, B1*0302) alleles were used by a computer program that searches a candidate protein to predict ligands with a relatively high probability of being processed and presented. This approach successfully identified both known T cell epitopes and eluted single peptides from the parent protein. Furthermore, the program identified ligands from proteins in which the binding motif of the HLA molecule was unable to do so. When the information from the nonbinding N- and C-terminal regions in the pool sequence was removed, the ability to predict several ligands was markedly reduced, particularly for the HLA-DQ alleles. This suggests a possible role for these regions in determining ligands for HLA class II molecules. Thus, the use of complete eluted peptide sequence data offers a powerful approach to the prediction of HLA-DQ and -DR peptide ligands and T cell epitopes. The Journal of Immunology, 1998, 161: 850–858.

A PCs process peptides from parent proteins and present them to T cells as epitopes bound to MHC molecules. A knowledge of these epitopes might be useful to design means of either enhancing or suppressing immune responses. Hence, much recent work has been directed at methods of identifying T cell epitopes (1–3).

Recently, it has been shown that T cell epitopes can be predicted by using the knowledge of the type of peptide ligand that a particular HLA molecule binds (4, 5). HLA molecules bind short peptides in an extended configuration along a groove on the surface of the protein, and using new methods including the sequencing of these naturally processed peptide ligands and in vitro binding assays, simple rules can be worked out for the binding motif of a particular HLA molecule (6).

Class I alleles bind peptides 8 to 10 amino acids long, and class II alleles longer peptides in the region of 12 to 20 amino acids (7). Analysis of the crystal structure of HLA-DR1 (8) and HLA-DR3 (9) has shown them to have binding grooves that are open at each end, allowing longer peptides of variable length to bind. There is a core binding region of nine amino acids, with certain key pockets in the groove accommodating peptide side chains in a similar fashion to the class I alleles (8). However, predicting the motif from sequencing the different length peptide ligands has proven to be far more difficult for class II alleles. Furthermore, the N- and C-terminal extensions, not bound in the groove, may reflect a preference of enzymes and other proteins involved in Ag processing.

An alternative approach to analyzing class II motifs involves the use of pool sequencing data (3, 10). All the eluted ligands are sequenced en masse in a pool after dominant peptides are removed. Motifs are identified by the presence of enrichment of certain amino acids during particular cycles of degradation and by matching specific known ligands to the proposed motif.

In this study we have used this technique of pool sequencing, but rather than trying to establish a specific motif to predict T cell epitopes, this step was bypassed. The pool sequence was organized into a table that takes into account the naturally diminishing yields of amino acids during the Edman degradation, as reported by Dav-enport et al. (11). This table was then used by a computer program to predict the likelihood that a particular peptide would be processed from a parent protein and presented by the HLA molecule of interest.

Initially, the validity of this approach was tested as follows. Using known HLA-restricted T cell epitopes and sequences of eluted ligands, we looked to see whether the program would in fact have predicted that these peptides would be processed and presented from the parent protein. For HLA-DR2(DR15), -DR3(DR17), -DQ2(A1*0501, B1*0201), and -DQ8(A1*0301, B1*0302), we found that the program was highly efficient in predicting these published peptides. For certain ligands, the information available from the pool sequence that covers the nonbinding N- and C-termini appears to be essential for prediction as well as the core binding region and may well reflect other mechanisms required for Ag processing.
We then looked to see whether this approach could identify ligands in a protein that did not appear to contain any sequences that would fit a binding motif. A-gliadin, a protein responsible for exacerbating celiac disease, was thought to be a suitable candidate for two reasons. Firstly, the disease is strongly associated with HLA-DQ2 and -DQ8 (12), and secondly, the gliadin protein is unusual in that it is extremely rich in proline and glutamine, with very few charged residues (13). Examining the A-gliadin sequence with the known motifs for HLA-DQ2 (14) and -DQ8 (15) does not reveal any candidate binding regions.

Screening the gliadin sequence with the program using the pool sequence data, two regions emerged that are predicted to be processed and presented by both HLA-DQ2 and -DQ8. Using a competitive binding assay against the class II invariant chain-derived peptide (CLIP)\(^3\) derived from the invariant chain, these regions from gliadin were shown to bind well to the HLA-DQ molecules. The significance of these ligands in the immunopathogenesis of celiac disease is discussed.

For class II HLA alleles, the use of complete eluted pooled sequence data offers a superior approach to that of HLA motifs alone in the identification of binding ligands and T cell epitopes, and offers clear evidence for the significance of the nonbound N- and C-terminal extensions in these class II-derived peptides.

**Materials and Methods**

**Preparation of class II alleles**

EBV-transformed B cell lines homozygous for HLA-DR2, -DR3/DQ2, and -DRA/DQB (LD2B, Cox, and BM14 respectively, obtained from the European Collection of Cell Culture, Salisbury, U.K.) were grown to 10\(^5\) cells in roller bottles with RPMI 1640 supplemented by 10% FCS, 1-glutamine, and antibiotics. The cells were then lysed with 3% Nonidet P-40 in PBS containing 1 μg/ml leupeptin, 1 μg/ml pepstatin, and 5 mM EDTA. The lysate was spun at 100,000 × g for 90 min. The supernatant was passed over a precolumn of Sepharose CL 4B followed by two affinity columns in series with cyanogen bromide-activated Sepharose beads linked to the class II protein (0.15 μg) and the class II protein (0.015 μg). The mixture was neutralized with 1 M Tris (pH 8.0) and transferred to wells precoated with anti-DQ Ab. The mixture was incubated for 1 h, and the wells were washed thoroughly with PBS containing 1% Tween, then with PBS alone.

The plate was incubated with avidin-horseradish peroxidase (Sigma, St. Louis, MO)/biotinylated antiavidin (ExtrAvidin, Sigma/avidin-horseradish peroxidase and developed with 100 μl of o-phenylene diamine (0.4 mg/ml) in phosphate-citrate buffer. The reaction was terminated with 12.5% sulfuric acid; the absorption was measured at 492 nm. The concentration of unlabelled peptide required to inhibit 50% of the binding (IC\(_{50}\)) of 1 was calculated as a micromolar concentration. All binding assays were repeated at least twice, and a mean IC\(_{50}\) was calculated.

**Measurement of T cell activation**

The biologic effect of the peptides on T cells was characterized ex vivo in three coellics (all HLA-DQ2) using single cell IFN\(γ\) release as a measure of effector function. This was achieved using an enzyme-linked immunospot assay (ELISPOT; Abs obtained from MARTECH, Stockholm, Sweden). PBMCs were purified and washed, and 5 × 10\(^5\) cells (in RPMI and 5% FCS) were added per well of a special polymer-backed 96-well plate (MAIP-S-45, Millipore, Bedford, MA) precoated with monoclonal anti-IFN\(γ\) (1-DIK). The peptides were tested in quadruplicate wells, having been added to a final concentration of 50 μg/ml, and compared with control wells with no peptide. The plate was incubated at 37°C in 5% CO\(_2\) for 18 h. The plate was then washed and incubated with a second layer of biotinylated anti-IFN\(γ\) (7B6-1) and developed with a streptavidin-alkaline phosphatase plus chromogenic substrate system. Activated T cells were enumerated at the single cell level by counting the number of spots per well.

**Statistics**

The correlation between predicted binders/nonbinders and the IC\(_{50}\) values was compared using both Spearman’s and Kendall’s coefficients of correlation. One-tailed significance levels are shown.

**Results**

**Pool sequence data and probability tables**

The pool sequence data were organized as described into a probability matrix. Table I shows the probability matrix for HLA-DR3, and Table II shows the matrix for HLA-DQ2. These tables reflect the pattern of enrichments of particular amino acids at each cycle of the Edman degradation of the pool of amino acids. In each pool, there are several thousand peptides, so this datum represents the summation.

Previous analysis of pool sequence data has shown that cycles 4 to 13 probably cover the core binding region (10). Table III outlines these patterns of enrichment for HLA-DR3 and reveals an excellent correlation with published binding motifs for HLA-DR3 derived by other means (18, 19). However, there are enrichments not only in other cycles within the binding core, possibly reflecting secondary anchors, but also in the early and late cycles that contain increased amounts of basic and polar amino acids. We have published similar observations (15) regarding pool sequence data and the binding motifs for the HLA-DQ2 and -DQ8 alleles (HLA-DR2; our manuscript in preparation).

**Assessment of the predictive power of the program, and the contribution of nonbinding data in the pool sequences**

To take into account all the information available from the pool sequencing data, the PPP scores were calculated for peptides 15

---

\(^1\) Abbreviations used in this paper: CLIP, class II invariant chain-derived peptide; PPP, probability of being processed and presented; IC\(_{50}\), concentration required to inhibit 50% of the binding; ELISPOT, enzyme-linked immunospot assay; hsp, heat shock protein; MBP, myelin basic protein.
amino acids in length. Proteins of interest giving rise to known T cell epitopes or eluted ligands for HLA-DR3 (18, 19, 21), HLA-DQ2 (22, 23), and HLA-DQ8 (24, 25) were then screened again, and the new calculated ranks are recorded in column 5 of Table IV. For two of the ligands, EBV gp30 592–606 and transferrin receptor 618–632, the ranking fell to below the top 5%. This suggests that although for some HLA-DR3 ligands the strength of the core binding data alone was enough, for others the additional information contained in these early and late cycles was essential for accurate prediction.

Table V shows the predictions using the pool sequence data for HLA-DR3 and -DQ8; again, overall, the program worked well. The highest PPP score was taken from the sequence of the epitope/ligand to be ranked in comparison to the others (Tables IV and V).

The epitopes/ligands for HLA-DR3 show an excellent correlation to the PPP ranked scores, with all of them falling into the top 5%. To examine further the contribution of amino acids from the nonbinding core, the probability table was reorganized, making the entries in columns 1 to 3 and 14/15 all equal to 1. The proteins were then screened again, and the new calculated ranks are recorded in column 5 of Table IV. For two of the ligands, EBV gp30 592–606 and transferrin receptor 618–632, the ranking fell to below the top 5%. This suggests that although for some HLA-DR3 ligands the strength of the core binding data alone was enough, for others the additional information contained in these early and late cycles was essential for accurate prediction.

Table V shows the predictions using the pool sequence data from HLA-DR3 and -DQ8; again, overall, the program worked well.
well. If the modified table was used, the deterioration in predictive power was even more marked, with six of the eight ligands/epitopes falling in rank, and four of these falling below the top 5%. To demonstrate the importance of this, consider the HLA-DQ8-restricted T cell epitope from p21\textsuperscript{ras} 51–67 (mutant Q to L at position 61). This was defined using cellular immunologic methods from a patient with medullary cell carcinoma of the thyroid (24). Figure 1 is a graph showing the scores of iterative 15mers for this protein. The peptide commencing at position 50 has the second highest score, ranking it at 2 of 157, i.e., 1.27%. If this screening is repeated with a modified table as described above, removing the data from the early and late cycles, this score falls markedly to 20 of 157 (12%).

The PPP ranking is superior to binding motifs alone in predicting epitopes

The sequences of the HLA-DR3 T cell epitopes listed in Table IV are hsp-65 (AKTIAYDEEARRG), hsp-70 (KNPLFLDEQLT), and groEL (SKLIEYDETARH). When these are compared with the binding motif shown in Table III, it is difficult to identify them as ligands if all the anchors are taken into consideration. This reflects the degeneracy of binding to class II molecules and the difficulty in predicting ligands using a binding motif alone. Screening the protein with a reduced number of anchors leads to multiple potential ligands. Exactly the same problem occurs with the other class II alleles.

Using the PPP ranking, these T cell epitopes are easily identified. The importance of the flanking regions has already been demonstrated, and the importance of using all the positions in the peptide is emphasized using the HLA-DR2-presented T cell epitope from myelin basic protein (MBP). The binding characteristics and T cell recognition of this peptide have been extremely well studied (20). Autoreactive T cells in patients with multiple sclerosis recognize the immunodominant MBP 84–102 DENPVYHFEKNIKT-PRTPP peptide presented by HLA-DR2 (dominant binding anchors are underlined). The epitope has been narrowed down further for certain clones to the peptides 85 to 98 and 86 to 98. As shown in Table IV, the PPP rank would predict these epitopes with remarkable accuracy, with the peptide starting at N86 gaining the highest score and top rank position. If the data from the flanking regions are removed, this peptide falls from 0.63 to 9.6%. The relative contribution of all the amino acids to the peptides’ score is shown in Figure 2. This emphasizes that although the anchor positions are important, the nonanchor and flanking amino acids are equally important in deciding whether a peptide will ultimately be processed as an epitope.

Identification of predicted ligands from A-gliadin

Figure 3 shows the sequence for A-gliadin (clone pTO-A10, which bears a close homology to the only reported sequence derived directly from protein sequencing (26)) with the motifs for HLA-DQ2

<table>
<thead>
<tr>
<th>Cycles</th>
<th>1–3</th>
<th>4/5</th>
<th>7/8</th>
<th>9/10</th>
<th>12/13</th>
<th>14/15</th>
</tr>
</thead>
</table>

Using the PPP ranking, these T cell epitopes are easily identified. The importance of the flanking regions has already been demonstrated, and the importance of using all the positions in the peptide is emphasized using the HLA-DR2-presented T cell epitope from myelin basic protein (MBP). The binding characteristics and T cell recognition of this peptide have been extremely well studied (20). Autoreactive T cells in patients with multiple sclerosis recognize the immunodominant MBP 84–102 DENPVYHFEKNIKT-PRTPP peptide presented by HLA-DR2 (dominant binding anchors are underlined). The epitope has been narrowed down further for certain clones to the peptides 85 to 98 and 86 to 98. As shown in Table IV, the PPP rank would predict these epitopes with remarkable accuracy, with the peptide starting at N86 gaining the highest score and top rank position. If the data from the flanking regions are removed, this peptide falls from 0.63 to 9.6%. The relative contribution of all the amino acids to the peptides’ score is shown in Figure 2. This emphasizes that although the anchor positions are important, the nonanchor and flanking amino acids are equally important in deciding whether a peptide will ultimately be processed as an epitope.

Identification of predicted ligands from A-gliadin

Figure 3 shows the sequence for A-gliadin (clone pTO-A10, which bears a close homology to the only reported sequence derived directly from protein sequencing (26)) with the motifs for HLA-DQ2

<table>
<thead>
<tr>
<th>Cycle</th>
<th>1–3</th>
<th>4/5</th>
<th>7/8</th>
<th>9/10</th>
<th>12/13</th>
<th>14/15</th>
</tr>
</thead>
</table>

Using the PPP ranking, these T cell epitopes are easily identified. The importance of the flanking regions has already been demonstrated, and the importance of using all the positions in the peptide is emphasized using the HLA-DR2-presented T cell epitope from myelin basic protein (MBP). The binding characteristics and T cell recognition of this peptide have been extremely well studied (20). Autoreactive T cells in patients with multiple sclerosis recognize the immunodominant MBP 84–102 DENPVYHFEKNIKT-PRTPP peptide presented by HLA-DR2 (dominant binding anchors are underlined). The epitope has been narrowed down further for certain clones to the peptides 85 to 98 and 86 to 98. As shown in Table IV, the PPP rank would predict these epitopes with remarkable accuracy, with the peptide starting at N86 gaining the highest score and top rank position. If the data from the flanking regions are removed, this peptide falls from 0.63 to 9.6%. The relative contribution of all the amino acids to the peptides’ score is shown in Figure 2. This emphasizes that although the anchor positions are important, the nonanchor and flanking amino acids are equally important in deciding whether a peptide will ultimately be processed as an epitope.

Identification of predicted ligands from A-gliadin

Figure 3 shows the sequence for A-gliadin (clone pTO-A10, which bears a close homology to the only reported sequence derived directly from protein sequencing (26)) with the motifs for HLA-DQ2

<table>
<thead>
<tr>
<th>Cycles</th>
<th>1–3</th>
<th>4/5</th>
<th>7/8</th>
<th>9/10</th>
<th>12/13</th>
<th>14/15</th>
</tr>
</thead>
</table>

Using the PPP ranking, these T cell epitopes are easily identified. The importance of the flanking regions has already been demonstrated, and the importance of using all the positions in the peptide is emphasized using the HLA-DR2-presented T cell epitope from myelin basic protein (MBP). The binding characteristics and T cell recognition of this peptide have been extremely well studied (20). Autoreactive T cells in patients with multiple sclerosis recognize the immunodominant MBP 84–102 DENPVYHFEKNIKT-PRTPP peptide presented by HLA-DR2 (dominant binding anchors are underlined). The epitope has been narrowed down further for certain clones to the peptides 85 to 98 and 86 to 98. As shown in Table IV, the PPP rank would predict these epitopes with remarkable accuracy, with the peptide starting at N86 gaining the highest score and top rank position. If the data from the flanking regions are removed, this peptide falls from 0.63 to 9.6%. The relative contribution of all the amino acids to the peptides’ score is shown in Figure 2. This emphasizes that although the anchor positions are important, the nonanchor and flanking amino acids are equally important in deciding whether a peptide will ultimately be processed as an epitope.

Identification of predicted ligands from A-gliadin

Figure 3 shows the sequence for A-gliadin (clone pTO-A10, which bears a close homology to the only reported sequence derived directly from protein sequencing (26)) with the motifs for HLA-DQ2
and -DQ8 beneath. No obvious peptide ligand candidates can be identified using these binding motifs.

The HLA-DQ2 and -DQ8 pool sequences were used to screen the A-gliadin. The PPP values for 15 mers are shown graphically for HLA-DQ2 in Figure 4, with the five highest scoring 15 mers indicated. Two of these top five predicted ligands for HLA-DQ2, 210 to 224 and 32 to 46, overlap closely with top predicted ligands for HLA-DQ8, as seen by comparing Tables VI and VII. Both these peptides have been strongly implicated as being toxic to patients with celiac disease (reviewed in Ref. 27).

To examine these predicted ligands further, the peptides were synthesized, and in vitro binding assays were conducted; the results are shown in Tables VI and VII. For both alleles there is a clear negative correlation between the IC50 and the predicted ability to bind. The difference was approximately 100-fold for HLA-DQ2 and 10-fold for -DQ8 between the predicted binders and non-binders (p < 0.05). This demonstrates that this approach can predict physically binding ligands from proteins in which the motif alone fails.

**T cell recognition of the gliadin peptides**

PBMCs from three patients with coeliac disease were used in an ELISPOT to assess T cell recognition of gliadin peptides. The results are shown in Figure 5. It is clear that the peptides G32, G150, and G210 are seen as T cell epitopes. G150 has not been described as an epitope previously.

**Discussion**

Predicting which sequences will be processed from a protein and presented by the HLA complex to a T cell has proven difficult. The peptide that is eventually presented from the parent protein will be the result of a series of events: the ability of enzymes, such as proteases, to act upon the protein and generate short peptides; the

---

**Table V. Correlation of known T cell epitopes or eluted ligands with the predicted rank calculated from the relative PPP scores from the source protein using data from HLA-DQ2 and -DQ8.**

<table>
<thead>
<tr>
<th>Parent Protein</th>
<th>Size Epitope/Ligand</th>
<th>Predicted Rank&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Predicted Rank&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-DQ2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSP-18</td>
<td>148</td>
<td>31–43&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3/134 (2.2)</td>
<td>9/134 (6.7)</td>
</tr>
<tr>
<td>HSP-65</td>
<td>540</td>
<td>243–255&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1/526 (0.2)</td>
<td>14/526 (2.6)</td>
</tr>
<tr>
<td>A-gliadin</td>
<td>269</td>
<td>210–224&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1/255 (0.4)</td>
<td>105/255 (41.2)</td>
</tr>
<tr>
<td>HLA-DQ8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p21 ras(Q to L61)</td>
<td>171</td>
<td>51–67&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2/157 (1.27)</td>
<td>20/157 (12)</td>
</tr>
<tr>
<td>DQ8 α-chain</td>
<td>232</td>
<td>75–87</td>
<td>19/218 (8.7)</td>
<td>45/218 (20.6)</td>
</tr>
<tr>
<td>RL-18</td>
<td>187</td>
<td>17–42</td>
<td>28/173 (16)</td>
<td>20/173 (11.5)</td>
</tr>
<tr>
<td>SSBP</td>
<td>148</td>
<td>18–46</td>
<td>2/134 (1.49)</td>
<td>19/134 (14.2)</td>
</tr>
<tr>
<td>Transferrin receptor</td>
<td>760</td>
<td>332–347</td>
<td>46/746 (6.16)</td>
<td>18/746 (2.4)</td>
</tr>
</tbody>
</table>

<sup>a</sup> Predicted rank is calculated from the position of the PPP score for the ligand/epitope out of the total number of theoretical 15 mers from the parent protein. Percentages shown in parenthesis.

<sup>b</sup> As for a, but using the modified probability table. Percentages shown in parenthesis.

<sup>c</sup> T-cell epitopes.
exposure of these peptides to a compartment containing the HLA molecules; and the ability of the generated peptides to physically bind to the groove in the HLA heterodimer.

The most successful approach to epitope prediction to date has been in the context of HLA class I molecules (4, 5). With these examples the binding motif of the HLA molecules alone successfully identified the T cell epitopes from the parent protein. Predicting T cell epitopes from proteins using the motif data for HLA class II molecules has proven difficult, even with reasonably well-defined motifs. This may be due to the exclusion of information derived from the N- and C-terminal extensions of the class II binding peptides. In this report we have demonstrated that by screening the protein with the data from the pool sequence, which gives a value to every amino acid in the theoretical ligand, 15 mers can be scored and ranked relative to each other. The majority of the known peptide ligands and epitopes fell within the predicted top 5%, and the program worked particularly well for known epitopes. This shows the application of the technique to particular proteins that are antigenic for a disease and not only avoids the constraints of a particular HLA binding motif, which only takes into account the core binding region, but includes information from the N- and C-terminal extensions. If the screening program was to be applied to a new protein de novo, then ligands ranked in the top 5% might be considered candidates for potential epitopes.

It is particularly interesting that when the information contained in the N- and C-termini is removed, the predictive power of the program diminishes for many of the known ligands. For HLA-DR3, 8 of the 10 ligands had small shifts in either direction, which did not shift the predicted rank from the top 5%. However, two of the ligands did shift out of the top 5%. This effect was more marked for the HLA-DQ alleles, where three-quarters of the listed ligands deteriorated, with half of these falling below the top 5%. Only one ligand actually improved to an extent where it moved into the top 5% of the ranked 15 mers.

Both class I and class II molecules are part of a large family of proteins that probably evolved from a common ancestral protein, and they show marked similarities in the structure, as revealed by x-ray crystallography. Furthermore, recent demonstrations of how

FIGURE 2. Graphs showing the relative contributions to the PPP score for two consecutive MBP peptides that are both recognized as T cell epitopes.
the TCR interacts with the trimolecular class I complex reveal that only a small amount of the bound core of the peptide actually interacts with the TCR. It seems likely that the TCR will interact with class II molecules in a similar way (28), which might suggest that the sequences of the N- and C-terminal extensions are not relevant.

However, we find that this information does not appear to be redundant, as these data enhance the prediction of certain epitopes from a parent protein. This might merely reflect the cleavage sites of proteases involved in the class II Ag processing pathway, and it is known that epitope clustering can occur in proteins (29). Interestingly, almost all ligands eluted from class II molecules are longer than 9 mers, which would not be expected if these extensions are merely a by-product of processing. There are clear examples of class II-restricted T cell clones that require peptides longer than the core binding 9 mer for activation (30). The extensions might nonspecifically enhance binding (15), but there are also examples of loss of T cell reactivity associated with loss of the noncore amino acids, which is not reflected by a diminution in binding (20). Whether the N- and C-terminal extensions of class II peptides reflect specificity imposed by Ag processing or by the TCR, this study shows that the sequences do not appear to be random. This information could be indicative of a fundamental difference in the biology of HLA class I and II molecules.

We chose to examine A-gliadin, the Ag in celiac disease, because of its highly unusual amino acid composition, rendering it difficult to identify binding regions to HLA molecules. Celiac disease is strongly associated with HLA-DQ2, being present in >90%
of affected cases; the remainder have HLA-DQ8 (12). As an exogenous Ag, it seems likely that high affinity ligands would be the main candidates for epitopes.

The A-gliadin sequence was screened with the probability tables for HLA-DQ2 and -DQ8. Ligands with both high and low PPP scores were synthesized, and binding affinities were measured in vitro. The results show a significant correlation between the ligands with relatively high PPP scores, and the high affinity binders indicated by low IC50 values. There has not been a published report on the binding of gliadin peptides to the HLA-DQ8 molecule, and only one for HLA-DQ2 (31). This study looked at the binding of peptides from only the N-terminal region of gliadin, and no high affinity ligands were identified.

Using this approach, not only did we identify binding ligands from the unusual gliadin sequence, but two of the ligands that coincide for HLA-DQ2 and -DQ8, 210 to 224 and 32 to 46, have been previously indicated to be of pathogenic significance. The top predicted ligand is from the region of A-gliadin 210 to 224. This contains the dodecapeptide sequence 210 to 221, which has some homology to a coat protein E1b from adenovirus type 12, and molecular mimicry after an enteric infection with the virus has been suggested as a possible contributing mechanism to the pathogenesis of the disease (32). This cross-reacting epitope may activate T cells, which then recognize the epitope from gliadin, leading to enteropathy after subsequent dietary exposure to the protein. There are three lines of evidence to support this theory: epidemiologic studies have found an increased exposure to virus in patients with celiac disease compared with controls (33, 34); there are increased T cell responses to both the viral peptide (35) and the dodecapeptide (36) in celiac patients; and in vivo feeding experiments in patients have shown a toxic effect with the synthesized dodecapeptide (22). In vivo feeding experiments have also suggested the peptide 32 to 46 is recognized as an epitope (37).

Using the sensitive biologic read-out of IFN-γ release from T cells, three of the gliadin peptides are recognized as epitopes, including the above two peptides, G32 and G210, as well as the peptide G150, which has not been implicated as an epitope previously. Determination of the roles of these peptides in disease pathogenesis awaits further studies.

Our findings show that complete eluted peptide sequence data can be employed to identify peptide ligands and T cell epitopes for both HLA-DR and -DQ alleles. It also highlights the importance of the information that is incorporated in the nonbinding regions of peptide ligands and suggests that these regions may have a unique role in the immunobiology of HLA class II molecules.

Acknowledgments

We are very grateful to Prof. Rammensee and his group in Tubingen, Germany, for providing the original data on pool sequences of HLA-DR3 and -DQ8, and to Nick Willcox and Awen Gallimore for useful comments on the manuscript.

References


