PEPTIDE SEQUENCES BINDING TO MHC CLASS I PROTEINS

MARGARET H. SMITH,* KIT S. LAM,† EVAN M. HERSH,‡ MICHAL LEBL† and WILLIAM J. GRIMES§

*Department of Biochemistry, University of Arizona, Tucson, AZ 85721, U.S.A.; †Arizona Cancer Center and Department of Medicine, Tucson, AZ 85724, U.S.A.; and ‡The Selectide Corporation, 10900 N. Stallard Place, Tucson, Arizona, U.S.A.

(First received 18 March 1994; accepted in revised form 11 July 1994)

Abstract—Motifs for peptides which bind specifically to the human class I major histocompatibility complex molecules HLA-A2 and B7 were determined by sequence analysis of class I-bound peptides selected from a random synthetic library of nonamers. Thirteen individual peptides were sequenced for HLA-A2, twelve individual and nine pooled peptides were sequenced for HLA-B7. Analysis of sequence alignment implicated four peptide positions in potential contact with the class I HLA-A2 molecule and three positions for the HLA-B7 molecule. The results demonstrate that a synthetic peptide library can be used to identify allele-specific motifs for class I molecules, providing information comparable to the results obtained from sequencing endogenous peptides. This method utilizes denatured class I heavy chains, and similar results were obtained using a class I protein purified from mammalian cells or by expression in Escherichia coli. This method has the potential to detect peptides which may not be generated physiologically, but due to their binding properties, may be valuable to predict or engineer immunomodulatory T cell epitopes.

Key words: major histocompatibility, peptides, antigens, T-cell.

INTRODUCTION

Crystallographic studies of HLA class I proteins identified a cleft formed by intrachain folding of the a-1 and a-2 domains of the heavy chain (Bjorkman et al., 1987; Saper et al., 1991; Madden et al., 1991). Electron density maps (Garrett et al., 1989) showed high variability in this region, and subsequent studies demonstrated that small peptides bound in this cleft are recognized by T cell receptors. Size and directionality of peptide binding is dictated by the interaction between the peptide and pockets in the cleft of the class I molecule, with a chain length of nine residues optimal. Recently, endogenous peptides were eluted and sequenced from five murine and human class I antigens (Falk et al., 1991; Jardetzky et al., 1991; Van Bleek and Nathenson, 1991; Hunt et al., 1992; Huczko et al., 1993; DiBrino et al., 1993). These findings support the hypothesis that the repertoire of class I-presented peptides is determined primarily by certain residues in key positions of the peptide (anchor residues), which interact specifically in the class I binding cleft. Molecular modeling of a few of these systems suggests that the side-chains of anchor residues protrude into unique allele-specific pockets within the binding cleft of the class I heavy chain. Mutational analysis of heavy chain encoded residues lining these pockets has demonstrated their importance in conferring peptide specificity (Hogan et al., 1988; Iatrou et al., 1997). A study by Madden et al. (1993) using X-ray crystallography to analyse five different viral peptides bound in the groove of HLA-A2 has shown that the peptide termini, residues in position 2 (P2), and the C-terminal anchor side-chains were bound similarly in all five cases. Although the ends of these various peptides were fixed in a similar manner, the conformation of each peptide varied in the center of the binding groove, consistent with the hypothesis that this portion of the peptide provides the primary antigenicity of the class I-peptide complex. Site-directed mutagenesis to "transplant" the B pocket from HLA-A2 into HLA-B27 demonstrated that a dominant polymorphic pocket in the heavy chain interacting with an anchor residue in the nonamer can permit the class I molecule to distinguish among peptides differing by a single amino acid (Colbert et al., 1993).

The identification of allele-specific motifs provides a starting point to search for gene products harboring T cell epitopes, and thereby a means to study the biology of this recognition system. We have pursued an alternative approach for determining class I allele-specific peptide motifs, which we hope will yield information complementary to that already obtained using endogenous peptides. Our method employs a library of randomly generated synthetic peptides bound to beads. Libraries of nonamers were constructed by a "split synthesis" approach (Lam et al., 1991; Furka et al., 1991). This technique produces peptide-bound beads where each individual solid-phase head expresses only one peptide entity, "the one-bead one-peptide concept", and is known as the "Selectide process" (Lam et al., 1991, 1993; Lam and Lebl 1992). Peptide-beads reacting with class I molecules during renaturation were identified by an enzyme-linked immunoassay with monoclonal antibody W6/32. This antibody recognizes an epitope found only in the native configuration of class I heavy chain, requiring association with β2m (Brodsky et al., 1979). Correct folding of class
I molecules on the surface of the appropriate peptide-bead is favored by the high concentration of peptides on the surface of the bead (Lam et al., 1993), and can be driven towards completion by the addition of excess β2-m (Elliott and Eisen, 1990). Positively stained beads were collected and subjected to microsequencing individually, or as pools of 4–5 beads sequenced simultaneously.

MATERIALS AND METHODS

Cell lines and antibodies

The human lymphoblastoid cell line JY was supplied by Dr. J. Strominger (Parham et al., 1977). Cells were maintained in RPMI 1640 (Sigma Chemical Company, St Louis, MO) supplemented with 10% FCS (Gemini Bioproducts, Calabasas, CA), 2 mM L-glutamine, 100 

| mg/ml streptomycin and 100 U/ml penicillin (Irvine Scientific, Santa Ana, CA). Cells were incubated at 37°C in a humidified atmosphere of 5% CO2.

Murine mAb W6/32 obtained from American Type Culture Collection (Rockville, MD) was purified from ascites fluid using protein A-Sepharose (Sigma). Purified mAb was conjugated to alkaline phosphatase using maleimide activated enzyme (Pierce, Rockford, IL), and screened for specificity by ELISA.

Purification of HLA antigens

Papain-solubilized HLA-A2.1 and HLA-B7 proteins were biochemically purified from JY cell membrane preparations as described (Parham et al., 1977; Turner et al., 1975). Purification was assessed by immunoprecipitation and SDS-PAGE analysis.

Peptide libraries and screening

Randomly generated libraries of nonamers were synthesized on resin support beads using the 19 common amino acids (cysteine was excluded) by a “split synthesis” approach described in detail elsewhere (Lam et al., 1991). Briefly, this method allows for the construction of a single peptide species on an individual bead. In the first cycle of synthesis, resin beads are aliquoted into separate reaction vessels each with a single amino acid under conditions driving the coupling reactions to completion. After the coupling reaction, beads from all the separate reactions are pooled and then randomly distributed again into separate reaction vessels each with a single amino acid. This process of repooling the beads after each coupling reaction followed by splitting the pool into separate reaction vessels for coupling is repeated until the required chain length is obtained. With this approach, each bead presents only a single peptide species and a large library of millions of beads can potentially present all possible random peptides in roughly equimolar proportions. Each bead contains 50–200 pmol of peptide. Two libraries were generated for these studies, the first utilized a polyethylene glycol linker to space the nonamer from the support bead, and the second library was constructed with an additional ββββ linker to provide greater flexibility for the nonamers.

For screening, approximately 200,000 randomly chosen linear nonapeptide beads were reacted with 50 μg class I protein in 50 mM sodium phosphate buffer (pH 7.0), 5 M NaSCN, 45 mg BSA and 90 μg of human β2-microglobulin (Calbiochem, La Jolla, CA). Reactions were mixed for 30 min at room temperature, and dialysed against PBS containing 0.1% Tween 20 for 24 hr. Beads were washed once with PBS and resuspended in PBS with 0.05% gelatin and 0.1% Tween 20. Alkaline phosphatase conjugated mAb W6/32 was then added to the reaction and allowed to incubate for 2 hr at room temperature. Resin beads were washed and reacted with BCIP colorometric enzyme substrate as described (Lam et al., 1991, 1993). After 1 hr, colored beads identified under a dissecting microscope were removed and sequenced as described previously (Lam et al., 1991).

RESULTS

Selection of HLA-A2 synthetic peptides

In the nonapeptide library there are 199 possibilities, however, only 200,000 randomly chosen peptide beads were screened in each experiment, as limited by the available quantities of purified class I proteins. Results from screening a portion of the library for peptides binding to HLA-A2.1 identified 13 positive beads, of which six peptides were sequenced as shown in Table 1a. A second independently generated peptide library was constructed using a β alanine- aminoacaproyic acid-β alanine- amino acid (ββββ) linker, to provide greater flexibility for interaction with the class I heavy and light chains. Selection based on this second library yielded 64 positive beads, seven of these were sequenced and are shown in Table 1b.

Comparison of HLA-A2 synthetic and endogenous peptides

Table 2 summarizes published data on the amino acid sequences of natural peptides either eluted from HLA-A2 molecules or from HLA-A2 restricted CTL epitopes obtained in various studies (Hunt et al., 1992; Utz et al., 1992; Tsomides et al., 1991; Parker et al., 1992a; Morrison et al., 1992; Wei and Cresswell, 1992; Henderson et al., 1992). Analysis of these endogenous peptides suggest the anchor residues for HLA-A2 consist of Leu or Met at position 2 and an aliphatic residue (Leu, Ile or Val) at position 9.

This binding motif is totally consistent with the 13 HLA-A2 specific peptides selected by our library screen (Table 1). However, we have identified three additional consensus residues in our motif that are absent in the peptides identified by other methods. In our peptides, 9 out of 13 contain Phe in position 1, 8 out of 13 have aromatic residues (4 Trp, 2 Tyr and 2 Phe) in position 3, and 9 out of 13 have a carboxyl side-chain (4 Asp and 5 Glu) in position 4. The reason for these differences is not totally clear, but could arise from the non-randomness of the natural situation, in which only a subset of potential peptides are available due to the nature of peptide processing and transport. The libraries we screened on the other hand, were totally random (Poisson Distribution). Additionally, we tend to select and sequence the darkest...
Class I binding peptides

<table>
<thead>
<tr>
<th>Sequence</th>
<th>t1/2</th>
</tr>
</thead>
<tbody>
<tr>
<td>F L Y E</td>
<td>730</td>
</tr>
<tr>
<td>F L W E</td>
<td>2890</td>
</tr>
<tr>
<td>F M L G</td>
<td>882</td>
</tr>
<tr>
<td>F M L D</td>
<td>202</td>
</tr>
<tr>
<td>M M Q D</td>
<td>132</td>
</tr>
<tr>
<td>K/W L N</td>
<td>172</td>
</tr>
</tbody>
</table>

Table 1. Synthetic peptides bound to HLA-A2.1

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Sequence</th>
<th>t1/2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td>F L Y E R/A V P</td>
<td>730</td>
</tr>
<tr>
<td></td>
<td>Q L</td>
<td></td>
</tr>
<tr>
<td></td>
<td>730</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2890</td>
<td></td>
</tr>
<tr>
<td></td>
<td>882</td>
<td></td>
</tr>
<tr>
<td></td>
<td>202</td>
<td></td>
</tr>
<tr>
<td></td>
<td>132</td>
<td></td>
</tr>
<tr>
<td></td>
<td>172</td>
<td></td>
</tr>
</tbody>
</table>

beads, or presumably the ones with highest binding affinity.

Our data suggest that Phe in position 1, aromatic residues in position 3, and carboxyl residues in position 4 may each contribute additional binding energy and therefore be selected preferentially. This premise is supported by the findings of Parker et al. (1992b) who measured binding affinities of synthetic peptides constructed with the minimal anchor residues of Leu and Val at positions 2 and 9, respectively, and Gly in all other positions. Their results showed that minimal peptides containing only two non-Gly residues do not bind, suggesting that side-chains from other residues contribute to the binding energy. Their most stable complexes formed with peptides containing Phe in position 3, and the peptide with the longest half-life for dissociation in their study was the HN pol 968-976 (shown in Table 2), which contains the aromatic residue Trp in position 3. Analysis

Table 2. HLA-A2 restricted CTL epitopes and endogenous peptides

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Reference</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>L/M - E/K - V - K V</td>
<td>Falk et al., 1991</td>
<td>Protein phosphatase</td>
</tr>
<tr>
<td>S I/L P S G G</td>
<td>Hunt et al., 1992</td>
<td></td>
</tr>
<tr>
<td>L L D V P T A A V</td>
<td>Hunt et al., 1992</td>
<td></td>
</tr>
<tr>
<td>G I/L V P F I/L V s V</td>
<td>Hunt et al., 1992</td>
<td></td>
</tr>
<tr>
<td>S I/L I/L P A I/L V E I/L</td>
<td>Hunt et al., 1992</td>
<td></td>
</tr>
<tr>
<td>S I/L I/L V R A I/L E V</td>
<td>Hunt et al., 1992</td>
<td></td>
</tr>
<tr>
<td>K I/L N E P V I/L I/L</td>
<td>Hunt et al., 1992</td>
<td></td>
</tr>
<tr>
<td>Y L L P A I V H I</td>
<td>Hunt et al., 1992</td>
<td>Human nuclear p68</td>
</tr>
<tr>
<td>T L W V D P Y F Y</td>
<td>Hunt et al., 1992</td>
<td>TIS21</td>
</tr>
<tr>
<td>F I A G N S A Y E</td>
<td>Utz et al., 1992</td>
<td>HCMV gB 618-628</td>
</tr>
<tr>
<td>L L F G Y P V Y V</td>
<td>Utz et al., 1992</td>
<td>HTLV Tax 11-19</td>
</tr>
<tr>
<td>I I K E P V H G V</td>
<td>Tsomides et al., 1991</td>
<td>HIV RT</td>
</tr>
<tr>
<td>I I D K K V E K V</td>
<td>Parker et al., 1992a</td>
<td>Hsp 84</td>
</tr>
<tr>
<td>G I L G F V F V T L</td>
<td>Morrison et al., 1992</td>
<td>flu M1 58-66</td>
</tr>
<tr>
<td>V L F R G G P P R G</td>
<td>Wei and Cresswell, 1992</td>
<td></td>
</tr>
<tr>
<td>L L D V P T A A V</td>
<td>Henderson et al., 1992</td>
<td>IP-30</td>
</tr>
<tr>
<td>M L L S V P L L L</td>
<td>Henderson et al., 1992</td>
<td>Calreticulin</td>
</tr>
<tr>
<td>S L Y N T V A T L</td>
<td>Parker et al., 1992b</td>
<td>HIV gag 7785</td>
</tr>
<tr>
<td>I L K E P V H G V</td>
<td>Parker et al., 1992b</td>
<td>HIV pol 510-518</td>
</tr>
<tr>
<td>L L W K K G E G A V</td>
<td>Parker et al., 1992b</td>
<td>HIV pol 965-976</td>
</tr>
</tbody>
</table>

Peptide sequences of HLA-A2 restricted T cell epitopes and endogenous nonamers described in the literature, shown here aligned by their 'anchor' residues.

Standard single letter aa code. Potential anchor residues are shown in bold. Falk et al. (1991), sequenced pools of peptides, and only the dominant residues are shown, while a hyphen indicates a position which could not be assigned. Peptides described by Hunt et al. (1992), were sequenced by tandem mass spectrometry which can not distinguish between Ile and Leu indicated by (I/L).

The protein source for the peptide is given when known.
of other peptides in their study showed that although Ile
is found in endogenous peptides at position 2, Leu in this
position leads to greatly enhanced stability. Both of these
findings suggest that our method selects for peptides with
greater binding affinities, since many of the HLA-A2
specific peptides we identified have Leu in position 2, and
contain aromatic residues in position 3.

X-ray crystallographic analysis of five different viral
peptides bound to HLA-A2 has shown that the ends of
the peptide are fixed in the class I binding groove (Madden
et al., 1993). The HLA-A2 specific peptides selected by
our method conform to the requirements for P2 and P9
to fix the peptide ends in the binding groove. Additionally,
peptides selected by our method show less variability in
P1, P3 and P4. These residues may be preferentially
selected because they contribute to the binding energy.
However, Madden et al. (1993) have clearly shown that
the side-chain of a residue in the same position in two
different peptides may have completely different
orientations in the binding site, demonstrating that the
overall conformation of the peptide is a function of its
entire sequence.

Theoretical dissociation rates for HLA-A2 synthetic
peptides

Extensive analysis of HLA-A2 nonamer binding by
Parker et al. (1994), has resulted in a method to predict
the relative binding strength for HLA-A2 specific
peptides. In their study, the rate of $\beta_m$ dissociation
was measured for a set of 154 peptides, and analysis of the
binding data suggested that each side-chain of the
nonamer contributes to the stability of the HLA-A2
complex independent of the peptide sequence.

Quantification of this data resulted in the generation of
a table of 180 coefficients (20 amino acid residues in
positions P1–P9), representing the contribution of a
particular amino acid residue at a specific position in the
nonamer. A theoretical binding stability is calculated by
multiplying the corresponding coefficients, and when
compared to experimentally derived values, were within
a factor of 5. This method is useful for correlating
sequence information with binding affinity. We have
calculated theoretical (half-lives) dissociation constants
for each of our HLA-A2 bound peptides using these
binding coefficients, and our values are shown in Table 1.
The normalization coefficient required to employ their
constants was directly determined to be 6.6 ± 0.1 from
their published data. Eleven of the thirteen HLA-A2
bound synthetic peptides from our selection were
calculated to have a theoretical half-life of dissociation of
$\beta_m$ ranging from 132 to 4800 min.

Two of our peptides are predicted to have a theoretical
dissociation rate of only 2.5 and 3.0 min. These peptides
each have a negatively charged residue in position 9 (P9),
which is contrary to most of the known HLA-A2
restricted CTL epitopes and endogenous peptides upon
which these calculations have been based. However, the
HLA-A2 endogenous peptide HCMV gB 618–628 shown
in Table 2 also contains a negatively charged Glu residue
in P9 (Utz et al., 1992), demonstrating its occurrence in
nature. These nonamers may fall into the category defined
by Parker et al., (1994), as peptides that violate the
side-chain independence rule. Both these peptides, shown
in Table 1b, were selected from the library constructed
with the additional $\beta$-al linker forming a covalent ester
linkage to the carboxyl end of the peptide. Under these
conditions, the normally free carboxylic acid charge is not
present, possibly enabling the linker to interact with the
binding pocket or provide flexibility to permit alternative
orientations of the negative charge on the Asp and Glu
residues within the binding cleft.

Selection of HLA-B7 synthetic peptides

We also screened these same libraries for peptides that
can bind to HLA-B7. Selection based on the first library
yielded 11 peptides, and seven of these were sequenced as
shown in Table 3a. A screen of the second library
identified 15 more peptides, of which five were sequenced
individually as shown in Table 3b. Two pools of 4–5 peptide beads each were also sequenced simultaneously and a strong signal for Pro in position 2, Arg in position 3 and mostly Leu or Val in position 9 was identified (data not shown). It is clear from our data that Pro at position 2 and Arg at position 3 are very important for binding to this class I allele. There is also a preference for small hydrophobic residues at position 9, typically Leu, Val or Ile, similar to requirements in position 9 for HLA-A2 peptides, which is not unexpected since the F pocket of A2 and B7 are quite similar.

Comparison of HLA-B7 synthetic and endogenous peptides

During the preparation of this manuscript, a motif for HLA-B7 endogenous peptides was published by Huczko et al. (1993). Fifteen endogenous peptides eluted from HLA-B7 were sequenced, and all but three contained Ala or Arg at position 1, 12 contained Pro in position 2, 11 had a positive charge at position 3 that was usually Arg, and they all carried a small hydrophobic residue in position 9.

Synthetic peptides selected from a random nonamer library by binding to HLA-B7 have identified the anchor residues as Pro in P2, Arg in P3 and Leu, Ile or Val in P4. These results are in complete agreement with the data obtained by Engelhard using endogenous peptides. Nine of the 15 endogenous peptides contained Ala at P1, while the remainder contained Met, Arg, Leu and Ser. Only three of our 12 synthetic peptides had Ala in P1, although five others also had Met, Arg or Leu. This difference may arise as a consequence of in vivo processing or transport.

Peptide binding cleft of HLA-B27 compared to HLA-B7

Examination of the X-ray crystal structure for HLA-B27 identified several residues in the heavy chain which contribute to the nature of individual pockets in the peptide binding groove. An essential feature of the HLA-B27 binding peptides is the anchor residue Arg in position 2, which fits into the “B” pocket in the bound form (Garrett et al., 1989; Saper et al., 1991). The B pocket is lined with the side-chains of four polymorphic MHC residues (positions 9, 24, 45 and 67). In HLA-B27, this pocket is deep and mostly hydrophobic with His9, Thr24, Cys67 and ends near a negatively charged Glu45 which is complementary to the long positively charged Arg side-chain (Madden et al., 1991). HLA-B7 also contains Glu45 but has Tyr at positions 9 and 67. The presence of Tyr in position 67 has been predicted to fill the B pocket essentially blocking the binding of long side-chains such as Arg (Madden et al., 1992). This prediction is supported by our findings of only Pro in peptide position 2. We also found Arg predominantly in position 3, which fits into the “D” pocket. This pocket has at least five polymorphic MHC residues, and a comparison of HLA-B27 with HLA-B7 shows only two differences, His114 and Leu156 in B27 are changed to Asp114 and Arg156 in B7. It is possible that Asp114 in the D pocket of B7 can neutralize the charges on Arg156 and the Arg in position 3 of the nonamer, to form a stable interaction.

**DISCUSSION**

We present a novel synthetic system for determining the unique peptide motifs of the HLA-A2 and B7 class I molecules. This report demonstrates the validity of the Selectide process to rapidly identify allele-specific binding motifs for MHC class I molecules, complementing data from studies of sequenced endogenous peptides. The peptide library screening process takes only 2 days with microsequencing of individual beads as the rate-limiting step. However, this process can be greatly facilitated by sequencing pooled positive beads simultaneously, as discussed above.

Although it may not be experimentally feasible to screen an entire nonapeptide library, results based on screens using only a small fraction of the library are clearly sufficient to determine the important allele-specific anchor residues. Once the anchors have been identified, a new library can be constructed with these positions fixed and all other positions randomized, thereby significantly reducing the size of the library for additional screens. The number of beads to screen all possible nonapeptides is $19^9$, however, in the case of HLA-B7, one could assign Ala, Pro, Arg and Leu to positions 1, 2, 3 and 9, respectively, while randomizing the remaining residues, thus decreasing the size of the library to $19^4$. This effectively reduces the number of beads in the library from $3 \times 10^{11}$ to $2.5 \times 10^6$. The identification of anchor residues allows proteins to be screened for potential antigenic nonamer sequences, and we are studying the use of limited libraries with allele-specific anchor residues for screening immune recognition.

This method has the potential for determining residues of a binding peptide that could be altered to increase affinity for class I molecules. Madden et al. (1993) have shown that sequence changes alter peptide conformation through the entire length of the peptide. It would be very interesting to take a known restricted antigenic peptide, alter residues that would increase affinity for class I HLA-A2, and then use specific CTL to determine whether such changes enhance or antagonize recognition. The melanoma/melanocyte peptide recently isolated from HLA-A2 melanoma cells that stimulates CTL recognition as described by Cox et al. (1994) could be used for these experiments. The *in vitro* assays might reveal peptides with a potentially therapeutic benefit.

The synthetic approach enables us to test peptides incorporating rare amino acids or D-configuration residues. Since denatured class I molecules lacking endogenous peptides are used, cloned class I molecules expressed in *E. coli* are suitable for our library screen. Preliminary results using HLA-A2 heavy chains expressed in *E. coli* have demonstrated that this method is also suited to studies with recombinant class I proteins. This approach may therefore be particularly advantageous for determining allele-specific peptide motifs for class I proteins that are normally expressed at very low levels on the cell surface, such as the HLA-C molecules (Neefjes and Ploegh, 1988).
Acknowledgements—We are grateful to Dr Jack Strominger for providing cell lines, Drs Viktor Krchnak and Farid Abdul-Latif and Ms Shelly L. Wade for technical assistance. This work was supported in part by National Institutes of Health Grant CA49743, the Arizona Disease Control Research Commission, and the Selectide Corporation. K.S.L. is a recipient of the Scholar Award from the Leukemia Society of America.

REFERENCES


Tsmonides T. J., Walker B. D. and Eisen H. N. (1991) An optimal viral peptide recognized by CD8+ T cells binds very tightly to the restricting class I major histocompatibility complex

