APPLICATION NOTE

Appl Bioinformatics 2006; 5 (1): 55-61
1175-5636/06/0001-0055/$39.95/0

MHCPred 2.0

An Updated Quantitative T-Cell Epitope Prediction Server

Pingping Guan, *Channa K. Hattotuwagama*, *Irini A. Doytchinova* and *Darren R. Flower*

Edward Jenner Institute for Vaccine Research, Compton, Berkshire, UK

Abstract

The accurate computational prediction of T-cell epitopes can greatly reduce the experimental overhead implicit in candidate epitope identification within genomic sequences. In this article we present MHCPred 2.0, an enhanced version of our online, quantitative T-cell epitope prediction server. The previous version of MHCPred included mostly alleles from the human leukocyte antigen A (*HLA-A*) locus. In MHCPred 2.0, mouse models are added and computational constraints removed. Currently the server includes 11 human HLA class I, three human HLA class II, and three mouse class I models. Additionally, a binding model for the human transporter associated with antigen processing (TAP) is incorporated into the new MHCPred. A tool for the design of heteroclitic peptides is also included within the server. To refine the veracity of binding affinities prediction, a confidence percentage is also now calculated for each peptide predicted.

Availability: As previously, MHCPred 2.0 is freely available at the URL http://www.jenner.ac.uk/MHCPred/

Contact: Darren R. Flower (darren.flower@jenner.ac.uk)

Background

Human major histocompatibility complex (MHC) molecules are polymorphic membrane proteins. One of the principal functions of the immune system is to recognise and eliminate foreign antigens in the body, such as viruses, bacteria and parasites.[1,2] Extracellular antigens can be recognised and destroyed by macrophages, T cells and B cells. The intracellular antigens, however, do not have direct contact with the immune system, and they have to be eliminated with the help of MHC proteins. The MHC proteins take up degraded intracellular antigenic fragments, or epitopes, and present them to T cells to induce an immune system response.[3-5]

The prediction of epitopes within a given protein sequence is an ongoing research topic. Many algorithms have been developed to predict T-cell epitopes:[6] motif search methods,[7] quantitative matrices,[8] sequence profiles,[9] structure-based approaches[10] and artificial neural networks,[11] etc. Many algorithms have been implemented as Internet-based servers, allowing users to predict T-cell epitopes in protein sequences. Internet-based epitope prediction is an effective means of efficiently disseminating immunoinformatics technology on a wide scale, as it helps laboratory-based scientists worldwide to use these methods successfully. For this purpose, an Internet implementation of the additive method,[12,13] called MHCPred, was produced.[14-16] MHCPred 2.0 includes many enhancements, notably an increased number of models, covering all human and mouse MHC alleles we have generated thus far.

Resource Description

Capabilities

The Algorithm

The additive method uses a modified Free-Wilson approach,[17] which assumes that each constituent makes an additive and independent contribution to the biological activity of the molecule.[12,13] The additive method considers three types of interactions that affect binding affinity: (i) the interaction between each amino acid and the binding site, (ii) the interaction between adjacent side chains (1-2 interactions) and (iii) the interaction between every second side chain (1-3 interactions). Two types of models are generated by the additive method. One is the single

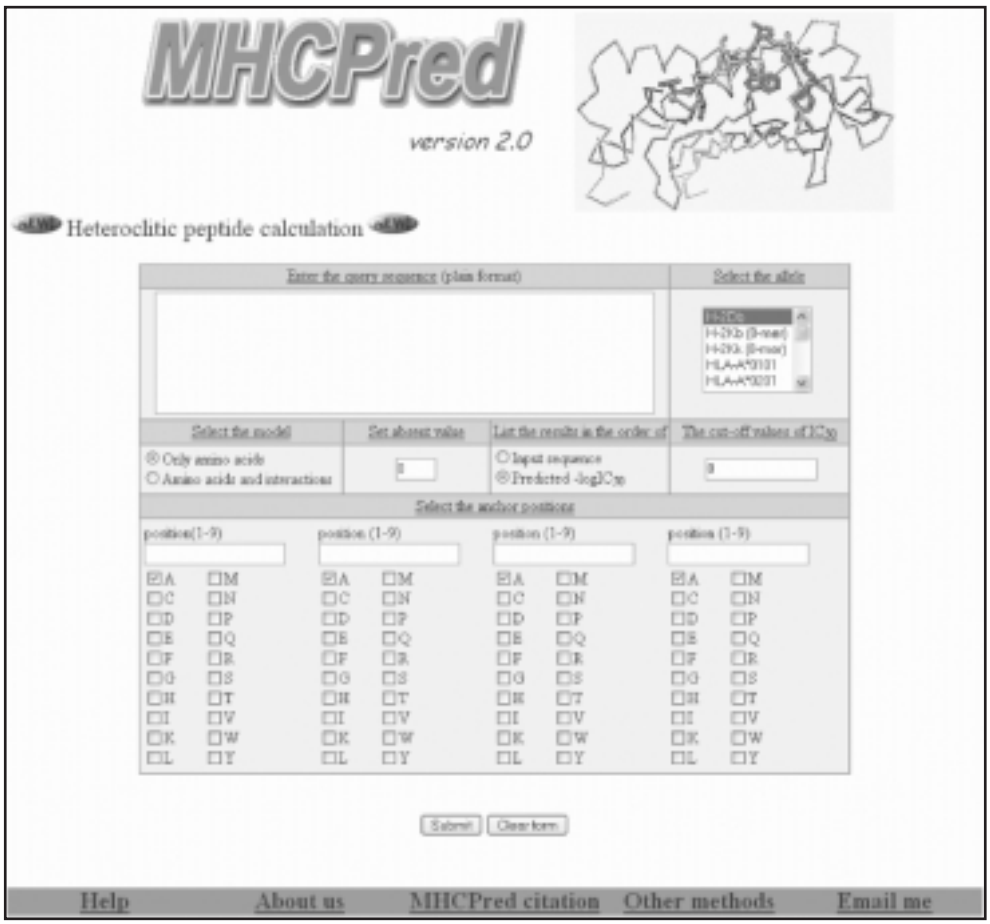


Fig. 1. Graphical user interface of MHCPred 2.0, from which the user can enter the query protein sequence and choose the allele and model used in the prediction. The inhibitory constant $(IC)_{50}$ threshold (nM) is used to restrict the output. If a value is entered, peptides with IC_{50} values higher than the value will not be listed in the output. Also, the user can choose to see an output with only peptides that have preferred residues at given positions. A total of four such positions can be selected.

amino acid model, which only accounts for the binding of each amino acid of the peptide. The other is the amino acid and interactions model, which considers both the contribution of individual amino acids and the interactions between adjacent and every second amino acids.

For a nonamer peptide, the single amino acid model and the amino acid and interactions model are given by equation 1 and equation 2:

$$pIC_{50} = const + \sum_{i=1}^{9} P_i \quad \text{(Eq. 1)}$$

$$pIC_{50} = const + \sum_{i=1}^{9} P_i + \sum_{j=i+1}^{8} P_{ij} + \sum_{k=i+2}^{7} P_{ik} \quad \text{(Eq. 2)}$$

where pIC_{50} is the binding affinity expressed in p-units ($-\log IC_{50}$, where IC is inhibitory constant), the *const* accounts for the peptide backbone contribution, $\sum_{i=1}^{9} P_i$ is the sum of amino acid contributions at each position, $\sum_{j=i+1}^{8} P_{ij}$ is the sum of adjacent peptide side-chain interactions, and $\sum_{k=i+2}^{7} P_{ik}$ is the sum of every second side-chain interaction.

Implementation and Developer Resources

The Server

To facilitate online T-cell epitope prediction, the additive method is implemented as a web server. MHCPred version 2.0 is freely available through the URL http://www.jenner.ac.uk/MHCPred/.[14,15] The web interface is shown in figure 1. Several improvements have been implemented in MHCPred 2.0. The first version of MHCPred only included data for human MHC classes I

and II. Currently, the server has models for a total of 17 human and mouse MHC alleles, of which 14 are human and three are mouse MHC models, as well as a model for transporter associated with antigen processing (TAP).[18] As research continues, models for other alleles will become available. For example, six mouse class II models will be added in due course. A summary of derived measures of statistical quality for each model is presented in table I.

To allow for the analysis of longer proteins, the input sequence length has been extended to 1000 amino acids. The server interface is written in HTML, and the common gateway interface (CGI) program in Perl. The calculation procedure of the CGI program is summarised in figure 2. Only sequences in plain text format are accepted by the server. Both upper and lower cases are accepted, and all non-amino acid characters are deleted before calculation. A pull-down box is used to select alleles (see table I for available choices). In version 1.0 of MHCPred, only one allele could be chosen at a time, whereas in MHCPred 2.0, multiple alleles can be selected.

Two types of models are generated by the additive method. The user can choose to use the single amino acid model, which only considers the interaction between the peptide and the binding site. Alternatively, if the user wants to take into account the interactions between adjacent amino acids of the peptide, the amino acid with interactions model should be used. The query sequence is chopped into nonamer peptides, except for the H-2Kd and H-2Kk model which uses octamers. If the user enters preferred residues and positions, MHCPred 2.0 filters all peptides that lack the specified residues at the specified positions. After processing the input sequence, the program reads the model coefficients. For each amino acid of the peptide, the program reads the corresponding value in the matrix and adds it to the constant value to give the final result. If the amino acid and interactions model is selected, the adjacent and 1-3 amino acid interactions are taken into account and their contributions are added to the result. A sample of the output page is shown in figure 3. The output is arranged in a table, and the input sequence is printed at the beginning of the results table.

Table I. Alleles included in the MHCPred 2.0 server

Class	Model	Allele	No. of peptides	q^2	NC	r^2
I	Human	A*0101	95	0.420	4	0.997
		A*0201	335	0.377	6	0.731
		A*0202	69	0.317	9	0.943
		A*0203	62	0.327	6	0.963
		A*0206	57	0.475	6	0.989
		A*0301	70	0.305	4	0.972
		A*1101	62	0.428	3	0.977
		A*3101	31	0.453	6	0.990
		A*6801	37	0.370	4	0.974
		A*6802	46	0.500	7	0.983
		B*3501	52	0.435	6	0.984
	Mouse	H2-Db	73	0.493	5	0.948
		H2-Kb	55	0.454	6	0.989
		H2-Kk	152	0.456	6	0.933
II	Human	DRB1*0101	90	0.808	8	0.994
		DRB1*0401	185	0.716	4	0.967
		DRB1*0701	84	0.649	7	0.999

NC = number of components that gave the optimal q^2 values; $\boldsymbol{q}^2$ = ability of the models to predict epitopes from sequences; $\boldsymbol{r}^2$ = proportion of variance explained by the model, or how well the models fitted the training data.

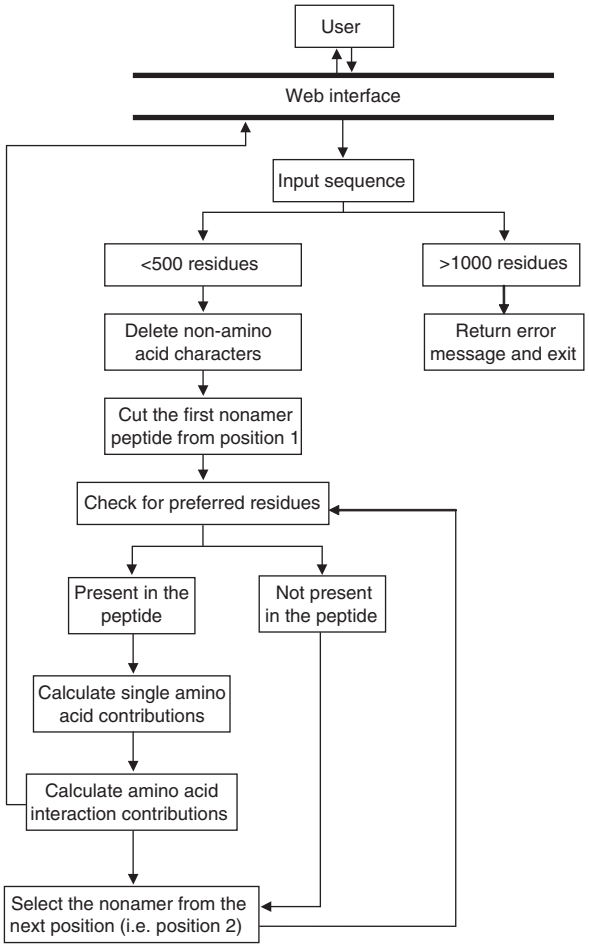
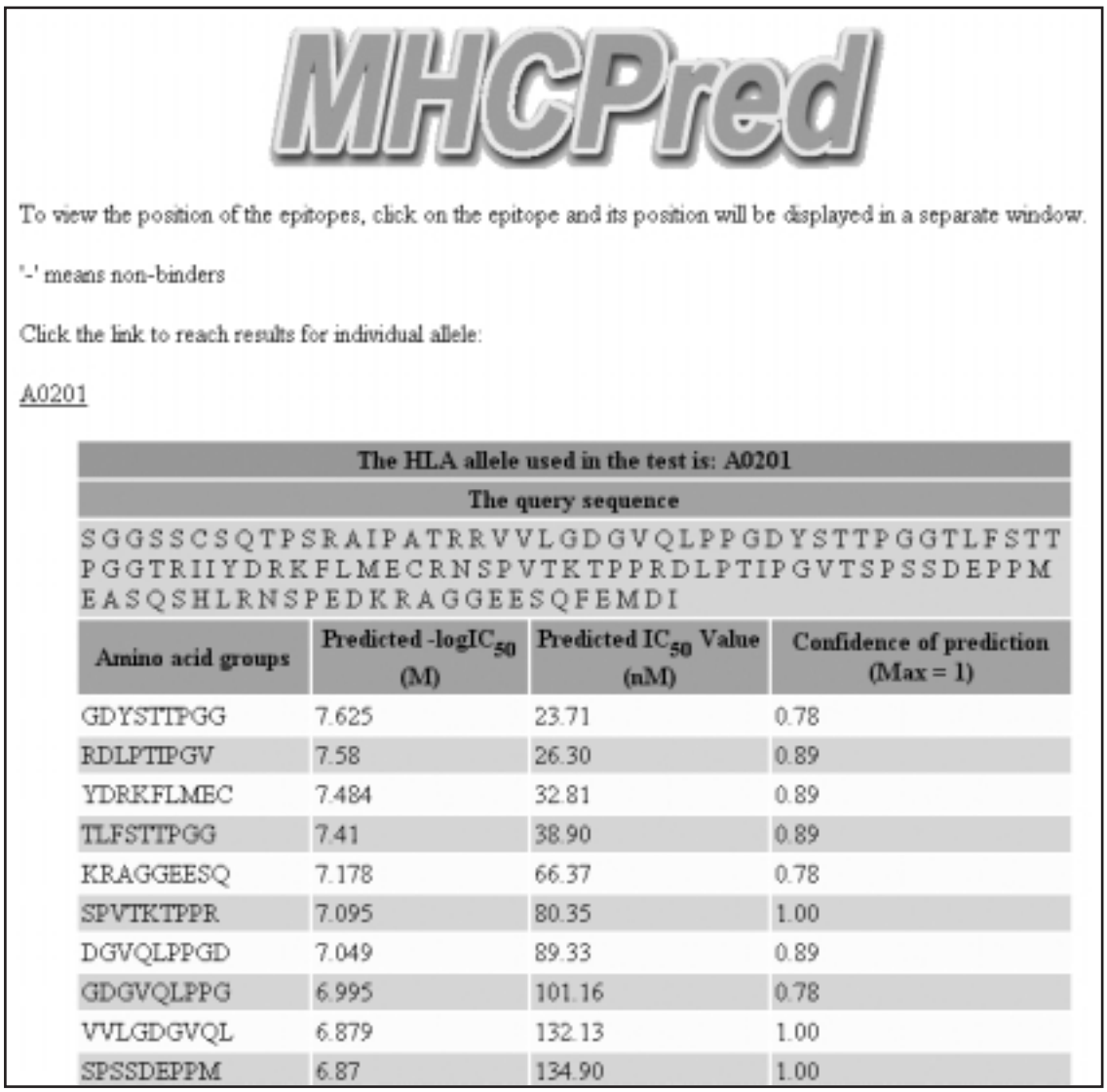


Fig. 2. A flowchart of how the common gateway interface (CGI) program works in MHCPred 2.0.



MHCPred

To view the position of the epitopes, click on the epitope and its position will be displayed in a separate window.

'-' means non-binders

Click the link to reach results for individual allele:

A0201

The HLA allele used in the test is: A0201

The query sequence

SGGSSCSQTPSRAIPATRRVVLGDGVQLPPGDYSTTPGGTLFSTT
PGGTRIIYDRKFLMECRNSPVTKTPPRDLPTIPGVTSPSSDEPPM
EASQSHLRNSPEDKRAGGEESQFEMDI

Amino acid groups	Predicted -logIC$_{50}$ (M)	Predicted IC$_{50}$ Value (nM)	Confidence of prediction (Max = 1)
GDYSTTPGG	7.625	23.71	0.78
RDLPTIPGV	7.58	26.30	0.89
YDRKFLMEC	7.484	32.81	0.89
TLFSTTPGG	7.41	38.90	0.89
KRAGGEESQ	7.178	66.37	0.78
SPVTKTPPR	7.095	80.35	1.00
DGVQLPPGD	7.049	89.33	0.89
GDGVQLPPG	6.995	101.16	0.78
VVLGDGVQL	6.879	132.13	1.00
SPSSDEPPM	6.87	134.90	1.00

Fig. 3. An example of the MHCPred 2.0 output. Peptides generated by the additive method are listed in the first column of the table, and their predicted $-\log IC_{50}$ values and IC_{50} (nM) values are listed in the second and third columns. Peptides that are potential binders are listed at the top of the table, and weak and non-binders are listed towards the end of the page. **IC** = inhibitory constant.

Three parameters can be used to control the format of the output page. Firstly, there are two ways to list the output peptides: (i) in ascending order of IC_{50} (nM) values or (ii) according to their position in the input sequences. Secondly, an IC_{50} cut-off value, where only those peptides that possess a predicted binding affinity lower than this threshold are listed, can be used to control the format. Usually peptides with predicted binding affinities <500nM are good binders, whereas those with binding affinities >5000nM are considered non-binders. If the user does not enter any value, all the peptides generated from the input sequence will be listed. The binding affinities of those with $IC_{50} > 5000$nM are not shown, and are replaced by '-'. Predicted $-\log IC_{50}$ values are also shown in the table output by the program.

The Peptide Library

An option added to the second version of MHCPred is the prediction of binding of mono- and di-amino acid mutations of a peptide. MHCPred 2.0 takes a single nonamer peptide as the input, substitutes the amino acid at a user-specific position with each of the 20 amino acids and calculates the binding affinities of the new peptides. The interface is shown in figure 4. This option is useful in comparing the binding affinities of heteroclitic analogues of the test peptide. It is also useful in examining the effect of different

amino substitutions on peptide affinity. We have recently used a similar approach to generate a series of high-affinity superbinders and anchorless epitopes for HLA-A*0201.[19]

Like the main MHCPred 2.0 webpage, the user can also choose to use either the single amino acid model or the amino acid with interactions model. The user can change one or two positions within the peptide. If the user decides to choose one position, then each of the nine positions of the peptide are substituted with each of the 20 amino acids in turn, generating a total of 180 peptides. If the user wants to change two positions, then two random positions are selected by the server. Each of the 20 amino acids will appear at each of the positions. The output cut-off has three options: (i) the input peptide affinity, (ii) peptides with affinities 5% lower than the input peptide or (iii) the user can select all and see all generated peptides. A sample output is shown in figure 5.

Discussion

Among T-cell epitope prediction servers, MHCPred 2.0 is the only server based on a quantitative structure-activity relationship

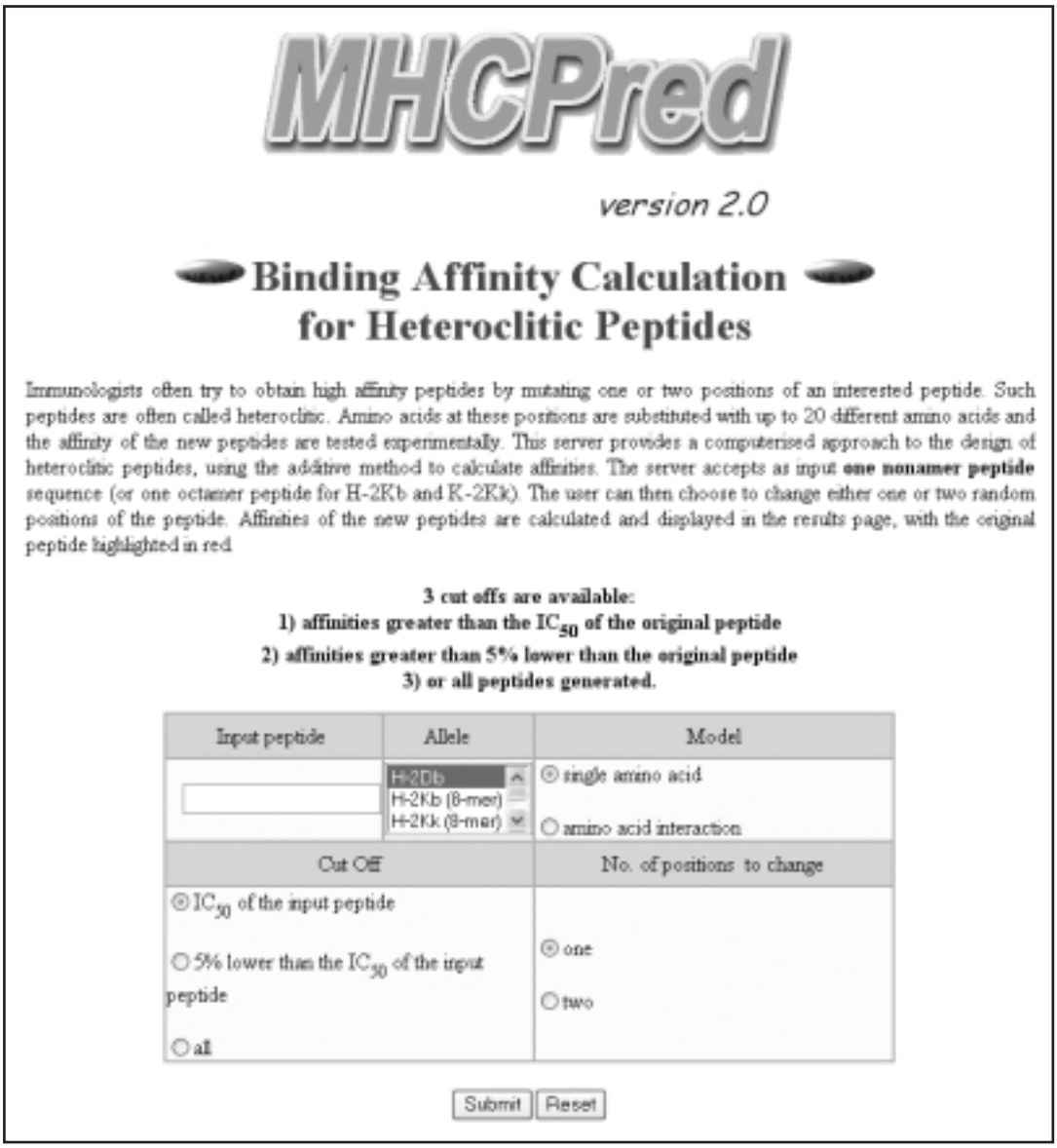


Fig. 4. The graphical user interface of the peptide library. A user can enter the query sequence, choose which allele the peptide is restricted to, and the additive model to use. For the output, the user can choose to modify one or two positions of the peptide and select the cut-off so that peptides below the cut-off will not be displayed.

original sequence:		
HLA allele: **H2Db**		
Predicted -logIC_{50} (nM)	IC_{50} (nM)	peptide
6.777	167.11	L
6.756	175.39	I
6.633	232.81	P
6.594	254.68	S
6.545	285.10	G
6.401	397.19	S
6.395	402.72	T
6.36	436.52	D
6.341	456.04	C
6.329	468.81	Y
6.247	566.24	F
6.229	590.20	N
6.216	608.14	K
6.195	638.26	V
6.189	647.14	N
6.174	669.88	Y

Fig. 5. Part of the output page of the peptide library. Mutated peptides are listed in ascending order of their binding affinities.

(QSAR) approach. The additive method uses literature peptides with measured affinities (IC_{50} values) as the training set. A statistical technique, partial least squares, is applied to the training set to generate models. The previous version of MHCPred only included alleles from the human HLA-A locus. As the mouse is the preeminent experimental animal used in immunology, we collected peptides restricted to three mouse class I alleles and generated a set of new models.[20] We believe these new models will prove valuable to immunologists, since only a few online servers include mouse models.

Additive method predictions use coefficients for each of the 20 amino acids at positions one to nine of the peptide and use these coefficients to predict affinities of other peptides. However, as our training set data are limited to published peptides, some amino acids are not present at certain positions. By default, these missing values do not contribute to the predicted affinity of peptides, although other values for missing coefficients can be used. The accuracy of prediction is clearly inversely proportional to the number of missing values. MHCPred 2.0 includes a new feature to calculate the confidence of prediction for each peptide (as a normalised percentage), according to the number of missing terms in the model. An example is to calculate the affinity of peptide WLFPGPMTV to A*0301 allele: Trp at position 1 and Val at position 9 of the peptide are not present in the model, other amino acids have corresponding coefficients in the A*0301 model, therefore the confidence of calculation is 0.78 (7/9). This feature helps the user to eliminate false-positive predictions and makes the prediction more reliable.

Conclusion

The additive method was developed for predicting T-cell epitopes and has been incorporated into the web prediction server MHCPred 2.0. MHCPred 2.0 is fast and easy to use. It has been shown to be effective in predicting epitopes binding to HLA-A alleles. In order to apply the additive method more broadly, we generated new models for class I MHC alleles and TAP, which are also included in MHCPred 2.0. Some new features are also added to MHCPred 2.0, such as multiple allele selection and prediction confidence calculation. We believe MHCPred 2.0 will prove a valuable tool for predicting T-cell epitopes.

Acknowledgements

We thank Mr Andrew Worth for his technical assistance. The Edward Jenner Institute for Vaccine Research wishes to thank its sponsors: GlaxoSmithKline, the Medical Research Council, the Biotechnology and Biological Sciences Research Council and the UK Department of Health.

The authors have no conflicts of interest that are directly relevant to the content of this article.

References

1. Jones EY. MHC class I and class II structures. Curr Opin Immunol 1997; 9: 75-9
2. Janeway CA. Immunobiology 5: the immune system in health and disease. New York: Garland, 2001
3. Ljunggren HG, Thorpe CJ. Principles of MHC class I mediated antigen presentation and T cell selection. Histol Histopathol 1996; 11: 267-74
4. Madnaka K, Jones EY. MHC superfamily structure and the immune system. Curr Opin Immunol 1999; 9: 745-53
5. Rau L, Cohan N, Robert J. MHC-restricted and -unrestricted CD8 T cells: an evolutionary perspective. Transplantation 2001; 72: 1830-5
6. Flower DR. Towards in silico prediction of immunogenic epitopes. Trends Immunol 2003; 24: 667-74
7. Rammensee H, Bachmann J, Emmerich NP, et al. SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 1999; 50: 213-9
8. Lauemoller SL, Holm A, Hilden J, et al. Quantitative predictions of peptide binding to MHC class I molecules using specificity matrices and anchor-stratified calibrations. Tissue Antigens 2001; 57: 405-14
9. Reche PA, Glutting JP, Reinherz EL. Prediction of MHC class I binding peptides using profile motifs. Hum Immunol 2002; 63: 701-9

10. Altuvia Y, Sette A, Sidney J, et al. A structure-based algorithm to predict potential binding peptides to MHC molecules with hydrophobic binding pockets. Hum Immunol 1997; 58: 1-11
11. Del Carpio CA, Hennig T, Fickel S, et al. A combined bioinformatic approach oriented to the analysis and design of peptides with high affinity to MHC class I molecules. Immunol Cell Biol 2002; 80: 286-99
12. Doytchinova IA, Blythe MJ, Flower DR. Additive method for the prediction of protein-peptide binding affinity: application to the MHC class I molecule HLA-A*0201. J Proteome Res 2002; 1: 263-72
13. Doytchinova IA, Flower DR. Towards the in silico identification of class II restricted T-cell epitopes: a partial least squares iterative self-consistent algorithm for affinity prediction. Bioinformatics 2003; 19: 2263-70
14. Guan P, Doytchinova IA, Zygouri C, et al. MHCPred: a server for quantitative prediction of peptide-MHC binding. Nucleic Acids Res 2003; 31: 3621-4
15. Guan P, Doytchinova IA, Zygouri C, et al. MHCPred: bringing a quantitative dimension to the online prediction of MHC binding. Appl Bioinformatics 2003; 2: 63-6
16. Hattotuwagama CK, Guan P, Doytchinova IA, et al. Quantitative online prediction of peptide binding to the major histocompatibility complex. J Mol Graph Model 2004; 22: 195-207
17. Free SM, Wilson JW. A mathematical contribution to structure-activity studies. J Med Chem 1964; 53: 395-9
18. Doytchinova I, Hemsley S, Flower DR. Transporter associated with antigen processing preselection of peptides binding to the MHC: a bioinformatic evaluation. J Immunol 2004; 173: 6813-9
19. Doytchinova IA, Walshe VA, Jones NA, et al. Coupling in silico and in vitro analysis of peptide-MHC binding: a bioinformatic approach enabling prediction of superbinding peptides and anchorless epitopes. J Immunol 2004; 172: 7495-502
20. Hattotuwagama CK, Guan P, Doytchinova IA, et al. New horizons in mouse immunoinformatics: reliable in silico prediction of mouse class I histocompatibility major complex peptide binding affinity. Org Biomol Chem 2004; 2: 3274-83

Correspondence and offprints: Dr *Darren R. Flower*, Edward Jenner Institute for Vaccine Research, Compton, Newbury, Berkshire RG20 7NN, UK. E-mail: darren.flower@jenner.ac.uk