Machine Intelligence approach for prediction of MHC binder from *Human papilloma virus- 16*

Gomase V.S. and Tripathi A.K.

Department of Bioinformatics, Padmashree Dr. D. Y. Patil University, Plot No-50, Sector-15, CBD Belapur, Navi Mumbai, 400614, MS, India, Mobile- +91-9987770696, Mail- virusgene1@yahoo.co.in

Abstract- Prediction of the binding ability of antigen peptides to major histocompatibility complex (MHC) class molecules is important in vaccine development from *Human papilloma virus*-16. The variable length of each binding peptide complicates this prediction. Such predictions can be used to select epitopes for use in rational vaccine design and to increase the understanding of roles of the immune system in infectious diseases. Antigenic epitopes of capsid protein L1 form *Human papilloma virus*-16 are important determinant for protection of many disorders form viral infection. This study shows active part in host immune reactions and involvement of MHC class-I and MHC II in response to almost all antigens. We also found the SVM based MHCII-IAb, MHCII-IAd, MHCII-IAg7 and MHCII- RT1.B peptide regions, which represented predicted binders from viral capsid protein (Table-1). These peptide nonamers are from a set of aligned peptides known to bind to a given MHC molecule as the predictor of MHC-peptide binding. Analysis shows potential drug targets to identify active sites against diseases.

Keywords- Antigen, Epitope, MHC, Peptide vaccine

1. Introduction

Human papillomavirus (HPV) is a papillomavirus, which infects the epidermis and mucous membranes of humans. *Human papillomavirus* can lead to cancers of the vulva, vagina, anus and cervix in women, also in men; it can lead to cancers of the anus and penis. However, 130 HPV types have been identified. Some HPV types can cause warts (verrucae), but those types don't cause cancer. Other types can cause cancer, but those types don't cause warts. Many other types have no symptoms and are harmless. Most people who become infected with HPV do not know they have it [1]. Prediction of Peptide fragments from *Human papillomavirus* involved multiple antigenic components to direct and empower the immune system to protect the host from infection. MHC molecules are cell surface proteins, which take active part in host immune reactions and involvement of MHC class-I & II in response to almost all antigens. The predicted binding affinity is normalized by the 1% fractil. The MHC peptide binding is predicted using neural networks trained on C terminals of known epitopes. In analysis predicted MHC/peptide binding is a log transformed value related to the IC50 values in nM units [2, 3].

2. Methodology

The Hopp-Woods, Welling and Parkers antigenicity scale was designed to predict the locations of antigenic determinants. Thus beta sheet show high antigenic response than helical region of this peptide, which is highly antigenic in nature [4, 5]. We also found the Sweet hydrophobicity, Kyte & Doolittle hydrophobicity, Abraham & Leo hydrophobicity, Bull & Breese hydrophobicity, Guy hydrophobicity, Miyazawa hydrophobicity, Roseman hydrophobicity, Cowan HPLC pH7.5 hydrophobicity, Rose hydrophobicity, Eisenberg hydrophobicity, Manavalan hydrophobicity, Black hydrophobicity, Fauchere hydrophobicity, Janin hydrophobicity, Rao & Argos hydrophobicity, Wolfenden hydrophobicity, Wilson HPLC hydrophobicity, von Heijne Hydrophilicity, Chothia hydrophobicity scales. Theses scales are essentially a hydrophilic index, with apolar residues assigned negative values [6, 7]. The MHC peptide binding of capsid protein is predicted using neural networks trained on C terminals of known epitopes. In analysis predicted MHC/peptide binding of protein is a log-transformed value related to the IC50 values in nM units. MHC2Pred predicts peptide binders to MHCI and MHCII molecules from protein sequences or sequence alignments using Position Specific Scoring Matrices (PSSMs). Support Vector Machine (SVM) based method for

*Correspondence: Dr. Virendra S. Gomase, Department of Bioinformatics, Padmashree Dr. D.Y. Patil University, Plot No-50, Sector-15, CBD Belapur, Navi Mumbai, 400614, India; Mobile- +91-9987770696; Mail- virusgene1@yahoo.co.in

prediction of promiscuous MHC class II binding peptides. SVM has been trained on the binary input of single amino acid sequence [**8-12**].

3. Results and Interpretation

Major capsid protein is 424 residues long, having antigenic MHC binding peptides. MHC molecules are cell surface glycoproteins, which take active part in host immune reactions and involvement of MHC class-I and MHC II in response to almost all antigens. BepiPrep Server antigenicity determinant shows epitopes present in the *Human papillomavirus* the desired immune response. The region of maximal hydrophilicity is likely to be an antigenic site, having hydrophobic characteristics, because C- terminal regions of major capsid protein is solvent accessible and unstructured, antibodies against those regions are also likely to recognize the native protein (Fig. 1, 2). For the prediction of antigenic determinant site of major capsid protein (*Human papillomavirus*), we got antigenic determinant sites in the sequence. We also found the SVM based MHCII-IAb peptide regions 311-FVTVVDTTR, 56-IHLPDPNKF, 239-RHLFNRAGT,

27-GHPYFPIKK, (optimal score is 1.124); MHCII-IAd peptide regions 20-TSRLLAVGH, 385-GGSGGEDPL,

19-GTSRLLAVG, 257-YIKGSGSTA, (optimal score is 0.760); MHCII-IAg7 peptide regions 110-TENASAYAA, 112-NASAYAANA, 114-SAYAANAGV, 15-YYHAGTSRL, (optimal score is 1.920); and MHCII- RT1.B peptide regions 69-TSFYNPDTQ, 404-KEKFSADLD, 132-YKQTQLCLI, 406-KFSADLDQF, (optimal score is 1.173) which represented predicted binders from protein. The predicted binding affinity is normalized by the 1% fractil. The average propensity for the major capsid protein is found to be above 1.0350. All residues having above 1.0 propensity are always potentially antigenic. Further this region form beta sheet. Thus beta sheet show high antigenic response than helical region of this peptide (Fig. 1, 2). A region preferably select peptides lying in loops connecting Secondary Structure (SS) motifs, avoiding peptides located in helical regions. CPHmodels novel method is used for fold recognition/homology modeling, in which a large sequence database is iteratively searched to construct a sequence profile. The predicted segments in major capsid protein identified through this approach tend to be high-efficiency binders, which is a lagers percentage of their atoms are directly involved in binding as compared to larger molecules.
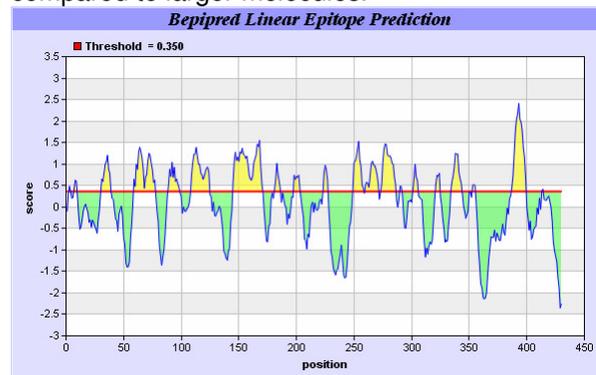

Fig. 1- predicting linear B-cell epitopes
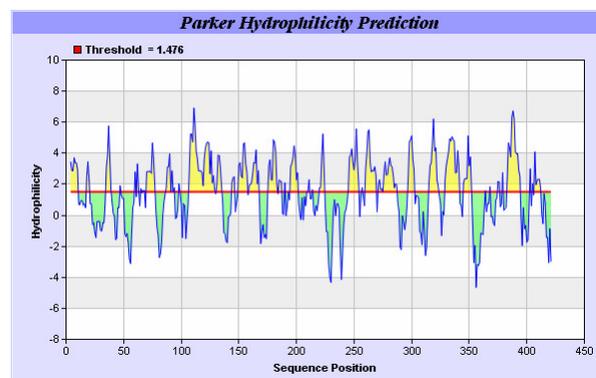

Fig. 2- Parker hydrophilicity prediction

4. Conclusion

Data mining techniques are used for development of peptide epitopes which are antigenic in nature. Predicted MHC binding regions acts like red flags for antigen specific and generate immune response against the parent antigen. So a small fragment of antigen can induce immune response against whole antigen. This theme is implemented in designing subunit and synthetic peptide vaccines. The sequence analysis method is allows potential drug targets to identify active sites which form antibodies against diseases. The method integrates prediction of peptide MHC class I binding; proteasomal C terminal cleavage and TAP transport efficiency.

5. References

[1]. Schiffman M., Castle P.E. (2003) *Arch Pathol Lab Med.* 127(8), 930-4.

[2]. López-Revilla R, Pineda MA, Ortiz-Valdez J, Sánchez-Garza M, Riego L. (2009) *Infect Agent Cancer* 4, 3.

[3]. Mohabatkar H. (2007) *Asian Pac J Cancer Prev*. 8(4), 602-6.

[4]. Gomase V.S. and Kale K.V. (2008) *Int. J. of Applied Computing*, 1(1), 39-46.

[5]. Gomase V.S., Kale K.V., Chikhale N.J., Changbhale S.S. (2007) *Curr. Drug Discov. Technol.*, 4(2),117-1215.

[6]. Gomase V.S., Kale K.V., Jyotiraj A. and Vasanthi R. (2007) *Medicinal Chemistry Research*, 15(1/6), 160.

[7]. Gomase V.S., Tandale J.P., Patil S.A., Kale K.V. (2006) *14th International Conference on Advance Computing & Communication, Published by IEEE Computer Society in IEEE Xplore USA* 614-615.

[8]. Gomase V.S., Kale K.V., Shyamkumar K. and Shankar S. (2008) *ICETET 2008, IEEE Computer Society in IEEE Xplore, Los Alamitos, California*, 629-634.

[9]. Bhasin M. and G.P. Raghava G.P. (2005) *Nucleic Acids Research*, 33, W202-207.

[10]. Reche P.A., Glutting J.P., Reinherz E.L. (2002) *Hum Immun.*, 63, 701-709.

[11]. Buus S., *et al.* (2003) *Tissue Antigens* 62, 378-384.

[12]. Nielsen M., *et al.* (2003) *Protein Sci.*, 12, 1007-1017.

TABLE 2- MHC class II binding peptide nonamers from capsid protein

MHC ALLELE	Rank	Sequence	Residue No.	Peptide Score
I-Ab	1	FVTVVDTTR	311	1.124
I-Ab	2	IHLPDPNKF	56	1.067
I-Ab	3	RHLFNRAGT	239	0.947
I-Ab	4	GHPYFPIKK	27	0.945
I-Ad	1	TSRLLAVGH	20	0.760
I-Ad	2	GGSGGEDPL	385	0.586
I-Ad	3	GTSRLLAVG	19	0.568
I-Ad	4	YIKGSGSTA	257	0.564
I-Ag7	1	TENASAYAA	110	1.920
I-Ag7	2	NASAYAANA	112	1.829
I-Ag7	3	SAYAANAGV	114	1.609
I-Ag7	4	YYHAGTSRL	15	1.576
RT1.B	1	TSFYNPDTQ	69	1.173
RT1.B	2	KEKFSADLD	404	1.153
RT1.B	3	YKQTQLCLI	132	1.087
RT1.B	4	KFSADLDQF	406	0.997