International Journal of Computational Biology, Vol. 1, Issue 1, 2010, pp-04-08

Support Vector Machine (SVM) based prediction of promiscuous MHC class binding peptides

Gomase V.S.* and Chitlange N.R.
School of Technology, S.R.T.M. University, Sub-Centre, Latur, 413531, MS, India, Mobile-09987770696, Mail- gomase.viren@gmail.com

Abstract- • *Shigella flexneri* is a species of Gram-negative bacteria in the genus Shigella that can cause diarrhea in humans. Peptide fragments of antigen protein can be used to select nonamers for use in rational vaccine design and to increase the understanding of roles of the immune system in infectious diseases. Analysis shows MHC class II binding peptides of antigen protein from *Shigella flexneri* are important determinant for protection of host form parasitic infection. In this assay, we used PSSM and SVM algorithms for antigen design and predicted the binding affinity of antigen protein having 574 amino acids, which shows 566 nonamers. Binding ability prediction of antigen peptides to major histocompatibility complex (MHC) class I & II molecules is important in vaccine development from *Shigella flexneri*.
Keywords- *antigen protein, epitope, PSSM, SVM, MHC, peptide vaccine*

Abbreviations: *Goldman, Engelberg and Steitz, (GES); major histocompatibility complex, (MHC); Position Specific Scoring Matrices, (PSSMs); Support Vector Machine, (SVM)*

I. Introduction

Shigella flexneri is a species of Gram-negative bacteria in the genus Shigella that can cause diarrhea in humans [1]. *Shigella flexneri* antigen peptides are most suitable for subunit vaccine development because with single epitope, the immune response can be generated in large population. This approach is based on the phenomenon of cross-protection, whereby a infected is protected against a more severe strain of the same. The phenotype of the resistant transgenic hosts includes fewer centers of initial infection, a delay in symptom development, and low accumulation. Antigen protein from *Shigella flexneri* is necessary for new paradigm of synthetic vaccine development and target validation [2-4].

II. Methodology

In this research work antigenic epitopes of antigen protein from *Shigella flexneri* lina is determined using the Gomase in 2007, Hopp and Woods, Bull & Breeze, Parker, Chou & Fasman and Deleage & Roux antigenicity [5-7]. The major histocompatibility complex (MHC) peptide binding of antigen protein is predicted using neural networks trained on C terminals of known epitopes. In analysis predicted MHC/peptide binding of antigen 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 [8-13]. In addition, we predict those MHC ligands from whose C-terminal end is likely to be the result of proteosomal cleavage [14].

III. Results and Interpretations

We found binding of peptides to a number of different alleles using Position Specific Scoring Matrix. An antigen protein sequence is 574 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 antigen protein sequence are as 11mer_H2_Db, 10mer_H2_Db, 9mer_H2_Db, 8mer_H2_Db and also peptide binders to MHCII molecules of antigen protein sequence as I_Ab.p, I_Ad.p, analysis found antigenic epitopes region in putative antigen protein (Table 1). We also found the SVM based MHCII-IAb peptide regions; MHCII-IAd peptide regions; MHCII-IAg7 peptide regions and MHCII- RT1.B peptide regions, which represented predicted binders from bacterial antigen protein (Table 2). 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 antigen

protein is solvent accessible and unstructured, antibodies against those regions are also likely to recognize the native protein (Fig. 1, 2, 3). It was shown that a antigen protein is hydrophobic in nature and contains segments of low complexity and high-predicted flexibility (Fig. 4, 5). Predicted antigenic fragments can bind to MHC molecule is the first bottlenecks in vaccine design.

IV. Conclusion

An antigen protein from *Shigella flexneri* peptide nonamers are from a set of aligned peptides known to bind to a given 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 bacterial antigen protein. These predicted of antigen protein antigenic peptides to MHC class molecules are important in vaccine development from *Shigella flexneri.*

V. References

[1] Ryan K.J., Ray C.G. (editors) (2004) *Sherris Medical Microbiology (4th ed. ed.). McGraw Hill*. ISBN 0-8385-8529-9.

[2] Joshi J. Tang, Kuzma M., Wagner J., Mookerjee B., Filicko J., Carabasi M., Flomenberg N., Flomenberg P. (2009) *J. Gen Virol.*, 90(Pt 1), 84-94.

[3] McDonald D., Stockwin L., Matzow T., Blair Zajdel M.E., Blair G.E. (1999) *Gene Ther.*, 6(9),1512-9.

[4] Gomase V.S., Kale K.V. and Shyamkumar K. (2008) *J. of Proteomics & Bioinformatics*, 1 (4), 188- 205.

[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. (2008) *Int. J. of Applied Computing*, 1(1), 39-46.

[7] Gomase V.S., Kale K.V. (2008) *Int. J. of Applied Computin*g, 1(1), 33-38.

[8] Gomase V.S., Kale K.V. (2008) Int. *J. of Information Retrieval*, 1(1), 11-15.

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

[10] 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.

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

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

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

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

MHC-I	POS.	N	Sequence	C	MW (Da)	Score	% OPT.
8mer_H2_Db	109	ELH	VNGNNLNI	LPT	838.9	14.615	27.84 %
8mer_H2_Db	273	TSS	PDYHGPQI	YFS	907.99	13.06	24.88 %
8mer_H2_Db	314	DVS	QIWHAFEH	EEH	1026.16	12.169	23.18 %
8mer_H2_Db	230	ALD	VSNNLLTS	LPE	828.91	9.693	18.46 %
8mer_H2_Db	118	NIL	PTLPSQLI	KLN	850.03	9.149	17.43 %
8mer_H2_Db	238	LTS	LPENIITL	PIC	894.08	8.937	17.02 %
8mer_H2_Db	170	ILR	IEGNRLTV	LPE	883.01	8.931	17.01 %
8mer_H2_Db	3	MK	PINNHSFF	RSL	957.06	8.094	15.42 %
8mer_H2_Db	253	NVN	ISGNPLST	RVL	769.85	7.238	13.79 %
8mer_H2_Db	51	NRI	QAVRLLKI	CLD	922.18	7.01	13.35 %
8mer_H2_Db	191	LFV	SGNRLQEL	PEF	897.99	6.951	13.24 %
8mer_H2_Db	287	MSD	GQQNTLHR	PLA	935.0	6.718	12.80 %
9mer_H2_Db	237	LLT	SLPENIITL	PIC	981.16	18.672	37.07 %
9mer_H2_Db	130	LNI	SFNRNLSCL	PSL	1035.19	17.594	34.93 %
9mer_H2_Db	109	ELH	VNGNNLNIL	PTL	952.06	17.236	34.22 %
9mer_H2_Db	96	ELI	SLPENSPLL	TEL	951.1	16.066	31.90 %
9mer_H2_Db	11	SFF	RSLCGLSCI	SRL	933.16	13.949	27.70 %
9mer_H2_Db	2	M	KPINNHSFF	RSL	1085.23	12.886	25.59 %
9mer_H2_Db	106	LLT	ELHVNGNNL	NIL	991.06	12.736	25.29 %
9mer_H2_Db	433	DKV	RTLHFVDEI	EVY	1111.27	12.469	24.76 %
9mer_H2_Db	87	IRE	LNISNNELI	SLP	1011.14	10.929	21.70 %
9mer_H2_Db	108	TEL	HVNGNNLNI	LPT	976.04	10.39	20.63 %
9mer_H2_Db	1		MKPINNHSF	FRS	1069.24	9.797	19.45 %
9mer_H2_Db	385	RVA	LTWNNLRKT	LLV	1104.29	9.013	17.90 %
10mer_H2_Db	136	RNL	SCLPSLPPYL	QSL	1071.32	17.645	29.98 %
10mer_H2_Db	149	QSL	SARFNSLETL	PEL	1119.25	14.451	24.55 %
10mer_H2_Db	459	QLS	TAVKEMRFYG	VSG	1183.39	13.388	22.75 %
10mer_H2_Db	385	RVA	LTWNNLRKTL	LVH	1217.45	12.967	22.03 %
10mer_H2_Db	20	SCI	SRLSVEEQCT	RDY	1133.25	12.471	21.19 %
10mer_H2_Db	322	FEH	EEHANTFSAF	LDR	1134.18	11.147	18.94 %
10mer_H2_Db	216	NQL	RRLSRLPQEL	LAL	1249.5	11.06	18.79 %
10mer_H2_Db	240	SLP	ENIITLPICT	NVN	1098.32	10.257	17.43 %
10mer_H2_Db	111	HVN	GNNLNILPTL	PSQ	1050.21	10.021	17.03 %
10mer_H2_Db	540	LKA	SGLSGDADAQ	REA	901.89	9.395	15.96 %
10mer_H2_Db	486	MVR	SREENEFTDW	FSL	1271.31	8.904	15.13 %
10mer_H2_Db	5	KPI	NNHSFFRSLC	GLS	1206.35	8.679	14.75 %
11mer_H2_Db	248	LPI	CTNVNISGNPL	STR	1113.24	16.495	20.75 %
11mer_H2_Db	64	DTR	EPVLNLSLLKL	RSL	1220.52	14.086	17.72 %
11mer_H2_Db	87	IRE	LNISNNELISL	PEN	1211.38	12.715	15.99 %
11mer_H2_Db	523	QKY	EMLENEYSQRV	ADR	1379.52	11.854	14.91 %
11mer_H2_Db	1		MKPINNHSFFR	SLC	1372.61	9.724	12.23 %
11mer_H2_Db	515	DRW	AQAEEQKYEML	ENE	1321.48	9.423	11.85 %
11mer_H2_Db	339	SDT	VSARNTSGFRE	QVA	1205.3	8.698	10.94 %
11mer_H2_Db	128	IKL	NISFNRNLSCL	PSL	1262.45	7.901	9.94 %
11mer_H2_Db	50	ENR	IQAVRLLKICL	DTR	1251.64	7.48	9.41 %
11mer_H2_Db	237	LLT	SLPENIITLPI	CTN	1191.44	6.529	8.21 %
11mer_H2_Db	228	LLA	LDVSNNLLTSL	PEN	1170.32	6.46	8.13 %
11mer_H2_Db	459	QLS	TAVKEMRFYGV	SGV	1282.52	6.043	7.60 %

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

MHC ALLELE	Rank	Sequence	Residue No.	Peptide Score
I-Ab	1	GVTANDLRT	471	1.572
I-Ab	2	PWHAVLKRT	501	1.314
I-Ab	3	RIQAVRLLK	49	1.201
I-Ab	4	PVLNLSLLK	65	0.989
I-Ad	1	ASGLSGDAD	539	0.649
I-Ad	2	TGALLSLGR	408	0.603
I-Ad	3	VSNNLLTSL	230	0.596
I-Ad	4	TANDLRTAE	473	0.573
I-Ag7	1	SFAVAADAT	368	1.715
I-Ag7	2	GDADAQREA	544	1.659
I-Ag7	3	SGVTANDLR	470	1.608
I-Ag7	4	SGDADAQRE	543	1.544
RT1.B	1	TFSAFLDRL	327	1.137
RT1.B	2	QTMLAEKLQ	448	1.121
RT1.B	3	SFAVAADAT	368	1.036
RT1.B	4	DRWAQAEEQ	512	0.858

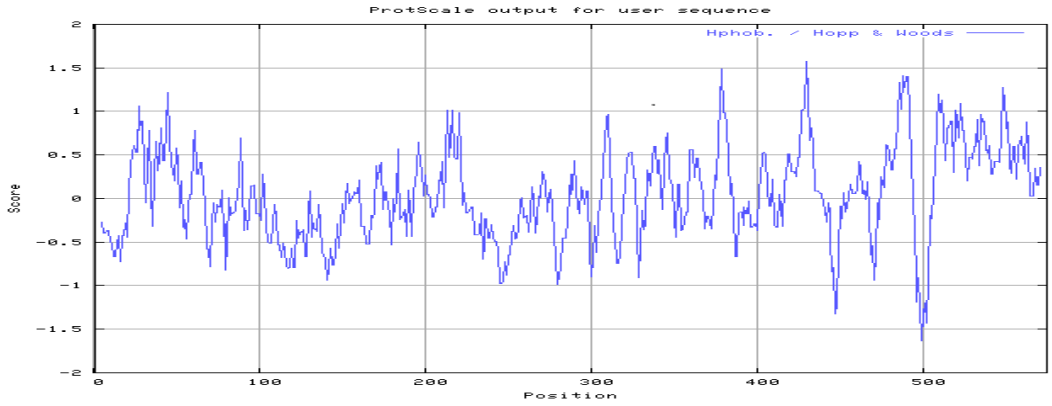

Fig.1-Hydrophobicity plot of antigen protein by Hphob./ Hopp & Woods scale

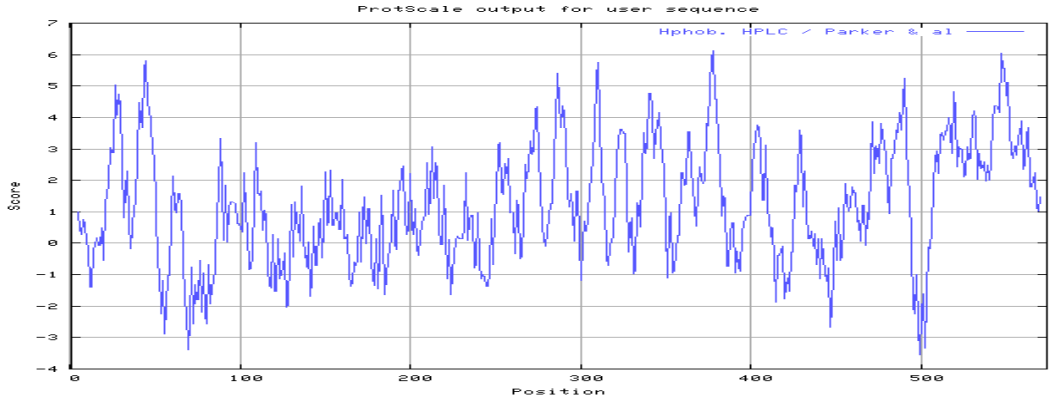

Fig. 2- Hydrophobicity plot of antigen protein by Hphob/ Parker, et al., scale

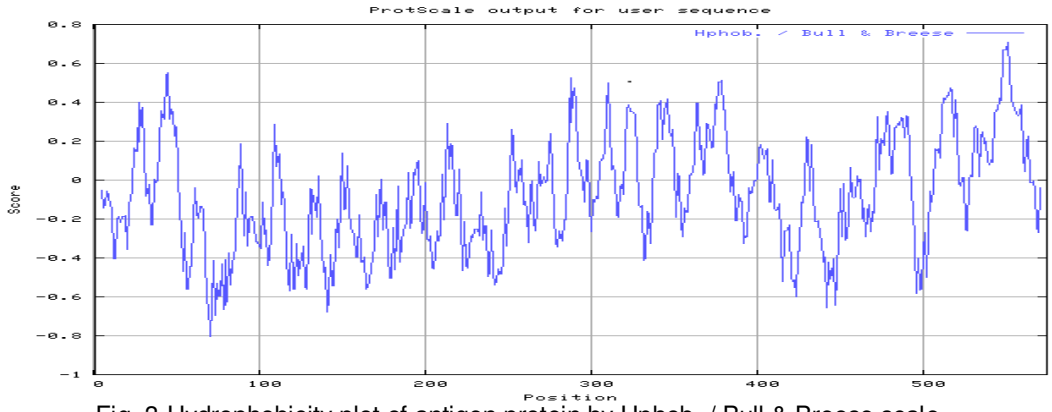

Fig. 3-Hydrophobicity plot of antigen protein by Hphob. / Bull & Breese scale

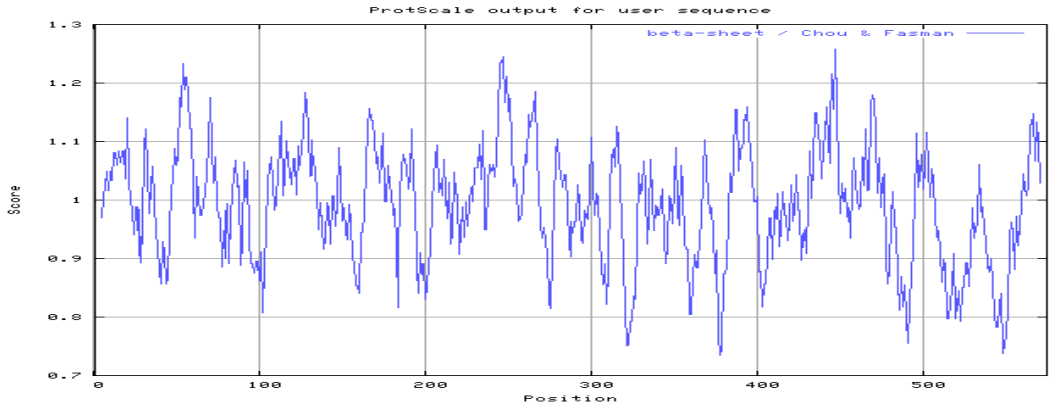

Fig. 4- Antigenicity plot of antigen protein by beta turn /Chou & Fasman., scale

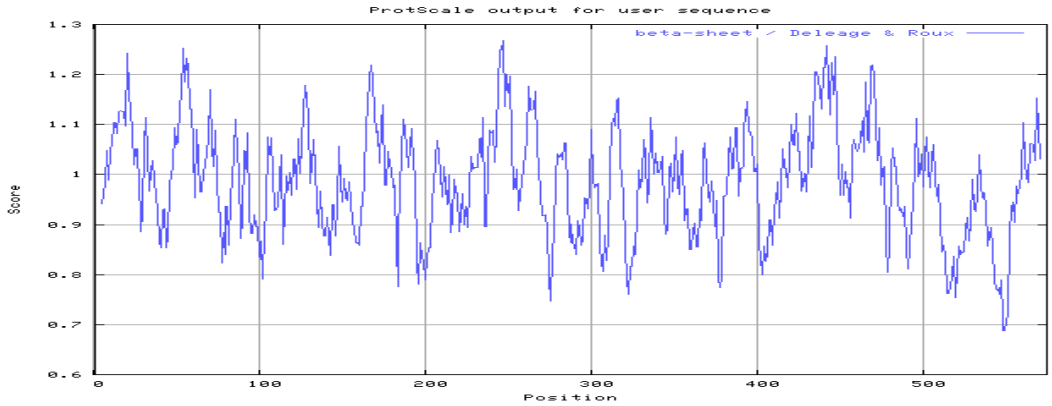

Fig. 5- Antigenicity plot of antigen protein by Deleage & Roux, scale