BIOINFORMATICS

Vol. 19 no. 8 2003, pages 1009–1014
DOI: 10.1093/bioinformatics/btg108

ProPred1: prediction of promiscuous MHC Class-I binding sites

*Harpreet Singh and G.P.S. Raghava**

Institute of Microbial technology, Chandigarh 160036, India

Received on July 18, 2002; revised on July 18, 2002; accepted on January 3, 2003

ABSTRACT

Summary: ProPred1 is an on-line web tool for the prediction of peptide binding to MHC class-I alleles. This is a matrix-based method that allows the prediction of MHC binding sites in an antigenic sequence for 47 MHC class-I alleles. The server represents MHC binding regions within an antigenic sequence in user-friendly formats. These formats assist user in the identification of promiscuous MHC binders in an antigen sequence that can bind to large number of alleles. ProPred1 also allows the prediction of the standard proteasome and immunoproteasome cleavage sites in an antigenic sequence. This server allows identification of MHC binders, who have the cleavage site at the C terminus. The simultaneous prediction of MHC binders and proteasome cleavage sites in an antigenic sequence leads to the identification of potential T-cell epitopes.

Availability: Server is available at http://www.imtech.res.in/raghava/propred1/. Mirror site of this server is available at http://bioinformatics.uams.edu/mirror/propred1/.

Supplementary information: Matrices and document on server are available at http://www.imtech.res.in/raghava/propred1/page2.html

Contact: raghava@imtech.res.in

INTRODUCTION

The altered proteins of self or pathogenic origin are fragmented to small peptides by the proteasomes. Some of these peptides bind to the class-I major histocompatibility complex (MHC) molecules. The resulting MHC–peptide complex is recognized by T-cells. Identification of peptides that will be processed and presented with MHC molecules is crucial for the development of peptide-based vaccines. The experimental identification of these peptides is very time consuming and costly, therefore efforts are being made to use computational methods in prediction of immunogenic regions in an antigenic sequence. The computer-based predictions can effectively reduce the labor involved in wet lab experiments in the identification of immunogenic regions (Hagmann, 2000).

Another interest of the immunologists is the identification of cross-reactive or promiscuous antigenic regions in sequence; these regions can bind to many MHC alleles (Sturniolo *et al.*, 1999). In this direction, we have previously developed ProPred (Singh and Raghava, 2001) for predicting the promiscuous MHC class-II binding peptides. A number of methods have already been developed for the prediction of binding site of various MHC class-I alleles e.g. (i) BIMAS for 41 alleles (Parker *et al.*, 1994). http://bimas.dcrt.nih.gov/molbio/hla_bind/); and (ii) SYFPEITHI for 19 MHC alleles (Rammensee *et al.*, 1999), http://syfpeithi.bmi-heidelberg.com/Scripts/MHCServer.dll/EpPredict.htm).

The MHC class-I molecules recognize the fragments of the antigenic protein after its processing by proteasome (Rock and Goldberg, 1999). Thus, proteasome plays a vital role in identification of potential T-cell epitopes (Niedermann *et al.*, 1999). It has been shown experimentally that proteasome generates the C terminus of MHC class-I binding peptide (Craiu *et al.*, 1997; Stoltze *et al.*, 1998). Recently, numerous methods have been developed for prediction of proteasome cleavage sites (Holzhutter *et al.*, 1999; Kuttler *et al.*, 2000; Nussbaum *et al.*, 2001) and attempt have been made to combine those with the prediction of MHC binding peptide like MAPPP (http://www.mpiib-berlin.mpg.de/MAPPP/).

In this paper we have described a matrix based web server ProPred1, developed for the prediction of MHC class-I binding peptides. The matrices used in ProPred1 have been obtained from BIMAS server and from the literature. The server gives the output of predicted MHC binder for selected MHC alleles in graphical or text format. These display formats help the user in easy detection of the promiscuous MHC binding regions in their query sequence.

We also have made an attempt for the simultaneous prediction of MHC binders and proteasome cleavage sites in a protein sequence. This server has implemented the matrices described by (Toes *et al.*, 2001) for the identification of proteasome (standard/constitutive proteasome and immunoproteasome) cleavage sites in an antigenic sequence.

*To whom the correspondence should be addressed at Bioinformatics Centre, Institute of Microbial Technology, Sector 39A, Chandigarh, India.

Table 1. The references of the quantitative matrices used in ProPred1 server

Name of MHC allele	Reference	Comment
HLA-A2.1	(Ruppert *et al.*, 1993)	Addition matrix
HLA-B*0702	(Sidney *et al.*, 1996)	Addition matrix
HLA-B51	(Sidney *et al.*, 1996)	Addition matrix
HLA-B*5301	(Sidney *et al.*, 1996)	Addition matrix
HLA-B*5401	(Sidney *et al.*, 1996)	Addition matrix
All other MHC alleles	Unpublished	Multiplication matrices (BIMAS Server)

Recently, (Kessler *et al.*, 2001; Ayyoub *et al.*, 2002) have demonstrated that MHC binders having proteasome cleavage site at their C terminus have high potency to become T-cell epitopes. These observations were implemented in ProPred1 in order to identify the potential T-cell epitopes. The ProPred1 allows identification of the predicted MHC binders who have predicted proteasome site at C terminus. In brief, the server assists users in identification of promiscuous MHC binders and potential T-cell epitopes in an antigenic sequence.

MATERIALS AND METHODS

Source of weight matrices

The most of the quantitative matrices have been obtained from BIMAS server and a few matrices were obtained from literature (See Table 1). The selection of matrices was not based on performance rather easily availability in the literature. The matrices obtained from BIMAS server are 'multiplication matrices,' where the score is calculated by multiplying scores of each position. For example, score of peptide 'PACDPGRAA' can be calculated by following equation.

$$\text{Score} = \text{P}(1) \times \text{A}(2) \times \text{C}(3) \times \text{D}(4) \times \text{P}(5) \times \text{G}(6) \times \text{R}(7) \times \text{A}(8) \times \text{A}(9) \quad (1)$$

Where $\boldsymbol{P}(\mathbf{1})$ is score of $\boldsymbol{P}$ at position **1**. The matrices obtained from the literature are 'Addition Matrices', where score is calculated by summing the scores of each position. For example, score for above peptide 'PACDPGRAA' is calculated as follows:

$$\text{Score} = \text{P}(1) + \text{A}(2) + \text{C}(3) + \text{D}(4) + \text{P}(5) + \text{G}(6) + \text{R}(7) + \text{A}(8) + \text{A}(9). \quad (2)$$

Threshold score

The selection of cut off threshold is crucial in matrix-based methods, because the stringency of prediction varies with the threshold. The threshold also provides the confidence to the user in their prediction. Thus, it is important to calculate the threshold score in advance for each allele so that binders and non-binders can be distinguished with confidence. In order to calculate the threshold there is a need of sufficient data of MHC binders and non-binders. Unfortunately, due to lack of data (particularly MHC non-binders), it is practically impossible to compute the threshold score for most of the alleles. To overcome the above problem, we adopted the following uniform procedure for the calculation of threshold score for each allele.

(i) All proteins were obtained from SWISSPROT databases for creating the overlapping peptides of length nine. For example, a protein of length $\boldsymbol{n}$ will have $(\boldsymbol{n} + \mathbf{1} - \mathbf{9})$ overlapping peptides.

(ii) The score of all natural 9mer peptides have been calculated using weight matrix of that allele. These peptides have been sorted on the basis of score in descending order and top 1% natural peptides have been extracted. The minimum score that we called threshold score was determined from these selected peptides. Similarly, threshold scores at 2, 3 ... 10% were calculated.

(iii) Step 1 and 2 were repeated for each MHC allele in order to calculate threshold score at different percent for each allele used in ProPred1.

Prediction of MHC binders

First, all possible overlapping 9mer peptides were generated for a given antigen sequence. The score of these 9mer peptides were calculated using quantitative matrix of selected MHC alleles. In next step, all peptides having score greater than selected threshold score (e.g. at 4%) were assigned as predicted binders for selected MHC allele.

Weight matrices for proteasomal cutters

The weight matrices used in ProPred1 for the prediction of standard proteasome and immunoproteasome, have been obtained from the work of (Toes *et al.*, 2001). The matrix for standard proteasome matrix was derived from Table 1 of (Toes *et al.*, 2001), where each value has been divided by one thousand, in order to rationalize the score. The derived matrix is an 'addition matrix' where score of a peptide is calculated by summing the score of each residue. Similar procedure has been adopted for deriving the matrix for immunoproteasome, from the Table 2 of (Toes *et al.*, 2001) The major difference between proteasomes matrices and MHC matrices is that proteasomes consider the peptide of length 12 instead of nine. In case of proteasome the cutting site is at the center of 12mer peptide.

Table 2. The percent coverage of binders for different MHC alleles using ProPred1 at default threshold 4%.

MHC Alleles (Total binder, % Coverage)		
HLA-A*0201(1221, 75%)	H2-Db (189, 74%)	HLA-B*0702(79, 92%)
HLA-A*0205(28, 61%)	H2-Dd (89, 74%)	HLA-B*2705(145, 98%)
HLA-A*1101(116, 80%)	H2-Kb (116, 78%)	HLA-B*3501(254, 84%)
HLA-A*3101(33, 70%)	H2-Kd (277, 83%)	HLA-B*5101(51, 92%)
HLA-A1 (128, 77%)	H2-Kk (28, 86%)	HLA-B*5102(33, 94%)
HLA-A2 (976, 69%)	H2-Ld (113, 60%)	HLA-B*5103(30, 97%)
HLA-A2.1 (77, 64%)	HLA-B*5401(60, 100%)	HLA-B8 (130, 75%)
HLA-A24 (60, 70%)	HLA-B61 (22, 95%)	HLA-B62 (29, 55%)
HLA-A3 (191, 64%)	HLA-B14 (81, 75%)	HLA-Cw*0401(20,80%)
HLA-B7 (134, 81%)	HLA-B*5301 (64, 95%)	

Computation of threshold score

The threshold scores for standard proteasome and immunoproteasome have been calculated at different percent by using the approach described above for calculation of threshold score for MHC alleles. The calculation of threshold score of proteasome matrices requires the 12mer overlapping peptides. The matrices and cutoff scores at different thresold $1, 2, \ldots 10\%$ are available at URL http://www.imtech.res.in/raghava/propred1/matrices/matrix.html.

Prediction of cleavage site in antigen sequence

In order to predict proteasome cleavage sites in an antigenic sequence. The overlapping 12mer peptides were generated for antigenic sequence and score of these peptides were calculated using weight matrix of proteasome. In next step, all peptides having score greater than selected threshold score (e.g. at 5%) are considered as peptides having proteasome cleavage site. The center positions of these peptides (6-position left and 6-position right) are considered as predicted proteasome cleavage site. Similar approach has been utilized for prediction of peptides having immunoproteasome cleavage site.

Simultaneous prediction of MHC binders and proteasome cleavage sites

The predicted MHC binders were filtered based on prediction of proteasome cleavage sites in an antigenic sequence. Firstly, the server computes the predicted MHC binders and their C terminus position for a selected MHC allele in an antigenic sequence. Secondly server predicts the cleavage sites of proteasome in an antigenic sequence at given threshold (e.g. at 5%). Finally, all predicted MHC binding peptides whose C terminal position coincides with proteasomes cleavage sites were filtered. In other words, server removes the MHC binders, who does not have cleavage site at C terminus.

RESULTS

Percent coverage

The fundamental question in MHC prediction is whether the prediction of binders is worth or not. In other words, whether the server can distinguish between binders and non-binders with significant accuracy. Thus, it is important to evaluate the performance of various matrices. We calculated the percent coverage (percent of binders correctly predicted as binder) for each allele for which sufficient amount of data was available. The data of binders and non-binders corresponding to each MHC alleles has been extracted from MHCBN database (Bhasin *et al.*, 2002 http://www.imtech.res.in/raghava/mhcbn/). The number of binders varies from 20 to 1200 (See Table 2). The default threshold score 4% (score at which sensitivity and specificity are nearly the same for most of the MHC alleles used in this study) was used for the prediction of the binders. The percent coverage has been calculated from predicted results. The value of percent coverage varies from 50 to 98% as shown in Table 2. These results clearly indicate that in most of the cases percent coverage is more than 80% which is reasonably good. Almost all alleles showed reasonable percent coverage, which means threshold criteria and matrices used in ProPred1 are beneficial for experimental scientists.

Performance of ProPred1

Nonetheless, the percent coverage is a useful measure to evaluate the ability of method for the identification of binders from a given sequence, but it does not provide any information about predicted false positive binders or accuracy of prediction etc. Following three parameters are commonly used to measure the performance of a method in the field of immunoinformatics

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100 \quad (3)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100 \quad (4)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (5)$$

The correlation coefficient (CC) is a rigorous parameter to measure the performance of a method, which can be defined as:

$$\text{CC} = \frac{(TP^{*}TN) - (FN^{*}FP)}{\sqrt{(TP + FN)^{*}(TN + FP)^{*}(TP + FP)^{*}(TN + FN)}}. \quad (6)$$

Where *TP* and *TN* are correctly predicted binders and non-binders respectively. *FP* and *FN* are wrongly predicted binders and non-binders respectively.

Table 3. The comprehensive performance of ProPred1 at different thresholds for MHC allele HLA-A*0201 (1220 binders & 56 non-binders) and H2-Kb (116 binders & 20 non-binders)

Thresholds	HLA-A*0201 Sensitivity (%)	Specificity (%)	Accuracy (%)	Correlation coefficient	H2-Kb Senstivity (%)	Specificity (%)	Accuracy (%)	Correlation coefficient
1%	36	98	38	0.1314	68	88	70	0.372
2%	57	93	58	0.1854	73	81	74	0.3775
3%	66	80	67	0.1783	78	81	78	0.4209
4%	75	78	75	0.2179	78	69	77	0.3367
5%	81	67	80	0.2151	82	62	80	0.3418

Table 4. The performance of Propred1 for HLA-A*0201 on data set of 128 peptides (19 high-affinity & 27 intermediate-affinity binders) of protein PRAME. These peptides were obtained from Table 1 of Kessler *et al.* (2001)

Threshold (%)	Correctly predicted high-affinity binders (out of 19)	Correctly predicted intermediate-affinity binders (out of 27)
1.0	4 (21%)	1 (4%)
2.0	9 (47%)	6 (22%)
3.0	10 (53%)	14 (52%)
4.0	11 (58%)	15 (56%)
5.0	12 (63 %)	21 (77%)
6.0	13 (68%)	22 (81%)
7.0	13 (68%)	23 (85%)
8.0	15 (79%)	24 (89%)
9.0	18 (95%)	26 (96%)
10.0	19 (100%)	27 (100%)

In this study, we compute all the above parameters for ProPred1 for its comprehensive evaluation. In order to evaluate a method one need sufficient data of experimentally proven MHC binders and non-binders. Unfortunately, most of the alleles have very limited number of binders and non-binders. Thus, the comprehensive evaluation of ProPred1 was performed only for two alleles (HLA-A*0201 & H2-Kb) for which sufficient number of binders and non-binders were available. The peptides for allele HLA-A*0201 (1220 binders & 56 non-binders) and H2-Kb (300 binders & 200 non-binders) were obtained from MHCBN database (Bhasin *et al.*, 2002). The performance of ProPred1 for these two MHC alleles at different percent threshold has been shown in Table 3.

A test for Propred1

MHC binder: The purpose of development of ProPred1 is to effectively reduce number of wet lab experiments involved in the identification of potential T-cell epitopes or suitable vaccine candidates. Recently, (Kessler *et al.*, 2001) have experimentally determined the MHC binders and T-cell epitopes from tumor associated antigenic protein, PRAME. We analyzed the performance of ProPred1 in the identification of experimentally proven MHC binders and T-cell epitopes of PRAME. The sequence of PRAME antigenic protein was obtained from SWISSPROT database. (Kessler *et al.*, 2001) tested 128 peptides and identified 19 as high-affinity binders and 27 intermediate-affinity binders. ProPred1 was used to predict these MHC 128 peptides of PRAME at various thresholds (See Table 4). As shown in Table 4, number of correctly predicted binders (intermediate/high affinity) depends on percent threshold. The ProPred1 predicted all binders correctly at 10% threshold. These results clearly indicate that server has capability to predict the binders with significant accuracy at 4% threshold (default threshold

Potential T-cell epitopes: It has been demonstrated experimentally that MHC binders having proteasomal cleavage site at C-terminus are mostly responsible for the activation of cytotoxic T lymphocytes (CTLs). (Kessler *et al.*, 2001) experimentally identified four regions having HLA-A*201 restricted T-cell epitopes. We tested these regions using ProPred1 server. Firstly, binding regions were predicted at default threshold (4%) in protein PRAME. Secondly, all proteasome sites were predicted at various thresholds. Finally, predicted binders having proteasome cleavage sites at C-terminus were identified. The number of peptides predicted by above falls in regions identified as T-cell epitopes by (Kessler *et al.*, 2001), is shown in Table 5.

It was observed that in the presence of standard proteasome filter at 7%, the server was able to predict the 50% of binding regions that are in agreement with experimentally proven binding regions as demonstrated by (Kessler *et al.*, 2001). Similarly, it has been observed that at 5% of threshold of immunoproteasome filter, the server was able to identify 75% of experimentally determined binding regions. The server was able to predict 75% of binding regions in simultaneous presence of either standard proteasome or immunoproteasome filters at 5% threshold.

Table 5. Propred1 was tested on four regions of PRAME (A: 90-116; B:133-159; C: 290-316; D: 415-441) which were identified as T-cell epitopes by Kessler *et al.* (2001). Propred1 was first used to predict the HLA-A*0201 restricted peptide at 4% threshold. The column 2 shows the number of predicted peptides and regions (in bracket), which agree with the experimentally identified epitopes

Name of filter	Correctly predicted T-cell epitopes in protein PRAME at different thresholds (out of 4)			
	2%	3%	5%	7%
Standard proteasome	0	1 (A)	1 (A)	2 (A,D)
Immunoproteasome	2 (A,D)	2 (A,D)	3 (A,C,D)	3 (A,C,D)
Immunoproteasome or standard proteasome	2 (A,D)	2 (A,D)	3 (A,C,D)	3 (A,C,D)

Hence, all the analysis clearly indicate that it is worth using ProPred1 for the identification of MHC binding regions having proteasomal cleavage site at their C terminus or potential T-cell epitopes. It is not possible to determine the default threshold in case of prediction of proteasomes cleavage site, because sufficient data is not available, however, we suggest default threshold 5% based on our experience.

Mapping of predicted binders on antigen sequence

One of the important aspects of MHC prediction is the representation of binding peptides found within the antigenic sequence. This can be achieved by developing a powerful web interface for the prediction method. The ProPred1 provides three major options to visualize results in user-friendly formats, including most popular tabular format. Following is the brief description of these options.

Graphical display: The graphical output represents the quantitative estimation of MHC binding propensity of the antigenic sequence. The server represents results in graphical format (X–Y Plot), where amino acid sequence is shown along the X-axis and peptide score is shown along the Y-axis. Each binder is represented as a peak crossing the dashed threshold line in the image. It allows user to locate the promiscuous regions in the query sequence by looking at the peaks in graphs for different MHC alleles.

Text or HTML format: This option of server presents the MHC binders within antigenic sequence in text or HTML format. It has two sub-options. The first sub-option displays the predicted MHC binders in separate lines along the antigen sequence. This option uses the separate lines for representing all the predicted overlapping binders within the sequence. This suboption is very useful for viewing the predicted overlapping binders. The second sub-option of the server represents predicted binder by different color i.e. blue. The first position of each binder is shown by red color so that user can easily distinguish the overlapping peptides. This option is useful in locating promiscuous MHC binders.

Tabular format: This is the most widely used option for the display of results in most of the web servers of MHC prediction. This option displays the peptides sorted in descending order of their score. The server creates a separate table corresponding to each selected allele.

IMPLEMENTATION

The common gateway interface (CGI) script of ProPred1 is written using PERL. It has been installed on a Sun Server (420E) under UNIX (Solaris 7) environment and launched using Apache web server. The protein sequences can be submitted to the ProPred1 by cut-and-paste technique or by directly uploading a sequence file. The server uses ReadSeq (developed by Dr Don Gilbert) to parse the input sequence, therefore it can accept most of the commonly used sequence formats. The server allows user to select the threshold for their prediction. The threshold plays a vital role in determining the stringency of prediction. Lower the threshold, higher is the stringency of prediction i.e. lowers rate of false positives and higher rate of false negatives in the prediction. In contrast, a higher threshold value (low stringency) corresponds to a higher rate of false positives and a lower rate of false negatives.

Limitations of ProPred1

All the matrices used in server were obtained from various servers and from the literature. The base of selection of matrices is on its availability from single source and not on the performance. Thus, it is not necessary that we are using best matrix for an allele if more than one matrix is available in the literature. In this server only 9mer peptide length are predicted not 8mer or 10mer. Thus it is possible that ProPred1 may miss potential 8mer and 10mer binders. The matrices for predicting ProPred1 were obtained from the paper of Toes *et al.* (2001), where their values were obtained for enolase-I protein. We have used these values for all predictions, it is not necessary that this generalization will work for all the proteins.

ACKNOWLEDGEMENTS

We are grateful to the anonymous referees for their valuable suggestions. Authors are thankful to Mr Manoj Bhasin for his help in data analysis. We are also thankful to Dr Balvinder Singh and Dr G. C. Varshney for checking the manuscript.

REFERENCES

Ayyoub,M., Stevanovic,S., Sahin,U., Guillaume,P., Servis,C., Rimoldi,D., Valmori,D., Romero,P., Cerottini,J.C., Rammensee,H.G. *et al.* (2002) Proteasome-assisted identification of a SSX-2-derived epitope recognized by tumor-reactive CTL infiltrating metastatic melanoma. *J. Immunol.*, **168**, 1717–1722.

Bhasin,M., Singh,H. and Raghava,G.P.S. (2002) MHCBN: a comprehensive database of MHC binding and non-binding peptides. *Bioinformatics*, **19**, 665–666.

Craiu,A., Akopian,T., Goldberg,A. and Rock,K.L. (1997) Two distinct proteolytic processes in the generation of a major histocompatibility complex class I-presented peptide. *Proc. Natl Acad. Sci. USA*, **94**, 10850–10855.

Hagmann,M. (2000) Computers aid vaccine design. *Science*, **290**, 80–82.

Holzhutter,H.G., Frommel,C. and Kloetzel,P.M. (1999) A theoretical approach towards the identification of cleavage-determining amino acid motifs of the 20S proteasome. *J. Mol. Biol.*, **286**, 1251–1265.

Kessler,J.H., Beekman,N.J., Bres-Vloemans,S.A., Verdijk,P., vanVeelen,P.A., Kloosterman-Joosten,A.M., Vissers,D.C.J., ten Bosch,G.J.A., Kester,M.G.D., Sijts,A., Drijfhout,J.W. *et al.* (2001) Efficient identification of novel HLA-A*0201-presented cytotoxic T lymphocyte epitopes in the widely expressed tumor antigen PRAME by proteasome-mediated digestion analysis. *J. Exp. Med.*, **193**, 73–88.

Kuttler,C., Nussbaum,A.K., Dick,T.P., Rammensee,H.G., Schild,H. and Hadeler,K.P. (2000) An algorithm for the prediction of proteasomal cleavages. *J. Mol. Biol.*, **298**, 417–429.

Niedermann,G., Geier,E., Lucchiari-Hartz,M., Hitziger,N., Ramsperger,A. and Eichmann,K. (1999) The specificity of proteasomes: impact on MHC class I processing and presentation of antigens. *Immunol. Rev.*, **127**, 29–48.

Nussbaum,A.K., Kuttler,C., Hadeler,K.P., Rammensee,H.G. and Schild,H. (2001) PAProC: a prediction algorithm for proteasomal cleavages available on the WWW. *Immunogenetics*, **53**, 87–94.

Parker,K.C., Bednarek,M.A. and Coligan,J.E. (1994) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J. Immunol.*, **152**, 163–175.

Rammensee,H., Bachmann,J., Emmerich,N.P., Bachor,O.A. and Stevanovic,S. (1999) SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics*, **50**, 213–219.

Rock,K.H. and Goldberg,A.L. (1999) Degradation of cell proteins and the generation of MHC class I presented peptides. *Annu. Rev. Immunol.*, **17**, 739–779.

Ruppert,J., Sidney,J., Celis,E., Kubo,R.T., Grey,H.M and Sette,A. (1993) Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules. *Cell*, **74**, 929–937.

Sidney,J., Southwood,S., Guercio,M.D., Grey,H.M., Chesnut,R.W., Kubo,R.T. and Sette,A. (1996) Specificity and degeneracy in peptide binding to HLA-B7 like class 1 molecules. *J. Immunol.*, **157**, 3480–3490.

Singh,H. and Raghava,G.P.S. (2001) ProPred: prediction of HLA-DR binding sites. *Bioinformatics*, **17**, 1236–1237.

Stoltze,L., Dick,T.P., Deeg,M., Pommerl,B., Rammensee,H.G. and Schild,H. (1998) Generation of the vesicular stomatitis virus nucleoprotein cytotoxic T lymphocyte epitope requires proteasome-dependent and independent proteolytic activities. *Eur. J. Immunol.*, **28**, 4029–4036.

Sturniolo,T., Bono,E., Ding,J., Raddrizzani,L., Tuereci,O., Sahin,U., Braxenthaler,M., Gallazzi,F., Protti,M.P., Sinigaglia,F. and Hammer,J. (1999) Generation of tissue-specific and promiscuous HLA ligand database using DNA microarrays and virtual HLA class II matrices. *Nat. Biotech.*, **17**, 555–561.

Toes,R.E., Nussbaum,A.K., Degermann,S., Schirle,M., Emmerich,N.P.N., Kraft,M., Laplace,C., Zwinderman,A., Dick,T.P., Muller,J. *et al.* (2001) Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. *J. Exp. Med.*, **194**, 1–12.