Brief Definitive Report

Precise Prediction of Major Histocompatibility Complex Class II–Peptide Interaction Based on Peptide Side Chain Scanning

By Juergen Hammer,* Elisa Bono,* Fabio Gallazzi,* Charles Belunis,† Zoltan Nagy,† and Francesco Sinigaglia*

From *Roche Milano Ricerche, I-20132 Milan, Italy; and the †Department of Inflammation and Autoimmune Diseases, Hoffmann-La Roche Inc., Nutley, New Jersey 07110-1199

Summary

We describe here a new method for predicting class II major histocompatibility complex-binding peptides, based on the preferences observed in a systematic series of peptide binding experiments where each position in a "minimal" peptide was replaced individually by every amino acid. The DRB1*0401 peptide binding preferences were determined and incorporated into a computer program that looks through sequences for potential epitopes and assigns each a score. These scores correlate well with previously determined T cell epitopes of foreign antigens and endogenous peptides from self proteins. Our findings hold implications for the design of subunit vaccines and in the identification of autoantigenic peptide regions within self proteins.

MHC class II molecules are highly polymorphic proteins that bind peptides derived from processing of antigens and present them to T cells (1). Methods to predict regions in protein sequences capable of binding to human MHC class II molecules would be very valuable for many immunological applications. A quantitative prediction of MHC binding sites, however, requires the understanding of the rules governing peptide–MHC class II interaction. The use of M13 peptide display libraries to identify allele-specific motifs (2–4), together with structural analyses of HLA-DR1–peptide complexes (5, 6) provided the basis to predict MHC class II–peptide interaction: (a) the identification of precisely spaced anchor positions in large HLA-DR-selected peptide pools and the x-ray crystallographic studies suggested an overall similar, sequence–independent peptide conformation (3, 5); (b) the fact that the anchor at position 1 (p1) was found obligatory for high affinity binding allowed us to define frames in HLA-DR bound peptides (4); (c) the position of a particular side chain, with respect to the p1-frame, determines whether a given amino acid behaves as anchor residue, inhibitory residue, or residue neutral for binding (4). This information led us to develop a new method for predicting class II MHC–binding peptides. The method is based on the hypothesis that peptide binding to HLA-DR correlates with the net result of all side chain effects in any given peptide, and that most side chain effects depend on their relative position to the p1 anchor, rather than on the remaining peptide sequence. P1-anchored designer peptide libraries were thus used for scanning the relative peptide positions 2–9 for the effects of each amino acid side chain on binding to the human class II molecule DRB1*0401. When values, individually obtained by side chain scanning, were assigned to the amino acid residues of the selected protein region, the sum of these values correlated with the MHC-binding affinity of the respective peptide. The processing of the side chain scanning data into a software allowed a rapid and nearly perfect prediction of DRB1*0401 binding regions in any protein sequence.

Materials and Methods

Preparation of HLA-DRB1*0401 Molecules. DRB1*0401 molecules were isolated from human lymphoblastoid cell lines BSM and PREISS as described (7).

Peptide Synthesis. P1-anchored designer peptide libraries were synthesized with multiple peptide synthesizers (models 350 and 396; Advanced ChemTech, Louisville, KY) using Fmoc chemistry. Double and triple coupling procedures and at least a 10-fold surplus of Fmoc amino acids as compared with the free NH2 termini of the resin were used to minimize the production of deletion mutants. Peptides were analyzed by analytical reverse-phase HPLC. Peptides used in this study were routinely >75% pure.

High-flux DR-peptide Binding Assay. Peptides were dissolved in DMSO to a final concentration of 2 mM and 50 µl of each peptide was transferred into the A-row wells of 96-well plates (dilution plates). 50 µl of PBS was added to the A-row wells and 180 µl of PBS/DMSO (1:1) to the B-F row wells of the dilution plates. Peptides were simultaneously diluted by serially transferring 20 µl from row A to row F. 10 µl of each well of the dilution plate was transferred to the corresponding wells of a second set of 96-well plates (binding plates). 30 µl of mix 1 (500 ng/ml 125I-labeled Flu

1 Abbreviations used in this paper: HA, haemagglutinin; IC, inhibitory concentration; MBP, myelin basic protein; p1, position 1.
hemagglutinin (HA) peptide 307–319 in PBS) and then 30 µl of mix 2 (30 µg/ml affinity-purified DRBI*0401 antigen in 1% NP-40; 1 mM PMSF; 1 mM EDTA; 10 µg/ml of each of the following protease inhibitors: soybean trypsin inhibitor, antipain, pepstatin, leupeptin, and chymostatin; 0.15 M NaCl; and 50–100 mM citrate-phosphate buffer, pH 7) were added to each well of the binding plates and incubated for 24 h at room temperature. 30 µl of mix 3 (150 µg/ml biotinylated anti-DR antibody in PBS) was added and incubated for another 3 h at room temperature. 200 µl of BSA-blocked streptavidin on 4% beaded agarose was added into each tube (row A to F) of cluster tube racks. 70 µl of each well from the binding plate was transferred to the corresponding tubes of the cluster tube racks and incubated for 10 min. The streptavidin solid phase was simultaneously washed four times with PBS using specifically constructed devices (8). The tubes were directly counted in a gamma counter. The high-flux peptide binding assay permitted the processing of a p1-anchored designer peptide library within 2 d. This corresponds to 960 single assays using six dilutions for each peptide.

**Prediction of HLA-DRBI*0401 Binding Regions.** The inhibitory concentration (IC₅₀) data derived from side chain scanning of p1-anchored designer libraries were processed as follows. The IC₅₀ data were normalized by dividing them by the IC₅₀ values obtained from the Ala-substitutions. The reverse values of the normalized data were determined, and last, the logarithms (log) of the reverse values were calculated. These values were directly incorporated in a simple prediction software, written with Microsoft QBasic for Macintosh (8). The score is thus the sum of log (IC₅₀-Ala/IC₅₀-residue) for each residue, minus 1 if P1 is aliphatic. The "aliphatic correction" reflects the ~10-fold lower binding affinity of designer peptides with aliphatic p1 anchors as compared with peptides with aromatic p1-anchors (4).

**Results and Discussion**

X-ray crystallographic studies indicate that different peptides bind with a similar conformation to HLA-DR molecules (6). This observation is supported by the analysis of
the large number of peptide sequences selected from M13 phage display libraries (2, 3). Although these peptides differ in their primary sequence, most of them share perfectly spaced anchor residues, suggesting an overall similar and sequence-independent peptide conformation. Experiments with designer peptides further support this hypothesis. Anchor and inhibitory residues lose their effect in different designer peptides if shifted by one position towards the NH2 or COOH terminus (4). This indicates a low degree of flexibility in the peptide structure once bound to the MHC molecule and suggests that once all the peptide side chain effects are known, peptide–MHC class II interaction can be predicted. To validate this hypothesis we performed a quantitative analysis of the side chain effects at each peptide position. To enable a detailed analysis it was important to prevent potential shifts within the peptide frame, and to supply a sensitive peptide binding assay to accurately measure side chain effects. To meet these requirements we synthesized pl-anchored designer peptide libraries. They consisted of short, pl-anchored and Ala-based designer peptides where all naturally occurring L-amino acids had been substituted at each position from 2 to 9 (Fig. 1 a). Nine residue–long peptides cover most of the potential side chain interactions within the MHC cleft, and show reduced main chain interaction compared with longer peptides (4, 6). The reduced length amplifies the effect of anchors and inhibitory residues and decreases the probability of shifts within the p1-frame.

A typical side chain scanning of the two basic library peptides YASAAAAAA and YRSMAAAAA with Glu and Arg on the HLA-DRB1*0401 molecule is shown in Fig. 1, b–c. Although the two peptides differ in their IC50 values ~100-fold, due to the two additional anchor residues within the latter (3, 4), the results of the side chain scanning were nearly identical, supporting the concept of position-specific side chain effects independent of the remaining sequence.

We next performed a side chain scanning of all natural amino acid residues (except Cys) at positions 2 to 9 of YASAAAAAA using a newly developed high-flux peptide binding assay as described in Materials and Methods. The results shown in Table 1 revealed new position-specific positive and negative side chain effects in addition to the anchor residues previously identified by phage library screening. The logarithm of the normalized and reverse IC50 values derived from the side chain scanning were then processed into a simple computer algorithm (Fig. 2). The program first scans a given protein sequence for the obligatory DRB1*0401–pl anchor residues. After their identification, amino acid residues at relative positions 2–9 are located. Next, values derived from the side chain scanning (Table 1) are assigned to each amino acid residue of the selected peptide or protein region. The sum of these values is defined as “peptide score.” Thus, the peptide score correlates directly with the biochemical data derived from the side chain scanning. If one assumes that the position-specific effects of amino acid side chains are independent of the remaining peptide sequence, the peptide score should in addition correlate with the binding affinity of the peptides and henceforth allow one to predict class II–peptide interaction.

### Table 1. Results of the Side Chain Scanning on DRB1*0401 with a p1-Anchored Peptide Library

<table>
<thead>
<tr>
<th>p1</th>
<th>p2</th>
<th>p3</th>
<th>p4</th>
<th>p5</th>
<th>p6</th>
<th>p7</th>
<th>p8</th>
<th>p9</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-1$\pm$0.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>-1.3$\pm$1.3</td>
<td>1.7</td>
<td>-0.2</td>
<td>0</td>
<td>-1.1</td>
<td>-1.1</td>
<td>-2.6</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>+0.1</td>
<td>-1.2</td>
<td>+0.8</td>
<td>-0.1</td>
<td>-1.2</td>
<td>-0.2</td>
<td>-0.2</td>
<td>-1.8</td>
</tr>
<tr>
<td>F</td>
<td>+0.8</td>
<td>+0.8</td>
<td>+0.8</td>
<td>-0.3</td>
<td>-1.3</td>
<td>-0.8</td>
<td>+0.1</td>
<td>-0.8</td>
</tr>
<tr>
<td>G</td>
<td>+0.5</td>
<td>+0.2</td>
<td>-1.5</td>
<td>+0.2</td>
<td>-1.1</td>
<td>-1.5</td>
<td>-0.5</td>
<td>-0.2</td>
</tr>
<tr>
<td>H</td>
<td>+0.8</td>
<td>+0.2</td>
<td>+0.8</td>
<td>-0.1</td>
<td>-1.6</td>
<td>-0.8</td>
<td>0</td>
<td>+0.3</td>
</tr>
<tr>
<td>I</td>
<td>-1.0</td>
<td>+1.1</td>
<td>+1.5</td>
<td>+0.8</td>
<td>+0.1</td>
<td>-0.2</td>
<td>-0.2</td>
<td>-0.1</td>
</tr>
<tr>
<td>J</td>
<td>+1.1</td>
<td>0</td>
<td>-2.2</td>
<td>+0.3</td>
<td>-2.3</td>
<td>1.2</td>
<td>+0.9</td>
<td>-0.9</td>
</tr>
<tr>
<td>L</td>
<td>-1.0</td>
<td>+1.0</td>
<td>+1.0</td>
<td>+0.6</td>
<td>-0.1</td>
<td>+1.3</td>
<td>+0.4</td>
<td>+0.6</td>
</tr>
<tr>
<td>M</td>
<td>-1.0</td>
<td>+1.1</td>
<td>+1.4</td>
<td>+1.4</td>
<td>+0.3</td>
<td>-1.3</td>
<td>+0.7</td>
<td>+0.4</td>
</tr>
<tr>
<td>N</td>
<td>+0.8</td>
<td>+0.5</td>
<td>+0.5</td>
<td>+0.2</td>
<td>+1.7</td>
<td>-0.1</td>
<td>+0.7</td>
<td>-1.1</td>
</tr>
<tr>
<td>P</td>
<td>-0.5</td>
<td>-0.3</td>
<td>-2.1</td>
<td>+0.5</td>
<td>+0.1</td>
<td>-0.3</td>
<td>-0.2</td>
<td>-1.6</td>
</tr>
<tr>
<td>Q</td>
<td>+1.2</td>
<td>0</td>
<td>+1.1</td>
<td>+0.1</td>
<td>-1.2</td>
<td>-0.5</td>
<td>+1.6</td>
<td>+0.7</td>
</tr>
<tr>
<td>R</td>
<td>+2.2</td>
<td>+0.7</td>
<td>-1.5</td>
<td>0</td>
<td>-2.2</td>
<td>+1.3</td>
<td>+0.4</td>
<td>+0.6</td>
</tr>
<tr>
<td>T</td>
<td>-0.3</td>
<td>+0.2</td>
<td>+1.1</td>
<td>+0.4</td>
<td>+1.7</td>
<td>-0.4</td>
<td>+0.6</td>
<td>+1.2</td>
</tr>
<tr>
<td>V</td>
<td>-1.0</td>
<td>+2.1</td>
<td>+0.5</td>
<td>+0.5</td>
<td>+0.4</td>
<td>+1.3</td>
<td>+0.5</td>
<td>+0.4</td>
</tr>
<tr>
<td>W</td>
<td>0</td>
<td>-0.1</td>
<td>0</td>
<td>-1.2</td>
<td>-0.1</td>
<td>-0.9</td>
<td>+1.3</td>
<td>-0.6</td>
</tr>
<tr>
<td>Y</td>
<td>+0.9</td>
<td>+0.8</td>
<td>-1.0</td>
<td>-0.2</td>
<td>+1.1</td>
<td>-1.7</td>
<td>+1.3</td>
<td>-1.5</td>
</tr>
</tbody>
</table>

The peptide YASAAAAAA was used as basic library peptide.

* The single letter code for amino acids is: A, Ala; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Glu; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. Cysteines were not included in the side chain scanning procedure.

† Dashes in the row corresponding to position 1 indicate that only aliphatic or aromatic residues were considered as p1 anchor (4).

‡ Relative binding values were determined as in Fig. 1.

§ –1 values were arbitrarily assigned to aliphatic p1 anchors. This corresponds to an ~10-fold lower binding affinity of designer peptides with aliphatic p1 anchors as compared with peptides with aromatic p1-anchors (4).

To validate our scoring system we first compared the distribution of peptide scores in a DRB1*0401 phase-selected peptide pool (3) with the score distribution in a nonselected peptide pool (Fig. 3 a). Fig. 3 b shows a significant shift of the DRB1*0401-selected peptides towards higher scores. 85% of the selected peptides had scores higher than 2 as compared with only 13% of the nonselected peptides. Three peptides from the DRB1*0401-selected pool had a low score (Fig. 3 b), and accordingly, these peptides were unable to bind to DRB1*0401, whereas seven out of seven peptides tested having high scores showed high affinity binding (data not shown).

To further evaluate the correlation between scores and peptide binding, we next synthesized 48 nonamer and 56 thirteener peptides derived from known protein sequences with scores ranging from –5 to +5. As shown in Fig. 3, c and d, most peptides exhibited a correlation between scores and binding affinities to DRB1*0401. It is important to note that peptides with the highest affinities also had the highest scores...
predict human MHC class II-peptide interaction. The side chain peptides having only one or two anchor residues, however, in frame (4, 9-11). The binding of the vast majority of the peptides or simple class II peptide binding motifs allowed only for limited predictions of peptides with three and four residues (Fig. 3 d), demonstrating the usefulness of this method to predict human MHC class II–peptide interaction.

It has so far been very difficult to predict with reasonable accuracy whether a particular peptide sequence will be able to bind to a particular mouse or human MHC molecule. Only very recently Reay et al. (8a), by using a synthetic peptide library, have been able to identify a predicting motif for the IE3 molecule. Previous approaches based on sequence alignments or simple class II peptide binding motifs allowed only for limited predictions of peptides with three and four residues in frame (4, 9–11). The binding of the vast majority of the peptides having only one or two anchor residues, however, could not be identified by these algorithms. The side chain scanning method is highly predictive. For the identification of DRBI*0401 binding regions it first selects all peptides with the obligatory p1-anchor residue, which account for as much as 30% of the possible nonamers in randomly chosen sequences (Fig. 3 a). However, only ~4% of all peptide frames have a score higher than 2, and even less bind with high affinity (Fig. 3, c and d). This demonstrates a considerably higher stringency of DRBI*0401-peptide binding than expected on the basis of previous DRBI*0401 motifs (3, 12).

The ability to predict class II–peptide binding, based solely on data obtained from side chain scanning, suggests that the conformation of most class II bound peptides is independent of their primary structure. This is in accordance with recent structural data from Stern et al. (6), indicating that conserved hydrogen bonds between HLA-DRBI*0101 residues and the peptide main chain may provide a universal mode of peptide binding. Although the scoring system seems to be highly predictive, we cannot exclude the possibility that in a few instances the peptide side chains might interact with one another, leading to changes within the conformation of the peptide. This could explain why the biochemical data of few of the peptides shown in Fig. 4, c and d did not conform to the predicted score.

The p1-anchor residues have been shown to be necessary for high affinity peptide-binding for many HLA-DR molecules (4, 13–15). The described procedure, therefore, seems applicable to many different HLA-DR molecules. In a few instances, however, peptides lacking the p1 anchor have been shown to bind with low affinity (13). For a complete prediction, which also includes these particular cases, it may be necessary to calculate the score for all peptide frames in a given protein.

An obvious application is the identification of T cell epitopes in natural polypeptides. Dominant T cell epitopes are often peptides that bind most avidly to MHC molecules. Hence, the possibility to predict HLA-DR–peptide interaction should help define potential dominant epitopes in individual proteins at high precision. For example, the scanning of the human myelin basic protein (MBP) molecule for regions with high peptide scores revealed a single peptide with a score higher than 2 (Fig. 4 a). This peptide (MBP 91–99)
belongs to a region of the MBP (MBP 87-106), previously described as immunodominant for the DRB1*0401-restricted T cell recognition (16, 17). Similarly, the region 309–317 of the influenza HA molecule shows the highest score of all possible HA peptides (Fig. 4 b). This peptide is known as the immunodominant T cell epitope of the HA molecule (18).

Sequences of naturally processed peptides, eluted from DRB1*0401 molecules were recently determined by a combination of mass spectrometry and Edman microsequencing (19). The predominant endogenous peptides were derived from MHC molecules. A few peptides derived from exogenous bovine serum proteins were also bound to the DRB1*0401 allele. Fig. 4, c and d shows the results of the screening of two of the above mentioned proteins using our scoring system. In each case the identified peptide is one of the sequences with scores higher than 2. This methodology and its predictive potential could now be used to identify the naturally processed and presented peptides from potentially autoimmunogenic proteins restricted by disease-linked class II MHC molecules. An interesting case could be, for example, the analysis of potential autoantigens in rheumatoid arthritis, an autoimmune disease genetically linked to the DRB1*0401 allele (for reviews see references 20, 21).

Figure 4. Immunodominant T cell epitopes or naturally processed peptides eluted from DRB1*0401 molecules show high peptide scores. Four proteins were scanned for peptide regions with high scores: (a) human MBP, (b) influenza HA, (c) bovine hemoglobin α chain (HBA), and (d) HLA-A2 β chain. Most of the p1-anchored nonamer peptides had scores between -2 and 1 as expected from the distribution of scores in randomly selected peptides (see Fig. 3 a), whereas only a few peptides had scores that correspond to high affinity binding. Among those were peptides (arrows) that were previously identified as immunodominant T cell epitopes (a and b) or as naturally processed peptides eluted from the DRB1*0401 molecule (c and d): MBP 91–99 (arrow, a), HA 309–317 (arrow, b), HBA 33–41 (arrow, c), and HLA-A2 58–66 (arrow, d).

We thank Ron Germain for critical reading of the manuscript and for helpful discussion.

Address correspondence to Dr. Francesco Sinigaglia, Roche Milano Ricerche, via Olgettina 58, I-20132 Milan, Italy.

Received for publication 23 May 1994 and in revised form 8 July 1994.

Note added in proof: The prediction program is now in press (reference 8) and is available from the authors.

References


7. Sinigaglia, F., P. Romagnoli, M. Guttinger, B. Takacs, and


