

# Prediction of MHC Binding Peptides and Epitopes from Coat Protein of Mungbean Yellow Mosaic India Virus-Ub05

Sarika<sup>1\*</sup>, Mohd. Akram<sup>2</sup>, M. A. Iquebal<sup>1</sup> and Naimuddin K<sup>2</sup>

<sup>1</sup>Scientist, Statistics and Computer Application Unit, Indian Institute of Pulses Research (ICAR), Kanpur 208024, Uttar Pradesh, India

<sup>2</sup>Senior Scientist, Division of Crop Protection, Indian Institute of Pulses Research (ICAR), Kanpur 208024, Uttar Pradesh, India

## Abstract

Mungbean yellow mosaic India virus (MYMIV) is recognized as one of the most economically important viruses that affects several leguminous crops like mungbean, urdbean, cowpea, pigeonpea etc. and occurs in Indian subcontinent. In the present study, coat protein of MYMIV is being used to find out highly suitable MHC binding peptides and epitopes. Thirty nine peptide regions were found to have high affinity to TAP binding peptides using cascade support vector machine (SVM). Few of these coat protein TAP transporters are 201- NRFFKVNNY with score 9.208, 108- KRFCIKSVY with score 8.817, 44-RWTNRPMWR with score 8.790, 134- NTVMFKLCR with score 8.672 and 41- KRRRWTNRP with score 8.498 where the scores are based on the average affinity of an amino acid at particular position. The SVM based method for prediction of promiscuous MHC Class II binders reported MHCII-IAB peptide regions, 30-PASAGGVPT, 127-IKSKNHTNT, 44-RWTNRPMWR, 6-YDTAFSTPI, (optimal score 1.220); MHCII-IAD peptide regions, 226-NALLYMAC, 30-PASAGGVPT, 32-SAGGVPTNM, 236- HASNPVYAT, (optimal score 0.620); MHCII-IAg7 peptide regions, 13-PISNARRRL, 212- YNHQEAAY, 221-ENHTENALL, 209-YVVYNHQEA, (optimal score 1.569) and MHCII- RT1.B peptide regions, 223- HTENALLY, 188- TGGQYACKE, 168- TVKNDLRDR, 4-RTYDTAFST, (optimal score 0.932) as possible predicted binders from coat protein. The most suitable predicted segments in coat protein of MYMIV virus for developing specific antibodies found in this study are 56-FYRLYRSPDVRGCEGPKVQSF-78, 206-VNNYVVYNHQ-215 and 108-KRFCIKSVYITG-119. Fragments identified through this approach tend to be high-efficiency binders, in which larger percentage of their atoms are directly involved in binding as compared to larger molecules. These fragments may, therefore, be used in cross protection and to develop begomovirus specific antibodies that can be exploited in sero-diagnostics.

**Keywords:** Cross protection; Epitope; MHC binders; Mungbean yellow mosaic India virus; Support vector machine

**Abbreviations:** MHC: Major Histocompatibility Complex; MYMIV: Mungbean Yellow Mosaic India Virus; SVM: Support Vector Machine

## Introduction

Yellow mosaic disease of many legumes in India and other South Asian countries is caused by whitefly (*Bemisia tabaci* Genn.) transmitted by geminiviruses belonging to the family *Geminiviridae* and genus *Begomovirus*. Four species viz., *Mungbean yellow mosaic virus* (MYMV), *Mungbean yellow mosaic India virus* (MYMIV), *Dolichos yellow mosaic virus* (DYMV) and *Horsegram yellow mosaic virus* (HYMV) are known to cause yellow mosaic disease in different leguminous species. All these viruses are bipartite begomoviruses, have geminate (twin) particles, 18-20 nm in diameter, 30 nm long, apparently consisting of two incomplete icosahedra joined together in a structure with 22 pentameric capsomeres and 110 identical protein subunits (Qazi et al., 2007). Estimation of actual losses due to yellow mosaic disease in farmers' field is difficult as these losses vary from year to year and from variety to variety. However, based on the incidence of yellow mosaic disease in mungbean, urdbean and soybean, an annual loss of over US \$ 300 million is estimated in these crops (Varma et al., 1992). Yellow mosaic disease occurs in a number of leguminous plants such as mungbean, urdbean, cowpea (Nariani, 1960; Nene, 1973), soybean (Suteri, 1974), horsegram (Muniyappa et al., 1975), lablab bean (Capoor and Varma, 1948) and French bean (Singh, 1979).

Despite the advances made in molecular plant pathology, a chemical that can kill a virus or suppresses its infection in plant system is still eluding scientists. Thus, once a plant is infected by virus there is neither effective cure nor treatment available. However, plant pathologists discovered cross protection, a phenomenon

similar to 'vaccine' concept in animals, for protecting plant against viral diseases. Cross Protection has been studied in various virus-host combinations and exploited to solve some of the crop production problems caused by plant viruses (Fraser, 1998). Numerous theories have been advanced to explain the possible mechanism of cross protection (De Zoeten and Fulton, 1975). The strongest evidence is for a central role of the coat protein or nucleocapsid protein in hindering the multiplication of challenging strain of a virus possibly by sequestering the nucleic acid, more likely by preventing its uncoating (Wilson and Watkins, 1986). Although the evidence for a central role of coat protein in cross protection is very strong, the mechanism may not be confined solely to inhibition of virus uncoating. There is, for example evidence that coat protein may interfere with replication process of the challenging virus.

With the advent of various prediction tools to find out small fragments of peptide from a known viral protein (eg. Coat protein), new vistas of creating host immune response are opened to scientific community. MHC molecules are cell surface glycoproteins, which take active part in host immune reactions. The involvement of MHC class-I in response to almost all antigens and the variable length of interacting peptides make the study of MHC Class I molecules very

\*Corresponding author: Sarika, Scientist Statistics and Computer Application Unit, Indian Institute of Pulses Research (ICAR), Kanpur 208024, Uttar Pradesh, India, Tel: +91-512-2572464, Ext 262; Fax: +91-512-2572582; E-mail: [ajjaiswal@gmail.com](mailto:ajjaiswal@gmail.com)

Received April 04, 2010; Accepted May 10, 2010; Published May 10, 2010

**Citation:** Sarika, Akram M, Iquebal MA, Naimuddin K (2010) Prediction of MHC Binding Peptides and Epitopes from Coat Protein of Mungbean Yellow Mosaic India Virus-Ub05. J Proteomics Bioinform 3: 173-178. doi:10.4172/jpb.1000136

**Copyright:** © 2010 Sarika, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



interesting. MHC molecules have been well characterized in terms of their role in immune reactions. They bind to some of the peptide fragments generated after proteolytic cleavage of antigen (Kumar et al., 2007). These binding sites are antigen specific and generate immune response against the parent antigen.

Prediction methods to find out the small peptides fragments from a protein which may represent the whole protein and excite the immune response are available (Gomase et al., 2008). The present paper deals with the possibilities of exploiting coat protein of MYMIV to find out the highly suitable MHC binding peptide and have high affinity to TAP binding peptides that can be used for inducing cross protection and as immunogen to produce antiserum for the development of sero-diagnostics for begomoviruses.

## Materials and Methods

### Protein sequence used

For recognition of immunologically relevant regions, hydrophilicity, antigenicity, solvent accessible regions and MHC class peptide binding of the coat protein sequence of *Mungbean Yellow Mosaic India Virus* (GenBank accession no. GQ387510) had been considered.

### Prediction of secondary structure of protein and its antigenicity

The secondary structure diagram based on Garnier algorithm provides additional information about possible sequence accessibility (Garnier et al., 1996). The aim of secondary structure prediction is to provide the location of alpha helices, and beta strands within a protein or protein family. 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 (Robson and Garnier, 1993; Gomase et al., 2008) are the important concepts involved in secondary structure prediction.

Antigenicity prediction tools adopted in this study predict those segments from coat protein that are likely to be antigenic by eliciting an antibody response using Hopp and Woods (Hopp and Woods, 1981), Welling (Welling et al., 1985), Parker (Parker et al., 1986), B-EpiPred Server (Larsen et al., 2006) and Kolaskar and Tongaonkar (Kolaskar and Tongaonkar, 1990).

### Targeting the location in solvent accessible regions

Protein antigenicity is a surface property. Antigenic epitopes can be located as those segments of primary structure that are markedly hydrophilic (Hopp and Wood, 1981). Hydrophilicity plots provide a measure of distribution of polar and apolar amino acid residues within the protein sequence. The Kyte-Doolittle scale (Kyte and Doolittle, 1982) provides a measure of hydrophobicity with each amino acids. Similarly Hopp-Woods scale was used to predict potential antigenic sites. This may be useful in predicting membrane-spanning domains, potential antigenic sites and regions that are likely exposed on the protein surface (Gomase, 2006; Janin, 1979; Abraham and Leo, 1987; Bull and Breese, 1974).

### Prediction of MHC binding peptide

Prediction methods for identifying binding peptides could minimize the number of peptides required to be synthesized and assayed, and thereby facilitate the identification of potential epitopes (Gomase et al., 2008). Several methods have been used to predict

MHC binding peptides, including those based on binding motifs (van Endert et al., 1995; Adams and Koziol, 1995), quantitative matrices (Bhasin and Raghava, 2003), artificial neural networks (ANNs) (Brusic et al., 1994; Brusic et al., 1995; Brusic et al., 1998) and support vector machine (SVM) (Donnes and Elofsson, 2002; Bhasin and Raghava, 2003; Ding and Dubchak, 2001). Binding motifs specify which residues at given positions within the peptide are necessary or favorable for binding to a specific MHC molecule (Rotzschke et al., 1992). In this study, prediction of MHC peptide binding is performed using neural networks trained on C terminals of known epitopes. Prediction of peptide binders to MHCI and MHCII molecules from protein sequences or sequence alignments is done using Position Specific Scoring Matrices (PSSMs). An elegant machine learning technique i.e. SVM based method is used for prediction of promiscuous MHC class II binding peptides. In SVM based method, the average accuracy is reported to be high as compared to other methods since SVM can handle noise or non linearity in data very well (Brown et al., 2000; Ding and Dubchak, 2001; Bhasin and Raghava, 2003). The predicted peptides from coat protein under study and their affinity to TAP binding peptides are determined by the scoring based on the average score /affinity of an amino acid at particular position and calculated as follows:  $A_{i,r}$  = Average affinity of peptides having residues r in position i, where  $A_{i,r}$  is the matrix entry of residue r in position i, r may be any natural amino acid and i varies from 1 to 9 (Bhasin and Raghava, 2003).

## Results

Study refers to the coat protein sequence of *Mungbean Yellow Mosaic India Virus* having 257 bp as described under material and methods.

### Determination of antigenic peptides

Parameters such as hydrophilicity, flexibility, accessibility, turns, exposed surface, polarity and antigenic propensity of polypeptides chains have been correlated with the location of continuous epitopes. Hydrophobicity (or hydrophilicity) plots are designed to display the distribution of polar and apolar residues along a protein sequence. In our study, antigenic determinants have been targeted by locating the positive peaks in hydrophilicity plots, thus identifying the regions of maximum potential antigenicity. Hopp-Woods scale (Hopp and Woods, 1981) was used for predicting potential antigenic sites of protein which is essentially a hydrophilic index, with apolar residues assigned negative values (Figure 1). Welling antigenicity plot (Welling et al., 1985) gives antigenicity value as the log of the quotient between percentage in a sample of known antigenic regions and percentage in average proteins (Figure 2). Parker (Parker et al., 1986), Kolaskar and Tongaonkar antigenicity methods (Kolaskar and Tongaonkar, 1990) and B-EpiPred Server (Larsen et al., 2006) were also studied (Figure 3, Figure 4 and Figure 5).

### Secondary alignment

For the protein under study, secondary structure was predicted using Garnier-Osguthorpe-Robson (GOR) method (Garnier et al., 1996). It assumes that the amino acids flanking the central amino acid also influence the secondary structure. Values for alpha helix, beta sheet, turns and coils are assigned for each residue (Figure 6). With the aid of these information parameters, likelihood of a given residue assuming each of the four possible conformations alpha, beta, reverse turn or coils can be calculated and the conformation with the largest likelihood may be assigned to the residue.



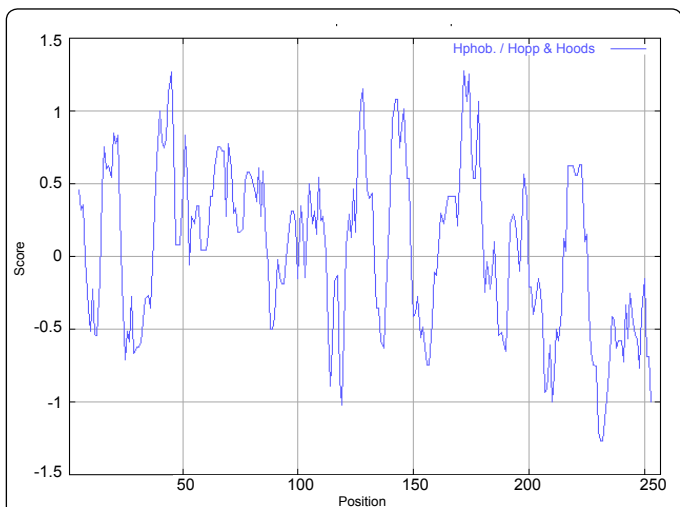


Figure 1: Hydrophobicity plot of Hopp & Woods for the coat protein.

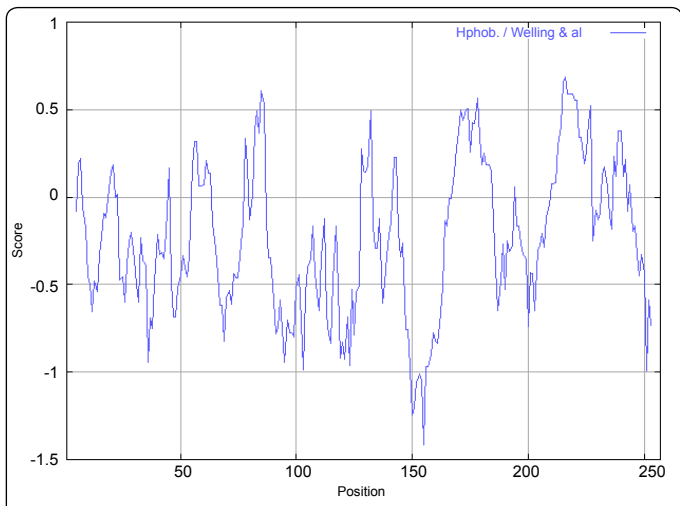


Figure 2: Hydrophobicity plot of Welling & al for the coat protein.

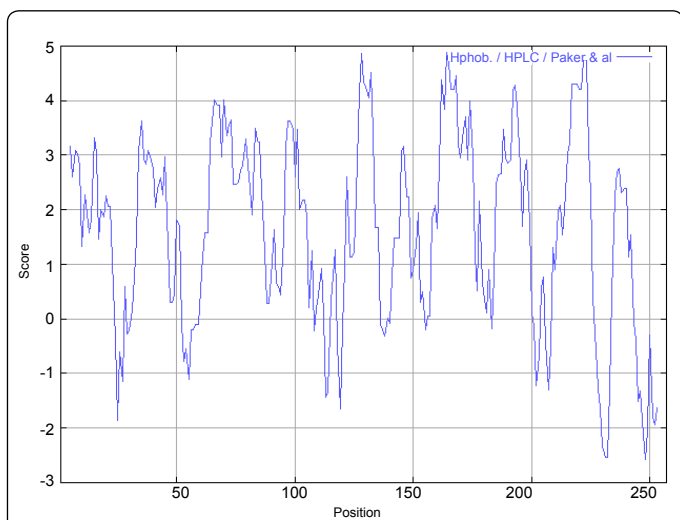


Figure 3: Hydrophobicity plot of HPLC / Parker & al for the coat protein.

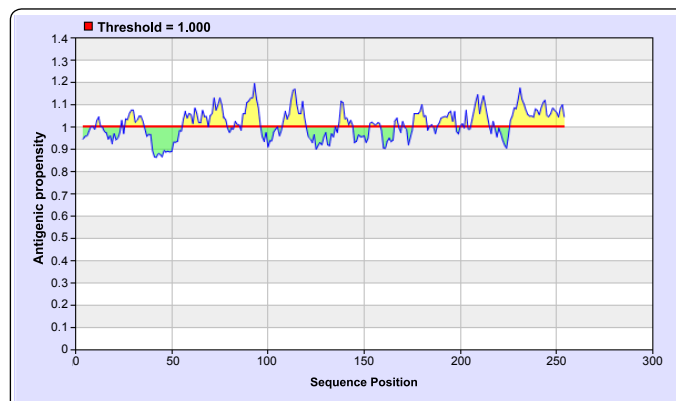


Figure 4: Kolaskar and Tongaonkar antigenicity sites for the coat protein.

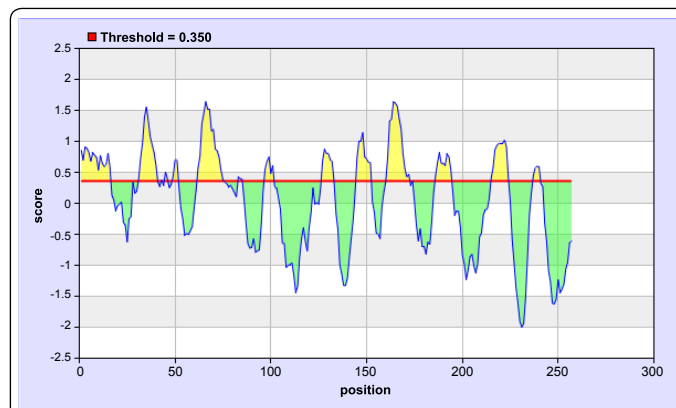


Figure 5: B-cell epitopes sites for the coat protein.

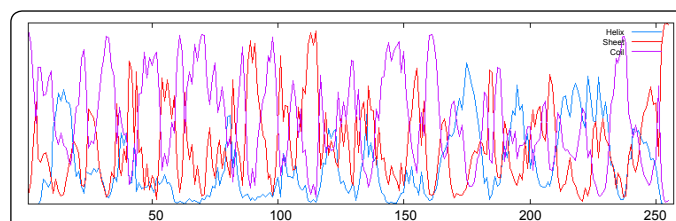


Figure 6: Secondary structure GOR plot of the coat protein.

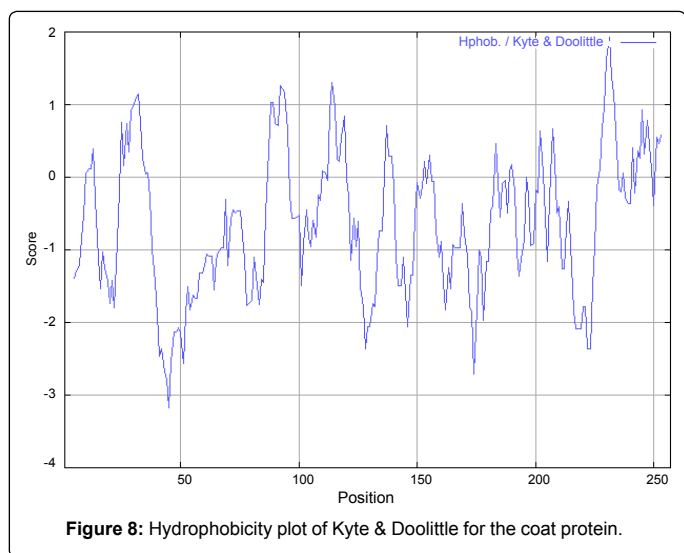
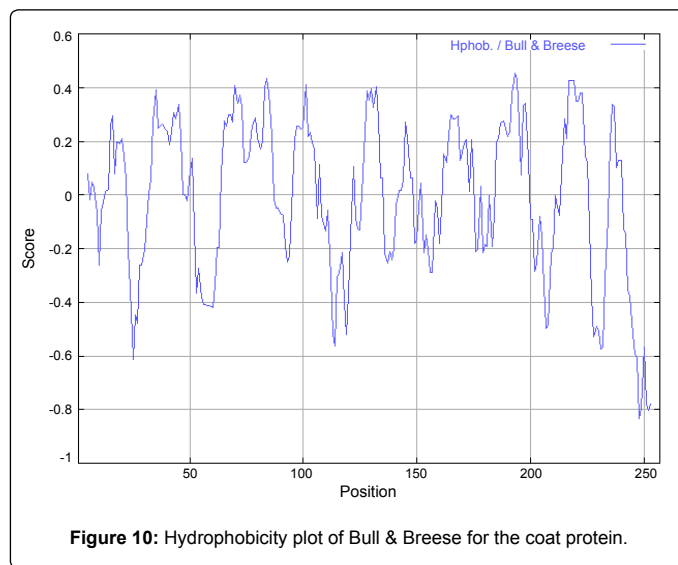
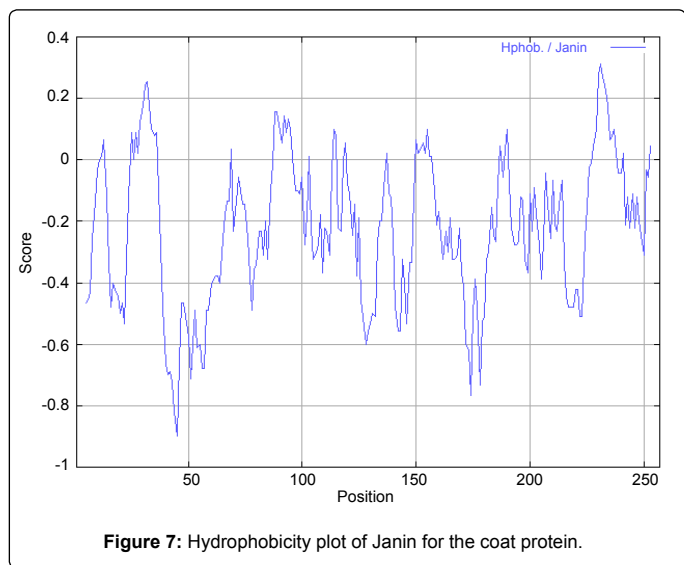
### Solvent accessible regions

To predict potential antigenic sites of globular proteins, which are likely to be rich in charged and polar residues, solvent accessible scales are developed which delineate hydrophobic and hydrophilic characteristics of amino acids. The protein under study was exposed to Janin, Kyte & Doolittle, Abraham & Leo and Bull & Breese methods to predict its nature and prediction flexibility (Figure 7, Figure 8, Figure 9 and Figure 10).

### Determination of MHC binding peptides

The binding between peptide epitopes and MHC protein(s) is an important event in the cellular immune response. SVMs are a class of learning based on non-linear modeling techniques with proven performance in a wide range of practical applications (Cristianini and Shawe, 2000). The prediction method used in our study is based on this elegant machine learning technique. The cascade support vector





Peptide Rank	Start Position	Sequence	Score	Predicted Affinity
1	201	NRFFKVNNY	9.208	High
2	108	KRFCKSVY	8.817	High
3	44	RWTNRPMWR	8.790	High
4	134	NTVMFKLCR	8.672	High
5	41	KRRRWTRNP	8.498	High
6	49	PMWRKPRFY	8.354	High
7	243	ATLKIRIYF	8.339	High
8	43	RRWTNRPMW	8.265	High
9	244	TLKIRIYFY	8.143	High
10	55	RFYRLYRSP	7.870	High

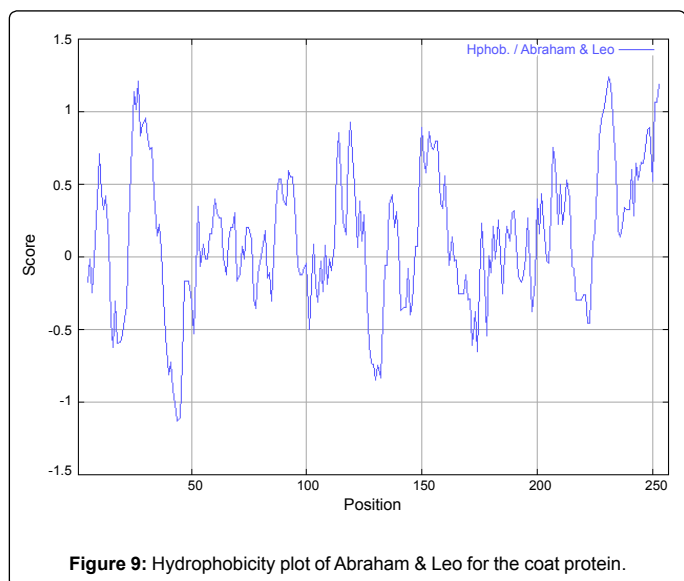
**Table 1:** TAP Peptide binders of coat protein.

machine approach based on amino acid sequence and properties was used to predict MHCI and MHCII binding regions. In this assay, prediction of the binding affinity of coat protein having 257 amino acids, showing 249 nonamers was performed. SVM was trained on the binary input of single amino acid sequence. The binding regions obtained are reported in Table 1 and Table 2.

Thirty nine peptide regions were found to have high affinity to TAP binding peptides. The data presented in Table 1 showed top ten peptide regions. Few of these were 201- NRFFKVNNY with score 9.208, 108- KRFCIKSVY with score 8.817, 44- RWTNRPMWR with score 8.790, 134- NTVMFKLCR with score 8.672 and 41- KRRRWTRNP with score 8.498, which are known as coat protein TAP transporters. The SVM based method for prediction of promiscuous MHC Class II binders are reported in Table 2. MHCII-IAb peptide regions, 30- PASAGGVPT, 127- IKSKNHTNT, 44- RWTNRPMWR, 6- YDTAFSTPI, (optimal score 1.220); MHCII-IAD peptide regions, 226- NALLYMAC, 30- PASAGGVPT, 32- SAGGVPTNM, 236- HASNPVYAT, (optimal score 0.620); MHCII-Iag7 peptide regions, 13- PISNARRRL, 212- YNHQEAAY, 221- ENHTENALL, 209- YVYNHQA, (optimal score 1.569); and MHCII- RT1.B peptide regions, 223- HTENALLY, 188- TGGQYACKE, 168- TVKNDLRDR, 4- RTYDTAFST, (optimal score 0.932) represent predicted binders from coat protein under study. Table 3 shows the predicted antigenic epitopes from MYMIV coat protein.

## Discussion

MYMIV is one of the most economically important viruses that affects several leguminous crops like mungbean, urdbean, cowpea, pigeonpea etc and requires attention. In the present study, B-EpiPred





23. Larsen JE, Lund O, Nielsen M (2006) Improved method for predicting linear B-cell epitopes. *Immunome Res* 2: 2. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
24. Muniyappa V, Reddy HR, Shivshankar C (1975) Yellow mosaic disease of *Dolichos biflorus* Linn. *Current Science* 4: 176. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
25. Nariani TK (1960) Yellow mosaic of mung (*Phaseolus aureus* L.). *Ind Phytopathology* 13: 24-29. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
26. Nene YL (1973) Viral diseases of warm weather pulse crops in India. *Plant Dis Rep* 57: 463-467. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
27. Parker JMR, Guo D, Hodges RS (1986) New Hydrophilicity Scale Derived from High-Performance Liquid Chromatography Peptide Retention Data: Correlation of Predicted Surface Residues with Antigenicity and X-ray Derived Accessible Sites. *Biochemistry* 25: 5425-5431. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
28. Qazi J, Ilyas M, Mansoor S, Briddon RW (2007) Legume yellow mosaic viruses: genetically isolated begomoviruses. *Mol Plant Pathology* 8: 343-348. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
29. Robson B, Garnier J (1993) Protein structure prediction. *Nature* 361: 506. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
30. Rotzschke O, Falk K, Stevanovic S, Jung G, Rammensee H (1992) Peptide motifs of closely related HLA class I molecules encompass substantial differences. *Eur J Immunol* 22: 2453-2456. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
31. Singh RN (1979) Natural infection of bean (*Phaseolus vulgaris*) by mungbean yellow mosaic virus. *Ind J Mycology Plant Pathology* 9: 124-126. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
32. Suteri BD (1974) Occurrence of soybean yellow mosaic virus in Uttar Pradesh. *Curr Sci* 43: 689-690. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
33. van Ender P M, Riganelli D, Greco G, Fleischhauer K, Sidney J, et al. (1995) The peptide-binding motif for the human transporter associated with antigen processing. *J Exp Med* 182:1883-1895. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
34. Varma A, Dhar AK, Mandal B (1992) MYMV transmission and control in India. In: *Mungbean Yellow Mosaic Disease* (Green SK and Kim D. eds) pp8-27. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
35. Welling GW, Weijer WJ, Van Der ZR, Welling WS (1985) Prediction of sequential antigenic regions in proteins. *FEBS Lett* 188: 215-218. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
36. Wilson TMA, Watkins AC (1986) Influence of exogenous viral coat protein in the cotranslation assembly of tobacco mosaic virus (TMV) particles in vitro. *Virology* 149: 132-135. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)

