Rev. Med. Virol. 2009; **19**: 77–96.
Published online 19 December 2008 in Wiley InterScience
(www.interscience.wiley.com)
DOI: 10.1002/rmv.602

REVIEW

An introduction to epitope prediction methods and software

Xingdong Yang and Xinglong Yu*
Department of Veterinary Medicine, Hunan Agricultural University, Changsha, Hunan, P.R. China

SUMMARY

In this paper, current prediction methods and algorithms for both T- and B cell epitopes are reviewed, and a comprehensive summary of epitope prediction software and databases currently available online is also provided. This review can offer researchers in this field a sense of direction and insights for future work. However, our main purpose is to introduce clinical and basic biomedical researchers who are not familiar with these biological analysis tools and databases to these online resources and to provide guidance on how to use them effectively. Copyright © 2008 John Wiley & Sons, Ltd.

Received: 16 October 2008; Accepted: 12 November 2008

INTRODUCTION

Epitopes are of particular interest to both clinical and basic biomedical researchers as they hold huge potential for vaccine design, disease prevention, diagnosis and treatment. Vaccine development in the past depends exclusively on biochemical and immunological experiment, such as phage display library, overlapping peptides, ELISA, NMR, immunoflurorescence, radioimmunoassay, Western blotting, immunohistochemistry, X-ray crystallography studies of antibody/antigen structure and attenuation of the wild type pathogens by random mutations and serial passages, which is very expensive, time-consuming, with low immunogenicity and reversible. Now, with the aid of epitope predictive software and databases, we are able to narrow protein of our interest, and drastically reduce the number of wet experiments. Effective individual vaccine designs are made cheaper and faster. Given the potential importance of epitope identification in developing vaccines against infectious, immune and other antigen-related diseases, epitopes are studied widely by researchers in various fields, and a large expansion of databases, predictive methods and software focussing on different types of epitopes has been witnessed. The average immunologists are overwhelmed with such a broad array of immunological analysis tools that are highly specific in use, not well understood or defined, tested on limited data and not publicly accessible [1]. There have been some reviews on computational methods and tools for MHC/peptide interaction studies [2–5] and B epitope prediction methods [5,6], but no work has been done on reviewing currently available methods, software and databases for both T- and B cell epitope prediction. Our primary purpose is to facilitate biomedical researchers' work by familiarising them with the essential updated internet resources on epitopes and assist them make informed selection among these diversely available tools.

Corresponding author: Xinglong Yu, Department of Veterinary Medicine, Hunan Agricultural University, Changsha, Hunan 410128, P.R. China. E-mail: xlyu999@yahoo.com.cn

EPITOPE PREDICTION METHODS

Antigenic determinants or epitopes are the regions of an antigen that bind to antigen-specific membrane receptors on lymphocytes or to secreted antibodies, thereby eliciting either cellular or humoral immune response. Epitopes can be classified into continuous (or linear) and discontinuous (or conformational) epitopes. Continuous epitopes are linear peptide fragments which are usually amphipathic helical 9–12 mers and recognised by Th cells or TCRs (T cell receptors). Prior to MHC binding and presentation to TCRs and Th cells, antigens must be processed into peptide fragments or epitopes through different antigen processing pathways, mostly being cleaved by proteasomes,

Copyright © 2008 John Wiley & Sons, Ltd.

although protein splicing is also observed [7], and there is about 0.05% chance of immunogenicity for every antigen [8]. Discontinuous epitopes are structurally more complicated, nonlinear and discrete 15–22 mers, aggregated together due to folding in natural protein, which can be recognised by both B cells and Th cells in their native structure. According to their respective receptors, epitopes are categorised into B- and T cell epitopes. B cell epitopes contain both continuous (~10%) and discontinuous (~90%) epitopes, while the majority of T cell epitopes are continuous. B cell epitopes are recognized by B cell receptors or antibodies in their native structure. Consequently, there has been a considerable amount of interest in studying Ag-Ab binding site topography and the amino acids residues in CDRs, or Specificity Determining Residues (SDRs) that directly interact with antigens [9,10]. The number of epitopes is enormous, with discontinuous epitopes estimated at about >> 10^{11}, continuous epitopes composed of nine amino acids > 10^{11}, and a comparable amount of T cell receptors [11], MHC haplotypes, combinatorial antibodies and B cell clonotypes [12]. B cell epitopes can also be divided into three categories based on their immunogenic potency (antibody level increase after administration of vaccines): immunodominant (two–threefold), immunogenic (onefold) and non-immunogenic (zerofold).

Epitope prediction dates back to 1981 when the first B cell epitope prediction method was developed by Hopp and Woods. Since then many more methods have been developed or adapted from other computational tools; for example B cell epitope prediction [13–16] and T cell epitope prediction [17–20]. Despite the early start, however, prediction systems for B cell epitopes are still in their infancy. Attention is, however, shifting to discontinuous B cell epitope prediction. For T cell epitope prediction, MHC I binding predictions are now very strong and have wide allelic coverage by integration with predictions of proteasomal cleavage and TAP binding sites. MHC II binding predictions are not as well developed as MHC I binding predictions, but are progressing at a fast pace.

General epitope prediction methods

Currently, four approaches are employed to predict epitopes: sequence-based methods, structure-based methods, hybrid methods and consensus methods. While prediction accuracy for hybrid methods is higher than sequential and structural approaches, consensus methods give the best prediction results.

Sequence-based epitope prediction

Sequence-based method utilises the notion that sequence dictates structure and identical structure in turn leads to identical functions. T cell epitopes have a common sequence pattern or motif, as well as MHC allele specificity determining subpatterns. The regions recognised by T cells on an antigenic protein are largely distinct from regions recognised by antibodies, and correspond to helices in the intact protein or can be modelled into amphipathic helices. To make useful, informative epitope prediction, epitope physicochemical properties are also used, such as exposed surface, accessibility, flexibility, hydrophilicity, charge, number of proline residues, the proximity of the segment towards the C- or N-terminal of the protein, etc. Due to the enormous number of physicochemical properties that are associated with epitopes, simpler quantitative descriptors of amino acid properties are sometimes used to simplify computation.

Techniques, such as binding motifs, quantitative matrices (QM), virtual matrices, machine learning algorithms (ANN, HMM, SVM), evolutionary algorithms [21], linear programming [22], etc. are used to identify the binding peptide. They all have their relative advantages and disadvantages [23]. For example, in a comparative study, Yu *et al*. suggested that motifs give the most accurate MHC-peptide binding predictions with a limited dataset, but as the data volume increases, machine learning predictions become more reliable [24]. Among machine learning methods, Bahsin and Ragahava have compared QM, SVM (support vector machine) and ANN (artificial neural network) techniques, and found that SVM techniques performed best [25]. SVM is particularly attractive to biological analysis, due to its ability to filter noise, large input spaces and good performance [27]. BLAST and FASTA are common sequence–sequence comparison methods used to find sequence homologues and assign amino acid scores by multiple alignment of query protein sequence with template sequences in a database.

However, prediction methods based on protein sequence analysis are not preferred because

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/rmv
Petitioner Merck, Ex. 1171, p. 78
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00024

some experiments have shown that proteins with high sequence identity may possess differential functions due to folding, interactions among variable regions in the sequence and different cellular compartmentisation [28]; and large datasets are required for training.

Structure-based epitope prediction

The structure-based prediction model bases on 3D protein structure to screen potential binders. Structural similarity between query protein and template proteins are used to predict epitopes of interest. Methods utilising structure in predicting sequential and discontinuous epitopes is new when compared to sequence-based approach. Existing structure-based epitope identifying techniques include mutagenesis, competition experiments, free energy scoring function, knowledge-based free energy scoring, protein threading [29,30], homology modelling [31], virtual pockets, rigid/flexible docking [32–36] and atomistic molecular dynamics simulations. The molecular dynamics modelling technique is fairly new and not many prediction methods and software have been based on it.

The structure-based epitope prediction approach has several advantages over the sequence-based one. Firstly, a smaller dataset is used for training. Secondly, it can predict candidate peptides for alleles that have not been extensively studied, and where sequence-based approaches have failed or were not attempted. Thirdly, it is only possible to predict discontinuous epitopes through structure. Lastly, even sequence-based approaches depend on structure to make reliable predictions. As a result, structural-based approaches give more accurate prediction. However, structure-based approach has been greatly limited due to high computational cost, development complexity and scarcity of 3D protein structures.

Hybrid prediction methods: combining sequential with structural analysis

Given the poor performance of epitope predictors based on sequence or structure analysis, it is clear that any single method cannot accurately predict epitopes. Consequently, some researchers turned to building predictive methods taking advantage of both sequential and structural information. For example, a new method which integrates 3D protein structure with physicochemical properties of amino acids using machine learning methods like Hidden Markov Model (HMM), supporting vector machine (SVM), ANN, etc. improved the prediction precision to a though small but significant degree [37]. Like structural-based approach, its further development is hampered by the limited availability of 3D structure data of antigens and true negative datasets, both to construct better predictors and evaluate the algorithms. There is also the possibility of false positives because different antibodies have overlapping binding sites.

Consensus from multiple methods

To gain even higher sensitivity and specificity, consensus method is utilised to combine strengths of various methods. Table 1 illustrates how a consensus method works. Theoretically, this is simple and easy to understand. In practice, many problems exist with this approach, such as computer computational limitations, compromised prediction speed, different output formats and the difficulty of integrating prediction results into the final consensus result. For example, a sequence segment is given a high score, but is predicted fewer times as an epitope, it is hard to decide whether high score or more prediction times should be included in the generation of consensus output. However, despite the increased difficulties in constructing a consensus method, a few consensus prediction software with increased epitope prediction accuracy are available online.

T cell epitope prediction

T cell epitopes are largely predicted indirectly by identifying MHC binders. MHC binding prediction consists of MHC I and II binding prediction. Prediction methods for MHC I and II are distinct from each other, as MHC I binding grove is closed, while the binding grove for MHC II is open at both ends which can accommodate peptides of variable length, typically 13–25 amino acids long (see Figure 1). As a result, MHC II binding prediction have much lower accuracy rate than that of MHC I binders. The predicted MHC II binding cores are often extended at both ends to obtain an effective T cell epitope. T cell epitope prediction methods and algorithms can be

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/rmv
Petitioner Merck, Ex. 1171, p. 79
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00024

Table 1. Consensus method for epitope prediction

Prediction methods	M 1	M 2	M 3	M 4	Consensus Method
Prediction result	Non-epitope	Non-epitope	Non-epitope	Non-epitope	0% epitope
	Epitope	Non-epitope	Non-epitope	Non-epitope	25% epitope
	Epitope	Epitope	Non-epitope	Non-epitope	50% epitope
	Epitope	Epitope	Epitope	Non-epitope	75% eitope
	Epitope	Epitope	Epitope	Epitope	100% epitope

roughly grouped into several distinct types: binding motifs, QM, free energy scoring functions (Fresno), machine learning techniques (ANN, HMM, SVM, etc.), MHC peptide threading, 3D-QSAR (three-dimensional quantitative structure and activity relations), and molecular modeling [38]. For a comparison of T cell epitope prediction methods, see Table 2. Although prediction accuracy are improving, these indirect methods still cannot give highly reliable predictions, as some MHC binders do not elicit immune response [26]. Moreover, not all predictors employing these methods have the uniform specificity and sensitivity, depending on the data selection for model building, training, validation, epitope types and users' ability to correctly use these predictors. In a bid to build better prediction tools and help users use these software, some researchers have made tentative attempts to establish a framework for building prediction tools from these methods [39].

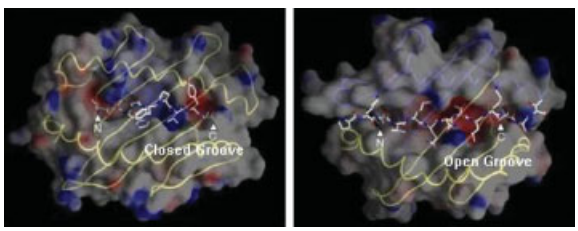

Figure 1. MHC I and II molecules. Left. HLA-A*0201 (MHC I molecule) in complex with a peptide LLFGYPVYV from HTLV-1 TAX protein (PDB: 1HHK). Right. HLA-DR1 (MHC II molecule) in complex with peptide PKYVKQNTLKLAT from Influenza A virus (PDB:1FYT). Note that the binding groove for MHC I molecule on the left is closed at both ends and therefore the peptides it is capable of presenting are short (nine residue-peptide in the closed groove above). On the right, MHC II binding groove is open at both ends which allows it to bind long peptides (13 residue-peptide in the open groove above), complicating MHC II binding prediction. The N and C termini of the binding peptides are marked by white triangular arrows

B cell epitope prediction

Existing methods used to predict B cell epitopes are: for continuous B cell epitopes, Hopp and Woods (1981) [40], Kyte and Doolittle (1982), Welling *et al*. (1985) [41], Parker *et al*. (1986) [42], Kolaskar and Tongaonkar (1990) [43], PEOPLE (Alix, 2000) [44], and for discontinuous B cell epitopes, Kolaskar and Urmila Kulkarni (1999), Disctope, etc. Continuous B cell epitope prediction is very similar to T cell epitope prediction, which has mainly been based on the amino acid properties, such as hydrophilicity, charge, exposed surface area, secondary structure, etc. Discontinuous B cell epitope prediction requires 3D structure of Ag-Ab complex. The binding of antigen to antibody is shown in Figure 2. Currently, only a few methods and prediction tools for discontinuous B cell epitopes are available, and their performances are very poor, with an area under the Receiver Operating Characteristic curve value (ROC) of approximately 0.6 [37]. B cell epitope prediction methods are compared in Figure 3. The poor performance of current prediction methods, which is a reflection of the generality of antigenicity, has prompted some researchers to doubt the existence of epitope as a discrete and intrinsic entity of protein [45]. Assuming epitopes exist, one reason for the intractable problem in predicting B cell epitopes is that B cell epitopes are mostly discontinuous, enormously diverse and existing computational tools for analysis of sequences are not applicable. Another reason is due to small training database. The prevalent low prediction accuracy rate for discontinuous B cell epitopes is not hard to explain when one considers the ratio of defined

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/rmv
Petitioner Merck, Ex. 1171, p. 80
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00024

Table 2. Comparison of T cell epitope prediction methods

Method	Training dataset			Prediction dataset		ROC analysis						Correlation coefficient
	Allele	Binder	Non-binders	Allele	Binders	PPV	NPV	AUC	Sensitivity	Specificity	Accuracy	
3D QSAR (Flower, 2001) [137]	A*0201	84	18	A*0201	50	—	—	—	—	—	—	$r^2 = 0.679$
3D Additive (Kangueane, 2003) [138]	8 alleles	—	—	> 50	2660	89%	18%	—	50–73%	52–58%	60%	—
ANN (Adams, 1995) [139]	A*0201	1038	—	A*0201	—	—	—	0.78	—	—	—	—
ANN + QM (Bhasin, 2007) [140]	30 alleles	—	—	—	—	94.0%	—	—	91.8%	94.9%	93.6%	—
BIMAS (Parker, 1994) [62]	HLA-A2	—	154	HLA-A2	—	—	—	—	—	—	—	—
CCLD (Zeng, 2001) [141]	A*0201	—	—	—	—	—	—	—	—	—	—	—
DynaPred (Antes, 2006) [142]	A*0201	—	—	A*0201	383	—	—	0.92	90%	83%	87%	—
EA-ANN (Brusic, 1998) [21]	DRB1*0401	338	312	B1*0401	62	—	—	—	—	—	—	—
EpiDock (Rognan, 2002) [143]	A*0201	—	—	A*0201, B*2705		—	—	—	80–100%	76–85%	75–85%	—
FESF (Rognan, 1999) [34]	A*0201, H-2^k			A*0204, H-2^k	24, 16	—	—	—	—	—	—	—
HMM (Brusic, 2002) [24]	HLA-A2	—	—	4 alleles	—	—	—	—	—	—	—	—
HMM-SMM (Kobayashi, 2002) [144]	DRB1*0101	305	472	DRB1*0101	—	—	—	0.89	—	—	—	—
Pockets (Hammer, 1999) [145]	11 HLA-DR	—	—	51 alleles	223	—	—	—	—	—	80%	—
SDS (Mallos, 1999) [146]	DRI	526	98	DRI	—	—	—	—	96.8%	73.5%	93.1%	—
SVM (Cul, 2007) [147]	30 alleles	7460	1591	30 alleles	1021	—	—	—	—	—	86–99%	—
SVMHC (Elofsson, 2002) [148]	26 alleles	—	—	6 alleles	—	—	—	—	—	—	91–100%	—
Threading (Margalt, 2000) [149]	10 alleles	123	—	—	106	—	—	—	—	—	71%	− 0.61 to 0.07

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/rmv
Petitioner Merck, Ex. 1171, p. 81
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00024

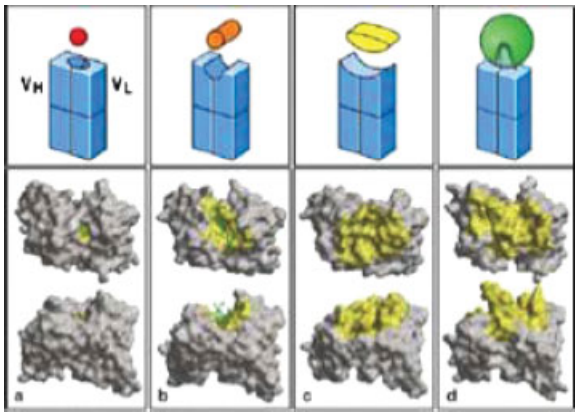

Figure 2. Antigens bind with antibody. a. Antigen binds to the pocket formed by the various regions of the heavy chain (V_H) and light chain (VL). b and c. Antigens bind to the grooves on the antibody. d. Antigens can also bind antibody through extended surface on the light and heavy chains. These binding properties for antibodies complicate B cell epitope prediction by requiring 3D protein structures. (Modified from Janeway, 2005)

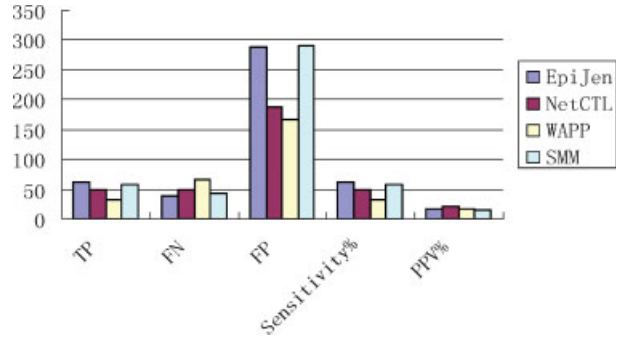

Figure 3. Comparison of B cell epitope prediction methods. The numbers on the *x*-axis refer to different B cell epitope prediction methods: 1 Hydrophilicity (Parker *et al.*, 1986); 2 Accessibility (Emini *et al.*, 1985); 3 Flexibility (Karplus and Schulz, 1985); 4 Surface (Janin and Wodak, 1978); 5 Polarity (Ponnuswamy *et al.*, 1980); 6 Turns (Pellequer *et al.*, 1998); 7 Antigenic Scale (Koloaskar and Tongaonkar, 1990); 8 [3] + [1]; 9 [3] + [1] + [5]; 10 [3] + [1] + [5] + [4]; 11 [3] + [1] + [5] + [4] + [6]; 12 [3] + [1] + [5] + [4] + [6] + [2]. [3] + [1] means method combining Flexibility and Hydrophilicity scales in their prediction. Prediction methods after 6 have better performance than the other six single scale-based methods. This shows that B cell epitope prediction performance can be increased by combining several methods. However, different epitope prediction software employing the same method from above may possess different performance due to training datasets, thresholds and other parameters chosen subjectively by both the software developers and users. (All methods above are evaluated on a single dataset by Sudipto Saha *et al.*)

epitopes to undefined epitopes. It has been estimated that there are about one trillion (10^{12}) antibodies in our body, excluding T cell receptors, hence a comparable number of comformational epitopes exist. But the current epitope database contains only several thousands of entries of Ag-Ab complexes, which are well defined and validated by experiments, and the various prediction methods are developed by analysing known epitope structures, or trained and validated using these databases.

Despite the doubt on the existence of epitopes, recently, there is a suggestion that continuous B cell epitopes can be classified into two classes (classes 1 and 2) based on the protectivity they elicit, which should be predicted separately [46]. Classes 1 and 2 protectivity are defined as inhibition of antigen function or pathogenic organism activity and destruction of organism, respectively. This is very similar to MHC I and II epitopes, in terms of their function. Molecular and mathematical modelling of antigen–antibody interaction, simulation of immune system and protein–protein interaction study are developing at a fantastic speed [53–55]. Large, special benchmarking databases of well-defined B cell epitopes are being constructed. These developments hold potential for further improving B cell epitope prediction sensitivity.

Performance evaluation of epitope prediction methods

To make prediction results valid, and to assess and compare the performance of various epitope prediction methods and software, ROC analysis is used. An ROC analysis includes PPV, NPV, AUC, specificity, sensitivity and accuracy. AUC analysis is often used. By selecting a threshold between the maximum and minimum value, the distribution of positive and negative prediction results are divided into two areas: area above and below the threshold. Positives above the threshold are considered true positives and false negatives are negatives in this area. Then an AUC curve is then obtained by plotting true positives on the *y*-axis against false positives on the *x*-axis. The area under the rectangular hyperbola curve represents specificity or sensitivity of prediction. To make evaluation more accurate, cross-validation or jacknife method is usually used [47], and then an average AUC value is calculated from many AUC values obtained. An AUC value of 0.5 is considered equal to random guessing, while an AUC value of 1.0 corresponds to

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/rmv
Petitioner Merck, Ex. 1171, p. 82
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00024

perfect prediction. To make prediction result useful, an empirical AUC value of at least 0.7 is required.

EPITOPE PREDICITON TOOLS

Antigenicity or epitope prediction approaches can be classified into the following categories: prediction of TAP binders, prediction of proteasomal cleavage sites, prediction of MHC binding regions and prediction of T- and B cell epitopes. Correspondingly, there are epitope predictors and databases focussing on one or several steps of this multi-step process of epitope generation. Epitope prediction complements wet bench experiment, and can greatly enhance research by reducing time, money and complexity of protein antigenicity study. So, it is desirable for researchers outside this field to possess adequate knowledge on how to use these tools to their maximum capacity. In this section, we intend to introduce the commonly used tools for T- and B cell epitope prediction.

T cell epitope predictive software and databases

With the number of MHC binding predictive software expanding rapidly, many studies are focussed on the comparison and evaluation of these software, for MHC I binding [24,48–51] and MHC II binding [52,56–58]. In short, these predictors achieved impressive prediction accuracies of 70–90% for binders but only 40–80% for non-binders, with limited range of MHC allele coverage and fixed peptide length [59]. T cell epitope predictive software with web interfaces publicly available online are listed in Table 3.

Basically, the majority of these software and databases are user friendly, and their outputs are straightforward. But to get maximum performance from these tools, one needs expert knowledge in this field. For those who are not specialised in this realm, tutorials are provided on these tools' accompanying websites and by effectively selecting and combining prediction results from various tools, a satisfactory prediction rate can be achieved. For example, to predict T cell epitopes from a known antigenic protein sequence, a comparison and combination of prediction results from software predicting MHC binding, TAP binding, proteasomal cleavage sites, even protein structure, will yield better results than a single or several randomly chosen prediction software. A study has shown that a heuristic-based method of combining the results of individual tools for predicting MHC I binders obtained more accurate prediction rate than any single tool [47].

Software utilising more than one method

Software that combine two or more methods include nHLAPred (QM + ANN for MHC I epitope), AntiBP (ANN + QM + SVM for antibacterial peptides), CTLpred (QM + ANN + SVM for the prediction of CTL epitopes instead of MHC I binders), IEDB binding I (ANN + ARB + SMM for MHC I binding peptides) IEDB binding II (consensus method + ARB + SMM + Sturniolo for MHC II binders), WAPP (QM + SVM + SVMHC for proteasomal cleavage, TAP binding and MHC binding prediction) and NetMHC (a combination of sparse encoding, Blosum encoding, and input derived from HMMs for MHC I epitope prediction).

Software for MHC I binding prediction

MHC I epitope prediction is very strong. The fact that most software includes this function is not unexpected. Software which are exclusively focussed on MHC I epitopes include BIMAS, CTLpred, IEDB, MHC I, MAPPP, ProPred-1, PREDEP, to name just a few. BIMAS is a software utilising a method to predict the relative binding strengths of all possible nonapeptides to the MHC I molecule HLA-A2 based on experimental peptide binding data. Side chain of the peptide contributes a certain amount to the stability of the HLA-A2 complex that is independent of the sequence of the peptide. To quantify these contributions, the binding data from a set of 154 peptides were combined together to generate a table containing 180 coefficients (20 amino acids $\times$ 9 positions), each of which represents the contribution of one particular amino acid residue at a specified position within the peptide to binding to HLA-A2. BIMAS and SYFPEITHI are considered to be two of the most commonly used T cell epitope prediction tools. IEDB binding I is a MHC I binding peptides predictor using ANN, ARB and SMM. In the IEDB binding I method, SMM (stabilised matrix methods) is a model used to prediction sequence recognition specificity of epitope generation process.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/rmv
Petitioner Merck, Ex. 1171, p. 83
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00024

Table 3. T-cell epitope related prediction servers and resources

Prediction server	URL	Description	References
AntiBP	http://www.imtech.res.in/raghava/antibp/	Prediction of bacterial peptides	[60]
ARB matrix	http://epitope.liai.org:8080/tools/matrix/iedb_input?matrixClass=I,II	MHC binding peptide prediction	[61]
BIMAS	http://www-bimas.cit.nih.gov/molbio/hla_bind/	Predicts HLA/peptide half time of disassociation	[62]
CTLpred	http://www.imtech.res.in/raghava/ctlpred	Prediction of CTL epitopes based on ANN and SVM	[25]
EpiDirect	http://www.epipredict.de/index.html	MHC II restricted T cell epitopes and ligands	n/a
EpiVaxb	http://www.epivax.com	MHC I and II conserved and promiscuous epitopes	n/a
FRAGPREDICT	http://www.mpiib-berlin.mpg.de/MAPPP/cleavage.html	Prediction of proteasome cleavage sites	[63]
IEDB binding	http://www.immuneepitope.org/analyze/html/mhc_processing.html	CD8+ T cell epitopes TAP binding and cleavage sites	[64,65]
IEDB, MHC I	http://tools.immuneepitope.org/analyze/html/	MHC I binding peptides using ANN, ARB and SMM	[61,66,67]
IEDB, MHC II	http://tools.immuneepitope.org/analyze/html/mhc_II_binding.html	MHC II binding prediction	[68–70]
IMTECH	http://www.imtech.res.in/raghava/mhc/page4.html	Prediction of the MHC binding core	[71]
MAPPP	http://www.mpiib-berlin.mpg.de/MAPPP/	MHC I antigenic peptide processing prediction	[72]
MHC bench	http://www.imtech.res.in/raghava/mhcbench/	Evaluation of MHC binding predictive algorithms	n/a
MHC2Pred	http://www.imtech.res.in/raghava/mhc2pred	SVM based method for promiscuous MHC II binders	n/a
MHCPred	http://www.jenner.ac.uk/MHCPred	MHC/peptide or TAP/peptide IC50 binding values	[73–75]
MHC-THREAD	http://www.csd.abdn.ac.uk/~gjlk/MHC-Thread/	Predicts potential MHC II binding peptides	[76]
MMBPred	www.imtech.res.in/raghava/mmbpred/	Mutated high affinity and promiscuous MHC I binders	[77]
MotifScan	http://www.hiv.lanl.gov/content/immunology/motif_scan	Find HLA anchor residue motifs	n/a
NetChop	http://www.cbs.dtu.dk/services/NetChop	Cleavage sites of the human proteasomes	[78,79]
NetCTL	http://www.cbs.dtu.dk/services/NetCTL	Predicts CTL epitopes in proteins	[80]
NetMHC	http://www.cbs.dtu.dk/services/NetMHC	Binding affinities for HLA-A2 and H-2Kk	[81–83]
nHLAPred	http://www.imtech.res.in/raghava/nhlapred/	Neural network based MHC I binding prediction	[84]
PAProC	http://www.paproc.de	Prediction algorithm for proteasomal cleavages	[85]
Pcleavage	http://www.imtech.res.in/raghava/pcleavage	SVM based method for proteasome cleavage prediction	[86]
PREDEP	http://margalit.huji.ac.il	MHC I epitope prediction	n/a
ProPred	http://www.imtech.res.in/raghava/propred/	MHC II binding peptide prediction server	[87]
ProPred-1	http://www.imtech.res.in/raghava/propred1	MHC I binding peptide prediction	[88]
RANKPEP	http://immunax.dfci.harvard.edu/Tools/rankpep.html	Prediction of MHC I and II binding peptides	[89–91]
REMUS	http://140.121.196.30/remus.asp	Unique segment from related protein sequences finder	[92,93]
SMM	http://cagt.bu.edu/page/SMM_submit	Prediction of high affinity HLA-A2 binding peptides	[94]
SVMHC	http://www-bs.informatik.uni-tuebingen.de/SVMHC/	Prediction of MHC classes 1 and 2 binding peptides	[95]
SYFPEITHI	http://www.syfpeithi.de	Epitopes, binding motifs, epitope alignments for MHC	n/a
TAP Pred	http://www.imtech.res.in/raghava/tappred	Predicts binding affinity of peptide to TAP transporter	[96]
TEPITOPEb	n/a	Prediction of promiscuous MHC class 2 epitopes	[97]
WAPP	http://www-bs.informatik.uni-tuebingen.de/WAPP	Proteasomal cleavage, TAP binding and MHC I binders	[98]

n/a indicates that the relevant information are unavailable.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/rmv
Petitioner Merck, Ex. 1171, p. 84
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00024

Software for MHC II binding prediction

MHC II binding peptide and epitopes are more difficult to predict than MHC I epitopes, owning to the inherent physical and chemical properties of MHC II molecules. A few software that are built specially for MHC II binding prediction exist, such as EpiDirect, MHC-THREAD, IEDB, MHC II, MHC2Pred and ProPred. IEDB binding MHC II is for the prediction of MHC II binding peptides. The prediction method list box allows choosing between four MHC class II binding prediction methods: consensus method, ARB (average relative binding), SMM, Sturniolo. According to their website, the performance of MHC II binding predictions rank as (1) consensus, (2) SMM, (3) Sturniolo and (4) ARB. By default, the overall best method (consensus) is selected. This software requires amino acid sequences being specified in single letter code and the input format can be either space separated sequences or one continuous sequence or FASTA format.

Software for MHC I and II binding prediction

Software which can be used to predict both MHC I and II binders are usually hybrid method based, or have incorporated several methods. RANKPEP, SVMHC, ARB Matrix, AntiBP and EpiVaxb are such examples. SVMHC is a software based on SVM for both MHC I and II epitope prediction. A singular feature of SVMHC is the ability to analyse the effects of SNPs (single nucleotide polymorphism) on MHC epitopes. RANKPEP uses PSSMs (position-specific scoring matrices) or profiles for the prediction of peptide–MHC I binding as a basis for CD8 and CD4 T cell epitope identification. A feature implemented in RANKPEP is a variability masking function which focusses on the conserved sequence core rather than highly variable ones. RANKPEP can therefore be used to discover promiscuous T cell epitopes. ARB matrix method identifies MHC I and II epitopes by directly predicting IC50 values, allowing combination of searches involving different peptide sizes and alleles into a single global prediction.

Software for proteasomal cleavage sites prediction

Proteasomal cleavage sites prediction is an alternative way to predict T cell epitopes, but prediction sensitivity is not very high, since proteasomal cleavage is just the first step in epitope generation process, and during later stages, the majority of peptides will be non-MHC binders. However, they complement predictive software in other steps. NetChop FRAGPREDICT, WAPP, PAProC and Pcleavage are just a few software that fall into this category. NetChop predicts proteasomal cleavage sites and MHC I ligands using neural networks. Pcleavage is an SVM-based software for the prediction of constitutive as well as immunoproteasome cleavage sites in antigenic sequences and is currently considered to be the best method for proteasomal cleavage site prediction. Proteasomal cleavage site prediction methods are not exclusive, and can give acceptable prediction results for MHC binders and epitopes prediction as well. PAProC is a prediction tool which predicts cleavages by human and yeast proteasomes. It is also useful for immunologists working on antigen processing and the prediction of MHC I molecule ligands and CTL epitopes.

Software for TAP binding prediction

TAP transport is an inconsequential step in epitope generation process in terms of epitope prediction, since many TAP binders are found to be not naturally presented by MHC molecules. As a consequence, exclusive TAP (transporter associated with antigen processing) binding prediction software are few, although such software do exist, for example, TAP Pred. Software that possess the functionality of predicting TAP binding peptides are IEDB Binding, MHCPred, TAP Pred, etc. TAP binding prediction alone is not very informative in identifying potential epitopes, they are usually used in conjunction with other software to obtain a more reliable epitope prediction result.

Software for promiscuous epitope prediction

To achieve maximum allele and pathogen coverage and to combat mutated pathogens, it is desirable to develop vaccines or drug based on conserved epitopes or peptide cores. For this purpose, a host of software proved to be useful by predicting promiscuous epitopes, namely epitopes which can bind to many MHC alleles or contain conserved sequence or structure. Such software including, but not limited to: TEPITOPEb,

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/rmv
Petitioner Merck, Ex. 1171, p. 85
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00024

MMBPred, MHC2Pred and EpiVaxb, etc. TEPITOPE method is based on the Gibbs sampling method, and locates the binding motif in a set of sequences and characterises the MHC I and II binding motif in terms of a weight-matrix, which is then used to identify epitopes. SMM combines both AIB method and general method, which takes into account independent contributions and pairwise interactions between two positions on the peptide sequence.

Software for epitope processing prediction

There are a few software which integrate the whole process of epitope generation into one method. The performance of software based on this approach is varied, depending on the specific software incorporated (See Figure 4). In theory, this approach should be superior to all single step-based software. WAPP is just such a software, which combines proteasomal cleavage sites, TAP, and MHC binding into one system. However, it is a pity that WAPP does not include the final step of epitope generation—epitope presentation or recognition—into the prediction method. IEDB binding is a MHC I epitope predictive method which combines proteasomal cleavage, TAP transportation and MHC class I binding. This kind of software, along with consensus software, is expected to further improve epitope prediction performance.

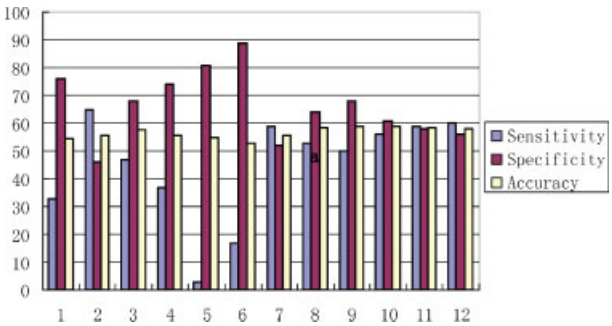

Figure 4. Performance comparison of integrated software: EpiJen, NetCTL, WAPP, SMM. These software integrate several steps of the T cell epitope generation process and are theoretically more accurate. However, from this figure, the false positive value for all four software is very high and not one software performed predominantly better than the other three. It has to be pointed out that the performance measurement data TP (true positive), FN (false negative), FP (false positive), sensitivity and PPV (positive predictive value) for each software are obtained by testing on different datasets by their respective developers. It might be interesting to compare their performance using a single benchmark dataset to eliminate biases

T cell epitope databases

T cell epitope databases are more abundant than B cell epitope databases as discontinuous epitopes are inherently more difficult to analyse. T cell epitope databases usually contain tens of thousands of epitopes, which is much larger than that of B cell epitopes. These databases can be searched or browsed for epitopes. Among these databases, IEDB (immune epitope databases and analysis resources) is one of several complete databases containing large number of epitopes, include about 63 332 peptide epitopes, 167 non-peptide epitopes and displays relatively complete information about each epitope while well cross-linked with other databases. IEDB covers a wide range of epitopes, subsuming both T- and B cell epitopes, and for a spectrum of host organisms, including humans. It can be browsed and searched by epitope structure, source and types. For epitope structure, you can input protein sequence, select protein or other non-protein epitopes. Epitope types search allows you to search T cell epitope, B cell epitope, MHC binders, MHC ligand elution, and by host organism and MHC alleles. Other T cell epitope databases can be found in Table 2.

B cell epitope prediction software and databases

Because B cell epitopes are mostly discontinuous, the prediction of which requires 3D protein structure and notoriously difficult to predict, predictive software and databases are with lower accuracy rate, in a smaller scale and fewer. Nevertheless, a few software and databases on discontinuous epitopes, including some specialised ones, are currently available. See Table 5. For more information on epitope prediction tools, website links are provided in Table 6.

Discontinuous B cell epitope prediction software

B cell epitope generation does not contain as many steps as T cell epitope generation, and they can be recognised by B cell epitope receptors on B cell surface or by antibody in their native

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/rmv
Petitioner Merck, Ex. 1171, p. 86
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00024

Table 4. T cell epitope related databases

Database	URL	Description	References
Allele frequencies	http://www.allelefrequencies.net	HLA frequencies and polymorphism frequencies	n/a
AntiJen	http://www.jenner.ac.uk/antijen	Quantitative binding data for MHC, TCR and TAP binder	[99]
dbMHC	http://www.ncbi.nlm.nih.gov/gv/mhc/main.cgi?cmd=init	DNA and clinical data related to HLA	n/a
dbMHC anthropology	http://www.ncbi.nlm.nih.gov/projects/mhc	HLA anthropology database with HLA frequencies	n/a
EPIMHC	http://immunax.dfci.harvard.edu/epimhc/	Database of MHC binding peptides and T cell epitopes	[100]
FIMM	http://sdmc.krdl.org.sg:8080/fimm	HLA, antigens, peptides and relevant disease associations	[101]
HCV immunology	http://hcv.lanl.gov/immuno/	CD8+ and CD4+ T cell HCV epitopes, proteome epitope maps	n/a
HIV immunology	http://www.hiv.lanl.gov/immunology	T cell HIV epitopes, antibody biding sites	[102]
HLA ligand/motif	http://hlaligand.ouhsc.edu/prediction.htm	HLA ligands and motifs	n/a
IEDB	http://epitope2.immuneepitope.org/home.do	Beta-version of biothreat pathogen T cell epitope database	[103]
IMGT/HLA	http://www.ebi.ac.uk/imgt/hla/allele.htm	Aligned and annotated HLA sequences by WHO nomenclature	[104]
IMGT/TR	http://imgt.cines.fr/textes/IMGTrepertoire	Aligned and annotated T cell receptor sequences	[105]
MHC Haplotype Project	http://www.sanger.ac.uk/HGP/Chr6/MHC	The haplotype of MHC-linked diseases, with genetic information	[106]
MHC/peptide	http://sege.ntu.edu.sg/wester/mhcp/query.htm	Structural characterisation of MHC protein–peptide complex	n/a
MHCBN	http://www.imtech.res.in/raghava/mhcbn	MHC and TAP peptide binders and non-binders, T cell epitopes	[107–109]
MHCPEP	http://wehih.wehi.edu.au/mhcpep/	MHC-presented epitopes	[110]
MotifScan	http://www.hiv.lanl.gov/content/immunology/motif_scan	HLA-specific primary anchor motifs, view motif libraries	n/a
MPID-T	http://surya.bic.nus.edu.sg/mpidt	MHC/peptide and TCR/MHC/peptide binding characterisation	[111]
PDB	http://www.rcsb.org/pdb/home/home.do	Structural database and MHC/peptide/TCR combinations	[112]
SYFPEITHI	http://www.syfpeithi.de	MHC binders, MHC-specific anchor and auxiliary motifs	[113]

n/a indicates that the relevant information are unavailable.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/rmv
Petitioner Merck, Ex. 1171, p. 87
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00024

Table 5. B-cell epitope prediction related resources

Database/server	URL	Description	References
ABcheck	http://www.bioinf.org.uk/abs/seqtest.html	Aligned antibody sequences to identify sequencing errors	n/a
ABCpred	http://www.imtech.res.in/raghava/abcpred/	Artificial neural network linear B cell epitope predictor	[114]
ABG	http://www.ibt.unam.mx/vir/structure/structures.html	Directory of antibody structures and sequence alignments	n/a
ANTIGENIC	http://bioinfo.bgu.ac.il/cgi-bin/emboss.pl?_action	Identification of B cell epitopes	[15]
AntiJen	http://www.jenner.ac.uk/antijen/	Quantitative binding data for proteins, including B cell epitopes	[99]
BCEPred	http://www.imtech.res.in/raghava/bcepred/	Linear B cell epitopes prediction by physico-chemical properties	[115]
BCIPEP	http://www.imtech.res.in/raghava/bcipep	Database of B cell epitopes of varying immunogenicity	[116]
BepiPred	http://www.cbs.dtu.dk/services/BepiPred	Linear B cell epitope prediction	[117]
BEPITOPE	jlpellequer@cea.fr	Sequence based tool for continuous B cell epitope prediction	[118]
BEPro(PEPITO)	http://pepito.proteomics.ics.uci.edu/	Discontinuous B cell epitope prediction	[119]
CED	http://202.41.70.74:8080/cgi-bin/cep.pl	Predicts conformational epitopes for proteins	[120]
CEP	http://bioinfo.ernet.in/cep.htm	Linear and conformational epitope prediction using 3D structure	[121]
CID	http://ludwig-sun5.unil.ch/CancerImmunomeDB/	Documents antibody eliciting antigens in cancer patients	n/a
ClusPro	http://nrc.bu.edu/cluster/	Rigid protein–protein docking (Fast-Fourier transform correlation)	[122,123]
ConSurf	http://consurf.tau.ac.il	Identification of functional regions in proteins of known structure	[124]
Discotope	http://www.cbs.dtu.dk/services/DiscoTope	Sequence/structure based discontinuous epitope prediction	[125]
DNAPLOT	http://vbase.mrc-cpe.cam.ac.uk	Align a rearranged V gene to closest V, D and J germ line genes	n/a
EMT	elro@novozymes.com	Phage-display based linear, conformational epitopes prediction	n/a
EPIMAP	mumey@cs.montana.edu	Phage-display based discontinuous epitopes prediction	n/a
Epitome	http://www.rostlab.org/services/epitome	Database of antigenic residues and interacting antibodies	[126]
HaptenDB	http://www.imtech.res.in/raghava/haptendb/	A listing of haptens, structural similarity searches	[127]
HIV Immunology	http://www.hiv.lanl.gov/immunology	B cell HIV epitopes, proteome linear epitope maps, citations	[102]
IEDB	http://epitope2.immuneepitope.org/home.do	T- and B cell epitopes and non-epitopes database	[103]
IEDB Ab Epitope	http://www.immuneepitope.org/tools/bcell/iedb_input	Sequence based tool for the identification of continuous epitopes	[103]
IMGT/IG	http://imgt.cines.fr/cgi-bin/IMGTlect.jv	Immunoglobulin structures and annotated sequences	[105]
JenPep	http://www.jenner.ac.uk/JenPep/	Binding data for immunological protein-peptide interactions	[128]
MMDB	http://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml	Listing of crystal structures, including antibodies, HLA, and TCRs	[129]
PatchDock	http://bioinfo3d.cs.tau.ac.il	Rigid-body protein-protein docking based on local shape feature	[130]
PIER	n/a	Binding interface prediction using structure and sequence properties	[131]
PPI-PRED	http://bmbpcu36.leeds.ac.uk/ppi_pred/index.html	Protein-protein interface prediction	[132]
ProMate	n/a	Binding interface prediction using structure and sequence properties	[133]
SACS	http://www.bioinf.org.uk/abs/sacs	Summary of antibody structures	[134]
SDAP	http://fermi.utmb.edu/SDAP/	Structural database of allergenic proteins	[135]

n/a indicates that the relevant information are unavailable.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/rmv
Petitioner Merck, Ex. 1171, p. 88
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00024

Table 6. Links to lists of links

Name	URL	Description
Syfpeithi	http://syfpeithi.bmi-heidelberg.com/Scripts/MHCServer.dll/Info.htm	Links to MHC relevant sites
CBS	http://www.cbs.dtu.dk/services/	Bioinformatics tools, servers, and databases
Informatik	http://www2.informatik.hu-berlin.de/~hakenber/links/software.html#bioinf	Bioinformatics and proteomics resources
IMGT	http://imgt.cines.fr/textes/Immunoinformatics.html	Links to immunoinformatic databases, tools, and resources.
IEDB	http://epitope2.immuneepitope.org/hyperlinks.do?dispatch=loadLinks	Bioinformatics resources
Cancer Immunity	http://www.cancerimmunity.org/links/tools.htm	T cell epitope prediction tools
ePitope Informatics	http://www.epitope-informatics.com/Links.htm	Protein analysis and annotation including epitope prediction
SAMSI	http://www.samsi.info/200405/compbio/workinggroup/magvad/MHC.html	Databases and prediction algorithms on MHC I binding

structure. Continuous B cell epitopes are easier to predict. However, the majority of B cell epitopes are discontinuous. This makes discontinuous B cell epitope prediction software development a major task for epitope prediction community. Therefore, software focussed on discontinuous B cell epitopes and their performances are expected to increase significantly in the future. The input for discontinuous B cell epitope prediction software is usually 3D protein structures and the output can also be viewed graphically as 3D structures. CEP is among the first few software to predict discontinuous epitopes or antibody binding sites of antigens. It uses an algorithm that utilises per cent accessible surface area (%ASA) and distance as criteria to map continuous and discontinuous epitopes from X-ray crystals of Ag-Ab complexes. Apart from predicting discontinuous epitopes, CEP also predicts antigenic determinants and continuous epitopes. Discotope predicts discontinuous B cell epitopes by calculation of surface accessibility (estimated in terms of contact numbers) and an epitope propensity amino acid score, which contribute to the final scores.

Continuous B cell epitope prediction software

Continuous B cell epitopes are easier to predict than their discontinuous counterparts, because they share many common characteristics with T cell epitopes. The input format for continuous B cell epitope prediction software can be either sequence or structure. Common predictive software on discontinuous B cell epitope include ABCpred, BCEPred, BEPITOPE and BepiPred. ABCpred is a predictive software based on standard feed-forward (FNN) and recurrent neural network (RNN) for predicting B cell epitopes in an antigenic sequence. BcePred predicts continuous B cell epitopes by a method which combines four amino acid properties, hydrophilicity, flexibility, polarity and exposed surface. IEDB antibody epitope prediction predicts continuous B cell epitopes, with the option of six amino acid property-based methods, Chou and Fasman Beta-Turn Prediction, Emini Surface Accessibility Prediction, Karplus and Schulz Flexibility Prediction, Kolaskar and Tongaonkar Antigenicity, Parker Hydrophilicity Prediction, Bepipred Linear Epitope Prediction.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/rmv
Petitioner Merck, Ex. 1171, p. 89
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00024

B cell epitope databases

CED is a discontinuous epitope database that has been developed by Jian Huang *et al.*, with all epitopes in the database manually curated from published literature [116]. This database is much smaller than other T cell epitope databases, but it provides a complement to other existing epitope databases. Most epitopes in CED can also be viewed interactively in the context of their 3D structures. In addition, the entries are also hyperlinked to various databases such as Swiss-Prot, PDB, KEGG and PubMed, providing wide background information. Apart from CED, many other databases which contain B cell epitopes are available as well, see Table 4.

These databases do not only provide comprehensive information about B cell epitopes, they are also useful tools for extracting and analysing this information. But, limitations of current epitope databases exist: (1) There is the obvious scarcity of discontinuous epitope databases. (2) Epitope entries in these databases are limited. (3) The information regarding the epitope is incomplete, for example, the potential to generate protective antibody. (4) Few databases contain all types of epitopes, binders and non-binders, cleavage sites, etc. (5) Integration of database with epitope prediction software is rare.

APPLICATIONS OF EPITOPE PREDICTION

The implication of epitope prediction in both public health and basic scientific research is vast. It is applicable to all epitope-related research, such as discovery of peptide candidate for subunit vaccines, autoimmune diseases study, allergy treatment, protein structural study, experiment design, etc. Developing epitope predictive methods and software to identify and map potential epitopes from an antigen protein is vital to contest the immune and infectious diseases. Drug development is the major financial drive for epitope prediction. Epitope-based vaccines have been shown to have promising results and confer protection to animal models in clinical trials [65,136], supporting the prophylactic, therapeutic and protective effects of these vaccines. The advantages of subunit vaccines over other types of vaccines are pronounced. Therefore, huge resources are being channeled into developing subunit vaccines against important, intractable diseases such as cancer, HIV/AIDS, HCV and many other infectious, viral and immune diseases.

However, despite its huge implications in public health, security and scientific arena, epitope prediction tools may be abused by terrorists to make biochemical weapons, and accelerate pathogen evolution and mutation. Another concern is that the application of epitope predictive software in discovering epitopes bias subsequent predictors, as researchers would normally narrow peptide targets by predicting possible epitopes first and then conduct experiments to discover epitopes, which in turn will be analysed to develop other epitope predictive software.

CONCLUSIONS

This review is intended as an introduction to the epitope predictive methods, software and databases publicly accessible online, which is expanding exponentially. We hope to assist researchers locate and take advantage of these tools and to provide insights into future development in this field. Although current prediction methods and tools are not perfect, especially so for B cell epitopes and discontinuous epitopes discovery, biomedical and immunological scientists still have a lot to gain through discreet usage, comparison and combination of different predictive software. As this field progresses, improvements over traditional methods and completely new ones are expected to make epitope prediction more reliable and drug discovery even less expensive and time-consuming.

ACKNOWLEDGEMENTS

Our profound admirations and respect go to researchers in this fascinating, fast developing and promising field, whose dedication and hard work make it all possible. This work was supported by National Natural Science Foundation of China (Grant No. 30571390) and Key Project of Science and Technology Foundation of Hunan Province (Grant No. 2006NK2003).

REFERENCES

1. De Groot AS. Immunomics: discovering new targets for vaccines and therapeutics. *Drug Discov Today* 2006; **11**(5/6): 203–209.
2. Pellequer JL, Westhof E, Van Regenmortel MHV. Predicting location of continuous epitopes in proteins from their primary structures. *Methods Enzymol* 1991; **203**(8): 176–201.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/rmv
Petitioner Merck, Ex. 1171, p. 90
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00024

3. Zhao B, Sakharkar KR, Lim CS, Kangueane P, Sakharkar MK. MHC-peptide binding prediction for epitope based vaccine design. *IJIB* 2007; **1**(2): 127–140.
4. Tong JC, Tan TW, Ranganathan S. Methods and protocols for prediction of immunogenic epitopes. *Brief Bioinform* 2007; **8**(2): 96–108.
5. Korber B, LaBute M, Yusim K. Immunoinformatics comes of age. *PLoS Comput Biol* 2006; **2**(6): 484–492.
6. Rick R. Review of immunoinformatic approaches to *in-silico* B-cell epitope prediction. Available from Nature Precedings http://dx.doi.org/10.1038/npre.2007.253.1 > (2007)
7. Hanada K, Yewdell JW, Yang JC. Immune recognition of a human renal cancer antigen through post-translational protein splicing. *Nature* 2004; **427**: 252–256.
8. Yewdell JT, Bennink JR. Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. *Ann Rev Immunol* 1999; **17**: 51–88.
9. Macallum RM, Martin ACR, Thornton JM. Antibody-antigen interactions: contact analysis and binding site topography. *J Mol Biol* 1996; **262**: 732–745.
10. Almagro JC. Identification of differences in the specificity-determining residues of antibodies that recognize antigens of different size. *J Mol Rec* 2004; **17**: 132–143.
11. Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P. A direct estimate of the human $_{\alpha\beta}$T cell receptor diversity. *Science* 1999; **286**(5441): 958–961.
12. Jerne NK. The nobel lectures in immunology. The nobel prize for physiology or medicine, 1984. The generative grammar of the immune system. *Scand J Immunol* 1993; **38**(1): 1–9.
13. Welling GW, Weijer WJ, van der Zee R, Welling-Wester S. Prediction of sequential antigenic regions in proteins. *FEBS Lett* 1985; **188**(2): 215–218.
14. Parker JM, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* 1986; **25**(19): 5425–5432.
15. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* 1990; **276**(1–2): 172–174.
16. Kolaskar AS, Kulkarni-Kale U. Prediction of three-dimensional structure and mapping of conformational epitopes of envelope glycoprotein of Japanese encephalitis virus. *Virology* 1999; **261**(1): 31–42.
17. Margalit H, Spouge JL, Cornette JL, Cease KB, Delisi C, Berzofsky JA. Prediction of immunodominant helper T cell antigenic sites from the primary sequence. *J Immunol* 1987; **138**: 2213–2229.
18. Rothbard JB, Taylor WR. A sequence pattern common to T cell epitopes. *EMBO J* 1988; **7**(1): 93–100.
19. Stille CJ, Thomas LJ, Reyes VE, Humphreys RE. Hydrophobic strip-of-helix algorithm for selection of T cell-presented peptides. *Mol Immunol* 1987; **24**(10): 1021–1027.
20. Sturniolo T, Bono E, Ding J, *et al*. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat Biotechnol* 1999; **17**(6): 533–534.
21. Brusic V, Rudy G, Honeyman G, Hammer J, Harrison L. Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network. *Bioinformatics* 1998; **14**: 121–130.
22. Murugan N, Dai Y. Prediction of MHC class II binding peptides based on an iterative learning model. *Immunome Res* 2005; **1**: 6.
23. Doytchinova IA, Taylor P, Flower DR. Proteomics in vaccinology and immunobiology: an informatics perspective of the immunone. *J Biomed Biotechnol* 2003; **5**(2003): 267–290.
24. Yu K, Petrovsky N, Schönbach C, Koh JYL, Brusic V. Methods for prediction of peptide binding to MHC molecules: a comparative study. *Mol Med* 2002; **8**(3): 137–148.
25. Bhasin M, Raghava GP. Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine* 2004; **22**: 3195–3204.
26. Zhang Q, Wang P, Kim Y, *et al*. Immune epitope database analysis resource (IEDB-AR). *Nucleic Acids Res* 2008; **36**: W513–W518.
27. Brown MPS, Grundy WN, Lion D, *et al*. Knowledge-based analysis on microarray gene expression data by using support vector machines. *Proc Natl Acad Sci* 2000; **97**: 262–297.
28. Rosenberg HF, Dyer KD. Eosinophil cationic protein and eosinophil-derived neurotoxin. Evolution of novel function in a primate ribonuclease gene family. *J Biol Chem* 1995; **270**(37): 21539–21544.
29. Altuvia Y, Schueler O, Margalit H. Ranking potential binding peptides to MHC molecules by a computational threading approach. *J Mol Biol* 1995; **249**: 244–250.
30. 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**: 1838–1846.
31. Michielin O, Luescher I, Karplus M. Modeling the TCR–MHC–peptide complex. *J Mol Biol* 2000; **300**: 1205–1235.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/rmv
Petitioner Merck, Ex. 1171, p. 91
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00024

32. Rosenfeld R, Zheng Q, Vajda S, DeLisi C. Computing the structure of bound peptides. Application to antigen recognition by class I major histocompatibility complex receptors. *J Mol Biol* 1993; **234**(3): 515–521.
33. Sezerman U, Vajda S, DeLisi C. Free energy mapping of class I MHC molecules and structural determination of bound peptides. *Protein Sci* 1996; **5**: 1272–1281.
34. Rognan D, Laumoeller SL, Holm A, Buus S, Tschinke V. Predicting binding affinities of protein ligands from three-dimensional models: application to peptide binding to class I major histocompatibility proteins. *J Med Chem* 1999; **42**: 4650–4658.
35. Desmet J, Maeyer MD, Spriet J, Lasters I. Flexible docking of peptide ligands to proteins. *Methods Mol Biol* 2000; **143**: 359–376.
36. Tong JC, Tan TW, Ranganathan S. Modeling the structure of bound peptide ligands to major histocompatibility complex. *Protein Sci* 2004; **13**(9): 2523–2532.
37. Greenbaum JA, Andersen PH, Blythe M, *et al*. Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. *J Mol Recognit* 2007; **20**(2): 75–82.
38. Doytchinova IA, Guan P, Flower DR. EpiJen: a server for multistep T cell epitope prediction *BMC Bioinformatics* 2006; **7**: 131.
39. Brusic V, Bajic VB, Petrovsky N. Omputational methods for prediction of T-cell epitopes—a framework for modelling, testing, and applications. *Methods* 2004; **34**: 436–443.
40. Hopp TP, Woods KR. Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci USA* 1981; **78**: 3824–3828.
41. Welling GW, Weijer WJ, van der Zee R, Welling-Wester S. Prediction of sequential antigenic regions in proteins. *FEBS Lett* 1985; **188**(2): 215–218.
42. Parker JM, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* 1986; **25**: 5425–5432.
43. Kolaskar AS, Tongaonkar PC. A semi empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* 1990; **276**(12): 172–174.
44. Alix AJ. Predictive estimation of protein linear epitopes by using the program PEOPLE. *Vaccine* 1999; **18**: 311–314.
45. Ponomarenko JV, Bourne PE. Antibody-protein interactions: benchmark datasets and prediction tools evaluation. *BMC Struct Biol* 2007; **7**: 64.
46. Sollner J, Grohmann R, Rapberger R, Perco P, Lukas A, Mayer B. Analysis and prediction of protective continuous b cell epitopes on pathogen proteins. *BMC Immunome Res* 2008; **4**: 1.
47. Zhang CT, Chou KC. An analysis of protein folding type prediction by seed-propagated sampling and jackknife test. *J Protein Chem* 1995; **14**: 583–593.
48. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics* 2007; **8**: 424.
49. Lin HH, Ray S, Tongchusak S, Reinherz EL, Brusic V. Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research. *BMC Immunol* 2008; **9**: 8.
50. Peters B, Bui HH, Frankild S, *et al*. A community resource benchmarking predictions of peptide binding to MHC-I molecules. *PLoS Comput Biol* 2006; **2**(6): 574–584.
51. Trost B, Bickis M, Kusalik A. Strength in numbers: achieving greater accuracy in MHC-I binding prediction by combining the results from multiple prediction tools. *Immunome Res* 2007; **3**: 5.
52. Rajapakse M, Schmidt B, Feng L, Brusic V. Predicting peptides binding to MHC class II molecules using multi-objective evolutionary algorithms. *BMC Bioinformatics* 2007; **8**: 459.
53. Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. *J Mol Biol* 1999; **285**(5): 2177–2198.
54. Jones S, Thornton JM. Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* 1996; **93**(1): 13–20.
55. Jones S, Thornton JM. Prediction of protein-protein interaction sites using patch analysis. *J Mol Biol* 1997; **272**(1): 133–143.
56. Gowthaman U, Agrewala JN. In silico tools for predicting peptides binding to HLA-class II molecules: more confusion than conclusion. *J Proteome Res* 2008; **7**: 154–163.
57. Borras-Cuesta F, Golvano J, Garcia-Granero M, *et al*. Specific and general HLA-DR binding motifs: comparison of algorithms. *Hum Immunol* 2000; **61**: 266–278.
58. Wang P, Sidney J, Dow C, *et al*. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS Comput Biol* 2008; **4**(4): e1000048.
59. Cui J, Han LY, Lin HH, *et al*. Prediction of MHC-binding peptides of flexible lengths from sequence-derived structural and physicochemical properties. *Mol Immunol* 2007; **44**(5): 866–877.
60. Lata S, Sharma BK, GPS Raghava. Analysis and prediction of antibacterial peptides. *BMC Bioinformatics* 2007; **8**: 263.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/rmv
Petitioner Merck, Ex. 1171, p. 92
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00024

61. Bui HH, Sidney J, Peters B, *et al*. Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications. *Immunogenetics* 2005; **57**(5): 304–314.
62. Parker KC, Bednarek MA, Coligan JE. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide sidechains. *J Immunol* 1994; **152**: 163.
63. Holzhutter HG, Frommel C, Kloetzel PM. A theoretical approach towards the identification of cleavage-determining amino acid motifs of the 20 S proteasome. *J Mol Biol* 1999; **286**: 1251–1265.
64. Tenzer S, Peters B, Bulik S, *et al*. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell Mol Life Sci* 2005; **62**: 1025–1037.
65. Peters B, Bulik S, Tampe R, Van Endert PM, Holzhutter HG. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *J Immunol* 2003; **171**: 1741–1749.
66. Nielsen M, Lundegaard C, Worning P, *et al*. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci* 2003; **12**: 1007–1017.
67. Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics* 2005; **6**: 132.
68. Wilson CC, McKinney D, Anders M, *et al*. Development of a DNA vaccine designed to induce cytotoxic T lymphocyte responses to multiple conserved epitopes in HIV-1. *J Immunol* 2003; **171**: 5611–5623.
69. 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.
70. Sturniolo T, Bono E, Ding J, *et al*. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat Biotechnol* 1999; **17**: 555–561.
71. Singh H, Raghava GPS. Detection of orientation of MHC Class II Binding peptides using Bionformatics tools. *Biotech Softw Internet Report*, 2002; **3**: 146.
72. Hakenberg J, Nussbaum AK, Schild H, *et al*. MAPPP: MHC class I antigenic peptide processing prediction. *Appl Bioinformatics* 2003; **2**(3): 155–158.
73. Guan P, Doytchinova IA, Zygouri C, Flower DR. MHCPred: bringing a quantitative dimension to the online prediction of MHC binding. *Appl Bioinformatics* 2003; **2**: 63–66.
74. Guan P, Doytchinova IA, Zygouri C, Flower DR. MHCPred: a server for quantitative prediction of peptide-MHC binding. *Nucleic Acids Res* 2003; **31**: 3621–3624.
75. Hattotuwagama CK, Guan P, Doytchinova IA, Zygouri C, Flower DR. Quantitative online prediction of peptide binding to the major histocompatibility complex. *J Mol Graph Model* 2004; **22**: 195–207.
76. Swain MT, Brooks AJ, Kemp GJL. An automated approach to modelling class II MHC alleles and predicting peptide binding. *Proceedings of the Second IEEE International Symposium on Bio-Informatics and Biomedical Engineering*. IEEE Computer Society Press, 2001; 81–88.
77. Bhasin M, Raghava GP. Prediction of promiscuous and high-affinity mutated MHC binders. *Hybrid Hybridomics* 2003; **22**(4): 229–234.
78. Nielsen M, Lundegaard C, Lund O, Kesmir C. The role of the proteasome in generating cytotoxic T cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* 2005; **57**(1–2): 33–41.
79. Kesmir C, Nussbaum A, Schild H, Detours V, Brunak S. Prediction of proteasome cleavage motifs by neural networks. *Prot Eng* 2002; **15**(4): 287–296.
80. Larsen MV, Lundegaard C, Lamberth K, *et al*. An integrative approach to CTL epitope prediction. A combined algorithm integrating MHC-I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur J Immunol* 2005; **35**(8): 2295–2303.
81. Buus S, Lauemoller SL, Worning P, *et al*. Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach. *Tissue Antigens* 2003; **62**: 378–384.
82. Nielsen M, Lundegaard C, Worning P, *et al*. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci* 2003; **12**: 1007–1017.
83. Nielsen M, Lundegaard C, Worning P, *et al*. Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach. *Bioinformatics* 2004; **20**(9): 1388–1397.
84. Bhasin M, Raghava GPS. A hybrid approach for predicting promiscuous MHC class I restricted T cell epitopes. *J Biosci* 2006; **32**: 31–42.
85. Nussbaum AK, Kuttler C, Hadeler KP, Rammensee HG, Schild H. PAProC: a prediction algorithm for proteasomal cleavages available on the WWW. *Immunogenetics* 2001; **53**(2): 87–94.
86. Bhasin M, Raghava GP. Pcleavage: an SVM based method for prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences. *Nucleic Acids Res* 2005; **33**: W202–W207.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/rmv
Petitioner Merck, Ex. 1171, p. 93
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00024

87. Singh H, Raghava GPS. ProPred: prediction of HLA-DR binding sites. *Bioinformatics* 2001; **17**(12): 1236–1237.
88. Singh H, Raghava GP. ProPred1: prediction of promiscuous MHC Class-I binding sites. *Bioinformatics* 2003; **19**(8): 1009–1014.
89. Reche PA, Glutting JP, Reinherz EL. Prediction of MHC class I binding peptides using profile motifs. *Hum Immunol* 2002; **63**(9): 701–709.
90. Reche PA, Glutting JP, Zhang H, Reinherz EL. Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. *Immunogenetics* 2004; **56**(6): 405–419.
91. Bhasin M, Reinhez EL, Reche PA. Modeling features of immunodominance into T-cell epitope identification. *2nd International Immunoinformatics Symposium*, Boston University, 7–9 March 2005.
92. Chang HT, Pai TW, Fan T, *et al*. A reinforced merging methodology for mapping unique peptide motifs in members of protein families. *BMC Bioinformatics* 2006; **7**: 38.
93. Pai TW, Chang MD, Chang HT, *et al*. Unique peptide identification of RNaseA superfamily sequences based on reinforced merging algorithms. *J Bioinform Comput Biol* 2006; **4**: 1.
94. 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**: 1765–1772.
95. Dönnes P, Kohlbacher O. SVMHC: a server for prediction of MHC-binding peptides. *Nucleic Acids Res* 2006; **34**(Web Server issue): W194–W197.
96. Bhasin M, Raghava GPS. Analysis and prediction of affinity of TAP binding peptides using cascade SVM. *Protein Sci* 2004; **13**(3): 596–607.
97. Bian H, Hammer J. Discovery of promiscuous HLA-II-restricted T cell epitopes with TEPITOPE. *Methods* 2004; **34**(4): 468–475.
98. Dönnes P, Kohlbacher O. Integrated modelling of the major events in the MHC class I antigen processing pathway. *Protein Sci* 2005; **14**(8): 2132–2140.
99. Toseland CP, Clayton DJ, McSparron H, *et al*. AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. *Immunome Res* 2005; **1**: 4.
100. Reche PA, Zhang H, Glutting JP, Reinherz EL. EPIMHC: a curated database of MHC-binding peptides for customized computational vaccinology. *Bioinformatics* 2005; **21**(9): 2140–2141.
101. Schönbach C, Koh JL, Sheng X, Wong L, Brusic V. FIMM, a database of functional molecular immunology. *Nucleic Acids Res* 2000; **28**: 222–224.
102. Frahm N, Linde C, Brander C. Identification of HIV-Derived, HLA Class I Restricted CTL Epitopes: Insights into TCR Repertoire, CTL Escape and Viral Fitness. *HIV Molecular Immunology*, Korber BTM, Brander C, Haynes BF, Koup R, Moore JP, Walker BD, Watkins DI (eds). Los Alamos National Laboratory, Theoretical Biology and Biophysics: Los Alamos, New Mexico, 2006/2007; 3–28, LA-UR 07-4752.
103. Peters B, Sidney J, Bourne P, *et al*. The immune epitope database and analysis resource: from vision to blueprint. *PLoS Biol* 2005; **3**(3): e91.
104. Robinson J, Waller MJ, Parham P, Bodmer JG, Marsh SGE. IMGT/HLA—a sequence database for the human major histocompatibility complex. *Nucleic Acids Res* 2001; 29: 210–213.
105. Giudicelli V, Duroux P, Ginestoux C, *et al*. IMGT/LIGM-DB, the IMGT comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. *Nucleic Acids Res* 2006; **34**: D781–D784.
106. Allcock RJ, Atrazhev AM, Beck S, *et al*. The MHC haplotype project: a resource for HLA-linkedassociation studies. *Tissue Antigens* 2002; **59**: 520–521.
107. Bhasin M, Singh H, Raghava GPS. MHCBN: a comprehensive database of MHC binding and nonbinding peptides. *Nucleic Acids Res* 2002; (online) http://www3.oup.co.uk/nar/database/summary/180
108. Bhasin M, Singh H, Raghava GPS. MHCBN: update 2002. *Nucleic Acid Res* 2003; (online, accepted).
109. Bhasin M, Singh H, Raghava GPS. MHCBN: a comprehensive database of MHC binding and nonbinding peptides. *Bioinformatics* 2003; **19**(5): 665–666.
110. Brusic V, Rudy G, Harrison LC. MHCPEP: a database of MHC-binding peptides. *Nucleic Acids Res* 1994; **22**(17): 3663–3665.
111. Tong JC, Kong L, Tan TW, Ranganathan S. MPID-T: database for sequence-structure-function information on T-cell receptor/peptide/MHC interactions. *Appl Bioinformatics* 2006; **5**(2): 111–114.
112. Berman HM, Westbrook J, Feng Z, *et al*. The protein data bank. *Nucleic Acids Res* 2000; **28**: 235–242.
113. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovi S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 1999; **50**(3–4): 213–219.
114. Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* 2006; **65**(1): 40–48.
115. Saha S, Raghava GPS. BcePred: Prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. In *Artificial Immune Systems*, Lecture Notes in Computer Science 3239, Nicosia G, Cutello V, Bentley PJ, Timis J (eds). Springer: Berlin, Heidelberg, 2004; 197–204.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/rmv
Petitioner Merck, Ex. 1171, p. 94
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00024

116. Saha S, Bahsin M, Raghava GPS. Bcipep: a database of B-cell epitopes. *BMC Genomics* 2005; **6**: 79.
117. Larsen JEP, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Res* 2006; **2**: 2.
118. Odorico M, Pellequer JL. BEPITOPE: predicting the location of continuous epitopes and patterns in proteins. *J Mol Recognit* 2003; **16**(1): 20–22.
119. Sweredoski MJ, Baldi P. PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. *Bioinformatics* 2008; **24**(12): 1459–1460.
120. Huang J, Honda W. CED: a conformational epitope database. *BMC Immunol* 2006; **7**: 7.
121. Kulkarni-Kale U, Bhosle S, Kolaskar AS. CEP: a conformational epitope prediction server. *Nucleic Acids Res* 2005; **33**(Web Server issue): W168–W171.
122. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* 2004; **20**(1): 45–50.
123. Mandell JG, Roberts VA, Pique ME, *et al*. Protein docking using continuum electrostatics and geometric fit. *Protein Eng* 2001; **14**(2): 105–113.
124. Glaser F, Pupko T, Paz I, *et al*. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* 2003; **19**(1): 163–164.
125. Andersen PH, Nielson M, Lung O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci* 2006; **15**: 2558–2567.
126. Schlessinger A, Ofran Y, Yachdav G, Rost B. Epitome: database of structure-inferred antigenic epitopes. *Nucleic Acids Res* 2006; **34**: D777–D780.
127. Singh MK, Srivastava S, Raghava GPS, Varsheny GC. HaptenDB: a comprehensive database of haptens, carrier proteins and anti-hapten antibodies. *Bioinformatics* 2006; **22**: 253-255.
128. Blythe MJ, Doytchinova IA, Flower DR. JenPep: a database of quantitative functional peptide data for immunology. *Bioinformatics* 2002; **18**: 434–439.
129. Marchler-Bauer A, Addess KJ, Chappey C, *et al*. MMDB: Entrez's 3D structure database. *Nucleic Acids Res* 1999; **27**: 240–243.
130. Schneidman-Duhovny D, Inbar Y, Polak V, *et al*. Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking. *Protein* 2003; **52**(1): 107–112.
131. Kufareva I, Budagyan L, Raush E, Totrov M, Abagyan R. PIER: protein interface recognition for structural proteomics. *Proteins* 2007; **67**(2): 400–417.
132. Bradford JR, Westhead DR. Improved prediction of protein-protein binding sites using a support vector machines approach. *Bioinformatics* 2005; **21**: 1487–1494.
133. Neuvirth H, Raz R, Schreiber G. ProMate: a structure based prediction program to identify the location of protein-protein binding sites. *J Mol Biol* 2004; **338**(1): 181–199.
134. Allcorn LC, Martin AC. SACS—self-maintaining database of antibody crystal structure information. *Bioinformatics* 2002; **18**(1): 175–181.
135. Ivanciuc O, Schein CH, Braun W. SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Res* 2003; **31**: 359–362.
136. Robinson HL, Amara RR. T cell vaccines for microbial infections. *Nat Med* 2005; **11**: S25–S32.
137. Doytchinova I, Flower DR. Towards the quantitative prediction of T-cell epitopes: CoMFA and CoMSIA studies of peptides with affinity to class I MHC molecule HLA-A*0201. *J. Med. Chem* 2001; **44**: 3572–3581.
138. Zhao B, Mathura VS, Rajaseger G, Moochhala S, Sakharkar MK, Kangueane P. A novel MHCp binding prediction model. *Human Immunology* 2003; **64**(12): 1123–1143.
139. Adams HP, Koziol JA. Prediction of binding to MHC class I molecules. *J Immunol Methods* 1995; **185**(2): 181–190.
140. Bhasin M, Raghava GPS. A hybrid approach for predicting promiscuous MHC class I restricted T cell epitopes. *J Biosci* 2007; **32**: 31–42.
141. Zeng J, Treutlein HR, Rudy GB. Predicting sequences and structures of MHC-binding peptides: a computational combinatorial approach. *Journal of Computer-Aided Molecular Design* 2001; **15**(6): 573–586.
142. Antes I, Siu SW, Lengauer T. DynaPred. *Bioinformatics* 2006; **22**(14): e16–e24.
143. Logean A, Rognan D. Recovery of known T-cell epitopes by computational scanning of a viral genome. *Journal of Computer-Aided Molecular Design* 2002; **16**(4): 229–243.
144. Noguchi H, Kato HT, Matsubara Y, *et al*. Hidden Markov model-based prediction of antigenic peptides that interact with MHC class II molecules. *Journal of Bioscience and Bioengineering* 2002; **94**(3): 264–270.
145. Sturniolo T, Bono E, Ding J, *et al*. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat. Biotechnol* 1999; **1**: 555–561.
146. Mallios RR. Class II MHC quantitative binding motifs derived from a large molecular database with a versatile iterative stepwise discriminant analysis meta-algorithm. *Bioinformatics* 1999; **15**: 432–439.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/rmv
Petitioner Merck, Ex. 1171, p. 95
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00024

147. Cui J, Han LY, Lin HH, *et al*. Prediction of MHC-binding peptides of flexible lengths from sequence-derived structural and physicochemical attributes. *Mol Immunol* 2007; **44**(5): 866–877.
148. Dönnes P, Elofsson A. Prediction of MHC class I binding peptides, using SVMHC. *BMC Bioinformatics* 2002; **3**: 25.
149. 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**: 1838–1846.

Copyright © 2008 John Wiley & Sons, Ltd.

DOI: 10.1002/rmv
Petitioner Merck, Ex. 1171, p. 96
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00024