Prediction of MHC-Peptide Binding: A Systematic and Comprehensive Overview

Esther M. Lafuente[¶] and Pedro A. Reche[¶§]*

[§]*Laboratory of ImmuneMedicine,* [¶]*Department of Immunology, Facultad de Medicina, Universidad Complutense de Madrid, Ave. Complutense s/n, Madrid 28040, Spain.*

Abstract: T cell immune responses are driven by the recognition of peptide antigens (T cell epitopes) that are bound to major histocompatibility complex (MHC) molecules. T cell epitope immunogenicity is thus contingent on several events, including appropriate and effective processing of the peptide from its protein source, stable peptide binding to the MHC molecule, and recognition of the MHC-bound peptide by the T cell receptor. Of these three hallmarks, MHC-peptide binding is the most selective event that determines T cell epitopes. Therefore, prediction of MHC-peptide binding constitutes the principal basis for anticipating potential T cell epitopes. The tremendous relevance of epitope identification in vaccine design and in the monitoring of T cell responses has spurred the development of many computational methods for predicting MHC-peptide binding that improve the efficiency and economics of T cell epitope identification. In this report, we will systematically examine the available methods for predicting MHC-peptide binding and discuss their most relevant advantages and drawbacks.

Keywords: Peptide, T cells, epitopes, MHC, HLA, prediction.

INTRODUCTION

Classical major histocompatibility complex (MHC) molecules play a pivotal role in the adaptive immunity mediated by the T cells. T cell immune responses are induced by the T cell receptor (TCR) mediated recognition of antigenic peptides that are bound to MHC molecules expressed on the surface of antigen-presenting cells (APCs) [1, 2]. T cells that recognize self-peptides are eliminated during the process of thymic selection; therefore, the primary targets of T cell immune recognition (reviewed in ref. [3]) are MHC molecules incorporating foreign peptides.

MHC molecules fall into two major classes –MHC class I (MHCI) and MHC class II (MHCII) [4]. MHC class I (MHCI) molecules are expressed on most nucleated cells, generally bind peptides processed from endogenously synthesized antigens, and are recognized by CD8 T cells. Activated CD8 T cells become cytotoxic T lymphocytes (CTLs) that recognize and kill host target cells that express MHCI-bound antigenic peptides (e.g., infected and tumor cells).

In contrast, MHC class II (MHCII) molecules are expressed mostly on professional APCs, bind peptides that are processed from extracellular antigens, and are recognized by CD4 T cells. Recognition of MHCII-peptide complexes stimulates the regulatory helper functions of CD4 T cells, which acquire T_H1 or T_H2 phenotypes that orchestrate the immune response [5, 6]. T_H1 immune responses activate phagocytes, NKs and CD8 T cells, and are essential for fighting a variety of intracellular infections. T_H2 responses are key in eradicating extracellular pathogens through the activation of B cells.

In humans, MHC molecules are known as HLAs – Human Leukocyte Antigens– and are encoded by the HLA region on chromosome 6p21.3. There are three types of classical HLA class I (HLA I) molecules (HLA-A, HLA-B, and HLA-C), as there are for classical HLA class II (HLA II) molecules (HLA-DP, HLA-DQ, and HLA-DR). HLA gene expression is codominant, and an individual typically expresses 6 different HLA I molecules and 12 or more different HLA II molecules. In the entire population, there are hundreds of different HLA alleles, rendering the HLA locus the most polymorphic gene system known [2]. The HLA IMGT/HLA database [7] currently includes 3528 HLA allelic sequences (2496 HLA I alleles and 1032 HLA II alleles) (release 2.25, 04/2009). Each individual HLA molecule binds a large repertoire of peptides estimated to range from 1,000 to 10,000 individual sequences [8], and because peptide binding specificity is determined by the HLA polymorphism [9], HLA allelic variants generally bind distinct sets of peptides.

The identification of both CD8 and CD4 T cell epitopes is important in understanding disease pathogenesis [10]. Moreover, it is the basis for the development of epitope-based vaccines against infectious agents [11] and treatments for allergic [12], autoimmune [13] and neoplastic diseases [14]. Traditionally, the identification of T cell epitopes required the synthesis of overlapping peptides that spanned the entire length of a protein, followed by experimental assays for each peptide, such as *in vitro* intracellular cytokine staining [15], to determine T cell activation. This method is economically viable only for single proteins or pathogens that consist of several proteins. As a result, alternative computational approaches have been developed

*Address correspondence to this author at Department of Immunology, Facultad de Medicina, Universidad Complutense de Madrid, Ave. Complutense s/n, Madrid 28040, Spain; Tel: +34-91-394-7229; Fax: +34-91-394-1641; E-mail: parecheg@med.ucm.es

1381-6128/09 $55.00+.00 © 2009 Bentham Science Publishers Ltd.

for the prediction of T cell epitopes, which have significantly decreased the experimental burden that is associated with epitope identification.

Peptide binding to MHC is a prerequisite for T cell recognition and is, in addition, the event that most selectively defines T cell epitopes [16]. Therefore, the prediction of T cell epitopes relies primarily on predicting MHC-peptide binding. Many models and methods have been developed to predict peptide binding to MHC molecules, using different types of peptide binding data and algorithms. Notably, the same algorithm trained on data of different nature (e.g. quantitative *versus* qualitative data) will yield contrasting models. Moreover, completely different techniques can result in models that, despite resembling formats, are intrinsically unrelated (e.g. motif matrices and quantitative matrices). This complexity makes the description and classification of the disparate methods for predicting MHC-peptide binding somewhat difficult, and has led to some confusion in the literature. To avoid such confusion, in this review, we will carry out a comprehensive overview of the different methods to predict MHC-peptide binding, following a classification schema that focuses primarily on the type of data used for training.

STRUCTURAL BASIS OF MHC-PEPTIDE BINDING

The three-dimensional (3D) structures of classical MHCI and MHCII molecules are strikingly similar despite sharing minimal sequence similarity (< 20% identity). MHCI molecules are heterodimers consisting of a single membrane-spanning α chain that is paired with the soluble β2 microglobulin (β2m) protein. The peptide-binding domain of the MHCI molecule, formed by the distal α1 and α2 segments of the α chain (α1α2 domain), consists of two α-helices that lie above a floor composed of eight antiparallel β-stranded sheets (Fig. **1A**).

MHCII molecules are also heterodimers, but they consist of two membrane-spanning chains, –the α and β chains. The distal membrane regions of the α and β chains (α1 and β1 segments) fold together to form the peptide-binding domain (Fig. **1B**), which closely resembles that of the MHCI [17, 18]. Yet, there are key structural differences between the peptide-binding domains of MHCI and MHCII molecules that allow them to bind peptides in completely different modes.

MHCI-peptide ligands are short (8-11 amino acids) and their N- and C-terminal ends are connected to conserved residues of the MHCI molecule through a network of hydrogen bonds [19-21] (Fig. **1A**). The MHCI binding groove contains deep binding pockets that are delineated by polymorphic side-chains [9] in which specific peptide side chains fit (anchor residues). Generally, positions 2 and 3, and the C-terminus of the peptides that bind to MHCI are anchor positions (Fig. **1A**) [21], although the preferences for specific residues at anchor positions vary between MHCI molecules.

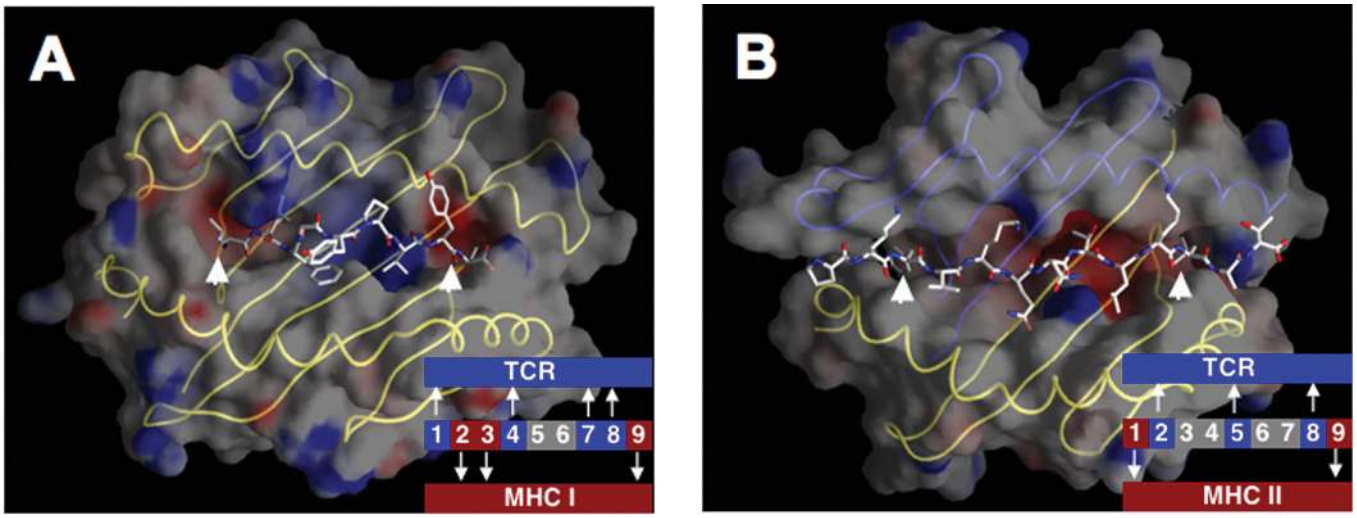

Fig. (1). *Binding of peptide ligands to MHCI and MHCII molecules.* This figure shows the top of the molecular surface of the antigen-presenting platform of representative human MHCI (A) and MHCII (B) molecules as viewed by the T cell receptor. The MHCI molecule corresponds to HLA-A*0201 in complex with the peptide LLFGYPVYV from HTLV-1 TAX protein (PDB: 1HHK). The MHCII molecule corresponds to HLA-DR1 in complex with the peptide PKYVKQNTLKLAT from influenza hemagglutinin protein (PDB:1FYT). Peptides bound to these molecules are represented by sticks to highlight the contours of the binding groove. Note how the peptide binding groove of the MHCI molecule is closed, and peptides bind in a manner such that both, the N-terminal and C-terminal ends of the peptide (indicated by arrow heads) are nestled into the MHCI binding groove, restricting their length to 8–11 residues. In contrast, the binding groove of the MHCII molecule is open, thereby imposing no limitation on the size of peptide ligands, whose N-terminal and C-terminal ends can extend beyond the binding groove. The side chains of N-terminal and C-terminal ends of the 9-mer peptide core that fits into the MHCII binding groove are indicated by arrow heads. The general binding mode of peptides to MHCI and MHCII is shown at the right corner of panels A and B. In this representation, peptide positions contacting the TCR and MHC are shaded in red and blue, respectively and are also indicated with opposing arrows. Positions shaded in grey can be anchor or TCR contact positions, depending of the specific MHC molecule. The figure was prepared using GRASP [122].

Because the geometry and chemical environment of the MHCI binding pockets impose tight constraints on the residues and overall peptide length to achieve binding, predicting MHCI-peptide binding is relatively simple and historically has progressed faster than that for MHCII molecules [22]. It is worth noting that peptides that have different sizes and bind to the same MHCI molecule often use alternative binding pockets [23]; therefore methods for predicting MHCI-peptide binding are specific for a fixed peptide length. However, most MHCI-peptide ligands have nine residues (they are 9-mers), making models for the prediction of 9-mer binders preferable.

In contrast to MHCI molecules, the peptide-binding groove of MHCII molecules is open, allowing the N- and C-terminal ends of a peptide to extend beyond the binding groove (Fig. **1B**). As a result, MHCII-bound peptides vary widely in length (9-22 residues), although only a core of nine residues (peptide-binding core) fits into the MHCII binding groove. The MHCII binding groove is shallower than that of MHCI, but it also has pockets that are outlined by polymorphic residues in which peptide side chains nestle. Typically, positions 1 and 9 of the peptide-binding core of MHCII-bound peptides are anchor positions [24] (Fig. **1B**). A major contribution to the overall binding of peptide to MHCII [18, 19, 25] is due to a set of conserved hydrogen bonds that lie between the backbone of the peptide core and the MHCII molecule [18, 19, 25]. With regard to these properties, MHCII molecules impose relatively weak limitations on peptide side chains, harboring broader peptide-binding repertoires than MHCI molecules. This quality along with the variable length of the MHCII-peptide ligands, complicates the identification of MHCII-peptide binding motifs and the generation of peptide binding models in general. Most MHCII-peptide binding models assume that the peptide-binding core alone determines the binding energy. Therefore, the generation of such models usually includes a preliminary step that identifies the binding core of the peptides. This step is an oversimplification, however, because there is evidence that the peptide-binding core flanking residues (PFRs) contribute to the overall binding [26].

PREDICTION OF MHC-PEPTIDE BINDING

There are myriad methods and models for predicting MHC-peptide binding, which differ in the techniques and algorithms that are applied as well as on the nature of the data used for training. However, the methods to predict MHC-peptide binding are generally described with regard to the techniques and algorithms used to develop them, which has led to some confusion in the literature. Instead, here we will review the different MHC-peptide binding prediction methods following a system that focus primarily on the type of data used for training (Fig. **2**). Under this scope, methods to predict MHC-peptide binding can be first classified in two overarching groups: those generated from sequences of peptide binders (sequence-based models or peptide-data driven models), and those that do not require any peptide binding data and are based solely on the sequence and structure of MHC molecules (modeling-based methods). Sequence-based models can also be divided in two types: those based on qualitative data that describe the ability of peptides to bind to MHC molecules (e.g. binders and non-binders) and those trained on quantitative data consisting of peptides whose binding affinity for MHC molecules has been determined. Qualitative data will only produce qualitative binding models that will predict, with some certainty, whether a test peptide binds to a given MHC molecule. Essentially, qualitative models try to recognize the binding pattern/feature that is present on the peptides used for training: therefore we will hereafter refer to them as binding pattern recognition models. In contrast, one can develop quantitative regression models from peptides that have known binding affinities to MHC molecules. These models target the prediction of the actual binding affinity of peptides to MHC; thus, we term these methods quantitative binding affinity models.

1. Binding Pattern Recognition Models

1.1. Sequence Motifs

A sequence motif is the simplest method of representing the peptide binding motif of a specific MHC molecule. It consists of a symbolic peptide string that lists the amino acid preferences of the MHC molecule at each position of the peptide. The first peptide binding motifs were obtained by pool sequencing of peptide ligands eluted from MHCI molecules [22]. In general, however, peptide binding motifs are identified by comparing sets of peptide sequences that are known to bind to MHC molecules [22, 27]. The largest collection of peptide binding motifs for MHCI and MHCII molecules is stored in the SYFPEITHI database [27].

MHC allele-specific peptide binding motifs were the first models to enable accurate identification of MHCI restricted T cell epitopes [28, 29], and they remain widely used (Table **I**). Yet, the use of sequence motifs is too simplistic, because residues that do not reside at anchor positions also contribute to binding [30, 31]. Moreover, sequence motifs are too rigid to predict MHC-peptide binding, yielding many false negatives.

1.2. Motif Matrices

Motif matrices represent an enhancement of sequence motifs. Examples of motif matrices were first developed by De Groot *et al.* [32] and Rammensee *et al.* [27] (Table **I**). These matrices consist of tables whose coefficients are associated with the position-specific amino acid frequencies found on peptides that bind to a specific MHC molecule. For any peptide sequence, the binding score is calculated by summing the relevant position and residue-matched matrix coefficients. Peptide biding scores that are generated using motif matrices are continuous; therefore, a binding threshold is required to distinguish the true binders.

Related to motif matrices are the profiles developed by Reche *et al.* [33-35] from sets of aligned peptides that bound to specific MHC molecules. These profiles are readily available for free public use at the RANKPEP site [33] (Table **I**). Profiles ––also known as position-specific scoring matrices (PSSMs)– were first introduced in 1987 by Gribskov *et al.* [36] to detect distantly related sequences and basically consist of log-odds matrices whose coefficients are

Table I. Freely Available Servers for Public Prediction of MHC-Peptide Binding

1. Binding Pattern Recognition Methods						
Server name	**URL**	**Technique/Algorithm**	**Class**	**nMHC(1)**	**Stypes(2)**	**Ref.**
MOTIF_SCAN	http://www.hiv.lanl.gov/content/immunology/motif_scan/motif_scan	Sequence Motifs	I and II	YES	YES	–
SYFFPEITHI	http://www.syfpeithi.de/Scripts/MHCServer.dll/EpitopePrediction.htm	Motif Matrices (MM)	I and II	YES	NO	[27]
RANKPEP	http://imed.med.ucm.es/Tools/rankpep.html	Profiles or PSSM	I and II	NO	NO	[33]
PEPVAC	http://imed.med.ucm.es/PEPVAC/	Profiles or PSSM	I	YES	YES	[112]
EPIMHC	http://imed.med.ucm.es/epimhc/	user made Profiles	I and II	NO	NO	[113]
NetMHC	http://www.cbs.dtu.dk/services/NetMHC/	Weight-Matrices	I	YES	NO	[38]
ANNPRED	http://www.imtech.res.in/raghava/nhlapred/neural.html	ANN	I	YES	NO	[114]
MULTIPRED	http://antigen.i2r.a-star.edu.sg/multipred/	ANN	I and II	YES	YES	[66]
MULTIPRED	http://antigen.i2r.a-star.edu.sg/multipred/	pHMM	I and II	YES	YES	[66]
SVMHC	http://www-bs.informatik.uni-tuebingen.de/SVMHC/	SVM	I and II	YES	NO	[50]
MHC2PRED	http://www.imtech.res.in/raghava/mhc2pred/	SVM	II	YES	YES	[51]
POPI	http://iclab.life.nctu.edu.tw/POPI/	SVM	I and II	YES	NO	[52]
KISS	http://cbio.ensmp.fr/kiss	SVM	I	YES	NO	[54]
2. Quantitative Binding Affinity Methods (regression models)						
Server name	**URL**	**Method/Algorithm**	**Class**	**nMHC(1)**	**Stypes(2)**	**Ref.**
BIMAS	http://www-bimas.cit.nih.gov/molbio/hla_bind/	Quantitative Matrices (QM)	I	NO	NO	[67]
IEDB	http://tools.immuneepitope.org/analyze/html/mhc_binding.html	ARB-QM	I	NO	NO	[115]
IEDB	http://tools.immuneepitope.org/analyze/html/mhc_binding.html	SMM-QM	I	NO	NO	[115]
netMHCII	http://www.cbs.dtu.dk/services/NetMHCII/	SMM-QM	II	YES	NO	[72]
PROPREDI(3)	http://www.imtech.res.in/raghava/propred1/	QM	I	YES	YES	[116]
PROPRED(4)	http://www.imtech.res.in/raghava/propred/	QM (Virtual QM)	II	YES	YES	[117]
MHCPRED	http://www.jenner.ac.uk/MHCPred/	QSAR	I and II	YES	NO	[77]
netMHC	http://www.cbs.dtu.dk/services/NetMHC/	ANN-regression	I	YES	NO	[79]
netMHCpan	http://www.cbs.dtu.dk/services/NetMHCpan/	ANN-regression	I	YES	NO	[81]
netMHCIIpan	http://www.cbs.dtu.dk/services/NetMHCIIpan/	ANN-regression	II	YES	NO	[82]
IEDB	http://tools.immuneepitope.org/analyze/html/mhc_binding.html	ANN-regression	I	NO	NO	[115]
SVRMHC	http://SVRMHC.umn.edu/SVRMHCdb	SVM-regression	I and II	NO	NO	[118]
HLABIND	http://atom.research.microsoft.com/hlabinding/hlabinding.aspx	Adaptive Double Threading	I	YES	NO	[84]
3. Modeling-Based Methods						
Server name	**URL**	**Method/Algorithm**	**Class**	**nMHC(1)**	**Stypes(2)**	**Ref.**
PREDEP	http://margalit.huji.ac.il/Teppred/mhc-bind/index.html	Threading	I	NO	NO	[89]
MHC-THREAD	http://www.csd.abdn.ac.uk/~gjlk/MHC-Thread/	Threading	II	NO	NO	[119]

(1) The server allows multiple (n) selections of MHC molecules for peptide binding predictions.
(2) The server allows the identification of promiscous MHC-binding peptides and/or prediction of peptides binding to supertypes
(3) PROPREDI QMs were compiled by the authors from the literature.
(4) PROPRED QMs correspond to the TEPITOPE virtual QMs.

logarithm ratios of observed amino acid frequencies with respect to the relevant background frequencies. Unlike basic motif matrices, profiles also correct for sequence redundancies (through sequence weights) and missing data (using pseudo-counts that are estimated from substitution matrices). RANKPEP MHCII-specific profiles were generated from peptide-binding cores that were identified from MHCII-peptide ligands using an expectation-maximization motif discovery program [37]. Virtually identical to these PSSMs are the weight-matrices latter described by Nielsen *et al.* [38].

All of the motif matrices described above were developed solely from positive examples that consisted of known MHC-binding peptides. In contrast, Mallios [39] described a motif matrix to predict peptide binding to HLA-DR molecules that was generated from positive and negative data using an stepwise discriminating analysis (SDA). Mallios [39] described this method as quantitative, because, in contrast to sequence patterns, it yielded continuous scores that discriminated peptides as binders and non-binders. The method, however, was not trained on quantitative binding affinity but on qualitative data; thus it cannot predict peptide binding affinities.

The prediction of MHC-peptide binding using motif matrices and sequence motifs assumes that peptide residues contribute independently to binding. Such an assumption, in general, is well supported by experimental data, but there is also evidence indicating that the contribution of peptide residues to MHC-binding is influenced by neighboring residues [40].

1.3. Machine Learning-Based Motifs

Machine learning algorithms (MLAs) can handle non-linear data and capture binding interferences between peptide residues. Using MLAs, peptide binding motifs are learned under a classification schema, in which MLAs are trained on positive (e.g., peptide binders) and negative (e.g., non-binders) sets of examples. Experimental data on non-binding peptides generally are unavailable, and randomly selected peptides from a large database often are used instead. MLAs can also learn more complex patterns under a multiclass classification, in which they are trained on MHC-binding peptides that are grouped into more than two classes (e.g., high binders, intermediate binders, weak binders, and non-binders).

Artificial Neural Networks (ANNs) are one of the MLAs most frequently applied to the recognition of MHC-peptide binding patterns. ANNs are connectionist models that commonly are used in biology for classification and pattern recognition [41]. ANNs were introduced first to predict peptide binding to MHCI – specifically to HLA-A*0201 [42, 43] and K^b [44] (a mouse MHCI molecule)– and latter extended to MHCII with HLA-DR4(DRB1*0401) [45, 46]. These models for peptide binding prediction to MHCII were trained on peptide-binding cores, which were identified prior to training the ANNs. To this end, Brusic *et al.* [46] devised a system that used an automated evolutionary algorithm to identify the peptide-binding core of MHCII-peptide binders, which were subsequently fed to an ANN. Predictions of peptide binding to MHC molecules using ANN-based motifs can be made at the MULTIPRED site (Table **I**).

Support vector machines (SVMs) have also been applied to recognize MHC-peptide binding. SVMs are a relatively new type of MLA that are increasingly being used in the life sciences to identify patterns [47-49]; they classify by constructing an *N*-dimensional hyperplane that separates the data optimally into categories. Using this approach, Donnes and Elofsson [50] developed the SVMHC method (Table **I**), which is available online to predict peptide binding to MHCI. Later, Bhasin and Raghava [51] used a similar SVM-based approach to predict peptide binding to HLA-DR4 (DRB1*0401).

Peptide input data used for training SVMs, and in general to any MLA, can be encoded into different formats, wherein the only limitation is that each peptide needs to be represented by a fixed length vector. Frequently, peptides are represented in binary format, but recently Tung and Ho [52] used the physicochemical properties of peptides known to bind to MHCI molecules as input data for SVMs, developing the POPI method (Table **I**). Similarly, Salomon and Flower [53] predicted MHCII-peptide binding using a kernel-based SVM that was trained on the similarity scores of MHCII allelic-specific peptide ligands. Notably, Salomon and Flower [53] did not preprocess the MHCII-peptide binders to identify the peptide-binding core; thus their method also can model the influence of PFRs (peptide-binding core flanking residues) on binding.

The unavailability of peptide binding data always limits the development of allelic-specific MHC-peptide binding predictors. Interestingly, Jacob and Vert [54] combined the routine SVM formulation, described by Donnes and Elofsson [50], with an user-defined measure of similarity between alleles, which allowed for the estimation of peptide binding to MHCI alleles with few known binders. Jacob and Vert's [54] predictive models are readily available at the KISS site (Table **I**).

Decision trees (DTs) are one of the most popular classification algorithms [55] and they also have been applied to predict MHC-peptide binding. In this approach, the DT is used to construct a graph model of the MHC-peptide binding motif, which subsequently can be used to decide whether a test peptide fits into that motif. Savoie *et al.* [56] were the first to report the use of BONSAI DTs to predict peptide binding for HLA-A*0201. A similar tree-structured technique was later reported by Segal *et al.* [57] to predict peptide binding to K^b (a mouse MHCI molecule). More recently, Zhu *et al.* [58] trained DTs using peptide binding data from different MHCI molecules simultaneously, which improved the prediction accuracy of peptides that bind to a specific MHCI molecule.

1.4. Hidden Markov Models Motifs

The Hidden Markov model (HMM) is the most widely used technique in speech and sequence pattern recognition [59, 60]. HMMs are based on parametric statistical models, wherein the system that is being modeled is assumed to be connected by a Markov chain of unknown hidden parameters that are obtained from the observable data. HMMs can cope

with non-linear data. However, unlike the MLAs that we have described, they can be trained only on positive data.

Three types of HMMs have been used to predict MHC-peptide binding: fully connected HMMs [61], structure-optimized HMMs [62, 63], and profile HMMs [63]. Fully connected HMMs (fcHMMs) were the first to be applied to predict MHC-peptide binding, specifically to HLA-A*0201 [61]. fcHMMs have the greatest potential to model non-linear data, recognizing different patterns in the peptide binders that are used for training, but they carry high computing costs. Profile HMMs (pHMMs) are derived from sets of aligned peptides using tools such as the HMMER and SAM packages [64]. pHMMs perform dramatically fewer computations than fcHMMs but are much weaker in modeling different patterns. In fact, pHMMs that are derived from sets of ungapped alignments (the case for peptides binding to MHC) are nearly identical to profile matrices [65]. In structure-optimized HMMs (soHMMs), the model's connectivity is greatly reduced with regard to fcHMMs, and yet they have some capacity to model non-linear data. HMMs used to predict peptide binding to MHCI and MHCII molecules are implemented in the MULTIPRED method [66] (Table **I**).

2. Quantitative Binding Affinity Models

2.1. Quantitative Additive Models

Quantitative matrices (QMs) are the most widely used additive models in predicting MHC-peptide binding. QMs resemble motif matrices, but, in contrast to motif matrices, they are generated from actual peptide binding affinity data. Therefore, peptide scores computed with QMs are intended to reflect actual binding affinities to MHC. The first QMs were derived by Parker *et al.* [67] for MHCI molecules, and they are implemented by the BIMAS method (BIMAS-QMs) (Table **I**). Parker *et al.'s* [67] QMs were generated from limited numbers of peptides.

QMs for predicting peptide binding to MHC molecules have also been developed using binding affinity data obtained from positional scanning combinatorial peptide libraries (PSCPLs) [68,69]. In this approach, all possible peptides of a given length are represented by sets of sublibraries, and in each sublibrary, one amino acid is fixed and the remaining positions contain mixtures of all amino acids. The coefficients of QMs that are generated using PSCPLs typically are logarithmic peptide concentrations relative to a reference peptide library.

Several other methods have been used to construct QMs from large sets of binding affinity data of peptides to MHC. Bui *et al.* [70] used the Average Relative Binding (ARB) method to derive QMs (ARB-QMs) to predict the binding affinity of peptides to MHCI and MHCII (Table **I**). Similarly, Peters *et al.* [71] have used a Stabilized Matrix Method (SMM) to generate QMs (SMM-QMs) from peptide binding affinity data, which was first applied to HLA-A*0201 [40]. For MHCII molecules, Nielsen *et al.* [72] used a novel SMM approach (SMM-align) to predict peptide binding affinity. Interestingly, Nielsen *et al.*'s [72] included the two most proximal PFRs to the peptide-binding core into their models, which improved their predictive performance.

Closely related with to QMs are the virtual QMs that were derived by Sturniolo *et al.* [73] to predict peptide binding affinity to HLA-DR molecules. These authors used peptide libraries to obtain a quantitative representation of the interaction of all natural amino acid residues with structurally identified binding pockets of HLA-DR molecules (pocket binding profile). Subsequently, they showed that the binding affinity of a peptide to any HLA-DR allele can be predicted by selecting the appropriated binding pocket profiles, using the TEPITOPE [74] system. TEPITOPE QMs are readily available for free public use at the PROPRED site (Table **I**).

As motif matrices, QMs assume an independent contribution of peptide side chains to MHC binding. In contrast, Doytchinova and colleagues [75, 76] used partial least squares (PLS), a robust multivariate statistical method, to derive quantitative structure activity relationship (QSAR) additive models, wherein the binding affinity of peptides to MHC is computed as the sum of amino acid contributions at each position plus the contribution of adjacent side chains interactions. PLS models for predicting peptide binding affinity to MHCI and MHCII molecules is implemented using the MHCPRED method [77] (Table **I**).

2.2. Machine Learning-Based Regression Models

Machine Learning regression models for predicting the binding affinity of peptides to MHC result from regression training of MLAs on peptides that have known binding affinities to MHC molecules. Both SVMs and ANNs have been used in this context. SVM regression has been applied by Liu *et al.* [78] to predict peptide binding affinity to MHCI molecules. These models are available at the SVRMHC site (Table **I**). ANNs were shown to perform sensitive predictions of peptide binding affinity to MHCI molecules by Buus *et al.* [79] and are implemented using the NetMHC method (Table **I**). The NetMHC method was latter improved by Nielsen *et al.* [80], who trained ANNs using combinations of novel input representations (sparse encoding, blosum encoding, and input that was derived from HMMs).

As noted, MHC binding predictions are limited by the availability of experimental peptide binding data. Ingeniously, by combining MHCI peptide binding residues and peptide binding affinity data as input information for training the ANNs, Nielsen *et al.* [81] developed the NetMHCpan method (Table **I**), which can predict peptide binding affinities to uncharacterized MHCI molecules. Nielsen *et al.* [82] also followed a similar approach in developing the NetMHCIIpan method (Table **I**), which predicts binding affinities of peptides to HLA-DR molecules with little or no binding data. The NetMHCIIpan predictive models were obtained from MHCII-peptide binding sequences of known binding affinity that were preprocessed with the SMM-align method and, in combination with the HLA-DR residues that make contact with the peptide, used as input data for training ANNs.

2.3. Structure-Based Quantitative Methods

Several groups have combined peptide binding data with 3D-structural information of MHC molecules to predict peptide binding affinities. To this end, Doytchinova and Flower [83] used a powerful molecular comparative

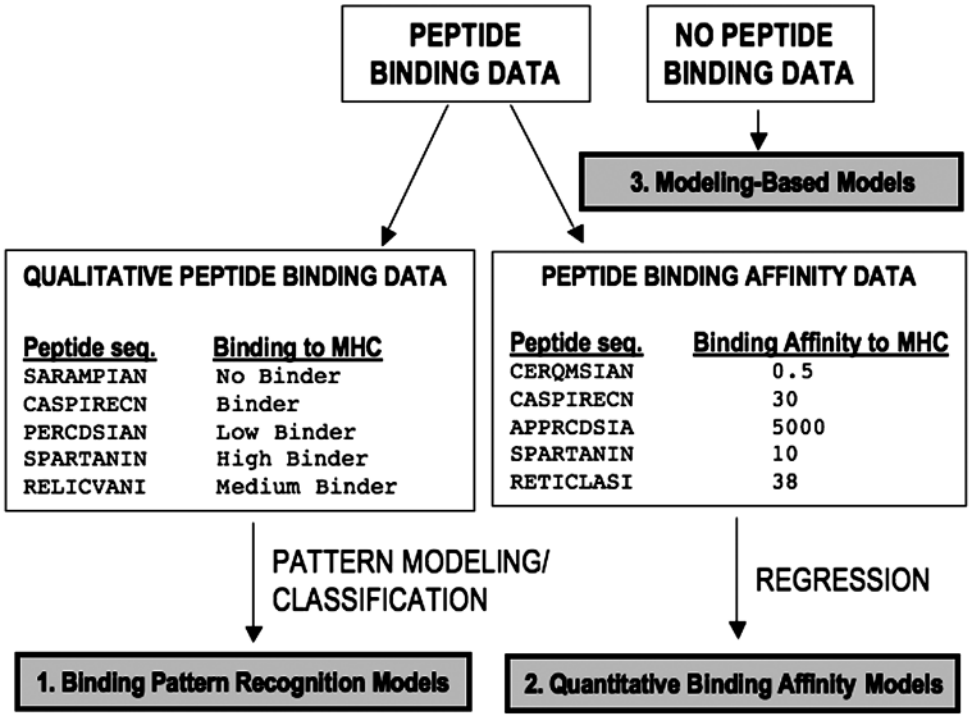

Fig. (2). *Classification of MHC-peptide binding prediction methods.* This figure shows a schematic overview of the different methods for predicting MHC-peptide binding according to the data used for training. Pattern Binding Recognition Models can also be generated using techniques that require only one class of data (e.g., peptides known to bind to a given MHC molecule).

similarity index analysis (CoMSIA) to predict the binding affinity of peptides to MHCI. The combination of 3D-structural information and peptide binding data can generate models that extrapolate peptide binding affinity predictions to MHC molecules for which very little binding data exist, as has been shown recently by Jojic *et al.* [84] using an adaptive double threading approach (Table **I**).

3. Modeling-Based Methods

In this category, we include methods that are not trained on any binding data and are solely based on the 3D structure of the MHC molecule; hence, they are often known as structure-based methods. However, we have classified them as modeling-based methods for two reasons: 1) as has been shown, there are other methods that use 3D-structural information of MHC molecules but in combination with peptide binding data; and 2) the 3D structure of most MHC molecules is not known but can be modeled or simulated. Modeling-based methods for predicting MHC-peptide binding are less accurate than data-driven models. Yet, they have the greatest potential, because they can be applied to any MHC molecule.

The first approach that was used to predict MHC-peptide binding was based on molecular dynamics simulations (MDSs) [85]. This method was computationally intensive and when it was tested on HLA-B*2705, it allowed only a crude discrimination between binders and non-binders. Altuvia *et al*. [86] introduced a structure-based algorithm to predict peptide binding to HLA-A*0201 that was considerably faster and more accurate. The algorithm consisted of fitting the peptide in the binding groove of the MHCI – peptide threading– and evaluating MHCI-peptide pairwise interactions using the knowledge-based potential matrix of Miyazawa and Jernigan [87]. This matrix, however, was only useful to predict the binding of peptides to MHCI molecules that had hydrophobic binding pockets but not to those with hydrophilic or charged pockets. In response to these limitations, Zhao *et al.* [88] and Schueler-Furman [89] developed novel knowledge-based potential matrices, which allowed to predict peptide binding to most MHCI molecules using to peptide threading. Recently, Mohanapriya *et al.* [90] also used peptide threading to predict peptide binding to MHCII molecules.

Docking or computer-simulated ligand binding has also been used to predict MHC-peptide binding. Docking methods use energy-scoring functions to calculate the binding energy of a series of ligand candidates. In this regard, Logean *et al*. [91] developed a tailor-made energy scoring function (FRESNO) to predict peptide binding to MHCI molecules. Peptide threading and docking methods do not perform exhaustive figurations of the MHC-peptide structure. Nevertheless, accurate prediction of MHC-peptide structures can certainly improve the accuracy of peptide binding predictions and relevant efforts have been made to model the appropriate geometry of MHCI- and MHCII-bound peptides [92-96].

MHC SUPERTYPES

The development of T cell epitope-based vaccines with wide population coverage is significantly hindered by HLA polymorphisms; they generally determine the peptide binding specificity of MHC molecules [9]. Therefore, grou-

ping MHC molecules into supertypes, based on similarities in peptide binding specificity, is useful to optimize the composition of epitope-based vaccines.

MHC supertypes were first identified by Sette and Sidney [97, 98], who, by inspecting HLA I allele-specific binding motifs, recognized that most HLA I alleles could be classified into nine supertypes. Further, they observed that 88% of the population expressed the alleles of just three supertypes (HLA-A2, HLA-A3, and HLA-B7). Subsequently, other more sophisticated computational approaches have emerged to classify HLA I and HLA II alleles, which have served to refine and expand the original supertypes defined by Sette and Sidney [97, 98]. Reche and Reinherz [99, 100] described a method for defining MHC supertypes by clustering MHC allelic-specific peptide binding repertories and Lund *et al*. [101] reported a similar method by clustering peptide binding matrices. These methods for identifying MHC supertypes are limited by the availability of allele-specific predictors of peptide binding. In contrast, Doytchinova *et al*. [102-104] have reported a system to classify into supertypes all HLA I and HLA II allelic variants, which is solely based on the 3D structure of MHC molecules. Similarly, Kangueane *et al.* [105] recognized that all HLA-A, -B, and -C alleles can be grouped into supertypes on the basis of several critical polymorphic residues.

COMPARING THE METHODS FOR PREDICTING MHC-PEPTIDE BINDING

Peptide-data driven models overwhelmingly outperform modeling-based methods. Despite the focus of many studies on evaluating the success of various MHC-peptide binding prediction methods, there is no consensus on the ideal method. This controversy arises because the predictive performance of the methods varies for different MHC molecules and evaluation test sets. Moreover, the value of comparative evaluations between pattern recognition models and regression models must be examined with caution, because a regression model can be used to predict whether a peptide binds to MHC, but a pattern recognition model cannot predict binding affinities. Therefore, peptide scores that are obtained using pattern recognition models should not be correlated with binding affinities.

Often predictive methods based on ANNs and SVMs are described as superior methods, because they can model binding interferences between peptide side chains. However, using a benchmark of peptides of known binding affinities to MHCI, Gulukota *et al.* [43] found that modeling peptide side chain interactions using ANNs only improved marginally the predictions obtained with QMs, which assume an independent binding contribution of each peptide residue [43]. In another study, Yu *et al*. [106] compared the efficacy of simple sequence motifs, SYFPEITHI motif matrices [27], ANN-based classification models, HMMs, and BIMAS-QMs to predict peptide binding to MHC, concluding that ANN-based classification models and HMMs were superior when trained on large datasets (more than 50 peptides). The HMMs that were used in this evaluation were profile HMMs generated from ungapped alignments of peptides. As noted, such HMMs are equivalent to profile PSSMs [107] and both should share the same detection and predictive powers, provided that they are generated from the same peptides. HMMs are superior to profile PSSMs only when they are derived from gapped alignments. Interestingly, Yu *et al*. [106] also found that despite the inferiority of simple sequence motifs to the other methods, they were the best predictors when the training sets consisted of merely several peptides.

Peters *et al*. [108] also performed a comparative validation using a benchmark that consisted of peptides of known affinities to different MHCI molecules. These authors found that SMM-QMs and ANN-based regression models performed better than ARB-QMs and BIMAS-QMs. They also reported that SYFPEITHI motif matrices [27] are poor predictors of peptide binding affinity, as one would expect. More recently, Lin *et al*. [109] evaluated several methods for predicting peptide binding affinity to MHCII using a benchmark of peptides of known binding affinities to MHCII wherein NetMHCIIpan was judged to be the best predictor, followed by PROPRED (TEPITOPE QMs).

WEB SERVERS AND DATABASES

The methods for predicting MHC-peptide binding that have been described in this article and are available for public use on the Internet are listed on Table **I**. All of the methods in Table **I** have been developed under cross-validation; therein, choosing the right tool for predicting MHC-peptide binding is not straightforward. However, one must know whether a specific method predicts binding affinities or merely qualitative binding (binding pattern recognition models).

Prediction rates obtained with any of these methods generally vary for the different MHC molecules. Moreover, not a single method has yet to yield reasonably accurate predictions for all MHC molecules. Therefore, users should make use of these web-based methods as frameworks to carry their own computational experiments and test the performance for the MHC molecule of interest. To this end, users should refer to their own sets of MHC-binding peptides, which can be obtained from public resources such as those listed in Table **II**. Web-based prediction tools should themselves facilitate computational experimentation, providing results that are easy to interpret and in formats that can be easily downloaded and processed for further analysis by users.

CONCLUDING REMARKS

The prediction of MHC-peptide binding is the basis for anticipating T cell epitopes. The identification of T cell epitopes requires laborious and costly experiments, and the prediction of MHC-peptide binding significantly decrease these experimental burdens. In recent years, there has been tremendous growth of methods that predict MHC-peptide binding, and significant progress has been reported with regard to their *in silico* predictive accuracy and the range of MHC molecules that can be targeted for peptide binding predictions.

In practice, however, the prediction of MHC-peptide binding is far from perfect. For example, in a recent study,

Table II. Databases of MHC-Binding Peptides and T-Cell Epitopes

Database	URL	Description	Ref.
SYFPEITHI	http://www.syfpeithi.de	MHC ligands and peptide motifs	[27]
HIV Database	http://hiv-web.lanl.gov/immunology	HIV T cell epitopes	–
EPIMHC	http://imed.med.ucm.es/epimhc/	MHC ligands	[113]
MHCBN	http://www.imtech.res.in/raghava/mhcbn/	MHC binding and non-binding peptides	[120]
ANTIJEN	http://www.jenner.ac.uk/antijen/	MHC and TAP binding peptides	[121]
IEDB	http://www.immuneepitope.org/	MHC ligands and MHC-binding peptides	[115]

Wang *et al*. [110] predicted 167 9-mer peptides from *influenza A virus* (IAV) as potential binders of various HLA I alleles, using NetMHC [79, 80]. Subsequently, these authors tested binding and only 89 peptides (53% of the pool) were classified as authentic binders; 39 failed to bind altogether, and another 39 bound very weakly. Of the successful binders, only 13 peptides elicited T cell responses (real T cell epitopes) (7.8% of their total predicted peptide pool). A similar discovery ratio of IAV-specific CD8 T cell epitopes restricted by K^b and D^b was reported by Zhong *et al*. [111] using SYFPEITHI motif matrices [27]. This low epitope-discovery rate is, however, deceiving, because MHC-peptide binding predictions do not substitute for experimental work but merely complement it. Thus, an alternative approach would have required to make and test over 4000 9-mers constituting the IAV proteome that, based on size, can bind to MHC. However, a reasonable question arises with regard to the use of MHC-peptide binding prediction methods: Do they overlook any true MHC-binding peptides?

Overall, the results that we have discussed indicate that predicting MHC-peptide binding is vital for epitope identification but remains challenging with ample room for improvement. Furthermore, the gap between the peptides that are predicted to bind to MHC and those that experimental turn out to bind, implies that new evaluation benchmarks and experimental data are required. Indeed, we believe the field will benefit from launching a Critical Assessment of Techniques for MHC-peptide binding/ Epitope Prediction, similar to the Critical Assessment of Techniques for Protein Structure (CAPS). Under this program, computational methods will be used for blind *de novo* prediction of peptides that bind to MHC from query proteins that, for evaluation purposes, have been experimentally screened for MHC-peptide binding. Given the synergy between peptide data and MHC-peptide binding predictions, this endeavor will undoubtedly serve to improve the accuracy of MHC-peptide binding prediction methods.

ACKNOWLEDGMENTS

This work and PAR were supported by Grant SAF2006-07879 from Ministerio of Educación y Ciencia (M.E.C) of Spain and by Grant CCG08-UCM/BIO-3769 from Universidad Complutense de Madrid (U.C.M). EML is supported by Grants SAF2007-60578 and CCG08-UCM/SAL-4259 from M.E.C and U.C.M., respectively.

REFERENCES

References 123-125 are related articles recently published.

[1] Zinkernagel M, Doherty PC. Restriction of *in vitro* T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. Nature 1974; 248: 701-702.

[2] Terasaki PI. A brief history of HLA. Immunol Res 2007; 38: 139-48.

[3] von Boehmer H. Positive and negative selection of the αβ T cell repertoire *in vivo*. Curr Opin Immunol 1991; 3: 210-5.

[4] Maenaka K, Jones EY. MHC superfamily structure and the immune system. Curr Opin Struct Biol 1999; 9: 745-53.

[5] Mosmann TR, Coffman RL. TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties. Annu Rev Immunol 1989; 7: 145-73.

[6] Reinherz EL, Schlossman SF. The differentiation and function of human T lymphocytes: a review. Cell 1980; 19: 821-7.

[7] Robinson J, Waller MJ, Fail SC, McWilliam H, Lopez R, Parham P, *et al*. The IMGT/HLA database. Nucleic Acids Res 2009; 37: D1013-7.

[8] Hillen N, Stevanovic S. Contribution of mass spectrometry-based proteomics to immunology. Expert Rev Proteomics 2006; 3: 653-64.

[9] Reche PA, Reinherz EL. Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. J Mol Biol 2003; 331: 623-41.

[10] Tchernev G, Orfanos CE. Antigen mimicry, epitope spreading and the pathogenesis of pemphigus. Tissue Antigens 2006; 68: 280-6.

[11] Reche PA, Keskin DB, Hussey RE, Ancuta P, Gabuzda D, Reinherz EL. Elicitation from virus-naive individuals of cytotoxic T lymphocytes directed against conserved HIV-1 epitopes. Med Immunol 2006; 5: 1.

[12] Akdis CA, Akdis M, Blesken T, Wymann D, Alkan SS, Muller U, *et al*. Epitope-specific T cell tolerance to phospholipase A2 in bee venom immunotherapy and recovery by IL-2 and IL-15 *in vitro*. J Clin Invest 1996; 98: 1676-83.

[13] Stienekemeier M, Falk K, Rotzschke O, Weishaupt A, Schneider C, Toyka KV, *et al*. Vaccination, prevention, and treatment of experimental autoimmune neuritis (EAN) by an oligomerized T cell epitope. Proc Natl Acad Sci USA 2001; 98: 13872-7.

[14] Lazoura E, Apostolopoulos V. Insights into peptide-based vaccine design for cancer immunotherapy. Curr Med Chem 2005; 12: 1481-94.

[15] Draenert, Altfeld M, Brander C, Basgoz N, Corcoran C, Wurcel AG, *et al*. Comparison of overlapping peptide sets for detection of antiviral CD8 and CD4 T cell responses. J Immunol Methods 2003; 275: 19-29.

[16] Jensen PE. Recent advances in antigen processing and presentation. Nat Immunol 2007; 8: 1041-8.

[17] Madden DR. The three-dimensional structure of peptide-MHC complexes. Annu Rev Immunol 1995; 13: 587-622.

[18] Stern LJ, Wiley DC. Antigen peptide binding by class I and class II histocompatibility proteins. Structure 1994; 2: 245-51.
[19] Madden D. The three-dimensional structure of peptide-MHC complexes. Annu. Rev. Immunol 1995; 13: 587-622.
[20] Matsumura M, Fremont D, Peterson PA, Wilson IA. Emerging principles for the recognition of peptide antigens by MHC class I molecules. Science 1992; 257: 927-34.
[21] Zhang C, Anderson A, DeLisi C. Structural principles that govern the peptide-binding motifs of class I MHC molecules. J Mol Biol 1998; 281: 929-47.
[22] Falk K, Rotzschke O, Stevanovic S, Jung G, Rammensee HG. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature 1991; 351: 290-6.
[23] Madden D, Garboczi DN, Wiley DC. The antigenic identity of peptide-MHC complexes: a comparison of the conformations of five viral peptides presented by HLA-A2. Cell 1993; 75: 693-708.
[24] Wang J-H, Reinherz E. Structural basis of T cell recognition of peptides bound to MHC molecules. Mol Immunol 2001; 38: 1039-49.
[25] Barber LD, Parham P. Peptide binding to major histocompatibility complex molecules. Annu Rev Cell Biol 1993; 9: 163-206.
[26] O'Brien C, Flower DR, Feighery C. Peptide length significantly influences *in vitro* affinity for MHC class II molecules. Immunome Res 2008; 4: 6.
[27] Rammensee HG, Bachmann J, Emmerich NPN, Bacho OA, Stevanovic S. SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 1999; 50: 213-9.
[28] Pamer EG, Harty JT, Bevan MJ. Precise prediction of a dominant class I MHC-restricted epitope of Listeria monocytogenes. Nature 1991; 353: 852-5.
[29] Rotzschke O, Falk K, Stevanovic S, Jung G, Walden P, Rammensee HG. Exact prediction of a natural T cell epitope. Eur J Immunol 1991; 21: 2891-4.
[30] Bouvier M, Wiley DC. Importance of peptide amino acid and carboxyl termini to the stability of MHC class I molecules. Science 1994; 265: 398-402.
[31] Ruppert J, Sidney J, Celis E, Kubo T, Grey HM, Sette A. Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules. Cell 1993; 74: 929-37.
[32] De Groot AS, Jesdale BM, Szu E, Schafer J, Chicz M, Deocampo G. An interactive web site providing major histocompatibility ligand predictions: application to HIV research and AIDS. AIDS Res Hum Retroviruses 1997; 13: 529-31.
[33] Reche PA, Glutting J-P, Reinherz EL. Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. Immunogenetics 2004; 56: 405-19.
[34] Reche PA, Glutting JP, Reinherz EL. Prediction of MHC class I binding peptides using profile motifs. Hum Immunol 2002; 63: 701-9.
[35] Reche PA, Reinherz EL. Prediction of peptide-MHC binding using profiles. Methods Mol Biol 2007; 409: 185-200.
[36] Gribskov M, McLachlan AD, Eisenberg D. Profile analysis: detection of distantly related proteins. Proc Natl Acad Sci USA 1987; 84: 4355-8.
[37] Bailey TL, Elkan C. The value of prior knowledge in discovering motifs with MEME. Proc Int Conf Intell Syst Mol Biol 1995; 3: 21-9.
[38] Nielsen M, Lundegaard C, Worning P, Hvid CS, Lamberth K, Buus S, *et al*. Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach. Bioinformatics 2004; 20: 1388-97.
[39] Mallios. Class II MHC quantitative binding motifs derived from a large molecular database with a versatile iterative stepwise discriminant analysis meta-algorithm. Bioinformatics 1999; 15: 432-9.
[40] Peters B, Tong W, Sidney J, Sette A, Weng Z. Examining the independent binding assumption for binding of peptide epitopes to MHC-I molecules. Bioinformatics 2003; 19: 1765-72.
[41] Presnell SR, Cohen FE. Artificial neural networks for pattern recognition in biochemical sequences. Annu Rev Biophys Biomol Struct 1993; 22: 283-98.
[42] Adams HP, Koziol JA. Prediction of binding to MHC class I molecules. J Immunol Methods 1995; 185: 181-90.
[43] Gulukota K, Sidney J, Sette A, DeLisi C. Two complementary methods for predicting peptides binding major histocompatibility complex molecules. J Mol Biol 1997; 267: 1258-67.
[44] Milik M, Sauer D, Brunmark AP, Yuan L, Vitiello A, Jackson MR, *et al*. Application of an artificial neural network to predict specific class I MHC binding peptide sequences. Nat Biotechnol 1998; 16: 753-6.
[45] Honeyman MC, Brusic V, Stone NL, Harrison LC. Neural network-based prediction of candidate T-cell epitopes. Nat Biotechnol 1998; 16: 966-9.
[46] Brusic V, Rudy G, Honeyman JH, Harrison LC. Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neuronal network. Bioinformatics 1998; 14: 121-30.
[47] Byvatov E, Schneider G. Support vector machine applications in bioinformatics. Appl Bioinformatics 2003; 2: 67-77.
[48] Bhasin M, Reinherz EL, Reche PA. Recognition and classification of histones using support vector machine. J Comput Biol 2006; 13: 102-12.
[49] Bhasin M, Zhang H, Reinherz EL, Reche PA. Prediction of methylated CpGs in DNA sequences using a support vector machine. FEBS Lett 2005; 579: 4302-8.
[50] Donnes P, Elofsson A. Prediction of MHC class I binding peptides, using SVMHC. BMC Bioinformatics 2002; 3: 25.
[51] Bhasin M, Raghava GP. SVM based method for predicting HLA-DRB1*0401 binding peptides in an antigen sequence. Bioinformatics 2004; 20: 421-3.
[52] Tung CW, Ho SY. POPI: predicting immunogenicity of MHC class I binding peptides by mining informative physicochemical properties. Bioinformatics 2007; 23: 942-9.
[53] Salomon J, Flower DR. Predicting Class II MHC-Peptide binding: a kernel based approach using similarity scores. BMC Bioinformatics 2006; 7: 501.
[54] Jacob L, Vert JP. Efficient peptide-MHC-I binding prediction for alleles with few known binders. Bioinformatics 2008; 24: 358-66. Epub 2007 Dec 14.
[55] Kingsford C, Salzberg SL. What are decision trees? Nat Biotechnol 2008; 26: 1011-3.
[56] Savoie CJ, Kamikawaji N, Sasazuki T, Kuhara S. Use of BONSAI decision trees for the identification of potential MHC class I peptide epitope motifs. Pac Symp Biocomput 1999: 182-9.
[57] Segal MR, Cummings MP, Hubbard AE. Relating amino acid sequence to phenotype: analysis of peptide-binding data. Biometrics 2001; 57: 632-42.
[58] Zhu S, Udaka K, Sidney J, Sette A, Aoki-Kinoshita KF, Mamitsuka H. Improving MHC binding peptide prediction by incorporating binding data of auxiliary MHC molecules. Bioinformatics 2006; 22: 1648-55. Epub 2006 Apr 13.
[59] Schuster-Bockler B, Bateman A. An introduction to hidden Markov models. Curr Protoc Bioinformatics 2007; Appendix 3A.
[60] Rose RC, Juang BH. Hidden Markov models for speech and signal recognition. Electroencephalogr Clin Neurophysiol Suppl 1996; 45: 137-52.
[61] Mamitsuka H. Predicting peptides that bind to MHC molecules using supervised learning of hidden Markov models. Proteins 1998; 33: 460-74.
[62] Noguchi H, Kato R, Hanai T, Matsubara Y, Honda H, Brusic V, *et al*. Hidden Markov model-based prediction of antigenic peptides that interact with MHC class II molecules. J Biosci Bioeng 2002; 94: 264-70.
[63] Zhang C, Bickis MG, Wu FX, Kusalik AJ. Optimally-connected hidden markov models for predicting MHC-binding peptides. J Bioinform Comput Biol 2006; 4: 959-80.
[64] Wistrand M, Sonnhammer EL. Improved profile HMM performance by assessment of critical algorithmic features in SAM and HMMER. BMC Bioinformatics 2005; 6: 99.
[65] Durbin R, Eddy S, Krogh A, Mitchison G. Biological sequence analysis: probabilistic models of proteins and nucleic acids, Cambridge University Press: Cambridge 1998.
[66] Zhang GL, Khan AM, Srinivasan KN, August JT, Brusic V. MULTIPRED: a computational system for prediction of promiscuous HLA binding peptides. Nucleic Acids Res 2005; 33: W172-9.
[67] Parker KC, Bednarek MA, Coligan JE. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individuial peptide side chains. J Immunol 1994; 152: 163-75.
[68] Stryhn A, Pedersen LO, Romme T, Holm CB, Holm A, Buus S. Peptide binding specificity of major histocompatibility complex class I resolved into an array of apparently independent subspeci-

ficities: quantitation by peptide libraries and improved prediction of binding. Eur J Immunol 1996; 26: 1911-8.

[69] Udaka K, Wiesmuller KH, Kienle S, Jung G, Tamamura H, Yamigishi H, *et al.* An automated prediction of MHC class I-binding peptides based on positional scanning with peptide libraries. Immunogenetics 2000; 51: 816-828.

[70] Bui HH, Sidney J, Peters B, Sathiamurthy M, Sinichi A, Purton KA, *et al.* Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications. Immunogenetics 2005; 57: 304-14.

[71] Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BMC Bioinformatics 2005; 6: 132.

[72] Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. BMC Bioinformatics 2007; 8: 238.

[73] Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, Sahin U, *et al*. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat Biotech 1999; 17: 555-61.

[74] Bian H, Hammer J. Discovery of promiscuous HLA-II-restricted T cell epitopes with TEPITOPE. Methods 2004; 34: 468-75.

[75] Doytchinova IA, Blythe MJ, Flower DR. Additive method for the prediction of protein-peptide binding affinity. Application to the MHC class I molecule HLA-A*0201. J Proteome Res 2002; 1: 263-72.

[76] Doytchinova IA, Flower DR. Towards the in silico identification of class II restricted T-cell epitopes: a partial least squares iterative self-consistent algorithm for affinity prediction. Bioinformatics 2003; 19: 2263-70.

[77] Guan P, Doytchinova IA, Zygouri C, Flower D. MHCPred: A server for quantitative prediction of peptide-MHC binding. Nucleic Acids Res 2003; 31: 3621-4.

[78] Liu W, Meng X, Xu Q, Flower DR, Li T. Quantitative prediction of mouse class I MHC peptide binding affinity using support vector machine regression (SVR) models. BMC Bioinformatics 2006; 7: 182.

[79] Buus S, Lauemoller SL, Worning P, Kesmir C, Frimurer T, Corbet S, *et al.* Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach. Tissue Antigens 2003; 62: 378-84.

[80] Nielsen M, Lundegaard C, Worning P, Lauemoller SL, Lamberth K, Buus S, *et al*. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci 2003; 12: 1007-17.

[81] Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, *et al.* NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. PLoS ONE 2007; 2: e796.

[82] Nielsen M, Lundegaard C, Blicher T, Peters B, Sette A, Justesen S, *et al.* Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. PLoS Comput Biol 2008; 4: e1000107.

[83] Doytchinova IA, Flower DR. Physicochemical explanation of peptide binding to HLA-A*0201 major histocompatibility complex: a three-dimensional quantitative structure-activity relationship study. Proteins 2002; 48: 505-18.

[84] Jojic N, Reyes-Gomez M, Heckerman D, Kadie C, Schueler-Furman O. Learning MHC I--peptide binding. Bioinformatics 2006; 22: e227-35.

[85] Rognan D, Scapozza L, Folkers G, Daser A. Molecular dynamics simulation of MHC-peptide complexes as a tool for predicting potential T cell epitopes. Biochemistry 1994; 33: 11476-85.

[86] Altuvia Y, Sette A, Sidney J, Southwood S, Margalit H. A structure-based algorithm to predict potential binding peptides to MHC molecules with hydrophobic binding pockets. Hum Immunol 1997; 58: 1-11.

[87] Miyazawa S, Jernigan RL. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. J Mol Biol 1996; 256: 623-44.

[88] Zhao B, Mathura VS, Rajaseger G, Moochhala S, Sakharkar MK, Kangueane P. A novel MHCp binding prediction model. Hum Immunol 2003; 64: 1123-43.

[89] Schueler-Furman O, Altuvia Y, Sette A, Margalit H. Structure-based prediction of binding peptides to MHC class I molecules: application to a broad range of MHC alleles. Protein Sci 2000; 9: 1838-46.

[90] Mohanapriya A, Lulu S, Kayathri R, Kangueane P. Class II HLA-peptide binding prediction using structural principles. Hum Immunol 2009; 30: 30.

[91] Logean A, Sette A, Rognan D. Customized versus universal scoring functions: application to class I MHC-peptide binding free energy predictions. Bioorg Med Chem Lett 2001; 11: 675-9.

[92] Tong JC, Tan TW, Ranganathan S. Modeling the structure of bound peptide ligands to major histocompatibility complex. Protein Sci 2004; 13: 2523-32.

[93] Bui HH, Schiewe AJ, von Grafenstein H, Haworth IS. Structural prediction of peptides binding to MHC class I molecules. Proteins 2006; 63: 43-52.

[94] Fagerberg T, Cerottini JC, Michielin O. Structural prediction of peptides bound to MHC class I. J Mol Biol 2006; 356: 521-46. Epub 2005 Dec 5.

[95] Knapp B, Omasits U, Frantal S, Schreiner W. A critical cross-validation of high throughput structural binding prediction methods for pMHC. J Comput Aided Mol Des 2009; 5: 5.

[96] Bordner AJ, Abagyan R. Ab initio prediction of peptide-MHC binding geometry for diverse class I MHC allotypes. Proteins 2006; 63: 512-26.

[97] Sette A, Sidney J. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. Immunogenetics 1999; 50: 201-12.

[98] Sette A, Sidney J. HLA supertypes and supermotifs: a functional perspective on HLA polymorphism. Curr Opin Immunol 1998; 10: 478-82.

[99] Reche PA, Reinherz EL. Artificial Immune Systems.; Catania, Italy, Springer-Verlag Berling Heidelberg 2004; vol. LNCS 3239: pp. 189-96.

[100] Reche PA, Reinherz EL. Definition of MHC supertypes through clustering of MHC peptide-binding repertoires. Methods Mol Biol 2007; 409: 163-73.

[101] Lund O, Nielsen M, Kesmir C, Petersen AG, Lundegaard C, Worning P, *et al.* Definition of supertypes for HLA molecules using clustering of specificity matrices. Immunogenetics 2004; 55: 797-810. Epub 2004 Feb 13.

[102] Doytchinova IA, Guan P, Flower DR. Identifiying human MHC supertypes using bioinformatic methods. J Immunol 2004; 172: 4314-23.

[103] Doytchinova IA, Guan P, Flower DR. Quantitative structure-activity relationships and the prediction of MHC supermotifs. Methods 2004; 34: 444-53.

[104] Doytchinova IA, Flower DR. In silico identification of supertypes for class II MHCs. J Immunol 2005; 174: 7085-95.

[105] Kangueane P, Sakharkar MK, Rajaseger G, Bolisetty S, Sivasekari B, Zhao B, *et al.* A framework to sub-type HLA supertypes. Front Biosci 2005; 10: 879-86.

[106] Yu K, Petrovsky N, Schonbach C, Koh JY, Brusic V. Methods for prediction of peptide binding to MHC molecules: a comparative study. Mol Med 2002; 8: 137-48.

[107] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997; 25: 3389-402.

[108] Peters B, Bui HH, Frankild S, Nielson M, Lundegaard C, Kostem E, *et al.* A community resource benchmarking predictions of peptide binding to MHC-I molecules. PLoS Comput Biol 2006; 2: e65. Epub 2006 Jun 9.

[109] Lin HH, Zhang GL, Tongchusak S, Reinherz EL, Brusic V. Evaluation of MHC-II peptide binding prediction servers: applications for vaccine research. BMC Bioinformatics 2008; 9: S22.

[110] Wang M, Lamberth K, Harndahl M, Roder G, Stryhn A, Larsen MV, *et al*. CTL epitopes for influenza A including the H5N1 bird flu; genome-, pathogen-, and HLA-wide screening. Vaccine 2007; 25: 2823-31.

[111] Zhong W, Reche PA, Lai CC, Reinhold B, Reinherz EL. Genome-wide Characterization of a Viral Cytotoxic T Lymphocyte Epitope Repertoire. J Biol Chem 2003; 278: 45135-44.

[112] Reche PA, Reinherz EL. PEPVAC: a web server for multi-epitope vaccine development based on the prediction of supertypic MHC ligands. Nucleic Acids Res 2005; 33: W138-42.

[113] Reche PA, Zhang H, Glutting JP, Reinherz EL. EPIMHC: a curated database of MHC-binding peptides for customized computational vaccinology. Bioinformatics 2005; 21: 2140-1.

[114] Bhasin M, Raghava GP. A hybrid approach for predicting promiscuous MHC class I restricted T cell epitopes. J Biosci 2007; 32: 31-42.

[115] Zhang Q, Wang P, Kim Y, Haste-Andersen P, Beaver J, Bourne PE, *et al*. Immune epitope database analysis resource (IEDB-AR). Nucleic Acids Res 2008; 36: W513-8.

[116] Singh H, Raghava GP. ProPred1: prediction of promiscuous MHC Class-I binding sites. Bioinformatics 2003; 19: 1009-14.

[117] Singh H, Raghava GP. ProPred: prediction of HLA-DR binding sites. Bioinformatics 2001; 17: 1236-7.

[118] Wan J, Liu W, Xu Q, Ren Y, Flower DR, Li T. SVRMHC prediction server for MHC-binding peptides. BMC Bioinformatics 2006; 7: 463.

[119] Swain MT, Brooks AJ, Kemp GJL. In Proceedings of the 2nd IEEE International Symposium on Bioinformatics and Bioengineering; IEEE Computer Society: Los Alamitos, 2001; pp. 81-8.

[120] Bhasin M, Singh H, Raghava GP. MHCBN: a comprehensive database of MHC binding and non-binding peptides. Bioinformatics 2003; 19: 665-6.

[121] Toseland CP, Clayton DJ, McSparron H, Hemsley SL, Blythe MJ, Paine K, *et al*. AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. Immunome Res 2005; 1: 4.

[122] Nicholls A, Sharp K, Honig B. Protein folding and association insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 1991; 11: 281-96.

[123] Engelhardt B. The blood-central nervous system barriers actively control immune cell entry into the central nervous system. Curr Pharm Des 2008; 14(16): 1555-65.

[124] Gueant JL, Gueant-Rodriguez RM, Gastin IA, Cornejo-Garcia JA, Viola M, et al. Pharmacogenetic determinants of immediate and delayed reactions of drug hypersensitivity. Curr Pharm Des 2008; 14(27): 2770-7.

[125] Kanduc D. Correlating low-similarity peptide sequences and allergenic epitopes. Curr Pharm Des 2008; 14(3): 289-95.

Received: May 6, 2009 **Accepted: May 7, 2009**