American Journal of Bioinformatics Research 2012, 2(1): 14-20
DOI: 10.5923/j.bioinformatics.20120201.03

# Major Histocompatibility Complex Class II Prediction

**Zeinab Abd El Halim*, Amr Badr, Khaled Tawfik, Ibrahim Farag**

Faculty of computers and information, CairoUniversity, Ahmed Zewiel Street, Giza, Egypt

**Abstract** Major Histocompatibility complex (MHC) molecules play an essential role in introducing and regulation immune system. The MHC molecules are divided into two classes, class I and class II which are differ in size of their binding pockets. Determining which peptides bind to a specific MHC molecule is fundamental to understanding the basis of immunity, and for the development of vaccines and immunotherapeutic for autoimmune diseases and cancer. Due to the variability of the locations of the class II binding cores, the process for predicting the affinity of these peptides is difficult.This paper investigates a new method for predicting peptides binding to MHC class II molecules and its affinity using genetic algorithms and metaheuristics. The algorithm is based on a fitness function that builds a scoring matrix for all suggested motifs in a specific iteration to test the motif ability to be one of the real motifs in the nature. The genetic algorithmpresented here shows increased prediction accuracy with higher number of true positives and negatives on almost of MHC class II alleles,about 80 percent of peptides were correctly classified when testing dataset from IEDB[26]. Generally, these results indicate that GA has a strong ability for MHC Class II binding prediction.

**Keywords** Major Histocompatibility complex (MHC), peptide Binding, Binders, NonBinders, Antigen presenting cells (APCs)

## 1. Introduction

Vaccines are the greatest single way for preventing infection diseases, with huge benefits to human wellbeing. Vaccine allows the immune system to develop antibodies and overcome the disease should we become infected with the real disease.

Developing a vaccine currently takes a long time but with the interconnectivity of the world there is a mounting fear of a disease quickly spreading as SARS and A (H1N1). To fight this, methods for designing vaccines are researched and implemented.

So the accurate and reliable prediction of MHC peptides binding is fundamental to the strong identification of T-cell epitopes and thus a successful design of peptide – protein based vaccine.

Predicting the peptides that bind to Major Histocompatibility Complex class II molecules can reduce the number of experiments for identifying T cell and play an important role in the process of designing vaccines.

A MHC molecule binds peptide that derived from an antigen, and then displays it on the cell surface for T cells recognition[10]. Thus determining which peptides bind to a specific molecule is important for treatment of autoimmune diseases and cancer. MHC molecules are classified into two major classes and MHC alleles are grouped according to their structure. For class I MHC alleles, the binding groove is closed at both ends, making it possible to predict exactly which residues is positioned in the binding groove. For class II MHC molecules, the binding groove is open at both ends, allowing peptides longer than 9-mers to bind[1]. However, it has been reported that a 9-mer core region is essential for MHC II binding. Because the precise location of the 9-mer region of the MHC II binding peptides is unknown, predicting MHC II binding peptides tends to be more challenging than MHC I binding peptides. Some of this reasons are: the variable length of binding peptides, the undermined core regions for individual peptides, the number of amino acids admissible as primary anchors and the experimental and reporting errors that depending on different methods[1].

Despite of the variability of the length of MHC II binding peptides, many computational methods exits and can be divided into two approaches: sequence based approach and structure based approach[15,17]. Peptide binding to MHC is allele specific and by looking at frequencies of different amino acids in different positions for a large number of known binders, sequence motifs can be seen. An example of sequence based approach is to create a scoring matrix for a specific MHC type and this can be done by looking at frequencies of different amino acids in different positions in the peptide. Another approach for prediction is based on structural information or crystal structure about MHC-peptide complexes and evaluates how well a new peptide fits in the binding groove of a MHC molecule[6].

Prediction has also been made by using machine learning approaches such as artificial neural network (ANN) and Hidden Markov Model (HMM)[3,4]. Each of all prediction

* Corresponding author:
zeinab_fci@yahoo.com (Zeinab Abd El Halim)
Published online at http://journal.sapub.org/bioinformatics

methods has its pros and cons.

This papershows an ideal prediction method which would integrate the strengths of these individual methods while minimizing their disadvantages.The aim of this paper is to predict MHC binding peptides and its affinity cores using genetic algorithms and metaheuristics. A genetic algorithm is an approach to solving certain kinds of search and optimisation problems, this approach involves maintaining a population of possibility solutions and then generating new solutions by the use of genetic operators such as reproduction, crossover and mutation[10]. A fitness function is a measure of the quality of the solution so as the genetic algorithm proceeds the binding matrices improve. In this project sequence data of peptides that bind different MHC types can be extracted from public databases. The sequences can then be trained in MHCPEP dataset[5]. The input will be a peptide sequences and the output will be yes or no (1 or 0) yes for each binder and no for each nonbinder. The genetic algorithm will then be able to predict if a peptide binds or not given its sequence.After that a comparative study done between the output of the genetic algorithm and the output of the test dataset (IEDB)[26,27]. The next step is to predict the best 9's through binder's classifier. The last step is to determine the accuracy of the algorithm. This paper also combined the results of experimental studies to represent the accuracy and utility of genetic algorithm in the prediction of peptide MHC II binding. These results are expected to be of practical interest to immunologists for efficient identification of peptides as candidates for immunotherapy.

# 2. Materialsand Methods

### 2.1. Data Collection

The data sets used for training and testing for binders and non-binders (BNB) were obtained from MHCBN (Bhasin*et al*., 2003), MHCPEP (Brusic*et al*., 1998b), and IEDB (Peters B, *et al*. 2005) for testing and predicting binding affinity (9-mers).

The MHCBN is a curated database consisting of detailed information about Major Histocompatibility Complex (MHC) binding, non-binding peptides and T cell epitopes. The MHCBNdatabase provides information about peptides interaction with TAP and MHC linked autoimmune diseases [16].

MHCPEP is a curated database comprising over 13000 peptide sequences for MHC molecules. Entries are compiled from published reports as well as from direct submission of experimental data. Each entry contains the source protein (when known), an estimate of binding affinity and critical anchor residues (if identified), and is fully referenced[5].

IEDB (Immune Epitopes Database) provides a catalog of experimentally characterized B and T cell epitopes, as well as data on Major Histocompatibility Complex (MHC) binding[26,27]. The IEDB database covers 99% of all publicly available information on peptide epitopes. With respect to MHC II the IEDB database provides a tool that employs a consensus approach to predict MHC class II epitopes and its 9-mers based upon different methods such as Sturniolo, ARB, and SMM_align[13].

In this study the MHCBN and MHCPEP are used as training datasets, these datasets contains many unique peptides known to bind or not bind to the MHC II molecules. The lengths of the peptides vary from 9 to 30 amino acids and have an average length of 15 amino acids.The structure of the peptides is a line containing the amino acids of the actual peptide; the first five peptides on the training dataset are shown below.

AAPYEKEVPLSALTNILSAQL
AEALERMFLSFPTTKTHLA
GMGWAGWLLSPRGSA
AGFKGEQGPKG
RPSWGPTDPRRRSRA

The IEDB is used as a testing dataset to evaluate the predictive performance of the genetic algorithm which used in this project. The output of this dataset is a table with many rows; each row corresponds to one peptide prediction. The columns contain the allele the prediction was made for, the position of the peptide in the input sequences, the core sequence, the predicted score and percentile rank for ARB, combinatorial library, SMM_align and Sturniolo. The last column is percentile rank for the consensus method such that top percentile means good binders.

### 2.2. Predictive Model (Algorithm)

A Genetic algorithm can be defined as a search heuristic that mimics the process of natural evolution. Genetic algorithms belong to the larger class of evolutionary algorithms (EA) which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover.

A GA that presented in this paper consists of 4 steps as follow:

(1). Representing input variables as individuals or chromosomes in population.

(2). Formulating the fitness (objective function) to evaluate individuals.

(3). Generating a new population by genetic operations (selection, crossover, and mutation) on the current population.

(4). Determining if the population has reached the optimal fitness.

The fitness function presented in this paper should produce a score for each peptide input evaluating how good the input is, the input consists of a number of sequences of amino acids.

Given a dataset containing n peptides, S1, S2, … Sn, the main goal of the genetic algorithm is to find the optimal alignment of the peptides to get a corresponding core of the motif(best 9mer) from each peptide. Each peptide is a sequence of amino acids and there are 20 different amino acids identified by a roman character in the alphabet $\sum$.

Example of input peptides can be seen in the following figure 1.

AA***PYEKEVPLS***ALTNILSAQL
***AEALERMFL***SFPTTKTHLA
GMGWA***GWLLSPRGS***A
A***GFKGEQGPK***G
RPSWGP***TDPRRRSRA***

**Figure 1**. Example of 5 Peptides Represented as Sequences of Amino Acids, The Highlighted Amino Acids in Each Peptide is The 9 mers which the Algorithm Should Determine and This is Called a Motif

To determine how to move in each peptide to get the start position for each motif the *Single Sequence Move*algorithm applied and implemented here. In the single sequence move a new starting point for the alignment of a sequence is selected at random. The motif for each peptide is defined as having length 9 and represented as Ti, for each peptide Si- it has a length defined as Mi and each motif has a start position Li this is can summarised below as equations.

∑={'A', 'D', 'G', 'N', 'P', 'R', 'S', 'C', 'E', 'H', 'I', 'L', 'K', 'M', 'F', 'T', 'W', 'Y', 'V', 'Q'} alphabet of amino acids.

N: number of sequences/ peptides.

S: representation for each peptide/sequence.

Si: a peptide i, i € [1,N].

Mi: length of each peptide i, Mi >9.

Ti: T (i, p) , …, T (i, 8), motif for each peptide Si , p € [1,9] as a position.

Li: start position of the motif in peptide i, Li € [1, Mi-8].

The fitness function has the same structure of Gibbs sampler fitness with other modifications. The fitness function f used by the algorithm has been designed and implemented by the author of this paper, and it is determined by the following equation 1.

$$\underset{a=1p=1}{\overset{209}{E}} = \left(\sum\sum Ca,\ p.Log2\left(Fa,p / Va\right)\right) / N \qquad (1)$$

In the above equationa is a representation of an amino acid, p for a position of amino acid in each suggested motifs given to the fitness function, C is a counted number of amino acid a found at position p in the motifs. Fa, p is the weighted frequency of which amino acid a found at position p in the motifs. V is the background frequency of which amino acid a is found at position p in the nature. And finally Nisthe number of motifs given to the function. The fitness function consists of small parts needed to be calculated to gets the fitness function value and it can be divided into 6 steps.

Step (1): Compute C from all suggested motifs Tis given to the function as follow:

$$Ca,\ p=|\{Tip|Tip = a,\ 1\leq i \leq N\}| \qquad (2)$$

Where p € [1, 9] and a € ∑, this is done for evaluating each amino acid.

Step (2): Calculate the sequence weight SWi for each sequence i, and this is used later when calculating the weighted frequencies. There are two methods to calculate the sequence weight either by Hobohm methodology or Henikoff and Henikoff sequence weighting methodology[7]. In this paper to calculate the sequence weight the Henikoff and Henikoff methodology is used to downweight every frequency according to the number of different amino acids at the given position[10], and it can be calculated as follow: the first two equations to give each amino acid a weight and compute the sequence weight as inthe equation 5.

$$Dp=|\{Ca,\ p|\ a \in \sum,\ Ca,\ p > 0\}| \qquad (3)$$

$$Ha,\ p=1/\ (Dp.\ Ca,\ p) \qquad (4)$$

$$SWi = \sum_{P=1}^{9}\{Ha,p \mid a = Ti,p\} \qquad (5)$$

Where i € [1,N] and p € [1,9].

Step (3): Compute the Weighted Frequency W from all sequence weights SWs, it can be calculated from the following equation 6:

$$Wa,p = \left(\sum_{i=1i=1}^{nn}\{SWi \mid Ti,p = a\}\right) / \left(\sum SWi\right) \qquad (6)$$

Where a € ∑ and p € [1,9].

Step (4): Compute the pseudocount correction G, this is done using BLOSUM62 frequency matrix which contains the frequency by which amino acid a isaligned to amino acid b. The pseudocount correction G can be calculated as the following equation 7.

$$Ga,\ p= \sum_b W(b,\ p).\ Q(b,\ a) \qquad (7)$$

Where Qb,a is a BLOSUM62 frequency value, a, b € ∑ and p € [1,9].

Step (5): Compute the effective frequency Fa, p, from the following equation:

$$Fa,p = \frac{\alpha.Wa,p + \beta.Ga,p}{\alpha + \beta} \qquad (8)$$

Where α and β are the weights put on sequence weighting and pseudocount correction respectively. B is a parameter to the algorithm optimized to 50 using Henikoff and Henikoff sequence weighting. α is the effective number of sequences minus one.

Step (6): Compute the value of the fitness function as all of its components is ready to be integrated and used in this step. The output of the fitness function is a value for variable E that describes how different the distribution of amino acids in the motifs are from the distribution of amino acids expected to be found in the nature, the higher the value of E the more likely that these amino acids suggested in the motifs given to the fitness function will bind to MHC-II molecules and by doing that the algorithm achieves the paper goal. The overall structure of this prediction system is showed in figure 2

### 2.3. The Parts of the Genetic Algorithm

A genetic algorithm consists of three main functions; Selection, Crossover and Mutation. These functions are used to convert an old generation of chromosomes into a new generation that fit the requirements of the fitness function.

The Selection process is responsible for selecting which chromosomes in the current generation are to be used in the new generation and which will be forgotten. The type of selection process used in this project is the Roulette selection.

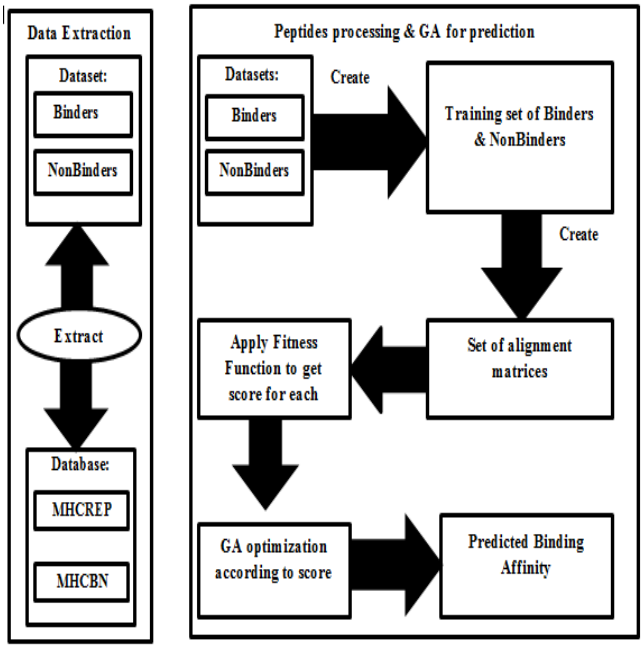


**Figure 2.** Overall structure of the prediction algorithm, in the data extraction stage, peptide sequences and their binding affinities are collected from a variety of sources. In the pre-processing stage, a GA algorithm generates alignment matrices which are then used to find a score for each peptide to be evaluated after that and get binding affinities for binders

The roulette selection is one of the simplest selection methods. The probability that a chromosome **xi** is picked is proportional to its fitness score. Assume X be a random variable representing the chromosome that is picked by the selection, the probability that the chromosome xi is chosen from a population of size S is defined as in the following equation 9:

$$P(X = xi) = \frac{f(xi)}{\sum_{i=1} f(xi)} \quad (9)$$

Crossover is a process for merging two solutions; it makes the genetic algorithm better than the randomized search. The main idea is that if we have two suitable chromosomes it might be possible to combine the good parts from both and extract them into a good and a bad chromosome. An example of the crossover process is shows in figure 3.

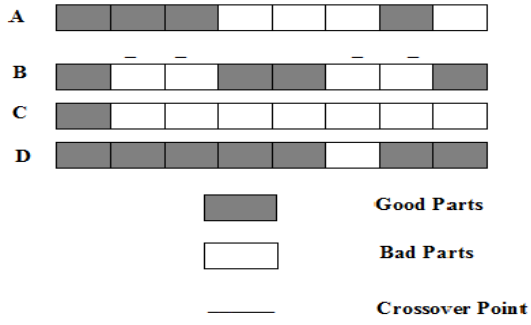


**Figure 3.** Crossover performed on the chromosomes A and B with the crossover points marked by the sign _ results in the chromosomes C and D

There are many methods for applying crossover process but this paper implements *Uniform Crossover*as shown in the above figure. In the uniform crossover every part of the chromosome is picked as a crossover point with a 50% chance. At every crossover point the parts of one chromosome is replaced with the parts from the other chromosome. As the parts which are to be replaced are chosen at random this type of crossover works well on all parts of the chromosome[10].

Mutation process is the changing in a single solution to get better one, one bit changed might result in new and better solutions, which will then be used to generate the next solutions.

*Shift Mutation* is implemented in this project, which moves the entire placement a random number of steps to one side or the other; it adds the same randomly chosen integer to all the start positions. To prevent the start position being located outside the allowed range the number of steps could be in the range [0, length -8].

This type of mutation is useful when the position of the motif in one sequence is dependent on the positions of the motifs in the other sequences. An example of shift mutation can be seen in figure 4.

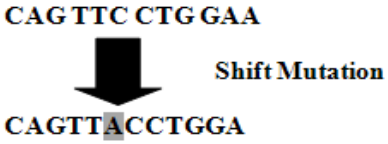


**Figure 4.** Shift Mutation is performed in the chromosome leading to leave the mutated amino acids to a large extent more different from its main shape (replicated from)

### 2.4. Evaluation

Evaluation done using the AUC (Area under Curve) gives a value indicating how much deference between the results of the algorithm and the real results that got from the labs. The evaluation starts after testing using IEDB dataset to know the binders, nonbinder, the output from the IEDB is to say if the peptide binds or not and with what value?.

In the AUC if the peptide binds it is said to be positive and represented as T and if it does not bind it is said to be negative and represented as N. And there are 4 categories:

TP: if a positive peptide is predicted to be binding it is considered a *true positive*.

FN: if a negative peptide is predicted to be binding it is considered a *false negative*.

TN: if a negative peptide is predicted not to be binding, it is considered a *true negative*.

FP: if a positive peptide is predicted not to be binding it is considered a *false positive*.

After the previous step of classification, the total number of actual positive peptides can be calculated as P=TP+FN and the total number of negative peptides calculated as N=FP+TN. After that the TP_Rate and FP_Ratecan be calculated as follow:

$$TP\_Rate = \frac{\text{Positives correctly classified}}{\text{Total number of positives}} = \frac{TP}{P}$$

$$FP\_Rate = \frac{\text{Negative incorrectly classified}}{\text{Total number of negatives}} = \frac{FP}{N}$$

A good prediction would be indicated by the point (0, 1). The TP_Rate=1 means that the number of positives correctly classified is equal to the number of positives (TP=P).

The worst prediction is located in (1,0) as this means that none of the positives has been correctly classified and all the negatives has been incorrectly classified. GA applied in this paper predicts binding peptides affinity with high accuracy; about 80 percent of peptides were correctly classified.

# 3. Results

The GA was applied to derive a position specific scoring matrix for predicting MHC-II binding affinities for the alleles in the dataset. The predictive performance of the GA was tested on IEDB datasetand compared with Gibbs sampler and SVRMHC method[28]. The binding of a peptide was calculated as the score of the highest scoring 9mer sub peptide.The predictive performance of the different methods was measured in terms of the area under the curve (AUC) [19].

**Table (1)**. Prediction Results (AUC values) for a Sample of MHC-II Alleles

| MHC-II Allele | AUC Value |
|---|---|
| DRB1*0101 | 0.768 |
| DRB1*0301 | 0.752 |
| DRB1*0401 | 0.761 |
| DRB1*0701 | 0.763 |
| DRB1*1101 | 0.770 |
| DRB1*1501 | 0.775 |

As seen in Table 1, GA achieved AUC values greater than 0.7 for all of MHC-II alleles provided by the IEDB dataset.

To compare the performance of paper's fitness function with other methods, a training datasetwas created by combining all the experimental datasets, motifs derived on the training dataset were tested on an IEDB dataset- a balanced set generated from this training set.

Comparison of performance of GA derived motifs with Gibbs sampler[1], SVRMHC[26], and ARB[23] is given in Table 2. As seen, GA shows comparable performance with Gibbs sampler.

**Table (2)**. Comparison for predictive results for GA, SVRMHC, Gibbs and ARB

| AUC | | | |
|---|---|---|---|
| SVRMHC | Gibbs | ARB | GA |
| 0.71 | 0.79 | 0.66 | 0.76 |
| 0.65 | 0.76 | 0.65 | 0.75 |
| 0.63 | 0.74 | 0.65 | 0.73 |
| 0.62 | 0.67 | 0.57 | 0.61 |
| 0.57 | 0.59 | 0.60 | 0.59 |

From these results, it is shown that Gibbs sampler and Genetic algorithm are very effective and have good results in solving MHC-II prediction problem.

But the question now is which one is better, Gibbs sampler or GA? The result may be dependent on the dataset, time taken to reach the effective motif and value (score) of the fitness function.

Finally, the GA model predicts binding affinity of peptides with high accuracy; about 80 percent of peptides were correctly classified. The prediction accuracy of this GA model is better than those of other methods, includingARB and SVRMHC.

# 4. Discussion

Determining which peptides bind a particular MHC molecule important for understanding the basis of immune responses, and has potential applications in the design of peptide vaccines and other issues. Tools to facilitate prediction of peptide MHC binding have practical utility in minimising the number of binding experiments in the laboratory. Many methods were implemented to do this job as predicting peptides binding to MHC class I and II and each one has its own performance and major. In this contribution to the field, GA used and a fitness function was developed for binding several MHC molecules using peptide data and other techniques required and used as classifier systems.

The main objective of this paper was to design a method for the prediction of MHC class II- binding peptides that could integrate experimental data and expert knowledge with the search and classification tools of the information science to be able to design vaccines for critical diseases like cancer. The results indicate that GA and its fitness function discussed in this paper succeeded in achieving this objective.GA predictions of peptides binding to MHC-II alleles are as good as or better than alternative methods.

Peptides are typically longer than the core motif, and correct alignment is a key for obtaining good prediction performance.

Prediction of MHC class II peptides is a difficult task, and the prediction accuracy of the method described is good.

As the genetic algorithm is as good as the Gibbs sampler and its ability to predict is very good, but the fitness function used here is simpler than the one used by the Gibbs sampler. However tests of the fitness function performed has shown that the fitness functions are identical with some exceptions like using Henikoff and Henikoff sequence weighting and the training and testing datasets. The following figure shows the comparable values for AUC between Genetic Algorithm and Gibbs sampler.

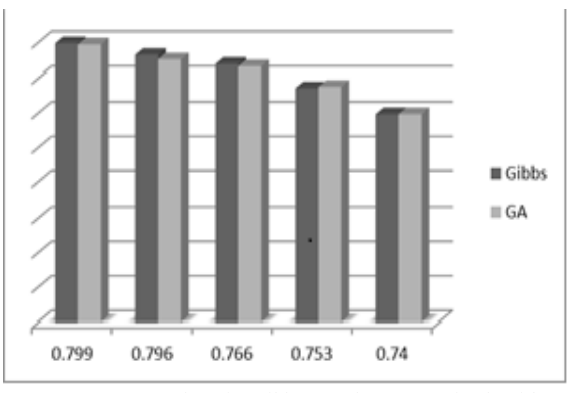


**Figure 5.** AUC values for Gibbs sampler & Genetic algorithm

Finally, this paper proposed a new approach for predicting binding affinity to MHC class II molecules by collecting peptides binding data from datasets. In this paper also a method for determining the start position of the placement (Single Sequence Move) is implemented in addition to the fitness function that worked well to achieve the goal.

The experimental results show that GA significantly improves the accuracy of predicting peptides binding to MHC-II molecules defined at IEDB dataset.

For MHC class -II allele's experimental methods to define the actual binding region of a peptide for every peptide would be very useful and effective. A database containing experimentally verified non-binding protein would also be very useful. Furthermore data would be needed to expand the predictions to more alleles and to test the final models.

## 5. Conclusions

This paper presents a Genetic algorithm for predicting peptides binding to MHC class-II molecules and finding its motifs. GA has successfully designed in modules to make it easy to change the types of elements (mutation, selection, etc.) and implementedin a simple way to optimise the fitness function. The problem of this paper was written using mathematical notation in order to clarify what has to be calculated. The experiment shows that the proposed GA is better than earlier methods in predicting binding sites on most alleles of class II MHC IEDB dataset. This shows the applicability of GA methodology to find binding motifs in a wide range of MHC alleles and thus can help biologists in designing vaccines for autoimmune diseases and cancer.

## REFERENCES


[1] Nielsen, M., Lundegaard, C., Worning, P., Hvid, C.S., Lamberth, K., Buus, S.,Brunak, S., Lund, O.: Improved prediction of mhc class i and class ii epitopes usinga novel gibbs sampling approach. Bioinformatics 20(9) (2004) 1388_1397

[2] Lund, O., Nielsen, M., Lundegaard, C., Kesmir, C., Brunak, S.: ImmunologicalBioinformatics. MIT Press (September 2005) ISBN 0-262-12280-4.

[3] Brusic,V., Rudy,G. and Harrison,L.C. (1994) Prediction of MHCbinding peptides using artificial neural networks. In Stonier,R.J.andYu,X.S. (eds), Complex Systems: Mechanism of Adaptation.IOS Press, Amsterdam, Holland, pp. 253–260.

[4] Brusic,V., Rudy,G., Honeyman,G., Hammer,J. and Harrison,L.(1998a) Prediction of MHC class II-binding peptides using anevolutionary algorithm and artificial neural network. Bioinformatics,14, 121–130.

[5] Brusic,V., Rudy,G. and Harrison,L.C. (1998b) MHCPEP, a databaseof MHC-binding peptides: update 1997. Nucleic Acids Res., 26,368–371.

[6] Chenna,R., Sugawara,H., Koike,T., Lopez,R., Gibson,T.J.,Higgins,D.G. and Thompson,J.D. (2003) Multiple sequencealignment with the Clustal series of programs. Nucleic Acids Res.,31, 3497–3500.

[7] Henikoff,S. and Henikoff,J.G. (1992) Amino acid substitutionmatrices from protein blocks. Proc. Natl Acad. Sci., USA, 89,10915–10919.

[8] Henikoff,S. and Henikoff,J.G. (1994) Position-based sequenceweights.J. Mol. Biol., 243, 574–578.

[9] Marshall,K.W., Wilson,K.J., Liang,J., Woods,A., Zaller,D. andRothbard,J.B. (1995) Prediction of peptide affinity to HLADRB1*0401. J. Immunol., 154, 5927–5933.

[10] HenrikEgebergMygind and Morten Mølgaard: Prediction of MHC class II epitopes using genetic algorithms and other metaheuristics, 2009

[11] Wan J, Liu W, Xu Q, Y R, Flower DR, Li T: SVRMHC predictionserver for MHC-binding peptides. BMC Bioinformatics 2006,7:463.

[12] Bui H, Sidney J, Peters B, Sathiamurthy M, Sinichi A, Purton K, MothéBR, Chisari FV, Watkins DI, Sette A: Automated generation andevaluation of specific MHC binding predictive tools: ARBmatrix applications. Immunogenetics2005, 57(5):304-314.

[13] Nielsen M, Lundegaard C, Lund O: Prediction of MHC class IIbinding affinity using SMM-align, a novel stabilization matrixalignment method. BMC Bioinformatics 2007, 8(238)

[14] Noguchi H, Kato R, Hanai T, Matsubara Y, Honda H, Brusic V, KobayashiT: Hidden Markov model-based prediction of antigenicpeptides that interact with MHC class II molecules. J BiosciBioeng2002, 94(3):264-270.

[15] MenakaRajapakse, Bertil Schmidt, Lin Feng and Vladimir Brusic: Predicting peptides binding to MHC class II molecules usingmulti-objective evolutionary algorithms, BMC Bioinformatics 2007, 8:459

[16] Bhasin M, Singh H, Raghava GP: MHCBN: a comprehensive databaseof MHC binding and non-binding peptides. Bioinformatics2003, 19(5):665-666.

[17] Noguchi H, Hanai T, Honda H, Harrison LC, Kobayashi T: Fuzzyneural network-based prediction of the motif for MHC classII binding peptides. J BiosciBioeng2001, 92(3):227-231.

[18] Jesper Salomon, Darren R Flower: Predicting Class II MHC-Peptide binding: a kernel based approachusing similarity scores, BMC Bioinformatics 2006, 7:501

[19] FatihAltiparmak, AltunaAkalin, HakanFerhatosmanoglu: Predicting the Binding Affinity of MHC class II Peptides,2001

[20] Arne Elofsson: Predicting of MHC class II binding peptides, using Support vector machines, Stockholm Bioinformatics center, 2003

[21] El-Manzalawy Y, Dobbs D, Honavar V: On evaluating MHC-II bindingpeptide prediction methods. PLoS One 2008, 3(9):e3268.

[22] Morten Nielsen, Claus Lundegaard and Ole Lund: Prediction of MHC class II binding affinity using SMM-align, a novelstabilization matrix alignment method, BMC Bioinformatics 2007, 8:238

[23] ARB[tools.immuneepitope.org/tools/matrix/iedb_input?matr

ix Class=II].

[24] Metropolis,N., Rosenbluth,A.W., Teller,A.H. and Teller,E. (1953)Equation of state calculation by fast computing machines.J. Chem. Phys., 21, 1087–1092.

[25] Goldberg,D.E. (1989) Genetic Algorithms in Search, Optimization,and Machine Learning. Addison-Wesley, Reading.

[26] Peters B, Sidney J, Bourne P, Bui HH, Buus S, et al. (2005) The immune epitopedatabase and analysis resource: from vision to blueprint.PLoSBiol 3: e91.doi:10.1371/journal.pbio. 0030091.

[27] Peters B, Sette A (2007) Integrating epitope data into the emerging web ofbiomedical knowledge resources. Nat Rev Immunol 7: 485 490.

[28] Donnes P, Kohlbacher O (2006) SVMHC: a server for prediction of MHCbindingpeptides. Nucleic Acids Res 34: W194–W197.

[29] Shanfeng Zhu, Keiko Udaka, John Sidney, Alessandro Sette, Kiyoko F. Aoki-Kinoshitaand Hiroshi Mamitsuka: Improving MHC binding peptide prediction by incorporating binding data of auxiliary MHC molecules, April 8, 2006

[30] Andrew J Bordner, Hans D Mittelmann: Prediction of the binding affinities of peptides toclass II MHC using a regularized thermodynamicmodel, BMC Bioinformatics 2010, 11:41