W194–W197 *Nucleic Acids Research, 2006, Vol. 34, Web Server issue*
doi:10.1093/nar/gkl284

SVMHC: a server for prediction of MHC-binding peptides

Pierre Dönnes* and Oliver Kohlbacher

Division for Simulation of Biological Systems, WSI/ZBIT, Eberhard Karls University Tübingen, Sand 14, D-72076 Tübingen, Germany

Received February 14, 2006; Revised March 2, 2006; Accepted April 5, 2006

ABSTRACT

Identification of MHC-binding peptides is a prerequisite in rational design of T-cell based peptide vaccines. During the past decade a number of computational approaches have been introduced for the prediction of MHC-binding peptides, efficiently reducing the number of candidate binders that need to be experimentally verified. Here the SVMHC server for prediction of both MHC class I and class II binding peptides is presented. SVMHC offers fast analysis of a wide range of alleles and prediction results are given in several comprehensive formats. The server can be used to find the most likely binders in a protein sequence and to investigate the effects of single nucleotide polymorphisms in terms of MHC-peptide binding. The SVMHC server is accessible at http://www-bs.informatik.uni-tuebingen.de/SVMHC/.

INTRODUCTION

The immune system provides an effective line of defense against invading pathogens and cancer. The adaptive part of the immune system, which is responsible for specific recognition of antigen and immunological memory, is highly dependent on the activation of T-cells. T-cells only recognize antigenic peptides bound to major histocompatibility (MHC) molecules on the surface of other cells. This makes MHC-peptide binding a prerequisite for T-cell activation. There are two major classes of MHC molecules. MHC class I molecules typically bind peptides that are 9 amino acids long and originate from intracellular proteins. Intracellular proteins are continuously degraded into smaller peptides that are displayed on the cells surface by MHC molecules, giving a kind of fingerprint of the cellular proteome. This mechanism ensures that virally infected cells or cancer cells can be detected, since virus or cancer-specific MHC-peptide complexes are displayed on the cell surface. Cytotoxic T-cells ($CD8^+$) of the immune system can recognize such abnormal cells and eliminate them. MHC class II molecules, on the other hand, bind peptides originating from extracellular antigens. These peptides are usually longer compared with MHC class I peptides (15–25 amino acids), however the main part of the MHC-peptide interactions is given by a binding core of 9 amino acids. MHC class II molecules are mainly presented on antigen presenting cells (APCs) and activate helper T-cells ($CD4^+$). In recent years, MHC-binding peptides have proven useful for immunotherapeutic purposes in studies concerning both different cancer types (1,2) and HIV infection (3). The aim of these approaches is to use antigen-specific peptides in order to activate the immune system. The first step here is to find a set of MHC-binding peptides given an antigen of interest. One challenge here is the extreme variability of the MHC molecules with many hundred allelic variants. However, typically only one in 100–200 potential peptides actually binds to a certain MHC allele (4). This has motivated computational approaches for modeling MHC allele-specific peptide preferences. Such methods can reduce the number of peptides that have to be verified experimentally.

The first prediction methods utilized simple sequence motif searches for identifying potential MHC class I binding peptides (5,6). These methods have since been refined into position-specific scoring matrix (PSSM) approaches (7–13). One drawback of these methods is that they assume an independent contribution of each amino acid in the peptides to the overall binding affinity, neglecting the effects of neighbouring residues. An obvious case where this might be a problem, is when two compete for the same space in a binding pocket. Several machine learning methods have been introduced that aim to model the MHC-peptide interaction in a non-linear fashion (14–18), potentially overcoming the limitation of PSSM-based methods. The above mentioned methods are all sequence-based, but a number of structure-based methods have also been presented (19–22).

Prediction of MHC class II peptides is more challenging, owing to the additional alignment needed to identify the binding core within the longer peptides. Once the sequences have been correctly aligned, the computational problem is very

*To whom correspondence should be addressed. Tel: +49 7071 29 70459; Fax: +49 7071 29 5152; Email: doennes@informatik.uni-tuebingen.de

similar to that of the class I case. Methods for MHC class II prediction include genetic algorithms coupled with neural networks (23) and Gibbs sampling (24), as well as the construction of PSSMs using virtual binding pockets (25). The predicted MHC class II binding cores are often extended at both ends to obtain an effective T-cell epitope.

Here, the SVMHC server for prediction of MHC class I and class II binding peptides is presented. In contrast to most other prediction servers, SVMHC offers several comprehensive results formats, easy access to data from protein databases and refinement of initial predictions. Furthermore, SVMHC enables analysis of the effects of single nucleotide polymorphisms (SNPs) in terms of MHC-peptide binding. A number of new prediction models for human and mouse MHC class I molecules have been added. Furthermore, MHC class II prediction can be done utilizing the matrices published by Sturniolo *et al*. (25).

THE SVMHC PREDICTION SERVER

Prediction models

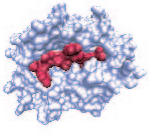
A support vector machine (SVM) approach is used for the prediction of MHC class I binding peptides. This approach has been described in detail in a previous publication (17) and is only briefly outlined here. MHC-binding peptides of different lengths were extracted from the MHCPEP (26) and SYFPEITHI (8) databases. The main difference between these two data sources is that MHCPEP contains both naturally processed and synthetic peptides, whereas SYFPEITHI exclusively contains naturally processed peptides. In order to construct the prediction models, each peptide was represented using binary sparse encoding. Different kernels and a grid search strategy were then employed to find optimal SVM parameters. Approximately 20 known binders are needed in order to construct prediction models with significant accuracy. For most alleles prediction models could only be generated for peptides with a length of nine amino acids due to the amount of data available. However, in some cases prediction models were also constructed for peptides with a length of eight or ten amino acids. In comparative studies against the prediction methods BIMAS (9) and SYFPEITHI (8), SVMHC showed improved performance for most MHC alleles (17). Prediction models are now available for 26 different human MHC alleles based on data from MHCPEP. Prediction models based on data from SYFPEITHI are available for 19 human and 5 murine MHC alleles.

Prediction of MHC class II binding peptides is based on the matrices published by Sturniolo *et al*. (25). By sequence similarity studies, they defined modular pockets in the MHC molecule involved in peptide interaction. These pockets are independent of the rest of the binding cleft and a limited number of pockets can be combined into virtual binding matrices for a wide range of MHC class II alleles. These matrices are also a part of the TEPITOPE prediction software and they have been used to identify candidate binding peptides for both HIV (3) and Tuberculosis (27) vaccines. Prediction is available for 51 different MHC class II alleles.

Whole protein prediction

The input required for analysis by SVMHC is a protein sequence and a specification of one or more MHC alleles. The protein sequence can either be directly pasted into the web interface or accessed directly by entering a database ID from the NCBI RefSeq (28) or Swiss-Prot (29) databases. Prediction is carried out for all possible peptides of the protein using a sliding window. Several different output formats are given in order to facilitate further analysis. The default output format is a list of putative binders, where the best binder is found at the top. A summary table is also generated for all peptides of a certain length. The summary table shows the results ordered according to peptide start position in the protein (see Figure 1 for an example).

Binders are highlighted, which enables fast identification of peptides likely to bind several MHC alleles, so called promiscuous epitopes. These are especially interesting for



SVMHC - PREDICTION RESULTS

Prediction
Information
Links
Back to SBS

Nonamer Peptides

Position	Sequence	A_01(s)	A_0201(s)	A_03(s)	A_24(s)	B_08(s)	B_2705(s)
1	VTVLDVGDA	-0.61	-1.17	-2.1	-0.82	-0.75	-0.97
2	TVLDVGDAY	0.02	-1.1	0.92	-0.63	-0.91	-0.96
3	VLDVGDAYF	-0.15	0.32	-0.8	-0.45	-0.81	-0.77
4	LDVGDAYFS	-0.92	-1.65	-1.84	-0.79	-2.18	-0.98
5	DVGDAYFSV	-0.94	0.22	-0.7	-0.74	-1.4	-1.15
6	VGDAYFSVP	-0.42	-1.3	-1.7	-0.83	-1.34	-1.1
7	GDAYFSVPL	-0.94	-0.77	-1.32	-0.6	-0.29	-0.55
8	DAYFSVPLD	-0.68	-1.38	-0.86	-0.55	-0.85	-0.75
9	AYFSVPLDE	-0.66	-1.7	-0.36	0.01	-1.01	-0.7
10	YFSVPLDED	-0.78	-1.05	-1.99	-0.7	-0.95	-0.92

Figure 1. A summary table produced by SVMHC showing peptide start position, sequence and allele-specific score. Predicted MHC binders are highlighted in red, enabling fast identification of peptides binding to several different MHC alleles.

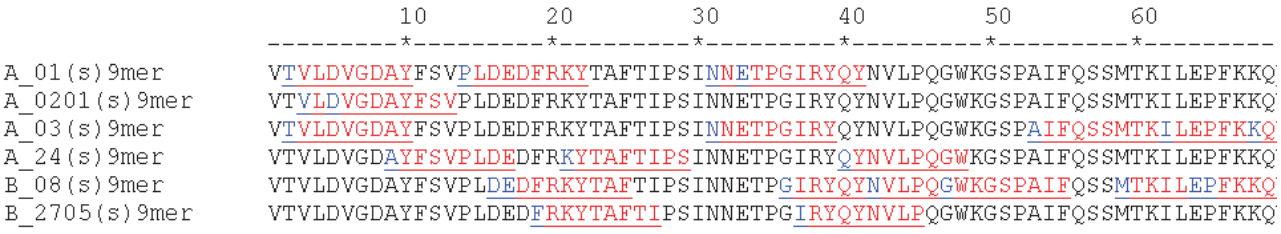


Figure 2. A graphical view of the predicted MHC-binding peptides. Predicted binding peptides are colored red, except for the first amino acid that is colored blue. This view enables a fast scan, even of long proteins, in order to identify promiscuous epitopes and epitope-rich regions.

Effects of the mutation **K27T** on **A_03 nonamer** peptide binding (model trained on SYFPEITHI data).

Mutation position	Normal sequence	Score	Mutated sequence	Score	Score difference
1	KYKLKHIVW	-1.23	TYKLKHIVW	-1.24	0.01
2	KKYKLKHIV	-1.4	KTYKLKHIV	-1.0	-0.4
3	KKKYKLKHI	-1.08	KKTYKLKHI	-1.11	0.03
4	GKKKYKLKH	-0.89	GKKTYKLKH	-0.74	-0.15
5	GGKKKYKLK	-0.0	GGKKTYKLK	0.0	-0.0
6	PGGKKKYKL	-1.97	PGGKKTYKL	-1.8	-0.17
7	RPGGKKKYK	0.22	RPGGKKTYK	0.25	-0.03
8	LRPGGKKKY	-1.03	LRPGGKKTY	-0.59	-0.44
9	RLRPGGKKK	1.5	RLRPGGKKT	0.19	1.31

Figure 3. Prediction results for analyzing the effect of the K27T mutation in the HIV matrix protein p17. From these results it can be seen that the peptide binding is substantially reduced, a possible explanation for immune escape. Predicted binders are highlighted green and if the difference in the predicted score between two binders is <0.5, the score difference is highlighted blue.

vaccine design since they cover a wider range of the population. A graphical overview is also given to further facilitate the identification of promiscuous epitopes (see Figure 2).

The initial prediction results can then be further refined by removing or adding alleles of interest. The complete prediction results can also be downloaded in tab-separated format, enabling further analysis in any spreadsheet-based program (e.g. Microsoft Excel).

Analyzing the effects of SNPs

Several studies have pointed out the importance of SNPs in terms of MHC-binding peptides (30–32). SVMHC allows for the analysis of SNPs in terms of MHC-peptide binding. For this analysis a protein sequence and a specified mutation (e.g. A23P, meaning that alanine in position 23 of the protein is changed to a proline) is required. All relevant peptides, with and without the mutation, are then generated and predicted by SVMHC. The results are presented in a comparative manner, highlighting the effects of the amino acid substitution. A good example for this type of analysis is the well-known HLA-A*03-restricted epitope RLRPGGKKK originating from the HIV matrix protein p17 (30,31). Studies have identified polymorphisms within this peptide, where the exchange of the lysine in position nine to a threonine, cause viral escape (31) (meaning that the mutated peptides are not recognized by the immune system). The whole p17 protein with the specified K27T mutation (corresponding to the mutation in position nine of the peptide of interest) was analyzed by SVMHC, see Figure 3.

The mutation substantially reduces the predicted MHC affinity of the peptide. SNPs can also influence amino acids of the peptide, which are less involved in MHC binding and rather important for T-cell recognition. Binders and non-binders are highlighted in green and red, respectively, in the result table (if at least one peptides is predicted as a binder). Furthermore, a blue highlighting is given if the difference between two peptides is >0.5. The dynamic coloring makes it easy to identify SNPs that are interesting for further analysis.

CONCLUSION

We present an updated and extended version of the SVMHC server for predicting MHC-binding epitopes. SVMHC combines high prediction accuracy with a wide range of both, MHC class I and MHC class II alleles. Compared with other prediction tools, it also provides a number of different output formats ranging from summary graphical views to detailed comparison tables. All data can also be exported for external analysis. Another singular feature of SVMHC is the ability to analyze the effects of SNPs on MHC epitopes. This type of analysis is interesting for viral epitopes (prediction of immune escape) and the analysis of minor histocompatibility antigens (miHAgs). SVMHC is updated regularly

and as new MHC-binding data becomes available it will be integrated into SVMHC. This ensures continual improvement of both prediction accuracy and allele coverage.

ACKNOWLEDGEMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (SFB 685). Funding to pay the Open Access publication charges for this article was provided by the Deutsche Forschungsgemeinschaft (SFB 685).

Conflict of interest statement. None declared.

REFERENCES

1. Hsu,F.J., Benike,C., Fagnoni,F., Liles,T.M., Czerwinski,D., Taidi,B., Engleman,E.G. and Levy,R. (1996) Vaccination of patients with B-cell lymphoma using autologous antigen-pulsed dendritic cells. *Nat. Med.*, **2**, 52–58.
2. Nestle,F.O., Alijagic,S., Gilliet,M., Sun,Y., Grabbe,S., Dummer,R., Burg,G. and Schadendorf,D. (1998) Vaccination of melanoma patients with peptide- or tumor lysate-pulsed dendritic cells. *Nat. Med.*, **4**, 328–332.
3. De Groot,A.S., Marcon,L., Bishop,E.A., Rivera,D., Kutzler,M., Weiner,D.B. and Martin,W. (2005) HIV vaccine development by computer assisted design: the GAIA vaccine. *Vaccine*, **23**, 2136–2148.
4. Yewdell,J.W. and Bennink,J.R. (1999) Mechanisms of viral interference with MHC class I antigen processing and presentation. *Annu. Rev. Cell Dev. Biol.*, **15**, 579–606.
5. Sette,A., Buus,S., Appella,E., Smith,J.A., Chesnut,R., Miles,C., Colon,S.M. and Grey,H.M. (1989) Prediction of major histocompatibility complex binding regions of protein antigens by sequence pattern analysis. *Proc. Natl Acad. Sci. USA*, **86**, 3296–3300.
6. Rötzschke,O., Falk,K., Stevanovic,S., Jung,G., Walden,P. and Rammensee,H.G. (1991) Exact prediction of a natural T cell epitope. *Eur. J. Immunol.*, **21**, 2891–2894.
7. Kondo,A., Sidney,J., Southwood,S., delGuercio,M.F., Appella,E., Sakamoto,H., Celis,E., Grey,H.M., Chesnut,R.W., Kubo,R.T. *et al.* (1995) Prominent roles of secondary anchor residues in peptide binding to HLA-A24 human class I molecules. *J. Immunol.*, **155**, 4307–4312.
8. Rammensee,H.-G., Bachman,J., Philipp,N., Emmerich,N., Bachor,O.A. and Stevanovic,S. (1997) SYFPEITHI: a database for MHC ligands and peptide motifs. *Immunogenetics*, **50**, 213–219.
9. 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.
10. Reche,P.A., Glutting,J.-P., Zhang,H. and Reinherz,E.L. (2004) Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. *Immunogenetics*, **56**, 405–419.
11. Sidney,J., Grey,H.M., Southwood,S., Celis,E., Wentworth,P.A., delGuercio,M.F., Kubo,R.T., Chesnut,R.W. and Sette,A. (1996) Definition of an HLA-A3-like supermotif demonstrates the overlapping peptide-binding repertoires of common HLA molecules. *Hum. Immunol.*, **45**, 79–93.
12. Sidney,J., Southwood,S., delGuercio,M.F., Grey,H.M., Chesnut,R.W., Kubo,R.T. and Sette,A. (1996) Specificity and degeneracy in peptide binding to HLA-B7-like class I molecules. *J. Immunol.*, **157**, 3480–3490.
13. Sidney,J., Southwood,S., Pasquetto,V. and Sette,A. (2003) Simultaneous prediction of binding capacity for multiple molecules of the HLA B44 supertype. *J. Immunol.*, **171**, 5964–5974.
14. Gulukota,K., Sidney,J., Sette,A. and DeLisi,C. (1997) Two complementary methods for predicting peptides binding major histocompatibility complex molecules. *J. Mol. Biol.*, **267**, 1258–1267.
15. Honeyman,M.C., Brusic,V., Stone,N.L. and Harrison,L.C. (1998) Neural network-based prediction of candidate T-cell epitopes. *Nat. Biotechnol.*, **16**, 966–969.
16. Mamitsuka,H. (1998) Predicting peptides that bind to MHC molecules using supervised learning of hidden markov models. *Proteins*, **33**, 460–474.
17. Dönnes,P. and Elofsson,A. (2002) Prediction of MHC class I binding peptides, using SVMHC. *BMC Bioinformatics*, **3**, 25.
18. Nielsen,M., Lundegaard,C., Worning,P., Lauemoller,S.L., Lamberth,K., Buus,S., Brunak,S. and Lund,O. (2003) Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci.*, **12**, 1007–1017.
19. Rognan,D., Scapozza,L., Folkers,G. and Daser,A. (1994) Molecular dynamics simulation of MHC-peptide complexes as a tool for predicting potential T cell epitopes. *Biochemistry*, **33**, 11476–11485.
20. Logean,A. and Rognan,D. (2002) Recovery of known T-cell epitopes by computational scanning of a viral genome. *J. Comput. Aided Mol. Des.*, **16**, 229–243.
21. Tong,J.C., Tan,T.W. and Ranganathan,S. (2004) Modeling the structure of bound peptide ligands to major histocompatibility complex. *Protein Sci.*, **13**, 2523–2532.
22. Schueler-Furman,O., Altuvia,Y. and Sette,A. (2000) Structure-based prediction of binding peptides to MHC class I molecules: application to a broad range of MHC alleles. *Protein Sci.*, **9**, 1838–1846.
23. Brusic,V., Rudy,G., Honeyman,G., Hammer,J. and Harrison,L. (1998) Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network. *Bioinformatics*, **14**, 121–130.
24. Nielsen,M., Lundegaard,C., Worning,P., Hvid,C.S., Lamberth,K., Buus,S., Brunak,S. and Lund,O. (2004) Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach. *Bioinformatics*, **20**, 1388–1397.
25. 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 databases using DNA microarrays and virtual HLA class II matrices. *Nat. Biotechnol.*, **17**, 555–561.
26. Brusic,V., Rudy,G., Honeyman,G., Hammer,J. and Harrison,L. (1998) Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network. *Bioinformatics*, **14**, 121–130.
27. De Groot,A.S., Bosma,A., Chinai,N., Frost,J., Jesdale,B.M., Gonzalez,M.A., Martin,W. and Saint-Aubin,C. (2001) From genome to vaccine: *in silico* predictions, ex vivo verification. *Vaccine*, **19**, 4385–4395.
28. Pruitt,K.D., Tatusova,T. and Maglott,D.R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **33**, D501–D504.
29. Boeckmann,B., Bairoch,A., Apweiler,R., Blatter,M.-C., Estreicher,A., Gasteiger,E., Martin,M., Michoud,K., O'Donovan,C., Phan,I. *et al.* (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.*, **31**, 365–370.
30. Altfeld,M., Allen,T.M., Yu,X.G., Johnston,M.N., Agrawal,D., Korber,B.T., Montefiori,D.C., O'Connor,D.H., Davis,B.T., Lee,P.K. *et al.* (2002) HIV-1 superinfection despite broad CD8+ T-cell responses containing replication of the primary virus. *Nature*, **420**, 434–439.
31. Allen,T.M., Altfeld,M., Yu,X.G., O'Sullivan,K.M., Lichterfeld,M., Gall,S.L., John,M., Mothe,B.R., Lee,P.K., Kalife,E.T. *et al.* (2004) Selection, transmission, and reversion of an antigen-processing cytotoxic T-lymphocyte escape mutation in human immunodeficiency virus type 1 infection. *J. Virol.*, **78**, 7069–7078.
32. Schuler,M., Dönnes,P., Nastke,M.-D., Kohlbacher,O., Rammensee,H.-G. and Stevanovic,S. (2005) SNEP: SNP-derived Epitope Prediction program for minor H antigens. *Immunogenetics*, **57**, 816–820.