Combinatorial Peptide Libraries and Biometric Score Matrices Permit the Quantitative Analysis of Specific and Degenerate Interactions Between Clonotypic TCR and MHC Peptide Ligands[1]

Yingdong Zhao,[2]* Bruno Gran,[2,3]† Clemencia Pinilla,‡ Silva Markovic-Plese,†
Bernhard Hemmer,†§ Abraham Tzou,† Laurie Ward Whitney,† William E. Biddison,†
Roland Martin,† and Richard Simon[4]*

The interaction of TCRs with MHC peptide ligands can be highly flexible, so that many different peptides are recognized by the same TCR in the context of a single restriction element. We provide a quantitative description of such interactions, which allows the identification of T cell epitopes and molecular mimics. The response of T cell clones to positional scanning synthetic combinatorial libraries is analyzed with a mathematical approach that is based on a model of independent contribution of individual amino acids to peptide Ag recognition. This biometric analysis compares the information derived from these libraries composed of trillions of decapeptides with all the millions of decapeptides contained in a protein database to rank and predict the most stimulatory peptides for a given T cell clone. We demonstrate the predictive power of the novel strategy and show that, together with gene expression profiling by cDNA microarrays, it leads to the identification of novel candidate autoantigens in the inflammatory autoimmune disease, multiple sclerosis. *The Journal of Immunology,* 2001, 167: 2130–2141.

The CD8[+] and CD4[+] T lymphocytes recognize short peptides of 8–10 and 12–16 aa in the context of self MHC class I and class II molecules, respectively (1, 2). During the last 15 years, this central process of cellular immune responses has received enormous attention and has been dissected using a vast array of different immunological and biochemical techniques. A quantitative analysis of the interaction between TCR and their MHC peptide ligands would be an important basis for the design of vaccines and therapeutic approaches to immune-mediated, infectious, and neoplastic diseases.

Because it has been difficult to describe the trimolecular complex in its entirety, experiments initially focused on the interaction between peptide and MHC molecules. Structural studies of MHC class I and class II molecules complexed with antigenic peptides disclosed that the latter bind in a linear fashion (3). Sequencing of peptide pools and of individual self peptides eluted from MHC molecules (4, 5) together with systematic binding analyses (6, 7) have provided experimental data for the definition of MHC-binding motifs (8–12) and the development of MHC peptide-binding models. A combination of positive and negative influences from amino acid side chains in the antigenic peptide has been shown to determine the interaction between peptide and MHC molecules (13). Indeed, the assumption of independent contribution of each amino acid side chain in the peptide sequence to MHC binding has been used to develop quantitative methods that predict peptide binding to MHC alleles (8, 14–16). More recently, elegant neural network approaches have been used to further refine the prediction of peptide binding to MHC (17–20). Based on the fact that a subset of MHC-binding peptides are also T cell epitopes (21, 22), MHC binding has been used to predict candidate T cell epitopes in bulk T cell populations, such as those contained in the peripheral blood (12, 19). However, to dissect and predict precisely the interaction of all three components of the trimolecular complex has until now been a difficult undertaking. Therefore, the quantitative study of MHC peptide recognition by single TCR has remained a largely unsettled issue.

The specificity of the trimolecular complex interaction has been studied using individual substitution analogues. Although initial studies showed that some amino acids in the antigenic peptide sequence are necessary for recognition by the TCR (primary TCR contacts) and others can tolerate conservative substitutions (secondary contacts) (23, 24), the systematic use of single and multiple amino acid-substituted peptides has shown that all amino acid side chains can contribute to peptide recognition in a largely independent manner (25). In extreme cases, this can lead to recognition of peptides with entirely different amino acid sequences by the same TCR (25).

The development of soluble- and bead-bound combinatorial peptide libraries in various formats representing millions to trillions of peptides has emerged as a powerful approach to both T

cell epitope determination and the analysis of TCR specificity and flexibility, as recently reviewed (26, 27). Recent studies (28–32) of T cell clones (TCC)[5] demonstrated the efficacy of using positional scanning synthetic combinatorial libraries (PS-SCL) for identifying target Ags and highly active peptide mimics. However, it was technically impossible to fully use this technology without the development of quantitative methods for predicting the stimulatory potential of peptides based on data from these complex libraries.

We report in this study a new strategy that combines data acquisition with PS-SCL and analysis with a quantitative scoring matrix to identify agonist peptides for clonotypic TCR of known and unknown specificity. Peptides can be identified from database searches with unprecedented efficiency and ranked according to a score that is predictive of their stimulatory potency. To our knowledge, this is by far the most efficient available approach to identify stimulatory peptides for individual TCR and predict their actual stimulatory potency with relatively high accuracy. While further improvements of this strategy will be pursued, we have developed a tool for the identification of potential T cell epitopes, the design of vaccines, and the quantitative analysis of TCR degeneracy. Finally, we demonstrate how the search results from the above prediction strategy can be related to tissue-specific expression profiles determined by cDNA microarray assays to identify candidate peptides that are derived from proteins that are overexpressed in a diseased tissue, i.e., the brain in multiple sclerosis (MS), and are thus available for the expansion of autoreactive T cells.

Materials and Methods

T cell clones

TCC were established from peripheral blood or cerebrospinal fluid (CSF) lymphomononuclear cells by a split-well technique, as previously described (33, 34). TCC GP5F11 was established from PBMC of a patient with MS using influenza virus hemagglutinin (HA) peptide (306–318) (PKYVKQNTLKLAT, single letter amino acid code) as an Ag. The TCC is restricted by DRB1*0404. TCC TL3A6 was established with myelin basic protein (MBP) from PBMC of a patient with MS and recognizes the immunodominant epitope $MBP_{87–99}$ (VHFFKNIVTPRTP) in the context of DR2a (DRα + DRB5*0101). The TCC has been extensively characterized for recognition of numerous altered peptides derived from $MBP_{87–99}$ as well as other molecular mimics (25, 31, 32, 35, 36). The TCR usage is TCRAV18 and TCRBV5S1. TCC CSF-3 was established with a lysate of *Borrelia burgdorferi* from the CSF of a patient with chronic Lyme disease, as described (34). The TCC recognizes several *B. burgdorferi*-derived as well as human peptides in the context of DR2b (DRα + DRB1*1501). The TCR usage is TCRAV13S2 and TCRBV14S1.

Peptides and peptide combinatorial libraries

Peptides were synthesized by the simultaneous multiple peptide synthesis method (37) and characterized using HPLC and mass spectrometry. A synthetic *N*-acetylated, C-amide L-amino acid combinatorial peptide library in a positional scanning format (PS-SCL; 200 mixtures in the OX_9 format, in which O represents one of the 20 L-amino acids, and X represents all of the natural L-amino acids, except cysteine) was prepared as described (38).

Proliferative assays

The proliferation of TCC in response to PS-SCL or individual peptides was tested by seeding in duplicate 2×10^4 T cells, 5×10^4 irradiated PBMC with or without mixtures from PS-SCL or peptide. Proliferation was measured by [³H]thymidine (Amersham, Arlington Heights, IL) incorporation (32).

Statistical analysis and model building

A positional scoring matrix was generated by assigning a value of the stimulatory potential to each of the 20 defined amino acids in each position. The score S_{ij} for each amino acid i at each position j was calculated as follows:

$$S_{ij} = \frac{L_{ij} - B}{\sqrt{(std(L_{ij}))^2 + (std(B))^2}}$$

where L equals the mean of replicate experimental measurements (cpm), B stands for background noise, $std(L_{ij})$ denotes the smoothed estimate of the SD for each measurement using a locally weighted regression smoothing technique (*S-plus* package) based on the assumption that the SD is dependent on level of response. We call this the Z-index score due to its similarity to statistical Z ratios of means divided by their SE values.

In an alternative score called stimulation index (S-index), we generated the score in each position by using the mean of duplicate cpm values in the presence of mixtures from the PS-SCL fractions divided by the mean of duplicate values in the absence of mixtures from the PS-SCL. The S-index score appeared preferable when the PS-SCL spectrum of the cpm value was more clearly defined.

Under the assumption of independent contribution to stimulation, the predicted stimulatory potential of given peptide is the sum of the scores in each position. A 10-mer peptide sequence can be represented by a 20×10 matrix of 0s and 1s (p_{ij}), where $p_{ij} = 1$ if the ith amino acid (using the same order as for the rows of the scoring matrix) is in position j. Let S_{ij} denote the components of the positional scoring matrix. Then the score for the peptide is:

$$S = \sum_{i=1}^{20} \sum_{j=1}^{10} p_{ij}S_{ij}$$

Database search

We wrote a Perl script to systematically search the GenPept database. A window with the same length of peptide as used in the PS-SCL was applied to slide over the available translated protein-coding sequences. The sum of the scores within the window was used as a ranking criterion. All peptides with scores higher than a threshold were output into a file. The threshold was chosen based on the statistical significance of the peptide score, compared with that for a random peptide. Those peptides were then sorted. Redundant peptides were removed. The database search can also be restricted to specific organisms (e.g., *Homo sapiens* or *Influenza* virus).

Statistical significance

We developed a statistical significance test of the hypothesis that the score for a peptide is no greater than would be expected if the peptide were obtained from 10 random draws of amino acids. Under the null hypothesis, it is not assumed that all amino acids are equally likely, but rather the relative frequencies $f_1, f_2, \ldots f_{20}$ are derived from the database being searched. Under the null hypothesis, the distribution of S will be approximately normally distributed. The mean and the variance of this null distribution can be expressed as

$$m = \sum_{i=1}^{20} f_i \sum_{j=1}^{10} S_{ij}$$

$$var = E[S^2] - m^2$$

The variance can be shown to equal:

$$var = \sum_{i=1}^{20} f_i \sum_{j=1}^{10} S_{ij}^2 + 2 \sum_{j=1}^{9} \sum_{j'=j+1}^{10} m_j m_{j'} - m^2$$
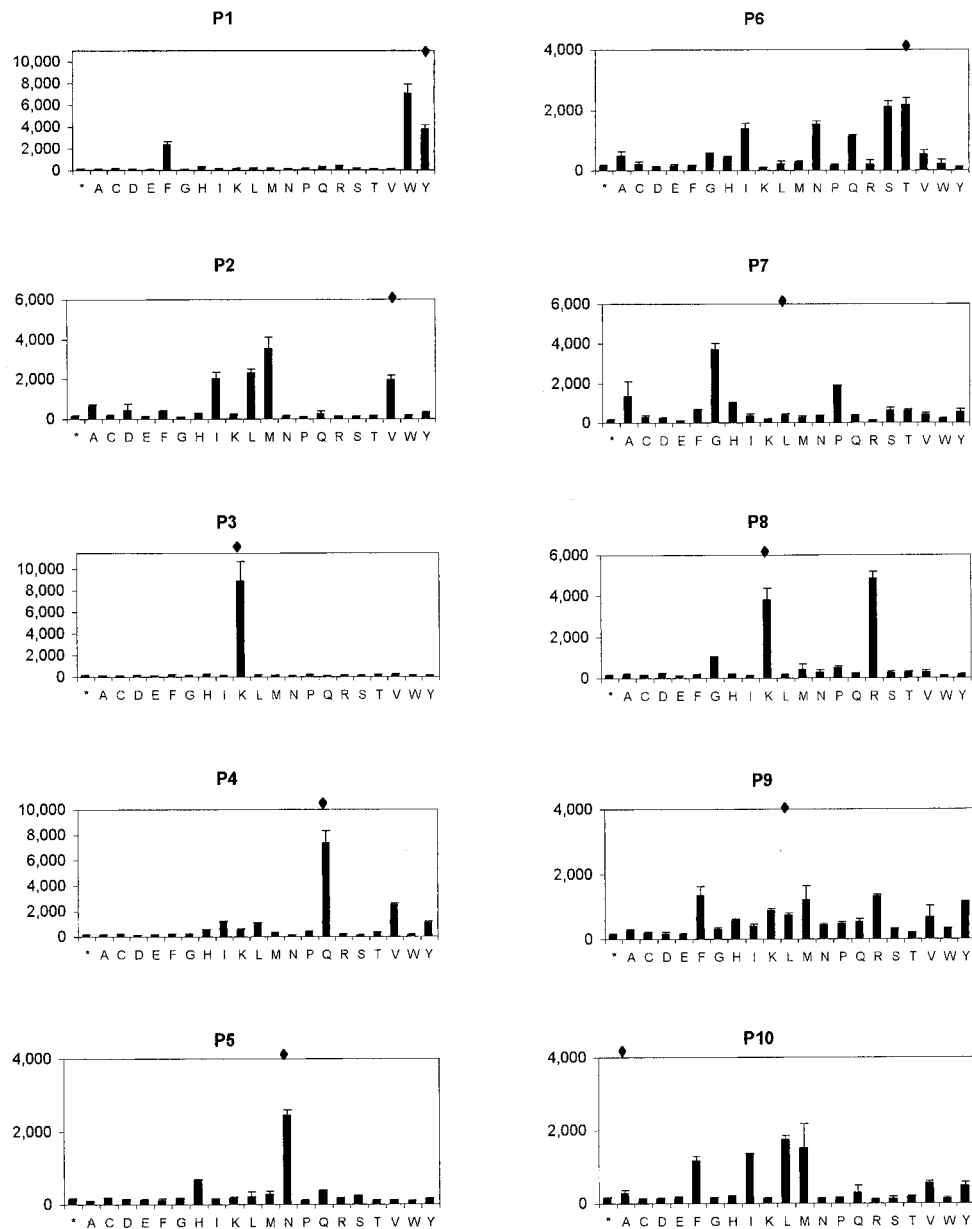
$$where \ m_j = \sum_{j=1}^{20} f_i S_{ij}.$$

The statistical significance of any score S can be approximated as

$$p = \phi\left(\frac{m - S}{\sqrt{var}}\right),$$

in which ϕ denotes the standard normal distribution function. However, this significance level does not account for the number of 10-mer sequences contained in the database.

FIGURE 1. Proliferative response of TCC GP5F11 (*A*) and TL3A6 (*B*) to the 200 mixtures of a decapeptide PS-SCL in which each position has one defined amino acid (20 for each of the 10 positions; the single letter amino acid code is used). Proliferation is shown as cpm induced by each mixture of the PS-SCL (mean and SD values of duplicate wells). *, Proliferation in the absence of peptide mixtures. TCC GP5F11 is specific for an influenza virus HA-derived peptide, Flu-HA$_{308-317}$; TCC TL3A6 is specific for a MBP-derived peptide. The corresponding sequences of HA$_{308-317}$, YVKQNTLKLA, and MBP$_{89-98}$, FFKNIVTPRT, are indicated by diamonds at the *top* of each panel. Proliferation in the absence of Ag was 124 + 42 cpm (*A*) and 1453 + 493 cpm (*B*).

Analysis of gene expression using cDNA microarrays

Brain tissue was obtained at autopsy from two MS patients. Patient W was a 46-year-old male with primary progressive MS (39); patient R was a 46-year-old female with relapsing-remitting MS. Normal white matter was dissected, postmortem, from three nondiseased brains. RNA extracted from these three normal white matter samples was pooled, in equal amounts, for use in hybridization experiments. Lesions were identified by H&E and Luxol fast blue-periodic acid Schiff staining of paraffin-embedded sections. Further characterization of lesions was performed using immunohisto-chemistry for cell-specific Ags. All staging of lesions was performed as previously described (40). From the first patient, patient W, one acute (W1) and one chronically active lesion (W2) were studied. From the second patient, R, 16 chronic lesions were studied. These lesions had inflammatory cells present, but the inflammatory cells were not participating in any form of ongoing demyelination.

The detailed methodology of cDNA microarray analysis has been described in detail elsewhere (41) Arrays for this study contained 2889 human cDNAs that were primarily derived from I.M.A.G.E. consortium cDNA libraries (42). A list of genes present on the arrays can be found at http://intra.ninds.nih.gov/Biddison/cDNA_microarray.asp. [^{33}P]dCTP-labeled cDNAs were produced by reverse transcriptase from RNAs obtained from individual MS lesions, pooled normal white matter, exper-imental allergic encephalomyelitis, and normal mouse brains, and hybrid-ized to the cDNA microarrays. Hybridizations of RNA obtained from MS lesions and experimental allergic encephalomyelitis brains were performed in two independent experiments, except for lesions R10, R11, and R16, in which enough RNA was obtained for only one hybridization. Quantitation of radioactivity bound to the arrays was performed on a Molecular Dynamics STORM PhosphorImager (Molecular Dynamics, Sunnyvale, CA) at 50 μm resolution. All data were analyzed from the PhosphorImager images using Pscan (Ref. 43, see also http://abs.cit.nih.gov/pscan). Pscan calculates spot intensities and compares spot intensities between samples, giving a ratio of gene expression between comparative samples. Using Pscan, spot intensities between arrays were automatically normalized to the
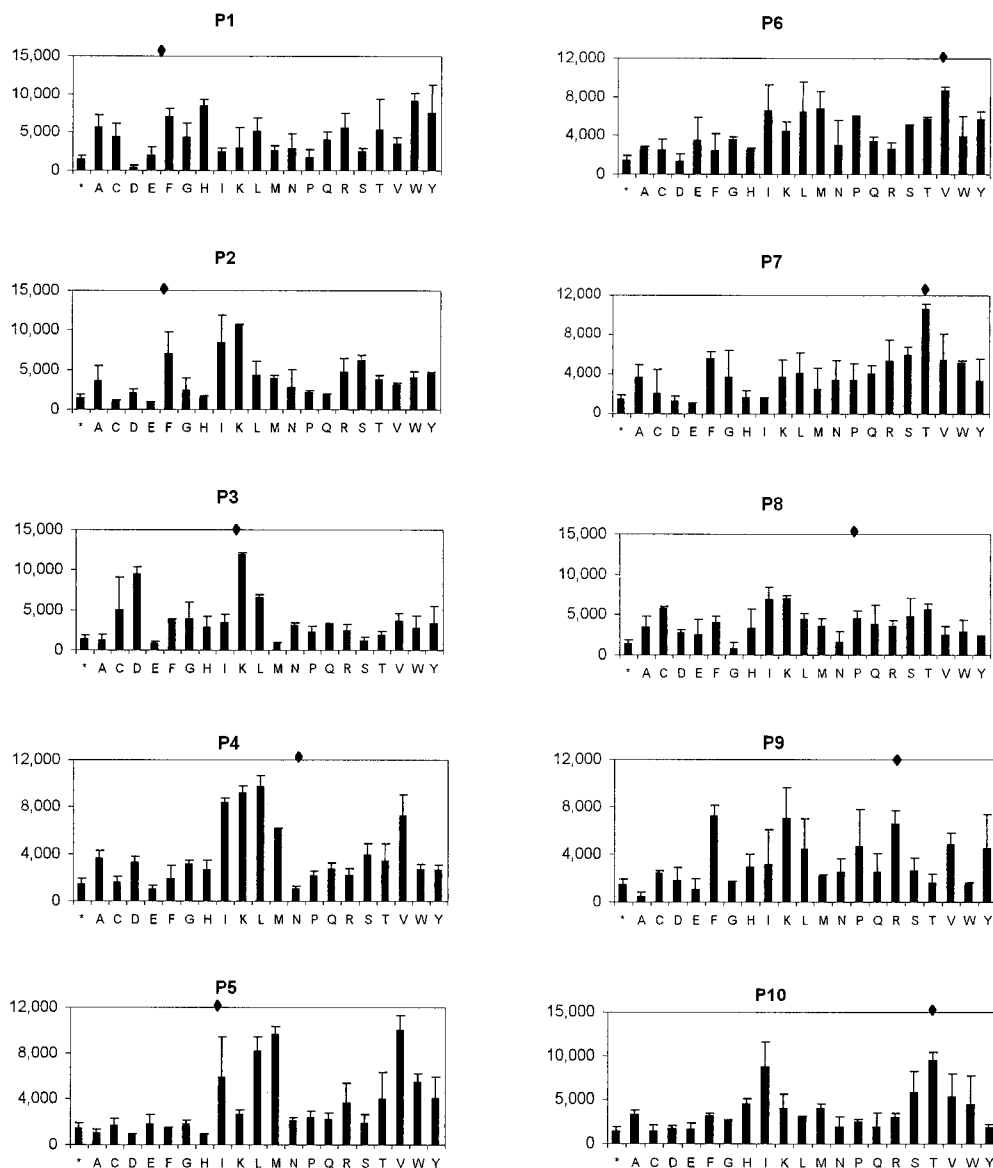
FIGURE 1. (*continued*)

median of all spot intensities on each individual array. Ratios of gene expression that were greater than 2-fold were considered significant based on a 99% confidence interval (44).

Results

Data obtained with combinatorial peptide libraries suggest different levels of TCR degeneracy for different CD4⁺ TCC

In this study, we sought to develop an approach that would combine the information generated from the screening of a decapeptide PS-SCL with all protein sequences in public databases. This strategy should allow the identification of the entire spectrum of stimulatory peptide ligands for a given TCC and the ranking of naturally occurring peptides with regard to predicted stimulation. The ultimate goal is to develop a methodology for identifying biologically relevant peptides for TCC of unknown specificity that have been isolated, e.g., from a tissue.

Three CD4⁺ TCC were tested in proliferative assays with the 200 mixtures of the decapeptide PS-SCL. Two TCC had known specificity, one specific for influenza HA (Flu-HA) (306–318) (TCC GP5F11), and one for MBP$_{83-99}$ (TCC TL3A6). We also

studied one clone of unknown specificity that recognizes *B. burgdorferi*, the causative organism of Lyme disease (TCC CSF-3).

Data obtained with combinatorial peptide libraries suggest different levels of TCR degeneracy for different CD4⁺ TCC. The stimulation profiles for TCC GP5F11 and TL3A6 are shown in Fig. 1, *A* and *B*, respectively. The profile for CSF-3 is shown previously (34). The profile of TL3A6 shows that more than one mixture in several positions of the PS-SCL generated a clear proliferative response. The amino acids of MBP$_{89-98}$ are marked by diamonds (FFKNIVTPRT). Although the target amino acids correspond to the defined amino acid in the most stimulatory mixtures in most positions, this is not observed in certain positions, such as N in position 4 and P in position 8. In contrast, the profiles for GP5F11 and CSF-3 show a very different pattern with fewer, but more differential activity between stimulatory and not stimulatory mixtures.

Limitations of motif searches

Motif searches are widely used to search protein databases in a nonquantitative manner. However, this approach was not successful for identifying the known target peptides of the TL3A6 and

Table I.　*Database search performed on SwissProt and GenPept to identify agonist peptides for TCC GP5F11*

		SwissProt		GenPept		
S-Index	Search Supermotif[a]	Target sequence	No. hits	Target sequence	No. hits, viral DB	No. hits, *H. sapiens* DB
>2	[WYFRH]-[MLIVADFYH]-**K**-[QVILYHKPTM]-[NHQM]-[TSNIQGVAHM]-[GPAHFSTYVNQLICM]-[RKGPMTNVS]-[FRMYKLVHQPNISWGA]-[LMIFVYQA]	Yes	513	Yes	560	177
>3	[WYFR]-[MLIVADF]-**K**-[QVILYHKP]-[NHQ]-[TSNIQGVAH]-[GPAHFSTYVNQL]-[RKGPM]-[FRMYKLVHQPNI]-[LMIFVY]	No	82	No	23	34

[a] Amino acids corresponding to Flu HA(308–317)(YVKQNTLKLA) are shown in bold underlined characters. SwissProt contains 83,857 protein sequences (3-3-00); GenPept viral database: 90,174 proteins (20,198,794 decamer peptides); *Homo sapiens* database: 43,795 proteins (13,879,822 decamer peptides).

GP5F11 clones. Motifs searches are generated from screening results of PS-SCL, and contained in each position are amino acids corresponding to mixtures with S-index greater than a specified threshold (see *Materials and Methods* for definition of S-index). Thresholds of 2 and 3 were used to generate the search motifs. The resulting motifs were then used to search the SwissProt and GenPept databases.

Tables I and II show the number of peptides that satisfied the motif searches, and indicates whether the target peptide was identified. The target peptide was not found with either of the motifs for TL3A6 in either database. The target peptide for GPF11 was identified only when the search criterion was so permissive/lax that over 500 other peptides were also selected. Furthermore, the inability of motif searches to rank peptides renders it almost impossible to identify the most likely epitopes in a rational way and without synthesizing and testing very large numbers of individual peptides.

Developing a score matrix-based approach for predicting T cell-stimulatory candidate peptides

It is clear that a more systematic approach that employs all the data generated from the screening of PS-SCL needs to be developed for the search of databases. Our strategy is outlined in the flow diagram (Fig. 2).

We recently demonstrated that each amino acid within a peptide contributes to recognition almost independently and in an additive fashion, so that amino acid substitutions that abrogate recognition can be compensated for by highly stimulatory substitutions in other positions (25). Thus, the overall stimulatory value of a peptide results from the combination of positive or negative effects of each of the amino acids. Based on these assumptions, we could show that peptides that shared no amino acid in corresponding positions of their sequences could still be recognized by the same TCR (25). Also, the findings that the specificity information derived from PS-SCL libraries is similar to that obtained with individual peptide analogues and the fact that highly active peptides can be identified allow the development of a new search algorithm.

Our algorithm provides a predicted stimulatory score for the peptide of the same length as used in PS-SCL libraries. Based on the above assumptions, the peptide score is the sum of position-specific scores of the component amino acids. The scoring is accomplished by calculation of a matrix in which the columns represent positions, and the rows the 20 aa used in PS-SCL libraries. The scoring matrix entry for a particular amino acid in a specific position is based on the stimulation assay results for the mixture of PS-SCL corresponding to that amino acid defined in that position (Fig. 3A). The scoring matrix entry can either use the S-index or use the Z-index, which takes into account the experimental errors (see *Materials and Methods*).

The matrix is then used to search for predicted stimulatory peptides in the public protein databases. By moving a decamer scoring window across the known protein sequences in 1-aa increments (Fig. 3B), a stimulatory score is calculated for all published 10-mer peptides, and then they are ranked accordingly. This strategy offers important advantages compared with motif searches: 1) all the information derived from the PS-SCL screening is used, and the selection based on a cutoff of activity is not required; 2) peptides are now ranked according to their predicted stimulatory score.
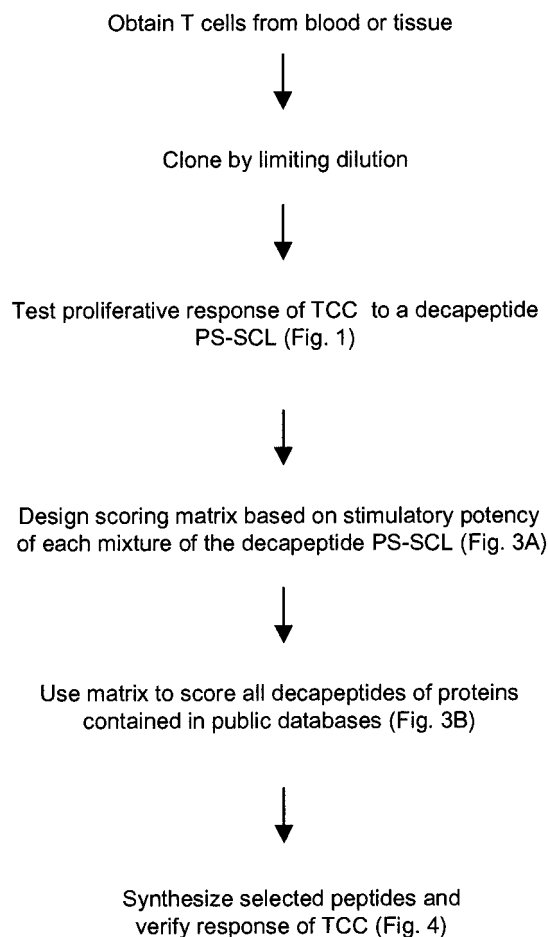
An example of a score matrix for one of the CD4[+] TCC (GP5F11) is shown in Fig. 3A. The amino acids of the Flu-HA$_{308–317}$ peptide are boxed. Note that the amino acids of the target peptide sequence L in position P7 and A in P10 are below an S-index value of 3, thus

Table II.　*Database search performed on SwissProt and GenPept to identify agonist peptides for TCC TL3A6*

		SwissProt		GenPept		
S-Index	Search Supermotif[a]	Target sequence	No. hits	Target sequence	No. hits, *H. sapiens* DB	No. hits, viral DB	No. hits, bacterial DB
>2	[WHYFARTLCGQVKN]-[KIFSRYLWMTAVN]-[KDLCGFVIYQNH]-[LKIMVSATDG]-[VMLIWYTR]-[VMILPTYSKWGEQNA]-[TSFVRWLQKGAPNY]-[KICTSPLFQMRAHW]-[FKRVPYLIH]-[TISVHWKMAFLR]	No	260,085	No	104,229	183,876	289,887
>3	[WHYFARTLCG]-[KIFSRYL]-[KDLC]-[LKIMV]-[VMLIW]-[VMILPTYSK]-[TSFVRW]-[KICTSPL]-[FKRVPYL]-[TISVHW]	No	797	No	285	502	776

[a] Amino acids corresponding to MBP(89–98)(FFKNIVTPRT) are shown in bold underlined characters. SwissProt contains 83,857 protein sequences (3-3-00); GenPept viral database: 90,174 proteins (20,198,794 decamer peptides); *H. sapiens* database: 43,795 proteins (13,879,822 decamer peptides); bacterial database: 111,807 proteins (32,604,667 decamer peptides).

Obtain T cells from blood or tissue

↓

Clone by limiting dilution

↓

Test proliferative response of TCC to a decapeptide
PS-SCL (Fig. 1)

↓

Design scoring matrix based on stimulatory potency
of each mixture of the decapeptide PS-SCL (Fig. 3A)

↓

Use matrix to score all decapeptides of proteins
contained in public databases (Fig. 3B)

↓

Synthesize selected peptides and
verify response of TCC (Fig. 4)

FIGURE 2. Flow diagram of the strategy used to quantitatively analyze TCR recognition of Ags by clonotypic T cells. Experimental data collected by measuring functional T cell responses to PS-SCL are then analyzed by a scoring matrix approach. This allows the identification and ranking of the spectrum of antigenic ligands for TCC of known and unknown specificity.

explaining the failure of the motif search to find the target influenza peptide. The principle of the sliding decamer scoring window that is moved across a protein sequence in 1-aa increments is shown in Fig. 3B. Three decamer peptides within the Flu-HA$_{304-321}$ sequence are scored by adding the stimulatory values of the respective 10 aa. Note the drastic changes in stimulatory scores when the scoring window is moved 1 aa to the left (score 51.98) or to the right (13.7) as compared with the optimal register that is shown in the middle (score 256.01). These changes of the scores indicate that, as soon as both MHC and TCR contact positions that contribute most of the stimulatory activity are out of the correct register, the peptide may lose binding to the MHC and/or fail to stimulate the clone because the TCR contacts are not positioned properly.

Testing the score matrix-based approach using clones with known specificity and with synthesized peptides

The effectiveness of this approach is demonstrated in Table III. When the score matrices for clones TL3A6 and GP5F11 were used to score all peptides in the GenPept database, both the target peptides (MBP$_{89-98}$ peptide for TL3A6, and Flu-HA$_{309-318}$ for GP5F11) were correctly identified. The GenPept database (ftp://ftp.ncifcrf.gov/pub/genpept) was searched because it is substantially larger than SwissProt (http://www.expasy.ch/sprot). The relative ranks obtained for the target peptides are given in Table III.

For GP5F11, the rank among viral peptides is given; for TL3A6, we show the rank among human peptides. Consistent with previous observations with another autoreactive clone (45), MBP$_{89-98}$ was far from optimal, i.e., it ranked only 202nd in the set of human peptides using the S-index matrix. In contrast, the target peptide Flu-HA$_{309-318}$ ranked as the sixth highest scoring peptide for GP5F11 among viral proteins, and 24th when not only viral, but also human proteins were scored. This also suggests that molecular mimics that are potentially more stimulatory than the native foreign peptide can be identified.

We assessed the predictive power of the algorithm using synthesized peptides tested for stimulation of the three clones (76 peptides for GP5F11, 144 peptides for TL3A6, and 88 peptides for CSF-3). For the two TCC of known specificity, TL3A6 and GP5F11, the peptide was considered stimulatory if its EC$_{50}$ (concentration that yields half-maximal stimulatory activity) was equal to or <10 times that of the target peptide (MBP$_{89-98}$ and *Influenza* virus HA$_{308-317}$, respectively). For CSF-3, the TCC of unknown specificity, the peptide was considered stimulatory if it activated the TCC with a Z-index >47.5 at any concentration between 0.001 and 100 μg/ml.

Table IV shows the relationship between stimulatory potential predicted by the scoring matrices and actual measurement of TCC stimulation. Thresholds for matrix score prediction were based on relative operating characteristic analysis (46) to balance sensitivity and specificity. For clone CSF-3, for example, of the 62 peptides predicted to be stimulatory (have scores above the threshold of 47.5), 58 did stimulate the TCC (a positive predictive value of 58/62, or 93.5%). Of the 26 peptides predicted to be nonstimulatory, only 5 stimulated the TCC (negative predictive value: 21/26, 80.8%). The sensitivity for predictions with this clone was 92%; that is, of the 63 peptides that actually stimulated the TCC, 58 were correctly predicted. The specificity was 84%; that is, of the 25 peptides that did not stimulate the TCC, 21 were correctly predicted. Although the sets of synthesized peptides are small compared with the number of peptides that would be predicted to be stimulatory, Table IV documents the excellent sensitivity, specificity, and negative predictive values for the three TCC.

Table V shows the information on the 10 highest scoring peptides derived from *B. burdorferi* database analysis for TCC CSF-3 with the half-maximal stimulatory value that was determined by dose-titration, proliferative experiments. Examples of the stimulatory activity of peptides predicted to activate TCC GP5F11 are shown in Fig. 4. Note that a predicted stimulatory peptide with optimal amino acids in each position (WMKQNIGRFL) and a higher score than the target peptide is in fact two orders of magnitude more potent than the target sequence. One of the shown peptides with a score of 132.40 ranks much lower than the putative stimulatory threshold for TCC GP5F11, and consequently it did not stimulate the clone. However, even a few high scoring peptides (data not shown) are not stimulatory from reasons that are currently under further investigation.

Combining scoring matrix predictions of TCC stimulation with cDNA microarrays to identify biologically relevant candidate peptide mimics

The novel strategy described in this work allows us to find peptides from every known source that have stimulatory activity for the clone that was tested with PS-SCL. This leads to the problem of how one identifies from this wealth of data which peptides may be biologically relevant. In cases in which the target Ag for the clone is not known or molecular mimics with potential relevance for an organ-specific disease are of interest, several strategies may be used.

FIGURE 3. *A*, Score matrix for TCC GP5F11. Data from a representative experiment of proliferative response of the TCC to a decamer PS-SCL experiment are used to generate the matrix. Each number represents the S-index (cpm in the presence of the mixture/cpm in the absence of the mixture) of each of the 200 mixtures of a decapeptide PS-SCL (20 aa, indicated by the single letter code, for each of the 10 positions of a decamer peptide, P1 to P10). In a model of independent contribution of each amino acid to peptide recognition, the stimulatory value of any decapeptide can be determined by summing the values of the individual amino acids in the score matrix. The example shown is a decamer peptide derived from influenza virus $HA_{308-317}$ that was used to establish the TCC. Boxed numbers correspond to the amino acid sequence of the peptide, and their sum represents the peptide score. Also shown are the maximum and minimum scores that can be assigned to any decamer peptides by this particular matrix. *B*, The scoring matrix can be used to score contiguous decamer peptides contained in all known protein sequences contained in public databases to find stimulatory peptides for a given TCC. The example shows a decamer scoring "window" moved in 1-aa increments along the sequence of influenza virus HA, recognized by TCC GP5F11. The matrix (Fig. 3) derived from a representative PS-SCL experiment (Fig. 1*A*) attributes the highest score to a decamer peptide (308–317) corresponding to the core of the 13-mer used to establish the TCC ($HA_{306-318}$). Dramatic changes can be shown by scoring the overlapping decamer peptides along the entire sequence (*B*). Remarkably, the highest score corresponds to the actual epitope recognized by the TCC.
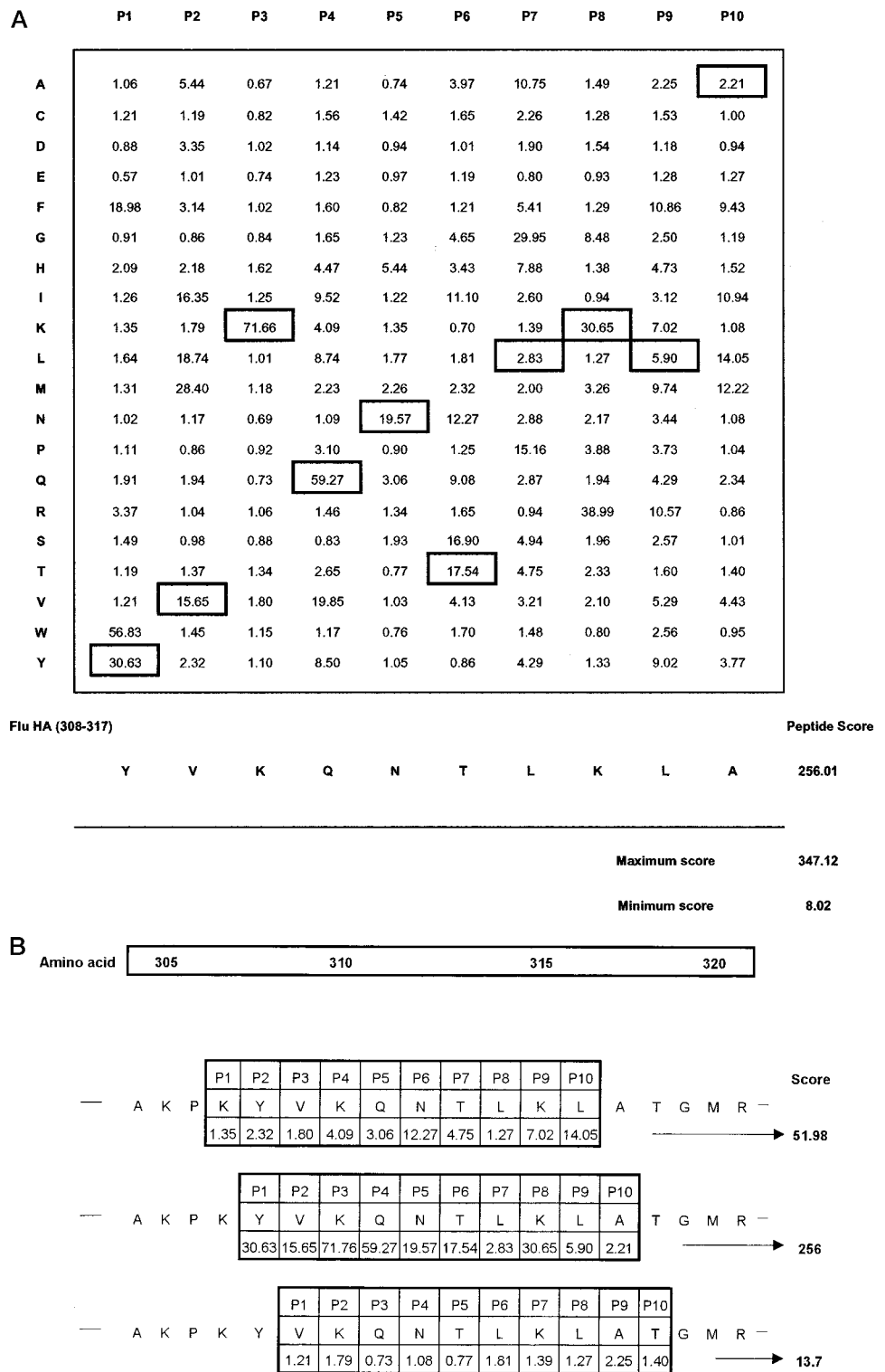
One approach to identify proteins involved in autoimmune diseases is to examine the expression of genes that are overexpressed in the target organ using cDNA microarray technology (41). We examined gene expression in 18 lesions from two MS patients and compared them with levels of gene expression in pooled normal white matter from three individuals with cDNA microarrays containing 2889 human genes. One of the genes that was overexpressed (>2-fold) in 17 of the 18 MS lesions examined was titin (Fig. 5*A*), a giant muscle protein (47). When we asked which genes that are overexpressed in MS plaques are

also identified as candidate epitopes/molecular mimics for CD4[+] TCC that were tested with the PS-SCL (Fig. 5*B*), we identified peptides derived from the same interesting candidate, titin, among the highest scoring peptides for both a CD4[+] TCC recognizing the immunodominant MBP peptide (83–99) in the context of the MS-associated DR allele DRB5*0101, but also for the *B. burgdorferi*-specific TCC CSF-3 (Fig. 5*C*). Titin, a giant muscle protein (47), is surprisingly overexpressed in MS brain tissue, and the identification of titin-derived peptides as candidate molecular mimics for two TCC that are potentially

Table III. *Database search performed on GenPept with a sum of S-index score matrix*

TCC	Target Sequence	Rank in Database
GP5F11	Yes	6[a]
TL3A6	Yes	202[b]

[a] A total of 90,174 proteins scored in viral database (20,198,794 decamer peptides).
[b] A total of 43,795 proteins scored in *H. sapiens* database (13,879,822 decamer peptides).

pathogenic in two different CNS inflammatory/autoimmune disorders, i.e., MS and chronic CNS Lyme disease, offers unique opportunities to study the involvement of such candidate Ags in the pathogenesis of these diseases.

Discussion

The experiments presented in this work have been conducted to better understand, measure, and predict both specific and degenerate interactions between clonotypic TCRs and MHC peptide ligands. For this purpose, an approach was devised that would allow us to 1) describe in a quantitative way the complex interactions of the trimolecular Ag recognition complex, and 2) identify the spectrum of stimulatory ligands for individual TCC with high predictive accuracy. We used combinatorial peptide libraries and biometric strategies in conjunction with large scale database searches to achieve this goal and could show for the first time that T cell recognition can be predicted in quantitative terms. This study builds on and expands previous investigations on the flexibility and degeneracy of TCR recognition of Ag. A role for degenerate T cell recognition has been postulated for such diverse immunological phenomena as thymic selection (48), peripheral T cell survival (49), protection from infectious diseases, and induction of autoimmunity (49, 50). It was previously shown that peptide combinatorial libraries in the positional scanning format can be used to define the spectrum of agonist ligands for clonotypic TCR (26, 49). In recent studies, we showed that functional responses elicited in CD4$^+$ TCC by PS-SCL could be used to build motifs for database searches and thus identify a spectrum of ligands of different potency for clonotypic TCR (45, 46). In the present study, we confirmed that functional T cell responses can be elicited by PS-SCL from certain CD4$^+$ TCC specific for both foreign (Fig.

1*A*) (34) and self (Fig. 1*B*) Ags. We then used a matrix-based methodology for the analysis of the experimental data generated with the PS-SCL (Fig. 2). This methodology is based on a model of independent and additive contribution of each amino acid in the peptide sequence to the interactions with both the TCR and the MHC molecule (25). Although numeric matrices (8) and other mathematical approaches based on independent amino acid contribution to antigenicity have been previously used to describe the interaction of antigenic peptides with specific MHC molecules (17, 18), the present study fills the important gap of applying a quantitative, matrix-based model to the interaction of an MHC peptide ligand (keeping the MHC molecule constant) with a specific, clonotypic TCR using the data generated from PS-SCL. The biometrical analysis described in this work systematically compares the information derived from a PS-SCL composed of trillions of decapeptides with all the decapeptides (13, 879, 822 for a *H. sapiens* database, and 20, 198, 794 for a viral database) contained in a public protein database to rank and predict the most stimulatory peptides for a given TCC. The predictions based on this methodology are so accurate (Tables III and IV, Fig. 4) (34) that they actually lend strong support to an additive, combinatorial model of peptide antigenicity. Available TCR crystal structures indeed suggest that peptides may modulate the preexisting affinity between MHC and TCR that is based on a large contact surface between these two components of the trimolecular complex (51, 52). It should be noted that this model does not contradict, but indeed extends and develops the concept of primary and secondary TCR contacts (23, 53). In fact, although complex substitutions of amino acids along the entire sequence of the peptide can lead to molecular mimicry in the absence of any sequence homology (25), the relative weight of different amino acids in each position of the peptides sequence is apparent from the experimental data (Fig. 1, *A* and *B*).

An important application of the above described model is that one can identify peptide ligands for a specific TCR by searching public database not only with MHC and TCR anchor motifs (54) or motifs obtained from PS-SCL data (34, 45, 49), but also using the scoring matrix derived from the screening of a PS-SCL composed of trillions of peptides (Fig. 3, *A* and *B*). We also illustrate the limitations of using motifs derived from PS-SCL screening to identify TCR agonist peptides. Such a strategy does not fully use the information generated by screening specific TCR with PS-SCL. Therefore, the native ligand may not be found if the motif is

Table IV. *Indices of the predictive power of the scoring matrix approach for the definition of the stimulatory potency of antigenic peptides*

	TCC CSF-3			TCC GP5F11			TCC TL3A6		
	Matrix score >47.5	Matrix score <47.5	Total	Matrix score >220	Matrix score <220	Total	Matrix score >45.2	Matrix score <45.2	Total
Experimental measurement									
Stimulatory	58	5	63	38	4	42	20	8	28
Nonstimulatory	4	21	25	2	32	34	18	98	116
Total	62	26	88	40	36	76	38	106	144
Sensitivity[a]		58/63 (92)			38/42 (90.5)			20/28 (71.4)	
Specificity[b]		21/25 (84)			32/34 (94.1)			98/116 (84.5)	
Positive predictive value[c]		58/62 (93.5)			38/40 (95.0)			20/38 (52.7)	
Negative predictive value[d]		21/26 (80.8)			32/36 (88.9)			98/106 (92.5)	
Overall accuracy[e]		79/88 (89.8)			70/76 (92.1)			118/144 (81.9)	

[a] Fraction of all stimulatory peptides that is correctly identified.
[b] Fraction of all nonstimulatory peptides that is correctly identified.
[c] Probability that a peptide predicted to stimulate actually does so.
[d] Probability that a peptide predicted to be nonstimulatory actually does not activate the TCC.
[e] Fraction of all predictions that is correct.

Table V. *Information on the 10 highest scoring peptides derived from* B. burdorferi *database analysis for TCC CSF-3*

Score	Sequence	Protein ID No.	Protein Description	EC_{50} $\mu g/ml^a$
54.82	N N I Y K K A L I S	AE001155	Hypothetical protein (section 41 of 70) of the complete genome	1
54.14	S N I I K S L S L F	AE001174	Hypothetical protein (section 60 of 70) of the complete genome	0.1–1
53.73	S N I I K K T S E D	AE001169	Similar to SP:P07017 (section 55 of 70) of the complete genome	1
53.70	F N I Y K R V V D N	AE001145	Hypothetical protein (section 31 of 70) of the complete genome	1
53.68	N N I D K K V Y T N	AE001135	(section 21 of 70) of the complete genome; similar to GB:Z32522	1–10
53.09	F F I K K R S L I I	AE000785	Hypothetical protein of plasmid Ip25	1
52.82	R N I F K K T V E N	AE001130	Similar to GB:L10328 (section 16 of 70) of the complete genome	>100
52.69	S N I K S K L I L V	AE001146	Similar to PID:1652132 (section 32 of 70) of the complete genome	1
52.63	Y N I I V S S L L L	AE001161	Hypothetical protein (section 47 of 70) of the complete genome	1–10
52.57	D N I F K K E T L I	AE001165	Similar to GB:L42023 (section 51 of 70) of the complete genome	1

^{a} Peptide concentration inducing half-maximal proliferation.

not sufficiently degenerate (Table I, S-index > 3; Table II, S-index > 3; S-index > 2), or if even one of the positions does not contain the amino acid that appears in the native sequence. Another advantage in the identification of T cell epitopes is that one can rank the predicted stimulatory peptides according to their score. This is of great practical value when the number of candidate peptides is very high (Table II) and one needs criteria to select which of the identified candidate peptides should be synthesized and actually tested with the TCC. In addition to identifying promptly the target peptide sequences (Table III), one can then synthesize and test a feasible number (hundreds) of candidate peptides to confirm their stimulatory activity (examples in Fig. 4; see also Table IV). Interestingly, we confirmed our previous observation that for autoreactive TCC, the ligand used to establish and expand the TCC is often a suboptimal one, consistent with the notion that high affinity self-reactive TCC are deleted in thymic selection (55). Whereas for autoreactive TCC we often found natural ligands derived from foreign or even self Ags whose potency was several orders of magnitude higher than that of the native peptide (45), for TCC GP5F11 and other TCC specific for foreign Ags (R. Martin, B. Gran, M. Nagal, E. Borras, S. Jacobsen, W. E. Biddison, R. Houghten, H. F. McFarland, and C. Pinilla, unpublished observations) the native ligand was much closer to the optimal one (Table III) (56, 57). Although more potent synthetic ligands could be designed based on the deconvolution of the PS-SCL data (26, 32) (e.g., peptide WMKQNIGRFL in Fig. 4), naturally occurring superagonists were rare. The fact that foreign Ag-specific TCC may recognize their antigenic peptides as highly potent ones is consistent with an efficient immune response required to eliminate infectious agents.

This study adds a new and important contribution to the definition and prediction of T cell epitopes using synthetic combinatorial libraries (26, 27). It should be noted that many of the previous approaches to the identification of T cell epitopes were based on the prediction of which peptides would be good binders for specific MHC/HLA molecules (8, 16). Because only a fraction of the potential MHC-binding peptides is a T cell epitope for an individual TCR, these approaches provide information that is specific for particular MHC molecules, but cannot predict which fraction of the peptides that bind a restriction element is actually stimulatory for a TCR with its unique structural features. Conversely, TCR ligands are not always high affinity MHC binders (58). The approach presented in this study takes into account the whole trimolecular complex of T cell activation by reading out a functional T cell response. This requires a certain degree of MHC peptide binding as well as the interaction of the MHC peptide ligand with a specific TCR. When both are considered, the overall accuracy of T cell epitope predictions is far superior to previously adopted methods (Table IV), although further improvements are currently being pursued. This is particularly helpful when the protein(s) recognized by a TCC is/are not known (34). Indeed, less than a third of the peptides that were identified and found to be stimulatory by the PS-SCL and scoring matrix approach would have been predicted to be good MHC binders based on a recently published MHC-binding prediction algorithm (12) (data not shown).

Finally, we show that combining the above-described methodology with the use of cDNA microarrays to assess differential gene expression in pathological and normal tissue of two patients with MS led to an interesting candidate molecule (titin, to date only known as an
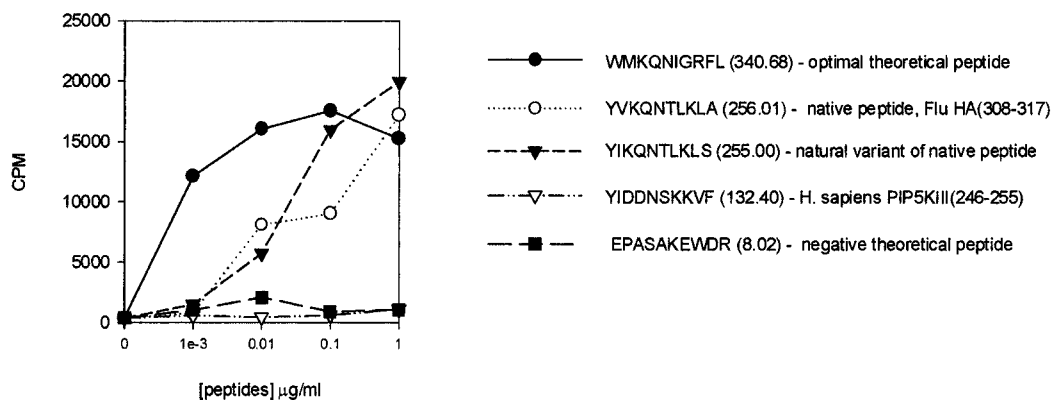
FIGURE 4. Proliferative response of the TCC GP5F11 to representative agonist peptides identified by the peptide library strategy. The potency is highest for a theoretical peptide that is predicted to be a potent one because it has a high score. The native peptide (influenza virus HA$_{308-317}$) and a double-substituted naturally occurring variant have intermediate potency. A low-scoring peptide derived from *H. sapiens* phosphatidylinositol-4-phosphate 5-kinase type III (PIP5KIII (246–255)) and a theoretical peptide predicted to be nonstimulatory because it has a very low score are indeed nonstimulatory.

A

FIGURE 5. *A*, Up-regulation of titin gene expression in lesions of two MS patients. Levels of titin expression in individual lesions from two MS patients (R and W). Bars represent ratios of expression of titin in the indicated 18 lesions relative to titin expression in pooled normal white matter. *B*, Identification of a potential autoantigen expressed in MS lesions by the integrated approach of peptide combinatorial libraries and cDNA microarray analysis. Two TCC reactive to myelin and microbial Ags were analyzed for their pattern of Ag recognition by the PS-SCL approach, and a numeric matrix was used to score and rank predicted stimulatory peptides for their potency (*left*). Gene expression in MS lesions and normal white matter was compared by cDNA microarray analysis, and a number of overexpressed genes was identified (*right*). The comparison of predicted stimulatory peptides and overexpressed genes identified interesting candidate target autoantigens such as the giant protein titin. *C*, Proliferative response of TCC CSF-3 to a titin-derived peptide. TCC CSF-3 was isolated from the CSF of a patient with chronic neuroborreliosis and recognizes a lysate of *B. burgdorferi* as well as a number of peptides derived from *B. burgdorferi*, human self Ags, and viral Ags (34). The proliferative response (in cpm) to titin (6205–6214) (GenBank accession no. X90569) is shown in one representative experiment. The background (no Ag) control proliferation was 198 cpm.

B

C
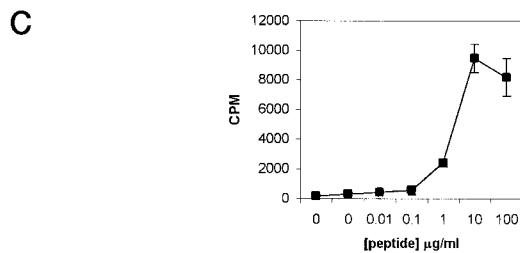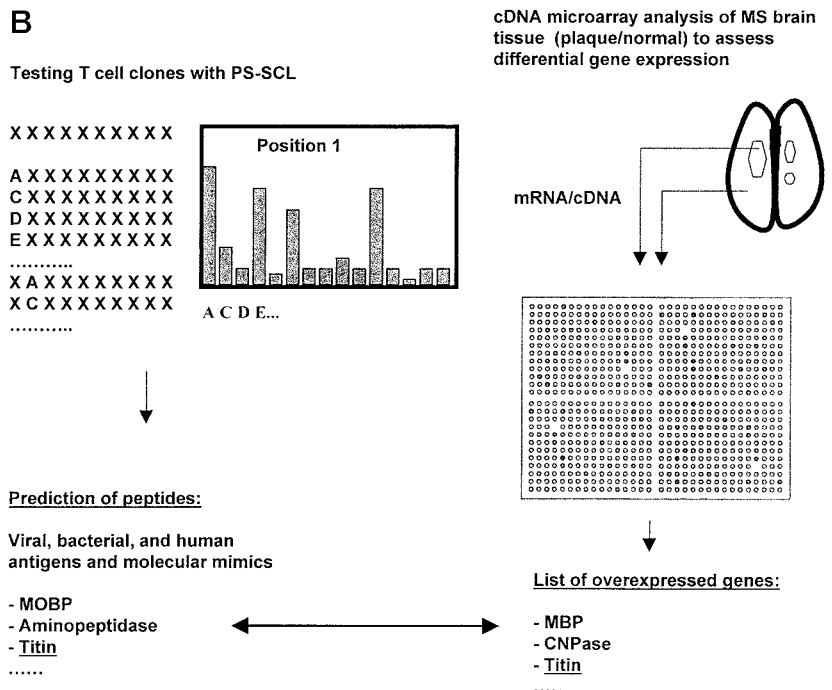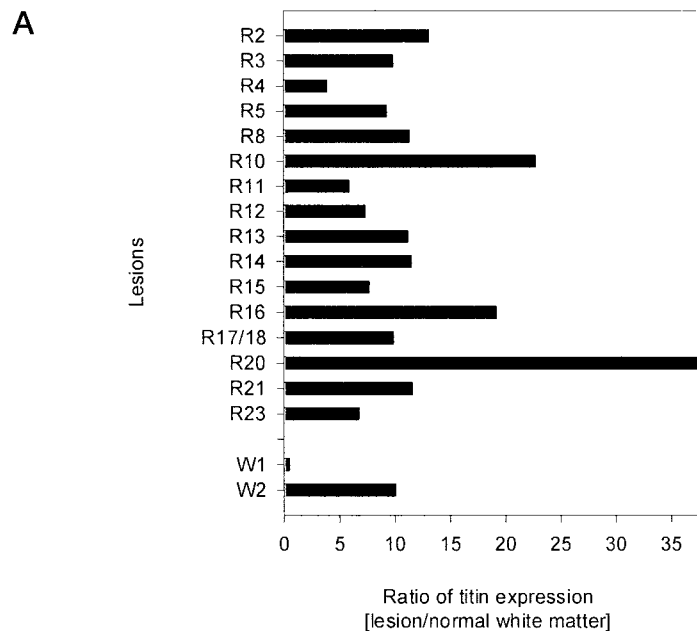
important component of skeletal muscle (47)) that is overexpressed in MS plaques and is recognized by a *B. burgdorferi*-specific TCC (Fig. 5). Preliminary pathological studies by immunohistochemistry indicate the expression of an isoform of this molecule in the pathologic, as opposed to normal white matter tissue, but further work to define its role is clearly needed. Thus, the combination of two powerful

methodologies can guide the discovery of candidate autoantigens that would otherwise not easily be identified by either approach.

In summary, we describe a methodology, PS-SCL-based biometrical analysis for ligand identification, which is consistent with a combinatorial model of TCR activation by antigenic peptides and allows the identification of T cell epitopes for both autoreactive

and foreign Ag-specific TCC with unprecedented efficacy. The same approach has also been successfully used for the prediction and identification of Ags by CD8[+] TCC (Ref. 59 and R. Martin, B. Gran, M. Nagai, E. Borras, S. Jacobson, W. E. Biddison, R. Houghten, H. F. McFarland, and C. Pinilla, unpublished results). For the first time, recognition of Ags by clones of unknown specificity can be decrypted. This is an important advance in the study of autoimmune disease, in which one tries to suppress specific immune responses, as well as for infectious and neoplastic diseases, in which a stimulation of specific responses by vaccines is pursued. Furthermore, it is important to note that this approach can be used to identify ligands within proteins in public database for any molecular interaction that has been or can be studied with PS-SCLs composed of L-amino acids.

Acknowledgments

References

1. Cresswell, P. 1994. Assembly, transport, and function of MHC class II molecules. *Annu. Rev. Immunol. 12:259.*
2. Engelhard, V. H. 1994. Structure of peptides associated with class I and class II MHC molecules. *Annu. Rev. Immunol. 12:181.*
3. Madden, D. R. 1995. The three-dimensional structure of peptide-MHC complexes. *Annu. Rev. Immunol. 13:587.*
4. Falk, K., O. Rotzschke, S. Stevanovic, G. Jung, and H. G. Rammensee. 1994. Pool sequencing of natural HLA-DR, DQ, and DP ligands reveals detailed peptide motifs, constraints of processing, and general rules. *Immunogenetics 39:230.*
5. Verreck, F. A., A. van de Poel, A. Termijtelen, R. Amons, J. W. Drijfhout, and F. Koning. 1994. Identification of an HLA-DQ2 peptide binding motif and HLA-DPw3-bound self-peptide by pool sequencing. *Eur. J. Immunol. 24:375.*
6. Rothbard, J. B., and M. L. Gefter. 1991. Interactions between immunogenetic peptides and MHC proteins. *Annu. Rev. Immunol. 9:527.*
7. Sette, A., J. Sidney, M. F. del Guercio, S. Southwood, J. Ruppert, C. Dahlberg, H. M. Grey, and R. T. Kubo. 1994. Peptide binding to the most frequent HLA-A class I alleles measured by quantitative molecular binding assays. *Mol. Immunol. 31:813.*
8. Hammer, J., E. Bono, F. Gallazzi, C. Belunis, Z. Nagy, and F. Sinigaglia. 1994. Precise prediction of major histocompatibility complex class II-peptide interaction based on peptide side chain scanning. *J. Exp. Med. 180:2353.*
9. Hammer, J., P. Valsasnini, K. Tolba, D. Bolin, J. Higelin, B. Takacz, and F. Sinigaglia. 1993. Promiscuous and allele-specific anchors in HLA-DR binding peptides. *Cell 74:197.*
10. Rammensee, H. G., T. Friede, and S. Stevanoviic. 1995. MHC ligands and peptide motifs: first listing. *Immunogenetics 41:178.*
11. Sette, A., S. Buus, E. Appella, J. A. Smith, R. Chesnut, C. Miles, S. M. Colon, and H. M. Grey. 1989. Prediction of major histocompatibility complex binding regions of protein antigens by sequence pattern analysis. *Proc. Natl. Acad. Sci. USA 86:3296.*
12. Sturniolo, T., E. Bono, J. Ding, L. Raddrizzani, O. Tuereci, U. Sahin, M. Braxenthaler, F. Gallazzi, M. P. Protti, F. Sinigaglia, and J. Hammer. 1999. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat. Biotechnol. 17:555.*
13. Hammer, J. 1995. New methods to predict MHC-binding sequences within protein antigens. *Curr. Opin. Immunol. 7:263.*
14. Mallios, R. R. 1994. Multiple regression analysis suggests motifs for class II MHC binding. *J. Theor. Biol. 166:167.*
15. Parker, K. C., M. A. Bednarek, and J. E. Coligan. 1994. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J. Immunol. 152:163.*
16. Southwood, S., J. Sidney, A. Kondo, M. F. del Guercio, E. Appella, S. Hoffman, R. T. Kubo, R. W. Chesnut, H. M. Grey, and A. Sette. 1998. Several common HLA-DR types share largely overlapping peptide binding repertoires. *J. Immunol. 160:3363.*
17. Brusic, V., G. Rudy, G. Honeyman, J. Hammer, and L. Harrison. 1998. Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network. *Bioinformatics 14:121.*
18. Gulukota, K., J. Sidney, A. Sette, and C. DeLisi. 1997. Two complementary methods for predicting peptides binding major histocompatibility complex molecules. *J. Mol. Biol. 267:1258.*
19. Honeyman, M. C., V. Brusic, N. L. Stone, and L. C. Harrison. 1998. Neural network-based prediction of candidate T-cell epitopes. *Nat. Biotechnol. 16:966.*
20. Milik, M., D. Sauer, A. P. Brunmark, L. Yuan, A. Vitiello, M. R. Jackson, P. A. Peterson, J. Skolnick, and C. A. Glass. 1998. Application of an artificial neural network to predict specific class I MHC binding peptide sequences. *Nat. Biotechnol. 16:753.*
21. Davenport, M. P., I. A. Ho Shon, and A. V. Hill. 1995. An empirical method for the prediction of T-cell epitopes. *Immunogenetics 42:392.*
22. Roberts, C. G., G. E. Meister, B. M. Jesdale, J. Lieberman, J. A. Berzofsky, and A. S. De Groot. 1996. Prediction of HIV peptide epitopes by a novel algorithm. *AIDS Res. Hum. Retroviruses 12:593.*
23. Kersh, G. J., and P. M. Allen. 1996. Structural basis for T cell recognition of altered peptide ligands: a single T cell receptor can productively recognize a large continuum of related ligands. *J. Exp. Med. 184:1259.*
24. Sloan-Lancaster, J., and P. M. Allen. 1996. Altered peptide ligand-induced partial T cell activation: molecular mechanisms and role in T cell biology. *Annu. Rev. Immunol. 14:1.*
25. Hemmer, B., M. Vergelli, B. Gran, N. Ling, P. Conlon, C. Pinilla, R. Houghten, H. F. McFarland, and R. Martin. 1998. Predictable TCR antigen recognition based on peptide scans leads to the identification of agonist ligands with no sequence homology. *J. Immunol. 160:3631.*
26. Pinilla, C., R. Martin, B. Gran, J. R. Appel, C. Boggiano, D. B. Wilson, and R. A. Houghten. 1999. Exploring immunological specificity using synthetic peptide combinatorial libraries. *Curr. Opin. Immunol. 11:193.*
27. Hiemstra, H. S., J. W. Drijfhout, and B. O. Roep. 2000. Antigen arrays in T cell immunology. *Curr. Opin. Immunol. 12:80.*
28. Gundlach, B. R., K.-H. Wiesmüller, T. Junt, S. Kienle, G. Jung, and P. Walden. 1996. Specificity and degeneracy of minor histocompatibility antigen-specific MHC-restricted CTL. *J. Immunol. 156:3645.*
29. Gundlach, B. R., K.-H. Wiesmüller, T. Junt, S. Kienle, G. Jung, and P. Walden. 1996. Determination of T cell epitopes with random peptide libraries. *J. Immunol. Methods 192:149.*
30. Udaka, K., K.-H. Wiesmüller, S. Kienle, G. Jung, and P. Walden. 1996. Self-MHC-restricted peptides recognized by an alloreactive T lymphocyte clone. *J. Immunol. 157:670.*
31. Wilson, D. B., C. Pinilla, D. H. Wilson, K. Schroder, C. Boggiano, V. Judkowski, J. Kaye, B. Hemmer, R. Martin, and R. A. Houghten. 1999. Immunogenicity. I. Use of peptide libraries to identify epitopes that activate clonotypic CD4[+] T cells and induce T cell responses to native peptide ligands. *J. Immunol. 163:6424.*
32. Hemmer, B., C. Pinilla, B. Gran, M. Vergelli, N. Ling, P. Conlon, H. F. McFarland, R. Houghten, and R. Martin. 2000. Contribution of individual amino acids within MHC molecule or antigenic peptide to TCR ligand potency. *J. Immunol. 164:861.*
33. Martin, R., U. Utz, J. E. Coligan, J. R. Richert, M. Flerlage, E. Robinson, R. Stone, W. E. Biddison, D. E. McFarlin, and H. F. McFarland. 1992. Diversity in fine specificity and T cell receptor usage of the human CD4[+] cytotoxic T cell response specific for the immunodominant myelin basic protein peptide 87–106. *J. Immunol. 148:1359.*
34. Hemmer, B., B. Gran, Y. Zhao, A. Marques, J. Pascal, A. Tzou, T. Kondo, I. Cortese, B. Bielekova, S. E. Straus, et al. 1999. Identification of candidate T-cell epitopes and molecular mimics in chronic Lyme disease. *Nat. Med. 5:1375.*
35. Vergelli, M., B. Hemmer, M. Kalbus, A. B. Vogt, N. Ling, P. Conlon, J. E. Coligan, H. McFarland, and R. Martin. 1997. Modifications of peptide ligands enhancing T cell responsiveness imply large numbers of stimulatory ligands for autoreactive T cells. *J. Immunol. 158:3746.*
36. Vergelli, M., B. Hemmer, U. Utz, A. Vogt, M. Kalbus, L. Tranquill, P. Conlon, N. Ling, L. Steinman, H. F. McFarland, and R. Martin. 1996. Differential activation of human autoreactive T cell clones by altered peptide ligands derived from myelin basic protein peptide (87–99). *Eur. J. Immunol. 26:2624.*
37. Houghten, R. A. 1985. General method for the rapid solid-phase synthesis of large numbers of peptides: specificity of antigen-antibody interaction at the level of individual amino acids. *Proc. Natl. Acad. Sci. USA 82:5131.*
38. Pinilla, C., J. R. Appel, and R. A. Houghten. 1994. Investigation of antigen-antibody interactions using a soluble, non-support-bound synthetic decapeptide library composed of four trillion (4×10^{12}) sequences. *Biochem. J. 301:847.*
39. Becker, K. G., D. H. Mattson, J. M. Powers, A. M. Gado, and W. E. Biddison. 1997. Analysis of a sequenced cDNA library from multiple sclerosis lesions. *J. Neuroimmunol. 77:27.*
40. Lassmann, H., C. S. Raine, J. Antel, and J. W. Prineas. 1998. Immunopathology of multiple sclerosis: report on an international meeting held at the Institute of Neurology of the University of Vienna. *J. Neuroimmunol. 86:213.*
41. Whitney, L. W., K. G. Becker, N. J. Tresser, C. I. Caballero-Ramos, P. J. Munson, V. V. Prabhu, J. M. Trent, H. F. McFarland, and W. E. Biddison. 1999. Analysis of gene expression in mutiple sclerosis lesions using cDNA microarrays. *Ann. Neurol. 46:425.*
42. Lennon, G., C. Auffray, M. Polymeropoulos, and M. B. Soares. 1996. The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics 33:151.*
43. Carlisle, A. J., V. V. Prabhu, A. Elkahloun, J. Hudson, J. M. Trent, W. M. Linehan, E. D. Williams, M. R. Emmert-Buck, L. A. Liotta, P. J. Munson, and D. B. Krizman. 2000. Development of a prostate cDNA microarray and statistical gene expression analysis package. *Mol. Carcinog. 28:12.*
44. Chen, Y., E. R. Dougherty, and M. L. Bittner. 1997. Ratio-based decisions and the quantitative analysis of cDNA microarray images. *Biomed. Optics 2:364.*
45. Hemmer, B., B. T. Fleckenstein, M. Vergelli, G. Jung, H. McFarland, R. Martin, and K. H. Wiesmüller. 1997. Identification of high potency microbial and self ligands for a human autoreactive class II-restricted T cell clone. *J. Exp. Med. 185:1651.*
46. Swets, J. A. 1988. Measuring the accuracy of diagnostic systems. *Science 240:1285.*
47. Labeit, S., and B. Kolmerer. 1995. Titins: giant proteins in charge of muscle ultrastructure and elasticity. *Science 270:293.*
48. Bevan, M. J. 1997. In thymic selection, peptide diversity gives and takes away. *Immunity 7:175.*

49. Hemmer, B., M. Vergelli, C. Pinilla, R. Houghten, and R. Martin. 1998. Probing degeneracy in T-cell recognition using combinatorial peptide libraries. *Immunol. Today 19:163.*

50. Gran, B., B. Hemmer, M. Vergelli, H. F. McFarland, and R. Martin. 1999. Molecular mimicry and multiple sclerosis: degenerate T-cell recognition and the induction of autoimmunity. *Ann. Neurol. 45:559.*

51. Garboczi, D. N., P. Ghosh, U. Utz, Q. R. Fan, W. E. Biddison, and D. C. Wiley. 1996. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature 384:134.*

52. Garcia, K. C., M. Degano, R. L. Stanfield, A. Brunmark, M. R. Jackson, P. A. Peterson, L. Teyton, and I. A. Wilson. 1996. An α/β T cell receptor structure at 2.5Å and its orientation in the TCR-MHC complex. *Science 274:209.*

53. Degano, M., K. C. Garcia, V. Apostolopoulos, M. G. Rudolph, L. Teyton, and I. A. Wilson. 2000. A functional hot spot for antigen recognition in a superagonist TCR/MHC complex. *Immunity 12:251.*

54. Wucherpfennig, K. W., and J. L. Strominger. 1995. Molecular mimicry in T cell-mediated autoimmunity: viral peptides activate human T cell clones specific for myelin basic protein. *Cell 80:695.*

55. Nossal, G. J. 1994. Negative selection of lymphocytes. *Cell 76:229.*

56. Bielekova, B., P. A. Muraro, L. Golestaneh, J. Pascal, H. F. McFarland, and R. Martin. 1999. Preferential expansion of autoreactive T lymphocytes from the memory T-cell pool by IL-7. *J. Neuroimmunol. 100:115.*

57. Hemmer, B., I. Stefanova, M. Vergelli, R. N. Germain, and R. Martin. 1998. Relationships among TCR ligand potency, thresholds for effector function elicitation, and the quality of early signaling events in human T cells. *J. Immunol. 160:5807.*

58. Muraro, P. A., M. Vergelli, M. Kalbus, D. Banks, J. W. Nagle, L. R. Tranquil, G. Nepom, W. E. Biddison, H. F. McFarland, and R. Martin. 1997. Immunodominance of a low-affinity major histocompatibility complex-binding myelin basic protein epitope (residues 111–129) in HLA-DR4 (B1*0401) subjects is associated with a restricted T cell receptor repertoire. *J. Clin. Invest. 100:339.*

59. Pinilla, C., V. Rubio-Godoy, V. Dutoit, P. Guillaume, R. Simon, Y. Zhao, R. A. Houghten, J. Cerottini, P. Romero, and D. Valmori. 2001. Combinatorial peptide libraries as an alternative approach to the identification of ligands for tumor-reactive cytolytic T lymphocytes. *Cancer Res. 61:5153.*