Gene Ther Mol Biol Vol 12, 31-38, 2008

Antigenic epitopes of viral polyprotein: an approach for fragment based peptide vaccines from Papaya Ringspot virus

Research Article

Virendra S Gomase[1,*] Karbhari V Kale[2]

[1]Department of Bioinformatics, Dr. D.Y. Patil Institute for Biotechnology and Bioinformatics, Padmashree Dr. D.Y. Patil University

[2]Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, 431004 (MS), India

***Correspondence**: Virendra S Gomase, Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, 431004 (MS), India; e-mail: virusgene1@yahoo.co.in

Key words: HMM, MHC, linear epitopes, cleavage sites

Abbreviations: Goldman, Engelberg and Steitz, (GES)grand average of hydropathicity, (GRAVY); Papaya ringspot virus, (PRSV)

Received: 28 May 2008; Revised: 2 June 2008
Accepted: 4 June 2008; electronically published: June 2008

Summary

Papaya ringspot is a destructive disease characterized by a yellowing and stunting of the crown of papaya trees and assay was designed to help assign putative genome polyprotein analysis of Papaya ringspot virus strain W. We used different methods for the prediction of linear epitopes using a combination of a hidden Markov model and a propensity scale method. Data set was collected from the literature, and data sets of epitopes in the genome polyprotein having twenty four antigenic determinants in 675 residues long sequence. The structural homology modeling method is allows potential drug targets to identify active sites i.e. linear epitopes, which form antibodies in host cells. The method integrates prediction of peptide MHC class I binding; proteasomal C terminal cleavage and TAP transport efficiency. The challenges for the future are to establish the function of all of protein structures. In this assay we use of multiple methods towards the accurate identification of antigenic epitopes. The proposed approach is useful not only for plant and viral biology but it covers the wide area of vaccines and antibodies for therapeutic purposes in humans.

I. Introduction

Carica papaya (*fam: Caricaceae*) is a herbaceous plant with a soft stem, which may grow as high as 8 meters. It produces male, female, and bisexual fl owers. The male plants do not normally produce fruits. It is a cross-pollinated plant widely grown in the tropics and subtropics. Papaya ringspot virus (PRSV), genus Potyvirus. PRSV is perhaps the most limiting factor in papaya production in many countries and occurs in the majority of papaya growing regions (Dahal et al, 1997; Xiao et al, 1997; Davis et al, 1999; Noe- Becerra et al, 1999). Virions are flexuous filamentous particles about 780 nm long and the virus produces inclusion bodies in the cytoplasm of host cells. Isolates of PRSV belong to one of two major types which are serologically related (Purcifull et al, 1984). Type P infects papaya and cucurbits and type W infects watermelon and other cucurbits but not papaya. Genetic diversity is reported to occur within type P with isolates from a region being more closely related than isolates from other regions. Type P and type W isolates from one region are generally more closely related to each other than to isolates of either type from other regions (Bateson et al, 1994; Brunt et al, 1996).

A. Description

Papaya ringspot is a destructive disease characterized by a yellowing and stunting of the crown of papaya trees, a mottling of the foliage, shoe-stringing of younger leaves, water-soaked streaking of the petioles (stalks), and small darkened rings on the surface of fruit (**Figure 1**). Other pest organisms, such as various species of mites and powdery mildew, may cause symptoms similar to PRV. Herbicides drifting onto developing papaya trees may also cause symptoms such as shoe stringing. In severe cases,

fruits may become distorted. Susceptible host species are *Carica papaya, Chenopodium amaranticolor, Chenopodium quinoa, Cucumis melo, Cucumis metuliferus, Cucumis sativus, Cucurbita maxima, Cucurbita moschata, Cucurbita pepo*.

B. Coat protein

Coat protein is named for their primary function; to encapsidate viral genomic nucleic acids. However, encapsidation is only one feature of an extremely diverse array of structural, functional, and ecological roles played during viral infection and spread (Callaway et al, 2001). The coat protein is multifunctional; in addition to having a role in encapsidation; it affects virus movement in plants, (Kaplan et al, 1998; Suzuki et al, 1991) transmission, symptom expression, and host range (Shintaku and Palukaitis, 1990). The predictive power of these bioinformatics approaches is strongest when information from several techniques is combined, including experimental confirmation of protein antigenicity predictions (Gomase and Changbhale, 2007; Gomase et al, 2007).

II. Materials and Methods

The protein sequences databases are used to store the vast amount of information issuing from the genome projects. We analysed the genome protein sequence of a viral genome polyprotein (Quemada et al, 1990; Urcuqui-Inchima et al, 2001). This program predicts those segments from within viral coat protein that are likely to be antigenic by eliciting an antibody response. Antigenic epitope is determined using the Hopp and Woods, Welling and Protrusion Index (Thornton) antigenicity methods (Welling et al, 1985; Thornton et al, 1986; Parker et al, 1994; IsHak et al, 2003; Gomase, 2006). Predictions are based on a table that reflects the occurrence of amino acid residues in experimentally known segmental epitopes, also we used BepiPred 1.0 server which predicts the location of linear B-cell epitopes using a combination of a hidden Markov model and a propensity scale method (Larsen et al, 2006). The important concepts in secondary structure prediction are identified as: residue conformational propensities, sequence edge effects, moments of hydrophobicity, position of insertions and deletions in aligned homologous sequence, moments of conservation, auto-correlation, residue ratios, secondary structure feedback effects, and filtering (Garnier et al, 1996 and Robson and Garnier, 1993). For setting the solvent accessible regions in protein, type of plot determine the hydrophobic scale and it is utilized for prediction. Sequence of coat protein was entered into program-Protein Hydrophobicity plot that characterize its hydrophobic and hydrophilic character, which may be useful in predicting membrane-spanning domains, potential antigenic sites and regions that are likely exposed on the protein surface (Manavalan and Ponnuswamy, 1978; Janin et al, 1978; Janin, 1979; von Heijne, 1981; Kyte and Doolittle, 1982; Fauchere and Pliska, 1983; Sweet and Eisenberg, 1983; Engelman et al, 1986).

III. Results and Interpretations

Sequence of viral genome polyprotein is as follows-

LVRKSCERLYEGRMGVWNGSLKAELRPAEKV
LAKKTRSFTAAPLDTLLGAKVCVDDFNNWFYSKN
MECPWTVGMTKFYKGWDEFLRKFPDGWVYCDAD
GSQFDSSLTPYLLNAVLSIRLWAMEDWDIGEQMLK
NLYGEITYTPILTPDGTIVKKFKGNNSGQPSTVVDNT
LMVLITMYYALRKAGYDTKTQEDMCVFYINGDDL
CIAIHPDHEHVLDSFSRSFAELGLKYDFTQRHRNKQ
NLWFMSHRGILIDDIYIPKLEPERIVAILEWDKSKLPE
HRLEAITAAMIESWGYGDLTHQIRRFYQWVLEQAP
FNELAKQGRAPYVSEVGLRRLYTSERGSMDELEAYI
DKYFERERGDSPELLVYHESRSTDDYQLVCSNNTH
VFHQSKNEAVDTGLNEKFKEKEKQKEKEKEKQKE
KEKDDASDGNDVSTSTKTGERDRDVNVGTSGTFTV
PRIKSFTDKMILPRIKGKSVLNLNHLLQYNPQQIDIS
NTRATQSQFEKWYEGVRNDYGLNDNEMQVMLNG
LMVWCIENGTSPDISGVWVMMDGETQVDYPIKPLI
EHATPSFRQIMAHFSNAAEAYIAKRNATERYMPRY
GIKRNLTDISLARYAFDFYEVNSKTPDRAREAHMQ
MKAAALRNTSRRMFGMDGSVSNKEENTERHTVED
VNRDMHSLLGMRN.

IV. Prediction of Antigenic peptides

In these methods we found the antigenic determinants by finding the area of greatest local hydrophilicity. The Hopp-Woods scale was designed to predict the locations of antigenic determinants in a protein, assuming that the antigenic determinants would be exposed on the surface of the protein and thus would be located in hydrophilic regions. Its values are derived from the transfer-free energies for amino acid side chains between ethanol and water. Welling antigenicity plot gives value as the log of the quotient between percentage in a sample of known antigenic regions and percentage in average proteins (**Figures 2-6**). A genome polyprotein

Figure 1. (**A**). Papaya leaf infected with Papaya ringspot virus; (**B**). close up of PRV infected papaya fruit showing ring spots.

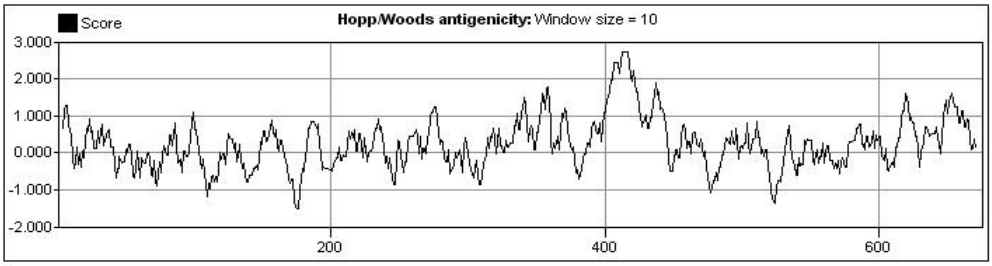

Figure 2 - Hopp-Woods antigenicity plot of genome polyprotein.

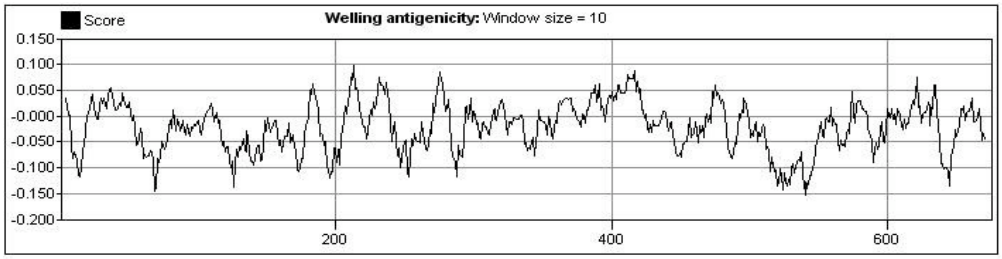

Figure 3 - Welling antigenicity plot of genome polyprotein.

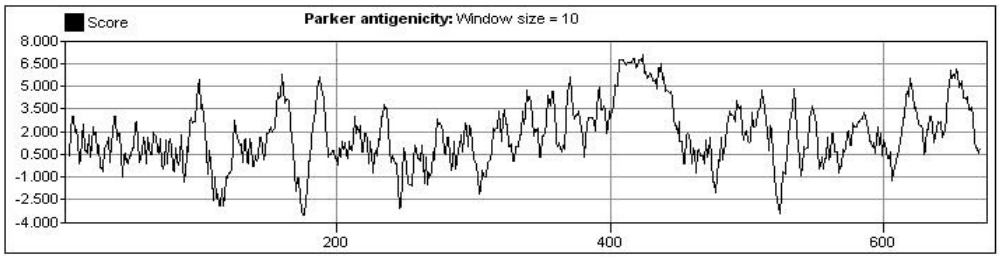

Figure 4 - Parker antigenicity plot of genome polyprotein.

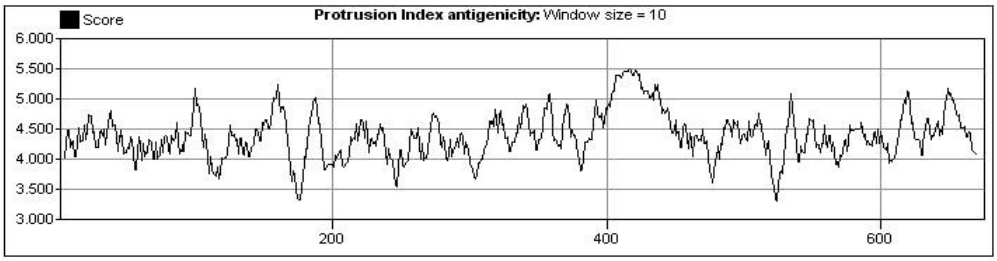

Figure 5 - Protrusion Index (Thornton) antigenicity plot of genome polyprotein.

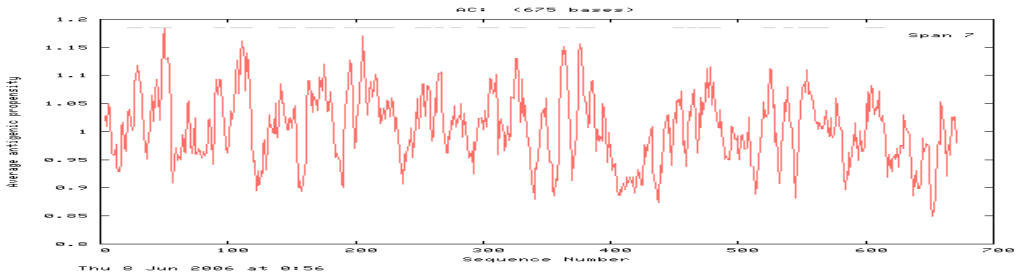

Figure 6 - Kolaskar and Tongaonkar antigenicity determinant plot. X-axis contains sequence number and y-axis contain average antigenic propensity.

sequence is 675 residues long, having twenty four antigenic determinants in sequence. It has a Molecular weight: 77925.1 KD; Aliphatic index is 72.52, grand average of hydropathicity (GRAVY) is -0.630 and

Theoretical pI is 5.98. Peptides found in the genome polyprotein are epitopes present in the papaya mosaic virus strain W eliciting the desired immune response. Predicted antigenic fragments can bind to MHC molecule is the first bottlenecks in vaccine design (**Table 1**).

V. Secondary alignment

The Robson and Garnier method predicted the secondary structure of genome polyprotein. Each residue is assigned values for alpha helix, beta sheet, turns and coils using a window of 7 residues. Empirical studies show that an amino acid exerts a significant effect on the conformational state of residues up to eight residues distant; Using these information parameters, the likelihood of a given residue assuming each of the four possible conformations alpha, beta, reverse turn, or coils calculated, and the conformation with the largest likelihood is assigned to the residue (**Figure 7**).

VI. Solvent accessible regions

Which used widely applied scale for delineating hydrophobic and hydrophilic characteristics of amino acids (**Figures 8-14**). This scale was developed for predicting potential antigenic sites of coat protein, which are likely to be rich in charged and polar residues. Scales shows a hydrophilic index, with apolar residues assigned negative values. It is suggest that the lack of rigid globular structure under physiological conditions might represent a considerable functional advantage for natively unfolded proteins, as their large plasticity allows them to

Table 1. There are 24 antigenic determinants in protein sequence**.**

Sr. no	Start Position	Sequence	End Position
1	25	LRP	27
2	37	RSFTAA	42
3	93	VYCDADGSQFDSS	105
4	125	WDI	127
5	144	ILTPDGT	150
6	156	KGNNSGQPSTVV	167
7	184	AGYDTKTQ	191
8	213	EH	214
9	235	RHRN	238
10	274	SKLP	277
11	294	GY	295
12	314	FNELAKQGRAPYV	326
13	337	SERGSMD	343
14	354	ERERGDSP	361
15	369	SRSTDD	374
16	390	KNEAVDTGLNEKFKEKEKQKEKEKEKQKEKEK DDASDGNDVSTSTKTGERDRDVNVGTSGTFT	452
17	484	PQQIDISNTRATQSQFE	500
18	504	EGVRNDYGLND	514
19	531	NGTSPDI	537
20	547	ETQVDYP	553
21	559	EHATP	563
22	584	RNATERYM	590
23	614	VNSKTPDRAREA	624
24	646	GSVSNKEENTERHTVEDVNR	665

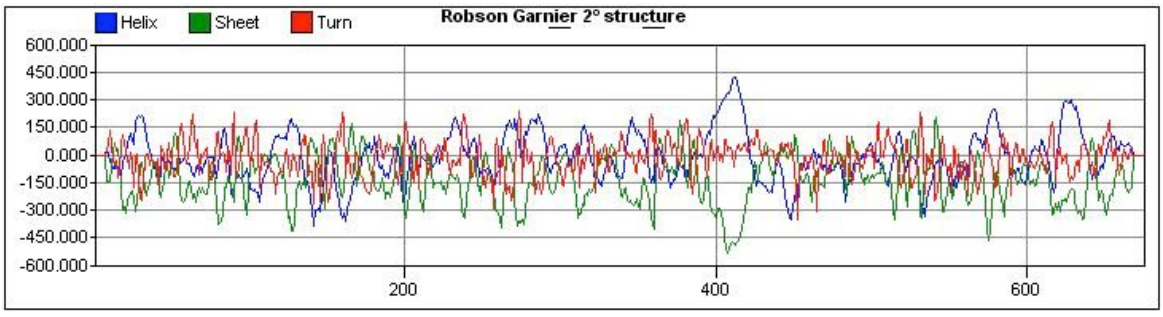

Figure 7. Secondary structure plot of genome polyprotein.

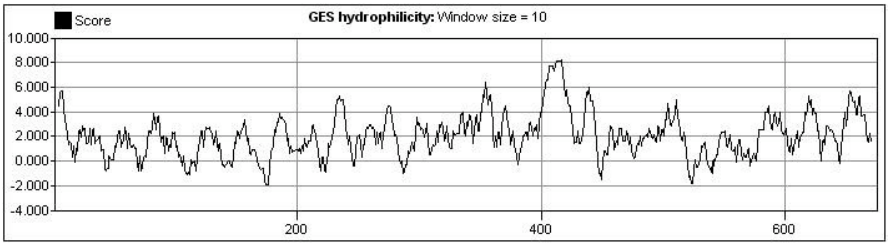

Figure 8. Goldman, Engelberg and Steitz (GES) hydrophilicity of genome polyprotein.

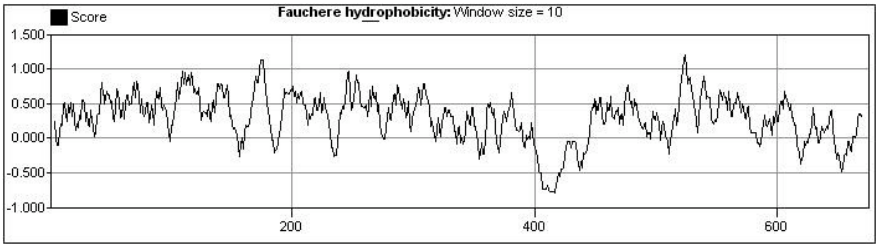

Figure 9. Fauchere hydrophobicity of genome polyprotein.

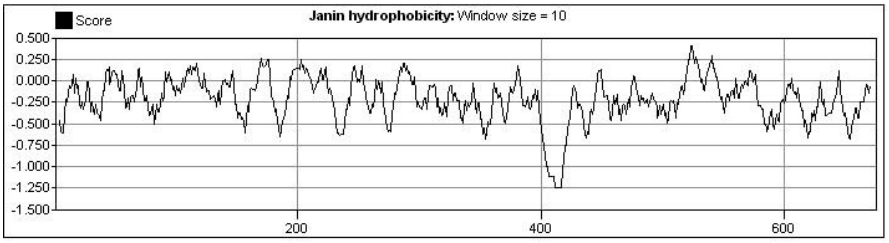

Figure 10. Janin hydrophobicity of genome polyprotein.

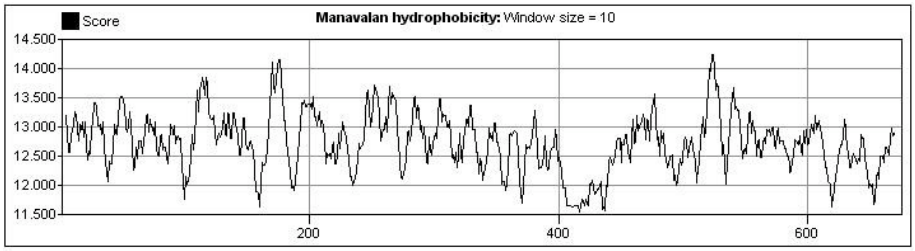

Figure 11. Manavalan hydrophobicity of genome polyprotein.

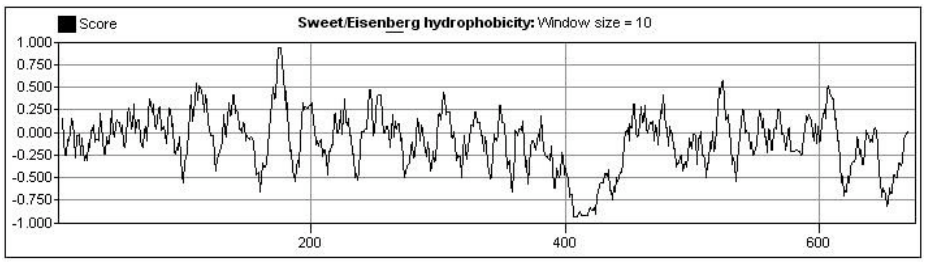

Figure 12. Sweet / Eisenberg hydrophobicity of genome polyprotein.

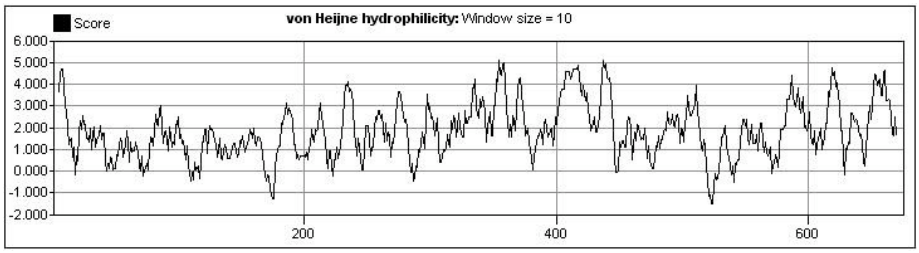

Figure 13. von Heijne hydrophilicity of genome polyprotein.

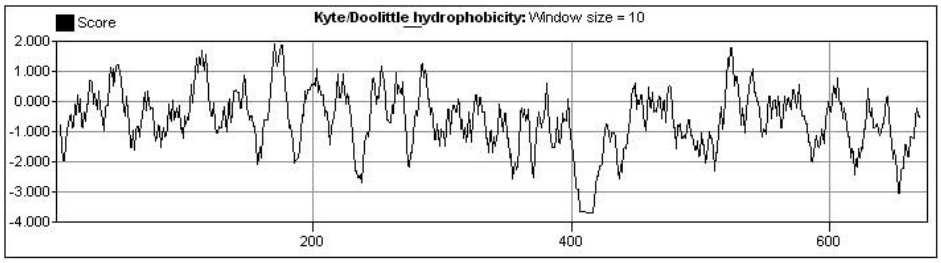

Figure 14. Kyte-Doolittle Hydrophobicity of Genome polyprotein.

interact efficiently with several different targets, as compared to a folded protein with limited conformational flexibility. According to BepiPred 1.0 Server, We have measured the performance in a non-parametric way by constructing ROC-curves and found epitopes (**Table 1**), which shows coat protein is hydrophobic in nature and contains segments of low complexity and high predicted flexibility.

VII. Prediction of MHC binding peptides

The MHC peptide binding is predicted using neural networks trained as described for the NetMHC server. In analysis predicted MHC/peptide binding is a log transformed value related to the IC50 values in nM units. Total numbers of peptides found are 667 and server predicted 20 MHC ligands (**Table 2**). 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 plant diseases. The method integrates prediction of peptide MHC class I binding; proteasomal C terminal cleavage and TAP transport efficiency.

VIII. Discussion

We found the antigenic determinants by finding the area of greatest local hydrophilicity. The Hopp-Woods, Welling and Protrusion Index (Thornton) antigenicity scale was designed to predict the locations of antigenic determinants (**Figures 2-6**). Further this region form beta sheet. Thus beta sheet show high antigenic response than helical region of this peptide (**Figure 7**). A genome polyprotein is highly antigenic in nature. We also consider Fauchere Hydrophobicity, Goldman, Engelberg and Steitz (GES) Hydrophilicity, Janin Hydrophobicity, Kyte / Doolittle Hydrophobicity, Manavalan Hydrophobicity, Sweet / Eisenberg Hydrophobicity, von Heijne Hydrophilicity scales, Theses scales are essentially a hydrophilic index, with apolar residues assigned negative values. The region of maximal hydrophilicity is likely to be an antigenic site, having hydrophobic characteristics (**Figures 8-14**). Because the N- and C- terminal regions of proteins are usually solvent accessible and unstructured, antibodies against those regions are also likely to recognize the native protein. These regions are antigenic in nature and form antibodies (**Tables-1-2**).

IX. Conclusion

BepiPred 1.0 server predicts the location of linear B-cell epitopes using a combination of a hidden Markov model and a propensity scale method. We have measured the performance in a non-parametric way by constructing ROC-curves. When tested on the validation data set this method performs significantly better than any of the other methods tested. Findings show that peptides presented in a genome polyprotein results in enhanced immune responses. Recombinant DNA vaccines involve targeting multiple antigenic components to direct and empower the immune system to protect the host from chemical reaction. Antigenic epitopes of coat protein are important antigenic determinants against the viral attack on papaya and other plants.

Table 2. prediction of peptide MHC class I binding, proteasomal C terminal cleavage and TAP transport efficiency.

Residue number	Sequence	Predicted MHC binding affinity	Rescale binding affinity	C terminal cleavage affinity	TAP transport efficiency
53	CVDDFNNWF	0.3545	1.5050	0.8791	2.5030
54	VDDFNNWFY	0.4408	1.8715	0.7449	2.4960
70	WTVGMTKFY	0.3862	1.6399	0.7561	2.8090
101	QFDSSLTPY	0.1484	0.6300	0.9160	2.9830
170	TLMVLITMY	0.2714	0.2714	0.9697	3.1290
171	LMVLITMYY	0.2669	1.1333	0.9771	3.1380
190	TQEDMCVFY	0.2664	1.1309	0.7258	2.9360
215	VLDSFSRSF	0.1226	0.5207	0.8979	2.4170
222	SFAELGLKY	0.1235	0.5242	0.9743	3.2940
286	AAMIESWGY	0.3171	1.3465	0.5146	3.1960
297	LTHQIRRFY	0.3677	1.5614	0.8043	2.7980
340	GSMDELEAY	0.3117	1.3234	0.2585	2.9830
344	ELEAYIDKY	0.3403	1.4448	0.9734	2.4570
371	STDDYQLVC	0.3884	1.6489	0.0468	0.0020
459	FTDKMILPR	0.2688	1.1412	0.0666	1.2460
502	WYEGVRNDY	0.1248	0.5300	0.9583	3.1190
571	HFSNAAEAY	0.1428	0.6064	0.7673	3.1510
585	ATERYMPRY	0.6661	2.8283	0.8763	3.0440
599	LTDISLARY	0.7724	3.2796	0.9628	2.8720
604	LARYAFDFY	0.1510	0.6411	0.5685	3.0840

References

Bateson MF, Henderson J, Chaleeprom W, Gibbs AJ, Dale JL (**1994**) Papaya ringspot potyvirus: isolate variability and the origin of PRSV type P (Australia). **J Gen Virol** 75: 3547-3553.

Brunt AA, Crabtree K, Dallwitz MJ, Gibbs AJ, Watson L (**1996**) Viruses of Plants. CAB International, Wallingford, UK, 1484.

Callaway A, Giesman-Cookmeyer D, Gillock ET, Sit TL, Lommel SA (**2001**) The multifunctional capsid proteins of plant RNA viruses. **Annu Rev Phytopathol**, 39, 419-60.

Dahal G, Lecoq H, Albrechtsen SE (**1997**) Occurrences of Papaya ringspot potyvirus and cucurbit viruses in Nepal. **Ann Appl Biol** 130: 491-502.

Davis MJ, Ying-Zhen T, Ying ZT (**1997**) Genetic diversity of Papaya ringspot virus in Florida. Proceedings of the 12th Annual Meeting of the Florida State Horticultural Society, Sturt, Florida, USA,194-196.

Engelman DM, Steitz TA, Goldman A (**1986**) Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. **Annu Rev Biophys Biophys Chem** 15, 321-353.

Fauchere JL, Pliska V (**1983**) Hydrophobic parameters pi from the partitioning of N-acetyl-amino-acid amides. **Eur J Med Chem** 18**,** 369-375.

Garnier J, Gibrat JF, Robson B (**1996**) GOR secondary structure prediction method version IV. Methods in Enzymology, R.F. Doolittle Ed., 266, 540-553.

Gomase VS (**2006**) Prediction of Antigenic Epitopes of Neurotoxin Bmbktx1 from Mesobuthus martensii. **Curr Drug Discov Technol** 3, 225-229.

Gomase VS, Changbhale SS (**2007**) Antigenicity Prediction in Melittin: Possibilities of in Drug Development from Apis dorsata. **Current Proteomics** 4, 107-114.

Gomase VS, Kale KV, Chikhale NJ, Changbhale SS (**2007**) Prediction of MHC Binding Peptides and Epitopes from Alfalfa mosaic virus. **Curr Drug Discov Technol** 4, 117-1215.

IsHak JA, Kreuze JF, Johansson A, Mukasa SB, Tairo F, Abo El-Abbas FM, Valkonen JP (**2003**) Some molecular characteristics of three viruses from SPVD-affected sweet potato plants in Egypt. **Arch Virol** 148, 2449-60.

Janin J (**1979**) Surface and inside volumes in globular proteins. **Nature** 277, 491-492.

Janin J, Wodak S (**1978**) Conformation of amino acid side-chains in proteins. **J Mol Biol** 125, 357-386.

Kaplan IB, Zhang L, Palukaitis P (**1998**) Characterization of cucumber mosaic virus. V. Cell-to-cell movement requires capsid protein but not virions. **Virology**, 246, 221-231.

Kyte J, Doolittle RF (**1982**) A simple method for displaying the hydropathic character of a protein. **J. Mol. Biol.**, 157, 105.

Larsen JE, Lund O, Nielsen M (**2006**) Improved method for predicing linear B-cell epitopes. **Immunome Res** 2, 2.

Manavalan P, Ponnuswamy PK (**1978**) Hydrophobic character of amino acid residues in globular proteins. **Nature**, 275, 673-674.

Noe-Becerra EE, Cardenas, Lozoya H, Mosquedo R (**1999**) Rhabdovirus in Papaya (Carica papaya L.) in the Southeast of Mexico. **Agronomia-Meroamericana**, 10, 85-90.

Parker KC, Bednarek MA, Coligan JE (**1994**) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. **J Immunol** 152, 163-75.

Purcifull DE, Hiebert E, Edwardson JR (**1984**) Papaya ringspot virus. CMI/ AAB Description of Plant Viruses No. 292 (revised). Wellsbourne, UK Association of Applied Biologists, 8.

Quemada H, L'Hostis B, Gonsalves D, Reardon IM, Heinrikson R, Hiebert EL, Sieu LC, Slightom JL (**1990**) The nucleotide

sequences of the 3'-terminal regions of papaya ringspot virus strains W and P. **J Gen Virol** 71 (PT 1), 203-210.

Robson B, Garnier J (**1993**) Protein structure prediction. **Nature** 361, 506.

Shintaku M, Palukaitis P (**1990**) Genetic mapping of Cucumber mosaic virus. Viral genes and plant pathogenesis, pp 156-164.

Suzuki M, Kuwata S, Kataoka J, Masuta C, Nitta N, Takanami Y (**1991**) Functional analysis of deletion mutants of cucumber mosaic virus RNA3 using an in vitro transcription system. **Virology**, 183, 106-113.

Sweet RM, Eisenberg D (**1983**) Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure. **J Mol Biol** 171, 479-488.

Thornton JM, Edwards MS, Taylor WR, Barlow DJ (**1986**) Location of 'continuous' antigenic determinants in the protruding regions of proteins. **EMBO J** 5, 409-413.

Urcuqui-Inchima S, Haenni AL, Bernardi F (**2001**) Potyvirus proteins: a wealth of functions. **Virus Res** 74 (1-2), 157-175.

von Heijne G (**1981**) On the hydrophobic nature of signal sequences. **Eur J Biochem** 116, 419-422.

Welling GW, Weijer WJ, van der Zee R, Welling-Wester S (**1985**) Prediction of sequential antigenic regions in proteins. **FEBS Lett** 188, 215-218.

Xiao H, Maoka T, Luo X (**1997**) Investigation and identifi cation of Papaya ringspot virus and Papaya leaf distortion mosaic virus in South China. **Journal of South China Agricultural University** 18, 52-53.