Machine Learning Techniques: Approach for Mapping of MHC Class Binding Nonamers

Gomase V.S.*, Yash Parekh, Subin Koshy, Siddhesh Lakhan
*Department of Bioinformatics, Padmashree Dr. D.Y. Patil University, Navi Mumbai, 400614, India, Mail- virusgene1@yahoo.co.in

Abstract- The machine learning techniques are playing a major role in the field of immunoinformatics for DNA-binding domain analysis. Functional analysis of the binding ability of DNA-binding domain protein antigen peptides to major histocompatibility complex (MHC) class molecules is important in vaccine development. 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 DNA-binding domain protein form *Human papilloma virus-31* are important determinant for protection of many host 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 used PSSM and SVM algorithms for antigen design, which represented predicted binders as MHCII-IAb, MHCII-IAd, MHCII-IAg7, and MHCII- RT1.B nonamers from viral DNA-binding domain crystal structure. 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 DNA-binding domain crystal structure, PSSM, SVM, MHC, epitope, peptide vaccine Abbreviations: Goldman, Engelberg and Steitz, (GES); major histocompatibility complex, (MHC); Position Specific Scoring Matrices, (PSSMs); Support Vector Machine, (SVM)

INTRODUCTION

Human papillomavirus (HPV) is a papillomavirus that infects the skin and mucous membranes of humans, epithelia of humans and animals, where they generally induce benign proliferation at the site of infection. Also, in some cases the lesions induced by certain papillomaviruses undergo malignant progression. There is a strong association between malignant progression of human genital lesions and certain human papillomavirus (HPV) types, such as HPV16. The cervical carcinomas contain and express HPV early genes, and these genes have been shown to have potent transforming and immortalizing activity in cultured cells. The virus is made up of many and varied strains that differ in a number of aspects including host range, symptoms, and means of transmission (Sénéchal, et al 2007; Davidson, et al 2004). The strains of DNA-binding domain protein [*Human papilloma virus-31*] are differentiated on the basis of symptoms on indicator host. Viral DNA-binding domain protein is central to the establishment of viral pathogenesis, and yet relatively little is understood about the structural and functional aspects of protein or about the host factors on which they depend. DNA-binding domain protein [*Human papilloma virus-31*] and the genetic material necessary for its production in; an exemption from the requirement of a tolerance is established for residues of the biological viral therapy (Wang, et al 2004; Gomase, et al 2008a; Gomase, et al 2007).

METHODOLOGY

Antigenic epitopes of DNA-binding domain protein from *Human papilloma virus-31* is determined using the Gomase (2007), Hopp and Woods, Welling, Parker and Protrusion Index (Thornton) antigenicity methods (Gomase and Kale 2008a,b,c; Gomase, et al 2008b). The MHC peptide binding of DNA-binding domain protein is predicted using neural networks trained on C terminals of known epitopes. In analysis predicted MHC/peptide binding of DNA-binding domain 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 prediction of promiscuous MHC class II binding peptides. SVM has been trained on the binary input of single amino acid sequence (Gomase, et al 2006; Reche, et al 2002; Buus et al 2003; Nielsen et al 2003; Bhasin and Raghava 2005). In addition, we predicted machine learning based MHC ligands from whose C-terminal end is likely to be the result of proteosomal cleavage.

RESULTS AND INTERPRETATIONS

Machine learning technique predicts binding of peptides to a number of different alleles using Position Specific Scoring Matrix (PSSM). DNA-binding domain protein sequence is 207 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. PSSM based server predict the peptide binders to MHCI molecules of DNA-binding domain protein sequence are as 11mer_H2_Db, 9mer_H2_Db, 8mer_H2_Db and also peptide binders to MHCII molecules of DNA-binding domain protein sequence as I_Ab.p, I_Ag7.p, I_Ad.p, analysis found antigenic epitopes region in DNA-binding domain protein (table - 1). We also found the SVM based MHCII-IAb; MHCII-IAd; MHCII-IAg7 and MHCII- RT1.B peptide regions, which represented predicted binders from viral DNA-binding domain protein. The predicted binding affinity is normalized by the 1% fractil. We describe an improved method for predicting linear epitopes (table -2). The region of maximal hydrophilicity is likely to be an antigenic site, having hydrophobic characteristics, because terminal regions of DNA-binding domain protein is solvent accessible (Fig. 9, 10) and unstructured, antibodies against those regions are also likely to recognize the native protein (Fig. 1-4). It was shown that DNA-binding domain protein is hydrophobic in nature and contains segments of low complexity and high-predicted flexibility (Fig. 5-8). Predicted antigenic fragments can bind to MHC molecule is the first bottlenecks in vaccine design.
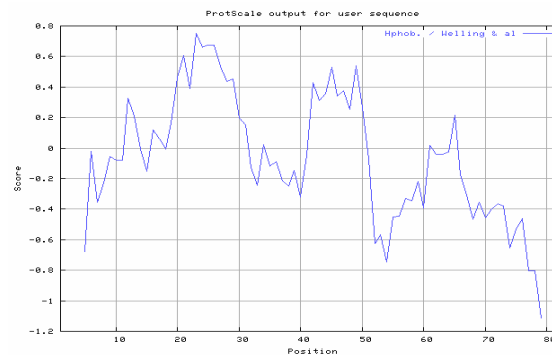

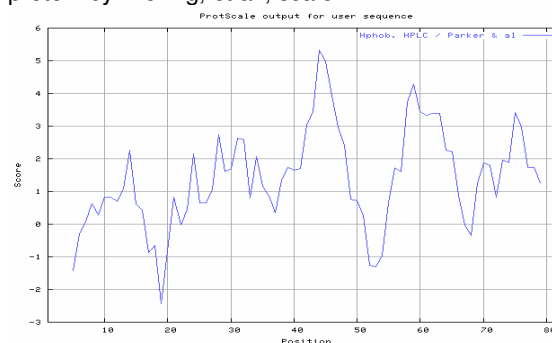Fig. 1- Antigenicity plot of DNA-binding domain protein by Welling, et al., scale

Fig. 2- Antigenicity plot of DNA-binding domain protein by HPLC / Parker, et al., scale
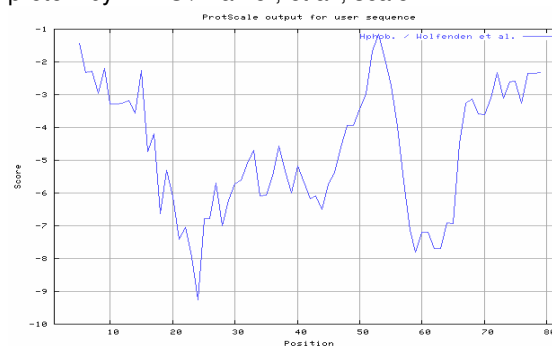

Fig. 3- Hydrophobicity plot of DNA-binding domain protein by Wolfenden, et al., scale
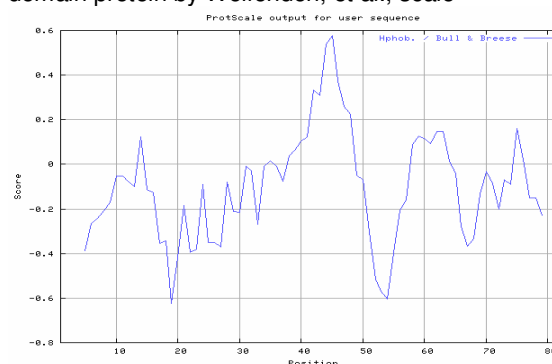

Fig. 4-Hydrophobicity plot of DNA-binding domain protein by Bull & Breese scale
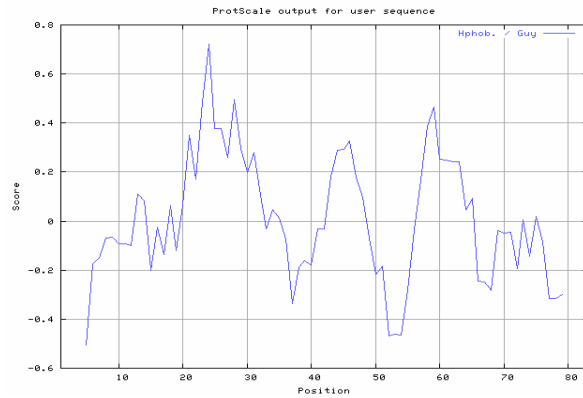
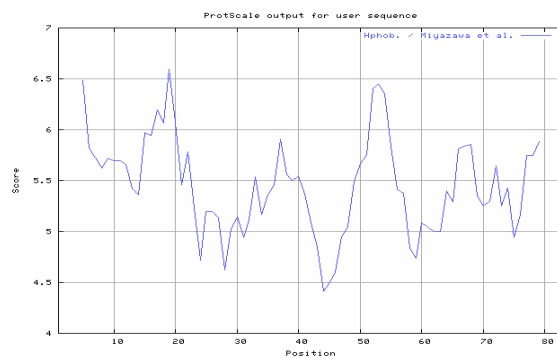Fig. 5- Hydrophobicity plot of DNA-binding domain protein by Guy scale

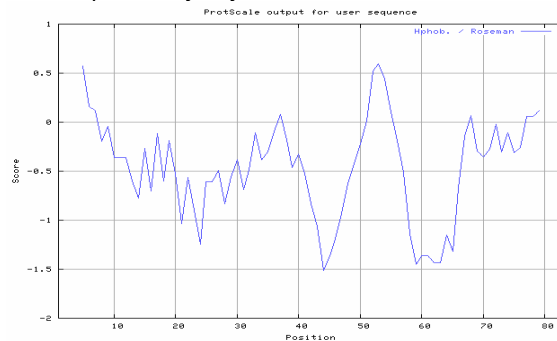Fig. 6- Hydrophobicity plot of DNA-binding domain protein by Miyazawa et al. scale

Fig. 7- Hydrophobicity plot of DNA-binding domain protein by Roseman scale
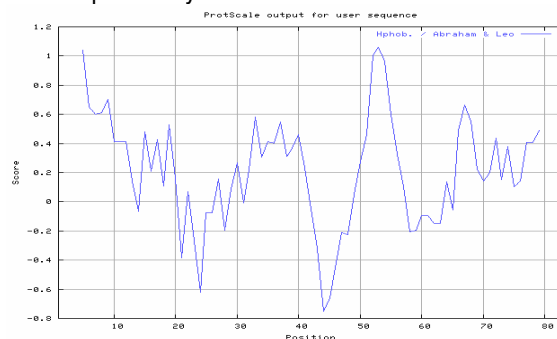

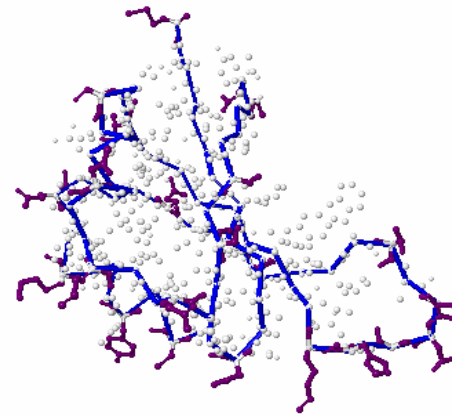Fig. 8- Hydrophobicity plot of DNA-binding domain protein by Abraham & Leo scale

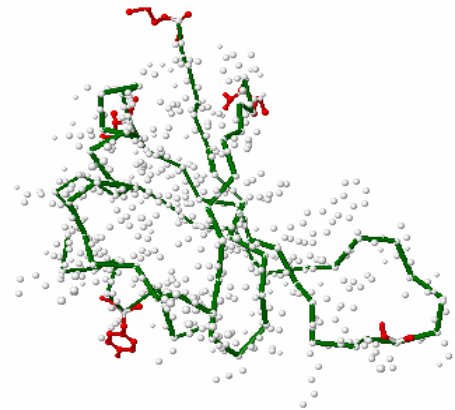Fig. 9- Structure showing accessible: accessibility > 75% in purple

Fig. 10-Structure showing accessible + hydrophobic in Red

CONCLUSION

DNA-binding domain protein from *Human papilloma virus-31* peptide nonamers are from a set of aligned peptides known to bind to a given major histocompatibility complex (MHC) molecule as the predictor of MHC-peptide binding. MHCII molecules bind peptides in similar yet different modes and alignments of MHCII-ligands were obtained to be consistent with the binding mode of the peptides to their MHC class, this means the increase in affinity of MHC binding peptides may result in enhancement of immunogenicity of viral DNA-binding domain protein. These predicted of DNA-binding domain protein antigenic peptides to MHC class molecules are important in vaccine development from *Human papilloma virus-31*.

REFERENCES

[1] Bhasin, M. and Raghava, G.P. (2005), "Pcleavage: an SVM based method for prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences". *Nucleic Acids Research*, 33, W202-207.

[2] Buus, S., *et al.*, (2003), "Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach". Tissue Antigens, 62, 378-384.

[3] Davidson, W., et al., (2004), "Characterization of the binding site for inhibitors of the HPV11 E1-E2 protein interaction on the E2 transactivation domain by photoaffinity labeling and mass spectrometry". *Anal Chem*., 76(7), 2095-102.

[4] Gomase, V.S. et al., (2006), "Automatic modeling of protein 3D structure Nucleoplasmin-like viral coat protein from Cucumber mosaic virus". *14th International Conference on Advance Computing & Communication, Published by IEEE Computer Society in IEEE Xplore USA* 614-615.

[5] Gomase, V.S. et al., (2007), "Prediction of MHC Binding Peptides and Epitopes from Alfalfa mosaic virus". *Curr. Drug Discov. Technol.*, 4(2),117-1215, 2007.

[6] Gomase, V.S. et al., (2008a), "Prediction of MHC Binding Peptides and Epitopes from Groundnut Bud Necrosis Virus (GBNV)". *J. of Proteomics & Bioinformatics*, 1 (4), 188- 205.

[7] Gomase, V.S. et al., (2008b), "Computer Aided Multi Parameter Antigen Design: Impact of Synthetic Peptide Vaccines from Soybean Mosaic Virus". *ICETET 2008, IEEE Computer Society in IEEE Xplore, Los Alamitos, California*, 629-634.

[8] Gomase, V.S. and Kale, K.V. (2008a), "In silico prediction of epitopes: a new approach for fragment based viral peptide vaccines". *Int. J. of Applied Computing*, 1(1), 39-46.

[9] Gomase, V.S. and Kale, K.V. (2008b), "Approach of proteomics system architecture in plant virus's database". *Int. J. of Applied Computin*g, 1(1), 33-38.

[10] Gomase, V.S. and Kale, K.V. (2008c), "Bioinformatics based sequence analysis of Nucleoplasmin like viral coat protein". Int. *J. of Information Retrieval*, 1(1), 11-15.

[11] Nielsen, M., *et al.*, (2003), "Reliable prediction of T-cell epitopes using neural networks with novel sequence representations". *Protein Sci.*, 12, 1007-1017.

[12] Sénéchal, H., et al., (2007), "Amino acid substitutions that specifically impair the transcriptional activity of papillomavirus E2 affect binding to the long isoform of Brd4". *Virology*, 358(1), 10-7.

[13] Reche, P.A. , (2002), "Prediction of MHC Class I Binding Peptides Using Profile Motifs". *Hum Immun.*, 63, 701-709.

[14] Wang, Y., et al., (2004), "Crystal structure of the E2 transactivation domain of human papillomavirus type 11 bound to a protein interaction inhibitor". *J Biol Chem.*, 279(8), 6976-85.

Table 1- PSSM based prediction of MHC ligands, from whose C-terminal ends are proteosomal cleavage sites

MHC-I	POS.	N	Sequence	C	MW (Da)	Score	% OPT.
8mer_H2_Db	76	TVS	VSTGYMTI		852.99	79.0	44.13 %
8mer_H2_Db	69	NTV	KIPNTVSV	STG	838.99	72.0	40.22 %
9mer_H2_Db	68	LNT	VKIPNTVSV	STG	938.12	94.0	64.38 %
9mer_H2_Db	75	NTV	SVSTGYMTI		940.07	71.0	48.63 %
10mer_H2_Db	10	IHL	KGDANILKCL	RYR	1056.28	110.0	49.77 %
10mer_H2_Db	59	IST	SQRDDFLNTV	KIP	1176.25	80.0	36.20 %
11mer_H2_Db	73	IPN	TVSVSTGYMTI		1140.3	82.0	38.68 %

Table 2- SVM based prediction of promiscuous MHC class II binding peptides from DNA-binding domain protein

MHC ALLELE	Rank	Sequence	Residue No.	Peptide Score
I-Ab	1	KGDANILKC	10	0.771
I-Ab	2	VTLTYISTS	51	0.770
I-Ab	3	MATTPIIHL	1	0.664
I-Ab	4	RLSKYKQLY	22	0.622
I-Ad	1	VSTGYMTI	76	0.468
I-Ad	2	HLKGDANIL	8	0.392
I-Ad	3	IIHLKGDAN	6	0.366
I-Ad	4	STSQRDDFL	57	0.336
I-Ag7	1	LNTVKIPNT	65	1.161
I-Ag7	2	LKGDANILK	9	1.091
I-Ag7	3	HKNAIVTLT	46	1.080
I-Ag7	4	KGDANILKC	10	0.908
RT1.B	1	SKYKQLYEQ	24	0.807
RT1.B	2	TSQRDDFLN	58	0.708
RT1.B	3	TLTYISTSQ	52	0.569
RT1.B	4	YISTSQRDD	55	0.269