Definition of Natural T Cell Antigens with Mimicry Epitopes Obtained from Dedicated Synthetic Peptide Libraries


*J Immunol* 1998; 161:4078-4082;
http://www.jimmunol.org/content/161/8/4078

References
This article cites 26 articles, 15 of which you can access for free at:
http://www.jimmunol.org/content/161/8/4078.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
Definition of Natural T Cell Antigens with Mimicry Epitopes Obtained from Dedicated Synthetic Peptide Libraries


Progress has recently been made in the use of synthetic peptide libraries for the identification of T cell-stimulating ligands. T cell epitopes identified from synthetic libraries are mimics of natural epitopes. Here we show how the mimicry epitopes obtained from synthetic peptide libraries enable unambiguous identification of natural T cell Ags. Synthetic peptide libraries were screened with Mycobacterium tuberculosis-reactive and -autoreactive T cell clones. In two cases, database homology searches with mimicry epitopes isolated from a dedicated synthetic peptide library allowed immediate identification of the natural antigenic protein. In two other cases, an amino acid pattern that reflected the epitope requirements of the T cell was determined by substitution and omission mixture analysis. Subsequently, the natural Ag was identified from databases using this refined pattern. This approach opens new perspectives for rapid and reliable Ag definition, representing a feasible alternative to the biochemical and genetic approaches described thus far. The Journal of Immunology, 1998, 161: 4078–4082.

A large number of autoimmunity and infection-related immunopathologies, as well as protective immune responses, are mediated via T cells. Therefore, knowledge about Ags that are recognized by disease or protection-related or tissue- or pathogen-specific T cells is essential for understanding the disease process. Three different methodologies for the identification of T cell-stimulating Ags have been reported. The first approach identifies T cell-stimulating peptide epitopes that are presented on APCs in the context of an HLA molecule by peptide elution (1, 2). This enables database searches to identify the Ag involved. The second procedure makes use of biochemical techniques to isolate and identify the Ag out of complex Ag mixtures that stimulate the T cell of interest (3, 4). A third possibility for Ag identification is based on screening expression libraries and subsequent database searches (5–7).

Alternatively, progress has been made recently in the use of synthetic peptide libraries to analyze the peptide specificity of T cell clones (8–11). The synthetic epitopes arising from synthetic peptide library screening are not natural epitopes, but mimics of natural epitopes. These mimicry epitopes may subsequently be used for the identification of natural Ags.

For Ag definition by database searching with mimicry epitope sequence information, a certain degree of sequence similarity between the mimicry epitope and the natural epitope is required. It has been reported that T cell clones can be activated upon stimulation by ligands that hardly share any sequence homology (12–14). This might impair database searches based on mimicry epitopes. Here we show that the natural Ag can be unambiguously identified using a synthetic library approach for three unrelated CD4+ T cell clones.

We describe a complete protocol for library screening, search pattern definition, and database searches that led to the identification of T cell Ags. DR3-restricted CD4+ T cell clones were used to screen two sublibraries with slightly different DR3-binding submotifs (15, 16). Two clones, MT1 (17) and MT2 (18), are Mycobacterium tuberculosis reactive. The third clone, HG (N. C. Schloot, O. M. C. Batstra, G. Duinkerken, R. R. de Vries, T. Dyberg, A. Chaudhuri, P. O. Behan, and B. O. Roep, unpublished observation), recognizes human 65-kDa glutamic acid decarboxylase (GAD65), a major autoantigen in insulin-dependent diabetes mellitus (19).

Materials and Methods

Peptide synthesis

Synthetic peptides were synthesized on an Abimed 422 multiple peptide synthesizer (Abimed Analyses-Technik, Langenfeld, Germany) using fluorenylmethoxycarbonyl (F-moc)-protected amino acids and Tentagel-SC resins (Rapp, Tübingen, Germany) as described (10). The purity of the peptides was determined by reversed-phase HPLC, and the integrity of the peptides was determined by matrix-assisted laser desorption ionization time of flight (MALDI-TOF) mass spectrometry on a Lasermat mass spectrometer (Finnigan-MAT; Hemel Hempstead, U.K.).

Synthetic peptide library design and synthesis

Hybrid TentaGel-AM resin (particle size, 90 μm; loading, 100 pmol/ bead, 16 pmol acid stable attached, 84 pmol acid labile attached) (Rapp) (20) was used to synthesize two random one-head/one-peptide 14-mer peptide libraries containing two different DR3-binding motifs (15, 16). The hybrid resin allows for a convergent library screening using the acid cleavable part of the peptide material attached to the resin (20), combined with

---

*Department of Immunohematology and Blood Bank, Leiden University Medical Center, Leiden, The Netherlands; †National Center for Computer-Aided Chemistry and Bioinformatics (CAOS/CAMM), Catholic University Nijmegen, Nijmegen, The Netherlands; and ‡Department of Medical Biochemistry, Sylvius Laboratory, Leiden University Medical Center, Leiden, The Netherlands.

Received for publication April 5, 1998. Accepted for publication June 11, 1998.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 This work was supported by the Leiden University Medical Center, the Dutch Leprony Relief Association, the Commission European Community, the Royal Netherlands Academy of Arts and Sciences, the Netherlands Organization of Scientific Research, the Macropa Foundation, and the Diabetes Fonds Nederland.

2 Address correspondence and reprint requests to Dr. Jan W. Drijfhout, Leiden University Medical Center, Department of Immunohematology and Blood Bank, Albinusdreef 2, P.O. Box 9600, 2300 RC Leiden, The Netherlands.

3 Abbreviation used in this paper: GAD65, 65-kDa glutamic acid decarboxylase.
highly efficient peptide identification by Edman sequencing, using the non-
cleavable part of the peptide material attached to the resin. The rationale of
the library design is the same as described before (10). The design of the two
libraries with two different DR3-binding motifs is summarized by the follow-
ning synthesis schemes: library 1, XXX(L,L,M,V,A,Y,F)X; XXXX
XXX-GABA; and library 2, XXX(L,L,M,V,A,Y,F)XXE/Q:S:T/X
(K,R,H,X)XXX-GABA, with X being one of 10 L-amino acids (all natural
amino acids except C, which was omitted for synthetic reasons), and GABA
being y-aminobutyric acid. Each library was synthesized using chemistry as
described above, following a mix and split protocol (21, 22), yielding two
one-head/one-peptide libraries, each with a complexity of $4 \times 10^6$.

**Synthetic library screening and bead sequencing**
Convergent peptide cleavage and screening was performed as described
(20). Shortly, libraries were divided into pools of 20,000 beads. Part of the
peptide material was released from the beads for testing in a T cell pro-
iferation assay (first round of screening). Beads of active pools were sub-
divided to pools of 70 beads. Again, peptides were released partially for the
second round of screening. For the third round of screening beads were
divided in limiting dilution. The remaining acid labile-attached peptide was
released and tested. Cleavage conditions for the hybrid resins (20) differ
from those described for Tentagel-AM (10). Peptide sequences were de-
termined by manual application of single beads (still containing sequence-
able amounts of peptide; 10–15 pmol) to a cartridge and subsequent se-
quencing using a Hewlett Packard (Palo Alto, CA) G1005A protein
sequencer.

**T cell clones**
MT1 (Rpl151-1) is a DR3-restricted CD4+ T cell clone that recognizes the
N-terminal part of a 65-kDa heat shock protein of mycobacteria (HSP65(2–15)) (17). MT2 (CAp151-3) (18) is a DR3-restricted CD4+
T cell clone that recognizes protein 85 (58b(55–68)) (23, 24) of M. terta-
buchlosis (28). HG (PM111) is a DR3-restricted CD4+ T cell clone that recognizes human 65-kDa glutamic acid decarboxylase (GAD65(339–352)) (N. C. Schloot et al., unpublished observations).

**Proliferation assay**
CD4+ T cell proliferation assays for testing library pools and synthetic
peptides were performed using $1 \times 10^5$ T cells and 5 × 105 irradiated
HLA-DR3-matched PBMCs per well in flat-bottom 96-well plates in com-
plete Iscove’s modified Dulbecco’s medium (150 μl) (Life Technologies,
Gaithersburg, MD) containing 10% pooled human serum. Phytomethemagglu-
tin (10 μg/ml), purified protein derivative of M. tuberculosis (10 μg/ml),
and IL-2 (T cell growth factor, 10% Lymphocult, Biotest Diagnostics, Dan-
ville, NJ) were used as positive controls for T cell proliferation. [3 H]Thy-
midine (0.3 μCi in 50 μl RPMI 1640) was added after 72 h, cells were har-
collected (Micro Cell Harvester, Skatron, Lier, Norway), and activity of
the T cell DNA was counted after another 18 h (Model 1205 Betaplate,
Liquid Scintillation Counter, LKB Instruments, Gaithersburg, MD).
Library pools were tested in a quantity of 7 μl per well, giving final test
concentrations of 5 nM for each individual peptide and 0.1% DMSO (v/v).

**Database searching with PeptideSearch**
The M. tuberculosis SHOTGUN database (TB_shotgun.dbs, December
1997) was retrieved from the Sanger Center (http://www.sanger.ac.uk/
Projects/M_Tuberculosis/) by ftp (ftp.sanger.ac.uk). The SHOTGUN da-
tabase, covering the complete genome of M. tuberculosis (strain H37Rv),
was converted by translating all possible open reading frames on both the
coding and the noncoding strand into hypothetical protein sequences. An
incompletely defined codon was translated to an X (±1% of the database).
Pattern searches were performed in the translated TB_shotgun.dbs using
PeptideSearch (25).

**Results**
**Mimicry epitopes from DR3 libraries for clones MT1, MT2, and HG**
Two distinct but related sets of peptides have been described to bind to
HLA-DR3 molecules (15, 16). The first set uses a hydrophobic amino acid at relative position 1 and A at relative position 4. The second set consists of peptides that need a third (basic) anchor at relative position 6 (K, R, or H), probably due to a weak anchor at position 4 (E, Q, N, S, T). Two 14-mer peptide libraries were syn-
thesized that reflect these two binding motifs. Both libraries had a
complexity of $4 \times 10^6$. The first library reflects $19^{12} \times 7 \times 1$ theo-
retically possible 14-mers. For the second library, this theoretical
complexity is $19^{11} \times 7 \times 5 \times 3$. This means that both libraries are
highly incomplete. Identification of mimicry epitopes from these li-
braries implies that T cells can recognize a large number of different
14-mer peptides. Because a mix and split protocol is used for syn-
thesis, random positions are not biased to certain synthetically pref-
erable combinations. Pools of each library, containing 20,000 indi-
vidual peptides, were screened with two DR3-restricted M. terta-
buchlosis-reactive T cell clones (MT1 and MT2) and a DR3-re-
stricted human GAD65-reactive T cell clone (HG) (Table I). Pools
were considered positive when counts exceeded five times the back-
ground counts. For clone MT1, activity was observed in one pool of
library 1. Convergent screening (10, 20) led to the identification of stimulatory peptide MT1-P1 (NSTVAYDEAMIFAQ) (Table II). For
cloned MT2, two active pools of library 1 were obtained, resulting in the
identification of MT2P1 (NSAIGIDPVPRRD) and MT2P2 (SHFVGXDIPVSLK) (Table II). For clone HG, activity was observed
in one pool of library 1 as well as in one pool of library 2, demonstrating that this clone is able to recognize peptides bearing two
distinct binding motifs. The stimulatory peptides were identified as
SIAMAFQIPMAA (HG-P1) and TDSLAFEPKVPRRQ (HG-P2)
(Table II). The stimulatory capacities of the identified peptides were
compared with their natural counterparts in a T cell proliferation assay
(Fig. 1). Retrospectively, it was shown that suboptimal concentrations of
5 nM (which equals the individual peptide concentrations during
library screening) are indeed stimulatory for all mimicry epitopes.

For mimicry epitopes MT1-P1 and HG-P1, the corresponding
natural epitopes could be identified using a homology search.
Searching with MT1-P1 (relative positions 1–7) in the M. terta-
buchlosis database yielded HSP65(2–15) as hit number 1, demonstrat-
ing that this clone is able to recognize peptides bearing two
distinct binding motifs. The stimulatory peptides were identified as
SIAMAFQIPMAA (HG-P1) and TDSLAFEPKVPRRQ (HG-P2)
(Table II). The stimulatory capacities of the identified peptides were
compared with their natural counterparts in a T cell proliferation assay
(Fig. 1). Retrospectively, it was shown that suboptimal concentrations of
5 nM (which equals the individual peptide concentrations during
library screening) are indeed stimulatory for all mimicry epitopes.

**Table I. Proliferative responses of T cell clones to active library fractions in the various screenings**

<table>
<thead>
<tr>
<th>T Cell Clone</th>
<th>Screen 1 (cpm)</th>
<th>Screen 2 (cpm)</th>
<th>Screen 3 (cpm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT1</td>
<td>785 (109)</td>
<td>1,502 (76)</td>
<td>5,392 (148)</td>
</tr>
<tr>
<td>MT2</td>
<td>6,444 (153)</td>
<td>23,196 (533)</td>
<td>25,738 (159)</td>
</tr>
<tr>
<td>HG</td>
<td>5,632 (153)</td>
<td>17,611 (533)</td>
<td>25,060 (159)</td>
</tr>
</tbody>
</table>

a Shown are counts of single library screening tests. In all library screenings,
individual peptide concentrations are approximately 5 nM. Background counts
are indicated between parentheses (duplicate tests for MT1 and MT2, triplicate tests
for HG).

**Table II. Alignment of natural epitopes with synthetic library mimicry epitopes**

<table>
<thead>
<tr>
<th>Epitope/Mimicry</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. tuberculosis</td>
<td></td>
</tr>
<tr>
<td>HSP65(2–15)</td>
<td></td>
</tr>
<tr>
<td>MT1-P1</td>
<td>AKTIAYDEARRGL</td>
</tr>
<tr>
<td>MT1-P2</td>
<td>TDSLAFEPKVPRRQ</td>
</tr>
<tr>
<td>MT2-P1</td>
<td>SIAIMFQIPMAA</td>
</tr>
<tr>
<td>MT2-P2</td>
<td>TDSLAFEPKVPRRQ</td>
</tr>
<tr>
<td>Human GAD65</td>
<td></td>
</tr>
<tr>
<td>GAD65(339–352)</td>
<td></td>
</tr>
<tr>
<td>HG-P1</td>
<td>TVYGGAFDPLLAVAD</td>
</tr>
<tr>
<td>HG-P2</td>
<td>TDSLAFEPKVPRRQ</td>
</tr>
</tbody>
</table>

a Epitopes and mimics or clones MT1, MT2, and HG. Shown are 14-mer se-
quences with N termini starting three amino acids before the first anchor
position (relative position 1). Due to problems with Edman degradation,
relative position three of MT2-P2 remained unclear (X in sequence of MT2-P2).
maximize the chance of only including positions specifically involved either in DR3 binding or TCR interaction (16).

Standard homology database searches with either the complete sequences of MT2-P1 and HG-P2 or part of the sequences of these mimicry epitopes did not lead to identification of the natural epitopes for clone MT2 and HG within the 100 best-matching hits involved either in DR3 binding or TCR interaction (16).

Amino acids L, I, M, V, A, G, W, Y, and F have been described as anchors for relative position 1 for HLA-DR3 (15, 16). To confirm the expectation that in MT2-P1 the I at relative position 1 can only be substituted by this set of amino acids without disturbing stimulatory capacity, an omission mixture was synthesized: NSAX1 GIDIPVARRD (X1 = all natural L-amino acids except C and the expected anchor residues). This mixture of 10 peptides was not able to induce a significant proliferation of clone MT2.

Although identification of the natural epitopes for clones MT2 and HG using mimicry epitopes MT2-P1 and HG-2 was performed similarly, only the procedure for clone MT2 is described in detail.

Promoter position 1, a mixture of 18 peptides was synthesized and tested for T cell stimulation (consensus sequence NSAO,GIDIPVARRD, where O represents all 20 natural L-amino acids except I, which was omitted because it is present at relative position 1 in MT2-P1, and C, which was omitted for synthetic reasons). The omission mixtures for relative position 2 and 5 induced no proliferative response (Table III). Therefore, it could be concluded that, in the context of the MT2-P1 sequence, a G at relative position 2 and an I at relative position 5 are essential for proliferation of clone MT2.

**Table III. Omission mixture analysis for positions 1–7 of MT2-P1**

<table>
<thead>
<tr>
<th>Omission Mixture&lt;sup&gt;a&lt;/sup&gt;</th>
<th>100 nM (cpm/1000)</th>
<th>10 μM (cpm/1000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSAIX1GIDIPVARRD</td>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>NSAI1XIDIPVARRD</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NSAIIGI1XIPVARRD</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>NSAI1IGIX2PVRD</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>NSAI1GID1X3PVRD</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>NSAI1GIDIX4VARRD</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>NSAI1GIDIX5PVRD</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

<sup>a</sup> X (identified in bold) means that the mixture contains all natural L-amino acids except C and is the amino acid present at the position of interest in MT2-P1. The peptide concentrations that are tested for proliferation induction, 100 nM and 10 μM, are total concentrations of all peptides in a mixture. Responses are the means of duplicate tests.

**Table IV. Omission mixture anchor analysis for positions 1, 4, and 6 of MT2-P1**

<table>
<thead>
<tr>
<th>Omission Mixture&lt;sup&gt;a&lt;/sup&gt;</th>
<th>100 nM (cpm/1000)</th>
<th>10 μM (cpm/1000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSAI1XGIDIPVARRD</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>NSAI1GIX1X2VARRD</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NSAI1GIX1X2VARRD</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<sup>a</sup> X1–X5 (identified in bold) mean that the position includes all natural L-amino acids except C and the amino acid present at the position of interest in MT2-P1. The peptide concentrations that are tested for proliferation induction, 100 nM and 10 μM, are total concentrations of all peptides in a mixture. Responses are the means of duplicate tests.
relative position 3 and 7. For these positions, all substitution peptides (except C substitutions for synthetic reasons) were synthesized and tested for T cell stimulation (Fig. 2). For relative position 3, L, R, S, T, N, D, W, and Y are allowed amino acids. For relative position 7, I, T, and K were shown to be allowed.

**Database searches and Ag definition (MT2)**

Based on the omission and substitution studies as described (Tables III and IV and Fig. 2), a search pattern was constructed for PeptideSearch. This resulted in a pattern at (relative) position 1 = (L,I,M,V,A,G,W,Y,F), position 2 = G, position 3 = (L,I,R,S,T,N,D,W,Y), position 4 = (D,N,S,T), position 5 = I, position 6 = (P,K,R,H), and position 7 = (V,T,I,K). The translated *M. tuberculosis* SHOTGUN database was searched with this pattern, yielding 19 hits (Table V). The natural epitope of clone MT2 (BCG85B(55–68)) was unambiguously identified. A comparable analysis and subsequently a pattern for searching the nonredundant database was performed for clone HG (Table II). This implies that all three clones have a strict recognition pattern. These data further suggest that recognition of ligands without any sequence similarity by one T cell clone may not be a very common phenomenon for epitopes that are recognized in low nanomolar concentration with a regular binding motif (14–16). This implies a probability for Ag definition by database searching using sequence information from synthetic mimicry epitopes, provided that the Ag is contained in the database.

**Discussion**

Molecular mimicry enables epitope identification from incomplete synthetic peptide libraries (10, 11). Mimicry epitopes for three unrelated HLA-DR3-restricted CD4+ T cell clones were identified from synthetic peptide libraries. For all three T cell clones studied here, the similarity (conservative substitutions or even identical amino acids) of synthetic mimicry epitopes and natural epitopes was high (Table II). This implies that all three clones have a strict recognition pattern. These data further suggest that recognition of ligands without any sequence similarity by one T cell clone may not be a very common phenomenon for epitopes that are recognized in low nanomolar concentration with a regular binding motif (14–16). This implies a probability for Ag definition by database searching using sequence information from synthetic mimicry epitopes, provided that the Ag is contained in the database.

In two cases, a simple homology search with the core of the sequence of a synthetic mimicry epitope obtained from the library screening led to immediate identification of the natural epitope as part of a large set of heterogeneous database hits. A homology search with the core sequences (relative positions 1–7) of mimicry epitopes (MT1-P1 and HG-P1) identified the corresponding natural Ags within the 100 best-matching database hits (*M. tuberculosis* database for clone MT1 and a human abstract of the Swiss Prot database for clone HG).

In two other cases (mimicry epitopes MT2-P1 and HG-P2), the size of the set of hits containing the natural epitope appeared to be too large to be investigated. Therefore we developed a generally applicable approach that increases the probability of successful identification of natural epitopes by pattern searching with a precisely defined pattern (Fig. 3). This approach is based on the observation that stimulating ligands can be predicted by studying individual amino acid positions (14). Using a precisely defined pattern for searching the *M. tuberculosis* SHOTGUN database, the natural epitope of clone MT2 (BCG85B(55–68)) was unambiguously identified. A comparable analysis and subsequently a pattern search in the nonredundant database was performed for clone HG using mimicry epitope HG-P2. GAD65 was indeed part of the human fraction of all pattern hits (data not shown). This indicates

**Table V. Hits from SHOTGUN database**

<table>
<thead>
<tr>
<th>Peptide</th>
<th><em>M. tuberculosis</em> SHOTGUN Database Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>LFLVGLSIPTGGPL</td>
</tr>
<tr>
<td>2.</td>
<td>SASMGDRKIVQVFQS</td>
</tr>
<tr>
<td>3.</td>
<td>DEIQLNIPFGPL</td>
</tr>
<tr>
<td>4.</td>
<td>SQPQIGNIKIIIAF</td>
</tr>
<tr>
<td>5.</td>
<td>LRFGGSTIPKIPFI</td>
</tr>
<tr>
<td>6.</td>
<td>MASIGLTIPTIALA</td>
</tr>
<tr>
<td>7.</td>
<td>LRFAGTIRILFAF</td>
</tr>
<tr>
<td>8.</td>
<td>ASI1VGRTPVAGAV</td>
</tr>
<tr>
<td>9.</td>
<td>SPSMGDRKIVQVFQS</td>
</tr>
<tr>
<td>10.</td>
<td>MWAAGWIPFSSS</td>
</tr>
<tr>
<td>11.</td>
<td>MWAAGWITIPGTAC</td>
</tr>
<tr>
<td>12.</td>
<td>SGRIGDITIRVLSA</td>
</tr>
<tr>
<td>13.</td>
<td>YPLVGTTIPIAGYT</td>
</tr>
<tr>
<td>14.</td>
<td>HPLFGSIIPTIPSF</td>
</tr>
<tr>
<td>15.</td>
<td>SXSMGDRKIVQVFQS</td>
</tr>
<tr>
<td>16.</td>
<td>SPSMGDRIPVFAXX</td>
</tr>
<tr>
<td>17.</td>
<td>HCRGYSIKTIPQQG</td>
</tr>
<tr>
<td>18.</td>
<td>DALEVGRDIKKLPTL</td>
</tr>
<tr>
<td>19.</td>
<td>RAAAGLTPVLLIR</td>
</tr>
</tbody>
</table>

*PeptideSearch hits for clone MT2. Peptides 2, 9, 15, and 16 induce proliferation of clone MT2. X means one of all natural L-amino acids. A mixture of all natural L-amino acids except C is synthesized at each X position. All hits completely match datasets III and IV and Fig. 2, a search pattern was constructed for PeptideSearch. This resulted in a pattern at (relative) position 1 = (L,I,M,V,A,G,W,Y,F), position 2 = G, position 3 = (L,I,R,S,T,N,D,W,Y), position 4 = (D,N,S,T), position 5 = I, position 6 = (P,K,R,H), and position 7 = (V,T,I,K). The translated *M. tuberculosis* SHOTGUN database was searched with this pattern, yielding 19 hits (Table V). The natural epitope of clone MT2 (BCG85B(55–68)) was unambiguously identified. A comparable analysis and subsequently a pattern for searching the nonredundant database was performed for clone HG (Table II). This implies that all three clones have a strict recognition pattern. These data further suggest that recognition of ligands without any sequence similarity by one T cell clone may not be a very common phenomenon for epitopes that are recognized in low nanomolar concentration with a regular binding motif (14–16). This implies a probability for Ag definition by database searching using sequence information from synthetic mimicry epitopes, provided that the Ag is contained in the database.

In two cases, a simple homology search with the core of the sequence of a synthetic mimicry epitope obtained from the library screening led to immediate identification of the natural epitope as part of a large set of heterogeneous database hits. A homology search with the core sequences (relative positions 1–7) of mimicry epitopes (MT1-P1 and HG-P1) identified the corresponding natural Ags within the 100 best-matching database hits (*M. tuberculosis* database for clone MT1 and a human abstract of the Swiss Prot database for clone HG).

In two other cases (mimicry epitopes MT2-P1 and HG-P2), the size of the set of hits containing the natural epitope appeared to be too large to be investigated. Therefore we developed a generally applicable approach that increases the probability of successful identification of natural epitopes by pattern searching with a precisely defined pattern (Fig. 3). This approach is based on the observation that stimulating ligands can be predicted by studying individual amino acid positions (14). Using a precisely defined pattern for searching the *M. tuberculosis* SHOTGUN database, the natural epitope of clone MT2 (BCG85B(55–68)) was unambiguously identified. A comparable analysis and subsequently a pattern search in the nonredundant database was performed for clone HG using mimicry epitope HG-P2. GAD65 was indeed part of the human fraction of all pattern hits (data not shown). This indicates

**Table V. Hits from SHOTGUN database**

<table>
<thead>
<tr>
<th>Peptide</th>
<th><em>M. tuberculosis</em> SHOTGUN Database Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>LFLVGLSIPTGGPL</td>
</tr>
<tr>
<td>2.</td>
<td>SASMGDRKIVQVFQS</td>
</tr>
<tr>
<td>3.</td>
<td>DEIQLNIPFGPL</td>
</tr>
<tr>
<td>4.</td>
<td>SQPQIGNIKIIIAF</td>
</tr>
<tr>
<td>5.</td>
<td>LRFGGSTIPKIPFI</td>
</tr>
<tr>
<td>6.</td>
<td>MASIGLTIPTIALA</td>
</tr>
<tr>
<td>7.</td>
<td>LRFAGTIRILFAF</td>
</tr>
<tr>
<td>8.</td>
<td>ASI1VGRTPVAGAV</td>
</tr>
<tr>
<td>9.</td>
<td>SPSMGDRKIVQVFQS</td>
</tr>
<tr>
<td>10.</td>
<td>MWAAGWIPFSSS</td>
</tr>
<tr>
<td>11.</td>
<td>MWAAGWITIPGTAC</td>
</tr>
<tr>
<td>12.</td>
<td>SGRIGDITIRVLSA</td>
</tr>
<tr>
<td>13.</td>
<td>YPLVGTTIPIAGYT</td>
</tr>
<tr>
<td>14.</td>
<td>HPLFGSIIPTIPSF</td>
</tr>
<tr>
<td>15.</td>
<td>SXSMGDRKIVQVFQS</td>
</tr>
<tr>
<td>16.</td>
<td>SPSMGDRIPVFAXX</td>
</tr>
<tr>
<td>17.</td>
<td>HCRGYSIKTIPQQG</td>
</tr>
<tr>
<td>18.</td>
<td>DALEVGRDIKKLPTL</td>
</tr>
<tr>
<td>19.</td>
<td>RAAAGLTPVLLIR</td>
</tr>
</tbody>
</table>
Generally applicable strategy for identification of T cell antigens

1. Identification of mimicry epitope from synthetic peptide library
2. Omission mixture analysis of anchor and non-anchor amino acids
3. Substitution analysis of non-anchor amino acids
4. Omission mixture analysis of anchor amino acids
5. Construction of search pattern
6. Database searching with pattern
7. Testing of database hits

FIGURE 3. Schematic representation of the procedure to define T cell Ags with synthetic peptide library mimicry epitopes.

that similarity rather than homology between mimicry epitopes and natural epitopes is required for natural Ag identification.

This pattern search approach is generally applicable for T cell clones with unknown Ag specificity. The interpretation of the output of a pattern search, compared with a homology search, is easier, because the match score is equal (100%) for all hits without any assumptions and statistical calculations involved. In addition, the use of pattern searches allows for approaches that are independent of the number of amino acid positions analyzed. The number of database hits using a pattern search depends on the size of the database, the number of positions that are defined in the pattern, the number of permitted amino acids for the defined position patterns, the nature of amino acids in the pattern, and the relative occurrence of each amino acid in databases. From a practical point of view, a balance must be found between the experimental effort needed to determine a very precise pattern, yielding only few database hits, and the effort that is needed to synthesize and test a large list of hits resulting from a less precise search pattern. In this study, the definition of the search pattern resulted in 19 hits, which is sufficiently limited for synthesis and testing.

One clone (HG) was able to recognize two similar mimicry epitopes containing two slightly different HLA-DR3-binding motifs (library 1 and library 2 motifs). In this example, our approach succeeded in unambiguous identification of human GAD65 using the sequence information of HG-P2, whereas a homology search with this peptide did not identify the natural Ag. Our approach might further contribute to the search for functional T cell-epitope mimicry in autoimmune disease that is not based on simple-sequence homology.

We conclude that mimicry epitopes identified from synthetic peptide libraries are highly similar to their natural counterparts. Therefore, in some cases the natural epitope can be defined by a simple homology database search, depending on the degree of homology and the size of the database of interest. Because the success probability of homology searching is limited, we present a generally applicable protocol based on similarity, which represents a convenient approach for Ag definition using sequence information obtained from synthetic mimicry epitopes (Fig. 3). This approach opens new perspectives for rapid and reliable Ag definition and represents a feasible alternative for the biochemical and molecular biology approaches described thus far. Defined Ags can be used for vaccine design in infectious disease and for the development of immune intervention strategies in autoimmune disease.

Acknowledgments

We thank Dr. F. Koning for critically reading the manuscript and Kees L. M. C. Franken for cloning and purification of M. tuberculosis 85B.

References


