

Computational methods for prediction of T-cell epitopes—a framework for modelling, testing, and applications

Vladimir Brusic^{a,b,*}, Vladimir B. Bajic^{a,c}, Nikolai Petrovsky^{b,d}

^a *Laboratories for Information Technology, 21 Heng Mui Keng Terrace, 119613, Singapore*

^b *Medical Informatics Centre, Division of Science and Design, University of Canberra, Bruce ACT 2617, Australia*

^c *South African National Bioinformatics Institute, University of the Western Cape, Cape Town, South Africa*

^d *National Health Sciences Centre, Canberra Clinical School, Woden ACT 2606, Australia*

Accepted 21 June 2004

Abstract

Computational models complement laboratory experimentation for efficient identification of MHC-binding peptides and T-cell epitopes. Methods for prediction of MHC-binding peptides include binding motifs, quantitative matrices, artificial neural networks, hidden Markov models, and molecular modelling. Models derived by these methods have been successfully used for prediction of T-cell epitopes in cancer, autoimmunity, infectious disease, and allergy. For maximum benefit, the use of computer models must be treated as experiments analogous to standard laboratory procedures and performed according to strict standards. This requires careful selection of data for model building, and adequate testing and validation. A range of web-based databases and MHC-binding prediction programs are available. Although some available prediction programs for particular MHC alleles have reasonable accuracy, there is no guarantee that all models produce good quality predictions. In this article, we present and discuss a framework for modelling, testing, and applications of computational methods used in predictions of T-cell epitopes.

© 2004 Elsevier Inc. All rights reserved.

1. Introduction

T cells of the immune system continually survey for the presence of unwanted foreign antigens that may indicate the presence of invading microorganisms (e.g., viruses, bacteria, fungi, and parasites), mutated own cells (tumours), or cells and tissues from other organisms (transplants). Short peptides produced by degradation of antigens are presented on the cell surface for recognition by T cells. T cells, through their receptors, recognizes foreign peptides on the cell surface [1,2]. This process involves major histocompatibility complex (MHC) molecules, which bind short peptides and display them to T cells. Peptides presented by MHC molecules originate from intracellular (MHC class I) or

extracellular (MHC class II) proteins. MHC class I bound peptides activate cytotoxic T-cells resulting in killing of target cells (e.g., infected, neoplastic, or transplanted tissues). MHC class II bound peptides serve mainly in regulation (initiation, enhancement, and suppression) of immune responses.

The ability of the immune system to respond to a particular antigen varies between individuals according to their different pattern of MHC genes. Each human individual expresses up to six HLA (human leukocyte antigen, i.e., human MHC) class I molecules and at least that many HLA class II molecules. MHC genes show extensive polymorphism. More than 800 variants of HLA class I, and more than 500 variants of HLA class II molecules have been characterized and named to date (September 2002) [3].

MHC molecules have a peptide-binding groove that binds peptides in a highly promiscuous manner. The

* Corresponding author. Fax: +65-6774-8056.

E-mail address: vladimir@i2r.a-star.edu.sg (V. Brusic).

number of peptides that can bind each individual HLA class I molecule was estimated at between 1000 and 10,000 individual sequences (S. Stevanovic, personal communication), or on average 0.1–5% of all overlapping 9- and 10-mer peptides spanning a protein [4]. T-cell epitopes are peptides first presented by MHC molecules, then recognized by T cells. Peptide binding to MHC molecules is a prerequisite for T-cell recognition, but in itself is not sufficient. For example, some peptides, although good binders in biochemical assays, never get properly processed for presentation by MHC molecules. Self-peptides get processed and presented by MHC molecules, but normally do not elicit immune response. T-cell epitopes are therefore a subset of MHC-binding peptides. Determining the peptides that bind to a particular MHC molecule is essential for understanding immune responses and for the design of peptide-based vaccines and immunotherapies [5]. Reported MHC-binding peptides or T-cell epitopes range in length from 7 to more than 30 amino acids [6]. Because of a high number of peptides that can be derived from pathogens and the combinatorial nature of human HLA phenotypes the systematic identification of candidate T-cell epitopes requires computational screening, combined with experimental validation.

2. Peptide binding to MHC molecules

The peptide-binding groove of MHC molecules consists of two α -helices supported by a β -sheet (for examples, see PDB entries 1DUZ and 1AQD at www.rcsb.org). Peptide binding is effected by a network of hydrogen bonds between the backbone of the peptide and the binding groove, and through interactions between the peptide side chains and pockets inside the binding groove [7,8]. The peptides presented by MHC class I molecules are mainly 8–11 amino acids long, with a small number of exceptions [6]. An MHC class I molecule accommodates the whole length of the binding peptide inside the binding groove [7]. The peptides presented by MHC class II molecules may be longer than 30 amino acids [6]. The binding groove of MHC class II molecules has open ends and accommodates the 9-mer binding core of the peptides inside, while peptide termini extend outside of the groove [8]. The polymorphic residues within the peptide-binding groove determine the repertoire of binding peptides. The diversity of HLA molecules increases the probability that a foreign antigen will contain MHC-binding peptides suitable as targets of immune responses. The amino acids forming the binding groove of the MHC molecule determine the specificity of peptide binding through preference for particular amino acids. The peptide-binding groove interaction is effected by primary and secondary anchors, which are the positions within the peptide that

provide the highest contribution to peptide binding [9]. Only a limited set of amino acids can act as anchors at a particular position within a peptide for any given MHC molecule. Anchor positions are, therefore, the key for defining the common patterns—binding motifs—within sets of peptides that bind a specific MHC molecule. Identification of a variety of MHC-binding motifs [10] and experimental characterization of thousands of allele-specific and promiscuous MHC binders and T-cell epitopes [6] provide a solid information base for computational prediction of binding peptides. This has triggered the development of a plethora of algorithms for prediction of MHC-binding peptides.

3. MHC-binding prediction methods

Artificial learning systems are usually realized as computer programs. These systems pass a phase of learning during which they extract information and necessary rules from the accumulated experience contained in known cases [11]. These systems learn to generate proper responses when particular input information is presented to them. One of the most important functions of learning systems is classification, in which case the system is called classifier. The classification of new information to an appropriate class is called prediction. The computer programs that recognize patterns and classify new information are called prediction systems. Representation of a real system or process in the form of a computer program is called computer model. Computer simulation is a computerized imitation of the behaviour of the real system or process.

The prediction systems developed for the study of MHC-binding peptides should accurately assign binding affinity of a query peptide to a specific MHC molecule. Classifiers can use various techniques such as pattern recognition, computer and mathematical modelling, or heuristics. Pattern recognition deals with identification and utilization of features that are important for characterizing a particular group of data, or for distinguishing several data groups. The three main branches of pattern recognition are statistical, structural, and neural approaches [12]. Statistical approaches use statistics to define statistical properties of data sets and use them as basis for classification algorithms. Structural approaches extract and quantify structural properties of data sets for the assessment of pattern similarity within the data, or of differences between the data representing different classes. Neural network based approaches use connectionist models, known as artificial neural networks (ANNs) [13], consisting of several interconnected units that can be activated. The training of statistical and neural systems usually requires the adjustment of parameters and coefficients in the algorithm, so that they represent the known cases well. A heuristic method

is an algorithm (a set of well-defined rules that were derived from experience) that includes a strategy or a set of rules that simplifies decision-making and the search for a solution to the problem [14]. Heuristics is usually used to narrow down the search space for the problem solution and to simplify otherwise a complicated process of finding an optimized solution. Thus, heuristics enables us to find some acceptable solutions to the problem, although these are not guaranteed to be optimal or even close to optimal.

The process of establishing a prediction system of peptide binding to MHC molecules involves the following steps:

1. Assembly of a list of peptides with known binding affinities to a particular MHC molecule (or a defined set of MHC molecules) and conversion of these data to a format suitable for input to a computer—generate training and test data for prediction system.
2. Defining the models (algorithms) and their parameters so that the computer model can produce the required output.
3. Training the model using the training data from Step 1.
4. Assessing the accuracy of the computer model by applying predictions to the completely independent test data from Step 1.
5. If the test results are not satisfactory, repeating steps 2–5 with redefined model until the model is acceptable.

After a prediction system is developed and tested, it is ready for application using new data from existing protein antigens to predict potential T-cell epitopes. Following predictions, the selected peptides are normally validated experimentally as MHC binders or T-cell epitopes. The commonly used methods for prediction of MHC-binding peptides and T-cell epitopes are shown in Table 1.

3.1. Sequence similarity

Heuristics rules have been used for identifying candidate T-cell epitopes. An obvious clue is the sequence

similarity between a known MHC-binding peptide and known T-cell epitopes or MHC binders (e.g., [15]). Sequence similarity is usually determined using standard alignment methods such as BLAST [16] or FASTA [17]. Standard comparison matrices available in these programs are not suitable for MHC-binding peptides and alternative matrices should be used [18]. Sequence similarity is the least accurate method for prediction of MHC-binding peptides and should be avoided if possible.

3.2. Binding motifs

A binding motif describes amino acids commonly occurring at particular positions within peptides that bind to a specific MHC molecule. The first allele-specific motifs were reported for mouse MHC class II molecules I-A^d and I-E^d [18] and were proposed for prediction of potential T-cell epitopes. Soon thereafter the pool sequencing method for experimental determination of naturally processed peptides was reported [9]. Pool sequencing determines, in the context of a specific MHC molecule, commonly occurring amino acids at particular positions within a large number of naturally presented peptide. The allele-specific motifs were used for building predictive methods using weighted motif scoring system [19–22]. Stauss et al. [19] compared experimentally determined mouse class I T-cell epitopes from human papilloma virus with those predicted by binding motifs [9] and found that 40% of actual MHC binders conformed to the proposed binding motifs. Nijman et al. [20] also reported in their study that 40% of binding-motif predicted peptides were experimental binders to *HLA-A*0201*.

The limitations in the predictive value of binding-motif predictions have been reported. It was suggested that factors other than size and anchor residues also determine peptide binding [23]. This led to an improved definition of MHC class I binding motifs that defined favourable binding residues at particular positions, as well as those that have negative effect. A similar definition was proposed for MHC class II motif HLA-DQ3.1 [24]. These refined motif definitions, together with weighting schemes of anchor positions, brought the concept of binding motifs closer to definition of quantitative matrices. The examples of binding motifs are shown in Fig. 1. The most comprehensive collection of more than 250 MHC binding motifs is contained in the SYFPAITHI database [25]. Binding motifs are simple to implement and easy to understand, but they are of modest accuracy. They are particularly useful for MHC alleles where not much experimental data are available. The predictive accuracy of binding motifs varies. In the study of five HLA class I alleles, positive predictive values (defined in Section 4.1) of predictions were reported in the range of 27–73% [26]. The lower number was

Table 1
General types of prediction systems and their applications to MHC-binding peptides

Prediction method	Type of MHC-binder prediction system
Statistical classifiers	Quantitative matrices, weighted motifs, hidden Markov models
Structural classifiers	Binding motifs, weighted motifs
Heuristics-based	Sequence similarity
Neural classifiers	Artificial neural networks
Computer modelling	Molecular modelling
Combined methods	Molecular modelling/hidden Markov models

A			T							M
			A							A
			V							L
			I	Y	S	Y				I
			M	F	T	F				V
Good binding		L	W	W	C	W		A		V
	position	1	2	3	4	5	6	7	8	9
No binding		D		D			R	D		
		E		E			K	E		
		P		R			H	R		
				K				K		
								H		

B										
	position	1	2	3	4	5	6	7	8	9
Good binding		L			D		K			Y
		I					R			L
		F					E			F
		M					Q			
		V					N			

Fig. 1. Examples of binding motifs. Proposed primary anchors are shown in bold. (A) Refined (expanded) binding motif for human MHC class I molecule *HLA-A*0201*, indicating amino acids/positions that are favourable or detrimental to peptide binding [20]. (B) Motif for human MHC class II molecule *HLA-DRB1*0301* [22].

reported for simple binding motifs comprising only primary anchors while the higher number was in respect of expanded motifs. Recent reports based on Edman sequencing and mass spectrometry [27] indicated that multiple and diverse binding motifs may be characteristic for a single MHC allele.

3.3. Quantitative matrices

Quantitative matrices provide coefficients for each amino acid and each position within the peptide. These coefficients can be used with appropriate formulae to calculate scores that predict peptide binding. The assumptions underpinning this method are: (a) each position within the peptide contributes independently to binding to an MHC molecule, and (b) a residue located at a given peptide position contributes an equal amount to binding, even within different peptides. Binding matrices were defined for several MHC class I molecules: *HLA-A*0201* [28,29], *HLA-A*2401* [30], *HLA-B*3501* [31], *HLA-B*2705* [29,32], and mouse H-2K^b, D^b, and -L^d [33]. Binding matrices were also defined for several HLA-DR class II molecules [34–38]. These matrices were defined directly from experimental peptide binding data [28,30,31,34–36], positional scanning of combinatorial libraries [33], or by combining peptide binding data from multiple sources and using search algorithms to fit matrix coefficients [29,32,36,37]. The TEPITOPE method for prediction of human class II molecules HLA-DR incorporates multiple quantitative matrices for identification of promiscuous binding peptides [38].

The quantitative matrix method represents an extension of binding motifs, and is simple and efficient to implement. The deficiency of this method is that

the independent-contribution-to-binding assumption ignores the contribution of overall peptide structure to binding. Quantitative matrices tend to over-fit data and bias towards the sets of peptides used to derive the matrix coefficients. This lowers the generalization properties of such methods. Quantitative matrices were reported as more accurate predictors than binding motifs [39].

3.4. Artificial neural networks

The artificial intelligence methods using ANNs are based on models that can capture complex relationships in the data sets and are not based on simplifying assumptions such as quantitative matrices. ANNs are connectionist models that consist of a number of interconnected units that can be activated by transmitting signals [11,13]. ANNs can tolerate a degree of erroneous data, and can classify nonlinear data, which makes them highly suitable for processing noisy biological information. Binding-peptide data contain ambiguities because of the variable lengths of peptides and uncertainty over the positions of binding-core regions. They thus need to be pre-processed before being processed with ANNs. Pre-processing involves alignments of peptides relative to the binding anchor positions. Because of well-defined anchor positions the alignment of MHC class I binding peptides is a relatively simple problem. However, MHC class II binding peptides have degenerate motifs and their alignment is more difficult, requiring application of sophisticated computational alignment techniques [39,40]. ANN applications have been described for predictions of MHC class I binding peptides [41–44] and for MHC class II peptides [39,40,45]. The prediction accuracy of ANN-based methods was reported to be close to 80% sensitivity and 80% specificity [39,42,45].

The advantages of ANNs are that they are adaptive and can self improve, they generalize well, are effective with nonlinear problems, and are tolerant to a certain level of erroneous data. Because they have larger number of parameters to be determined ANNs require larger amounts of binding data than simpler prediction methods. Also, the types of ANNs currently used for prediction of MHC-binding peptides require pre-processing of data (peptide alignment) before these data can be used by ANN. With increasing amount of peptide data, ANN-based models will become more highly trained and are expected to further improve in prediction accuracy. ANNs were reported as more accurate predictors than quantitative matrices [46].

3.5. Hidden Markov models

HMMs are statistical models that can capture complex relationships in data sets. An HMM is a finite state machine that governs transition of a system-process

between states using associated probability distributions [47]. They can model sets of variable length sequences making them suitable for modelling biological sequences. HMMs were reported as high accuracy predictors of MHC-binding peptides [48]. A single HMM was used for modelling MHC–peptide interactions for multiple alleles of a HLA-A2 super-type of MHC class I peptides [49].

HMMs perform with a similar accuracy as ANNs and require larger data sets for training than quantitative matrices. The main advantage of a HMM is that it does not require pre-processing (alignment) of peptides before training.

3.6. Molecular modelling

Molecular modelling utilizes detailed knowledge of the crystal structure of MHC molecules [50] and of protein–peptide interactions [51]. Comparative modelling is used if known crystal structures and protein–peptide interactions are available as templates for building 3-D models. *Ab initio* modelling, which uses atomic simulations and residue statistics, is used when initial structural data are not available. Molecular modelling is a complementary approach to other data-driven approaches. The early work used molecular dynamics to model binding of peptides to *HLA-B*2705* [52]. This work was extended to prediction of peptide binding affinities using free energy scoring functions [53]. The crystal structures of 23 solved peptide–MHC structures were used for the development of a modelling algorithm [54]. The known peptide structure in the groove was used as a template for threading peptide candidates and their binding potential was analysed by statistical pair-wise potential for a broad range of MHC class I alleles [55]. Quantitative structure–function relationship studies were performed to identify physicochemical requirements of peptide binding to MHC molecules using large numbers of peptides [56,57].

Molecular modelling provides a detailed insight into specific 3-D structures and interactions. However, it is computationally demanding and therefore less suitable for large-scale screening. Molecular modelling can provide information for building complex data-driven methods, i.e., for prediction of promiscuous MHC-binding peptides using HMMs [49].

3.7. Applications

The discovery of novel T-cell epitopes through use of computational predictions has been reported in cancer immunity [58–60], autoimmunity [45], infectious diseases [61–63], and allergy [64]. Computational predictions have also been used to assist identification of promiscuous T-cell epitopes [65,66]. A promiscuous peptide will bind a variety of MHC molecules. Because

they are potentially effective in large populations, promiscuous peptides represent excellent targets for the design of vaccines and immunotherapies. Most recently, computational predictions have been applied in large-scale screening of T-cell epitopes in HIV [67] and West Nile virus [68]. We are entering a new era of genomic vaccinology and computational prediction methods enable systematic screening of complete genomes of pathogens for vaccine targets [69].

Computational predictions of MHC-binding peptides have been proven to minimize the time and the cost of T-cell epitope mapping. However, before theoretical prediction methods can be used as standard methodology for T-cell epitope discovery and mapping, it is essential to first know their accuracy, coverage, and the potential biases they may introduce.

4. Using prediction systems

4.1. Assessment of the prediction accuracy

For maximum benefit, the use of computational results must be treated as experiments analogous to standard laboratory procedures. This involves due care in both the design of simulated experiments and in the interpretation of results. Computational models that are valid, relevant, and properly assessed for accuracy can be used for planning of complementary laboratory experiments [4]. Several measures are available for the statistical estimation of the accuracy of prediction models [70]. The common statistical measures are sensitivity (SE), specificity (SP), positive predictive value (PPV), and negative predictive value (NPV) (see definition in Table 2). The SE indicates the “quantity” of predictions, i.e., the proportion of real positives correctly predicted. The SP indicates the “quality” of predictions, i.e., the proportion of true negatives correctly predicted. The PPV indicates the proportion of true positives in predicted positives—‘the success rate,’ while NPV is the proportion of true negatives in predicted negatives. Currently, the best predictive performances of the reported systems for prediction of MHC-binding peptides are of the order of SE = 80% and SP = 80%.

4.2. Data and model quality issues

The main issues for building systems that predict MHC-binding peptides are the quality, quantity, and sufficient statistical diversity of data available for the model development, the complexity of the selected predictive model relative to the natural complexity of the peptide–MHC interaction, and the training and testing of the predictive model [4].

A good quality data set is critical for building an accurate prediction system. The available data sets in

Table 2
Definition of common measures for the accuracy of prediction systems

	Experimental positives	Experimental negatives
Predicted positives	True positives (TP)	False positives (FP)
Predicted negatives	False Negatives (FN)	True negatives (TN)
Accuracy measure	Formula	Pairs with
Sensitivity	$SE = TP/(TP + FN)$	SP
Specificity	$SP = TN/(TN + FP)$	SE, sometimes with PPV
Positive predictive value	$PPV = TP/(TP + FP)$	NPV
Negative predictive value	$NPV = TN/(TN + FN)$	PPV
Accuracy	$Acc = (TP + TN)/(TP + FP + TN + FN)$	—

databases and in the literature usually contain significant biases, because many peptides have been pre-selected for experimental testing using binding motifs. It was shown that data sets for some MHC alleles are poor and require data cleaning [40]. In such cases, computational filtering techniques and data cleaning procedures are necessary for building improved predictive models of MHC-binding peptides.

Data quantity is important for selecting an appropriate prediction method. Broad guidelines can be suggested based on a recent comparative study [71] as follows: (1) if no binding data are available, molecular modelling should be used; (2) with small numbers of peptides, (i.e., below 50), binding motifs should be used; (3) data sets of 50–100 peptides should be used for building quantitative matrices or HMMs; (4) larger data sets should be used for building HMM- or ANN-based prediction models; (5) with sufficiently

large data sets, ANNs are useful for high specificity predictions, at the cost of slightly lower sensitivity than HMM predictions. This property is useful for genome-wide screening for T-cell epitopes [40]. Alternatively, HMMs are useful for high sensitivity predictions, such as screening a single protein for all binding peptides.

Models should be tested before being used—as described in the examples of testing strategies [40,71]. Computational methods for testing involve internal cross-validation, where data sets are split into training and testing parts. These data sets should be used in the process of building systems for generating predictions. These predictions can be assessed against the previously available data. Test sets should comprise examples that were not used for building the model. However, for a new prediction model, the best testing strategy is experimental validation.

Table 3
Databases of MHC-binding peptides, T-cell epitopes, and MHC structures

Database	URL	Description
IMGT/MHC	www.ebi.ac.uk/imgt/mhc	MHC sequences—various species
MHCPEP	wehih.wehi.edu.au/mhcpep	MHC-binding peptides (static since 1998)
SYFPEITHI	www.syfpeithi.de	MHC ligands and peptide motifs
FIMM	research.i2r.a-star.edu.sg	HLA, antigens, peptides, and disease
HIV molecular immunology	hiv-web.lanl.gov/immunology	HIV-1 T-cell epitopes
MHCDB	www.hgmp.mrc.ac.uk/Registered/Option/mhcdb.html	HLA genetic and physical data
MHCBN	www.imtech.res.in/raghava/mhcbn	MHC binding and non-binding peptides
JENPEP	www.jenner.ac.uk/JenPep	MHC and TAP binding peptides
HLA ligand/motif	hlaligand.ouhsc.edu	HLA ligands and motifs
MPID	surya.bic.nus.edu.sg	Structural information on MHC/peptide interactions

Table 4
WWW servers for prediction of MHC-binding peptides and proteasome cleavage sites

Server	URL	Predictions
SYFPEITHI	www.syfpeithi.de	MHC class I and II ligands
BIMAS	bimas.dcrn.nih.gov/molbio/hla_bind	MHC class I ligands
ProPred	www.imtech.res.in/raghava/propred	HLA-DR
NetMHC	www.cbs.dtu.dk/services/NetMHC	HLA-A2 and H-2K ^k
MAPPP	www.mpiib-berlin.mpg.de/MAPPP	MHC class I and proteasome cleavage
PAProC	www.paproc.de	Proteasome cleavage
NetChop	www.cbs.dtu.dk	Proteasome cleavage

4.3. Selection of optimal prediction strategies

The selection of optimal strategies utilizing publicly available resources include's the following steps:

- (a) Build a diverse and representative test data set and assess the accuracy of the web-accessible prediction method.
- (b) Use the best method to predict MHC-binding peptides, or even combine multiple prediction systems.
- (c) Combine MHC class I binding predictions with proteasome cleavage predictions and TAP (transporter associated with antigen processing) to identify peptides that are likely to be processed in the MHC class I pathway.

Developers of prediction methods should provide detailed testing of predictive models for each MHC-binding allele. The cyclical refinement of predictive models [62] is a useful technique that can be used for improvement of predictive models with newly generated data.

5. Databases and web servers

Several databases of MHC sequences, MHC-binding peptides, and T-cell epitopes are accessible via the web (Table 3). Users are well advised to access multiple data sources to analyse and cross-check entries with literature sources.

Servers for prediction of MHC-binding peptides are listed in Table 4. Although predictive models for some MHC alleles are reasonably accurate across servers, there is no guarantee that all models produce high quality predictions. To assure the quality of predictions, users are advised to create test sets of known peptides (from literature and databases) to test the performance of the prediction system for the MHC allele of their interest before using predictions for further research.

6. Conclusion

Computational prediction is an important immunoinformatic technology supporting the determination of MHC-binding peptides and T-cell epitopes. Computational analyses and predictions complement, but cannot replace, laboratory experimentation. However, computational screening can help minimize the number of required experiments and allow researchers to focus on critical experiments. Because of the increasing amounts of data available, computational techniques for advanced analyses and translation of basic discoveries into novel vaccines and immunotherapies will grow in importance and are expected to become a standard complement to laboratory experimentation.

References

- [1] H.G. Rammensee, K. Falk, O. Rotzschke, *Annu. Rev. Immunol.* 11 (1993) 213–244.
- [2] P. Cresswell, *Annu. Rev. Immunol.* 11 (1994) 259–293.
- [3] J. Robinson, A. Malik, P. Parham, J.G. Bodmer, S.G. Marsh, *Tissue Antigens* 55 (2000) 280–287.
- [4] V. Brusic, J. Zeleznikow, *Lett. Pept. Sci.* 6 (1999) 313–324.
- [5] J.A. Berzofsky, J.D. Ahlers, I.M. Belyakov, *Nat. Rev. Immunol.* 1 (2001) 209–219.
- [6] V. Brusic, G. Rudy, L.C. Harrison, *Nucleic Acids Res.* 26 (1998) 368–371.
- [7] D.R. Madden, D.N. Garboczi, D.C. Wiley, *Cell* 75 (1993) 693–708.
- [8] L.J. Stern, J.H. Brown, T.S. Jardetzky, J.C. Gorga, R.G. Urban, J.L. Strominger, D.C. Wiley, *Nature* 368 (1994) 215–221.
- [9] K. Falk, O. Rotzschke, S. Stevanovic, G. Jung, G. Jung, H.G. Rammensee, *Nature* 351 (1991) 290–296.
- [10] H.G. Rammensee, J. Bachmann, N.P. Emmerich, O.A. Bachor, S. Stevanovic, *Immunogenetics* 50 (1999) 213–219.
- [11] S.M. Weiss, C.A. Kulikowski, *Computer Systems that Learn* Morgan Kaufman Publishers, San Mateo, 1991.
- [12] R.J. Schalkoff, *Pattern Recognition: Statistical, Structural, and Neural Approaches*, John Wiley & Sons, New York, 1992.
- [13] J.M. Zurada, *Introduction to Artificial Neural Systems*, PWS Publishing Company, Boston, 1992.
- [14] M.W. Firebaugh, *Artificial Intelligence: A Knowledge Based Approach*, PWS-Kent Publishing, Boston, 1989, pp. 119–129.
- [15] E. Celis, J. Larson, L. Otvos Jr., W.H. Wunner, *J. Immunol.* 145 (1990) 305–310.
- [16] W.R. Pearson, D.J. Lipman, *Proc. Natl. Acad. Sci. USA* 85 (1998) 2444–2448.
- [17] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, *J. Mol. Biol.* 215 (1990) 403–410.
- [18] A. Sette, S. Buus, E. Appella, J.A. Smith, R. Chesnut, C. Miles, M. Colon, H.M. Grey, *Proc. Natl. Acad. Sci. USA* 86 (1989) 3296–3300.
- [19] H.J. Stauss, H. Davies, E. Sadovnikova, B. Chain, N. Horowitz, C. Sinclair, *Proc. Natl. Acad. Sci. USA* 89 (1992) 7871–7875.
- [20] H.W. Nijman, J.G. Houbiers, M.P. Vierboom, S.H. van der Burg, J.W. Drijfhout, J. D' Amaro, P. Kenemans, C.J. Melief, W.M. Kast, *Eur. J. Immunol.* 23 (1993) 1547–1553.
- [21] M.L. Disis, J.W. Smith, A.E. Murphy, W. Chen, M.A. Cheever, *Cancer Res.* 54 (1994) 1071–1076.
- [22] G.A. Meister, C.G. Roberts, J.A. Berzofsky, A.S. De Groot, *Vaccine* 13 (1995) 581–591.
- [23] J. Ruppert, R.T. Kubo, J. Sidney, H.M. Grey, A. Sette, *Behring Inst. Mitt.* 94 (1994) 48–60.
- [24] J. Sydney, C. Oseroff, M.F. del Guercio, S. Southwood, J.I. Krieger, G.Y. Ishioka, K. Sakaguchi, E. Apella, A. Sette, *J. Immunol.* 152 (1994) 4516–4525.
- [25] H.G. Rammensee, J. Bachmann, N.P. Emmerich, O.A. Bachor, S. Stevanovic, *Immunogenetics* 50 (1999) 213–219.
- [26] W.M. Kast, R.M. Brandt, J. Sidney, J.W. Drijfhout, R.T. Kubo, H.M. Grey, C.J. Melief, A. Sette, *J. Immunol.* 152 (1994) 3904–3912.
- [27] K.R. Prilliman, M. Lindsey, K.W. Jackson, J. Cole, R. Bonner, W.H. Hildebrand, *Immunogenetics* (1998) 89–97.
- [28] K.C. Parker, M.A. Bednarek, J.E. Coligan, *J. Immunol.* 152 (1994) 163–175.
- [29] J.R. Schafer, B.M. Jesdale, J.A. George, N.M. Kouttab, A.S. De Groot, *Vaccine* 16 (1998) 1880–1884.
- [30] A. Kondo, J. Sidney, S. Southwood, M.F. del Guercio, E. Apella, H. Sakamoto, E. Celis, H.M. Grey, R.W. Chesnut, R.T. Kubo, A. Sette, *J. Immunol.* 155 (1995) 4307–4312.
- [31] C. Schönbach, K. Miwa, M. Ibe, H. Shiga, K. Nokihara, M. Takiguchi, *J. Immunol.* (1996) 5951–5985.

- [32] V. Brusic, C. Schönbach, M. Takiguchi, V. Ciesielski, L.C. Harrison, *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 4 (1997) 75–83.
- [33] K. Udaka, K.H. Weismuller, S. Kienle, G. Jung, H. Tamamura, H. Yamagishi, K. Okamura, P. Walden, T. Suto, T. Kawasaki, *Immunogenetics* (2000) 816–826.
- [34] J. Hammer, E. Bono, F. Gallazzi, C. Belunis, Z. Nagy, F. Sinigaglia, *J. Exp. Med.* 180 (1994) 2353–2358.
- [35] J.B. Rothbard, K. Marshall, K.J. Wilson, L. Fugger, D. Zaller, *Int. Arch. Allergy Immunol.* 105 (1994) 1–7.
- [36] S. Southwood, J. Sidney, A. Kondo, M.F. del Guercio, E. Apella, S. Hoffman, R.T. Kubo, R.W. Chesnut, H.M. Grey, A. Sette, *J. Immunol.* 160 (1998) 3363–3373.
- [37] R.R. Mallios, *Bioinformatics* 17 (2001) 942–948.
- [38] T. Sturniolo, E. Bono, J. Ding, L. Radrizzani, O. Tuereci, U. Sahin, M. Braxenthaler, F. Gallazzi, M.P. Protti, F. Sinigaglia, J. Hammer, *Nat. Biotechnol.* 17 (1999) 433–534.
- [39] V. Brusic, G. Rudy, M.C. Honeyman, J. Hammer, L. Harrison, *Bioinformatics* 14 (1998) 121–130.
- [40] V. Brusic, J. Zeleznikow, T. Sturniolo, E. Bono, J. Hammer, in: *Proceeding of ICONIP99*, IEEE, 1999, pp. 603–609.
- [41] V. Brusic, G. Rudy, L.C. Harrison, in: R. Stonier, X.H. Yu (Eds.), *Complex Systems: Mechanism of Adaptation*, IOS Press, Amsterdam, 1994, pp. 253–260.
- [42] H.P. Adams, J.A. Koziol, *J. Immunol. Meth.* 185 (1995) 181–190.
- [43] K. Gulukota, J. Sidney, A. Sette, C. DeLisi, *J. Mol. Biol.* 267 (1997) 1258–1267.
- [44] M. Milik, D. Sauer, A.P. Brunmark, L. Yuan, A. Vitiello, M.R. Jackson, P.A. Peterson, J. Skolnick, C.A. Glass, *Nat. Biotechnol.* 16 (1998) 753–756.
- [45] M.C. Honeyman, V. Brusic, N.L. Stone, L.C. Harrison, *Nat. Biotechnol.* 16 (1998) 966–969.
- [46] F. Borrás-Cuesta, J. Golvano, M. García-Granero, P. Sarobe, J. Riezu-Boj, E. Huarte, J. Lasarte, *Hum. Immunol.* (2000) 266–278.
- [47] P. Baldi, P.S. Brunak, in: *Bioinformatics. The Machine Learning Approach*, MIT Press, Cambridge, 1998.
- [48] H. Mamitsuka, *Proteins* 33 (1998) 460–474.
- [49] V. Brusic, N. Petrovsky, G. Zhang, V. Bajic, *Immunol. Cell Biol.* 80 (2002) 280–285.
- [50] L.J. Stern, D.C. Wiley, *Structure* (1994) 245–251.
- [51] R.L. Stanfield, I. Wilson, *Curr. Opin. Struct. Biol.* 5 (1995) 103–113.
- [52] D. Rognan, L. Scapozza, G. Folkers, A. Daser, *Biochemistry* 33 (1994) 11476–11485.
- [53] D. Rognan, S.L. Laumoller, A. Holm, S. Buus, V. Tschinke, *J. Med. Chem.* 42 (1999) 4650–4658.
- [54] O. Schueler-Furman, R. Elber, H. Margalit, *Fold. Des.* 3 (1998) 549–564.
- [55] O. Schueler-Furman, Y. Altuvia, A. Sette, H. Margalit, *Protein Sci.* 9 (2000) 1838–1846, I.A..
- [56] Doytchinova, D.R. Flower, *Immunol. Cell Biol.* 80 (2002) 270–279.
- [57] C.A. Del Carpio, T. Hennig, S. Fickel, A. Yoshimori, *Immunol. Cell Biol.* 80 (2002) 286–299.
- [58] S. Manici, T. Sturniolo, M.A. Imro, J. Hammer, F. Sinigaglia, C. Noppen, G. Spagnoli, B. Mazzi, M. Bellone, P. Dallabona, M.P. Protti, *J. Exp. Med.* 189 (1999) 871–876.
- [59] J.L. Vissers, J.J. De Vries, M.W. Schreurs, L.P. Engelen, E. Oosterwijk, C.G. Figdor, G.J. Adema, *Cancer Res.* 59 (1999) 5554–5559.
- [60] H.M. Zarour, J.M. Kirkwood, L.S. Kierstead, W. Herr, C.L. Slingluf Jr., J. Sidney, A. Sette, W.J. Storkus, *Proc. Natl. Acad. Sci. USA* 97 (2000) 400–405.
- [61] R. Khanna, S.R. Burrows, J. Nicholls, L.M. Poulsen, *Eur. J. Immunol.* 28 (1998) 451–458.
- [62] V. Brusic, K. Bucci, C. Schönbach, N. Petrovsky, J. Zeleznikow, J.W. Kazura, *J. Mol. Graph. Model* 19 (2001) 405–411.
- [63] X. Jin, C.G. Roberts, D.F. Nixon, J.T. Safrit, L.Q. Zhang, Y.X. Huang, N. Bhardwaj, B. Jesdale, A.S. De Groot, R.A. Koup, *AIDS Res. Hum. Retroviruses* 16 (2000) 67–76.
- [64] C. De Lalla, T. Sturniolo, L. Abbruzzese, J. Hammer, A. Sidoli, F. Sinigaglia, P. Panina-Bordignon, *J. Immunol.* 163 (1999) 1725–1729.
- [65] M. Panigada, T. Sturniolo, G. Besozzi, M.G. Boccieri, F. Sinigaglia, G.G. Grassi, F. Grassi, *Infect. Immun.* 70 (2002) 79–85.
- [66] H.M. Zarour, B. Maillere, V. Brusic, K. Coval, E. Williams, S. Pouvelle-Moratille, F. Castelli, S. Land, J. Bennouna, T. Logan, J.M. Kirkwood, *Cancer Res.* 62 (2002) 213–218.
- [67] C. Schönbach, Y. Kun, V. Brusic, *Immunol. Cell Biol.* 80 (2002) 300–306.
- [68] A.S. De Groot, C. Saint-Aubin, A. Bosma, H. Sbai, J. Rayner, W. Martin, *Emerg. Infect. Dis.* 7 (2001) 706–713.
- [69] A.S. De Groot, H. Sbai, C.S. Aubin, J. McMurphy, W. Martin, *Immunol. Cell Biol.* 80 (2002) 255–269.
- [70] V.B. Bajic, *Brief. Bioinform.* 1 (2000) 214–228.
- [71] K. Yu, N. Petrovsky, C. Schonbach, J.Y. Koh, V. Brusic, *Mol. Med.* 8 (2002) 137–148.