Determination of MHC Binding Peptides and Epitopes from Non-Structural Movement (NSm) Protein of Groundnut Bud Necrosis Virus

Sarika*[,1], M. Akram[2], M.A. Iquebal[1] and K. Naimuddin[2]

[1]*Scientist (Agricultural Statistics),* [2]*Senior Scientist (Plant Pathology), Indian Institute of Pulses Research, Kalyanpur, Kanpur-208 024, Uttar Pradesh, India*

Abstract: *Groundnut bud necrosis virus* (GBNV) is recognized as one of the most economically important viruses and is known to affect several crops including peanut, potato, tomato and soybean. For managing plant virus diseases, determination of their causal agents' identity at an early stage of crop is a pre-requisite. In the present study, NSm protein of GBNV has been used to predict out MHC binding peptides and epitopes that are highly suitable for antigenicity. Eighteen peptide regions were found to have high affinity. Few of these NSm protein TAP transporters are 126- RRYMHISRL with score 11.638, 125- NRRYMHISR with score 10.280, 46- AIMNKAKTL with score 7.762, 120- PTWNSNRRY with score 7.632 and 171- ASLKDPMCF with score 7.277. The support vector machine (SVM) based approach predicted MHCII-IAb peptide regions, 45- SAIMNKAKT, 151- ASLIDPNKM, 23- PAVKKENNR, 229- PIAAENNTC, (optimal score 0.938); MHCII-IAd peptide regions, 208- YAKGVGFAS, 101- NDSLVGNGN, 55- NGKQYVSSG, 63- GDSSVLGTY, (optimal score 0.852); MHCII-IAg7 peptide regions, 277- LQKAAERLA, 145- SKNNVKASL, 228- TPIAAENNT, 276- SLQKAAERL, (optimal score 1.640); and MHCII- RT1.B peptide regions, 193- TPKQCMQLN, 195- KQCMQLNLT, 246- KVIQSAALI, 166- IISRQASLK, (optimal score 0.800) as binders from NSm protein. The most suitable predicted segments in NSm protein of GBNV virus found in the study are 164- KII-ISRQASLKDPMCFIFHLNWS-186 and 237-CDVVPINRAKVIQSAALIEACKLMIP-262. These two fragments, obtained from non-structural movement protein with average propensity 1.016, are high-efficiency binders and may, therefore be used in cross protection to provide resistance against GBNV and develop GBNV specific antibodies that can be exploited in sero-diagnostics.

Keywords: Cross protection, Groundnut bud necrosis virus, epitope, MHC binders, non-structural movement protein, support vector machine.

1. INTRODUCTION

Tospoviruses cause considerable loss in yield and quality of produce from vegetables, legumes and ornamental crops worldwide. Infection at early stages of crop growth often results in death of plant causing a substantial decrease in plant stand which lead to significant yield loss [1]. Tospoviruses belong to the genus *Tospovirus*. This name is derived from *Tomato spotted wilt virus* (TSWV), the first member of the group described and its type member. Based on its genome structure and organization, the genus *Tospovirus* is placed within the family *Bunyaviridae*. The genome consists of three RNAs designated as large (L), medium (M) and small (S). The L RNA is in negative-sense while the M and S RNAs are ambisense. The L RNA codes for RNA-dependent RNA polymerase (RdRp), and the M RNA for precursor of two glycoproteins (GN and GC) and a non-structural movement protein (NSm). The majority of tospoviruses cause systemic infection in most of the crop plants [2].

Tospoviruses are exclusively transmitted by several thrips species in a circulative and propagative manner [3]. While there are more than 5000 thrips species, so far only 10 are known vectors of tospoviruses, suggesting marked co-evolution for transmission specificity between tospoviruses and these thrips vector species [4]. Of the 19 tospovirus species recorded worldwide, Asia has by far the greatest diversity with 14 identified so far infecting a wide range of crop plants. TSWV, the most widely occurring member of the tospovirus group, damages crops in Middle Eastern countries causing severe disease in vegetables such as tomato, pepper, lettuce and cabbage [5]. *Groundnut bud necrosis virus* (GBNV) is recognized as one of the most economically important viruses and affects several crops including peanut, potato, tomato and soybean in parts of China, India, Iran, Nepal, Sri Lanka and Thailand causing annual loss of over US $89 million in Asia [6]. In India GBNV causes 70–90% losses in peanut [7] and up to 29% in potato due to stem necrosis [8].

Detection of plant viruses is a pre-requisite for the management of diseases caused by them. Various methods are available for the detection of plant viruses such as enzyme linked immunosorbant assay (ELISA), PCR, RT-PCR etc. However, ELISA still remains most widely used detection method in most of the laboratories, quarantines and also by the growers of planting materials. Of the various ingredients used in ELISA, the primary antibodies, which are produced against the specific virus (antigen) in warm blooded animals

*Address correspondence to this author at the Scientist (Agricultural Statistics), Indian Institute of Pulses Research, Kalyanpur, Kanpur-208 024, Uttar Pradesh, India; Tel: +91-512-2572464, Ext. 262, +91-9795180631;
Fax: +91-512-2572582; E-mail: aijaiswal@gmail.com

1570-1638/10 $55.00+.00

such as mouse, rabbit, goat etc. are the most important. There are two main approaches for the production of antibodies against plant viruses. One is to inject the purified virus in to the animals and second is to get the viral protein through recombinant technology. Both of these techniques are costly and require specific expertise. Tools to predict the small peptides (from a large protein) that possess as good antigenic property as the whole protein are available on web server [9-12]. The peptides fragments thus predicted would make the production of virus specific antibodies easier and economical.

The antigenic specificity of a protein resides in restricted areas of the molecule, known as antigenic determinants or epitopes, which are recognised by the combining sites of paratopes of certain immunoglobulin molecules. Once an immunoglobulin has been shown to bind to an antigen, it becomes an antibody specific for that antigen. Methods are available for predicting the small peptides fragments from a protein which may represent the whole protein and excite the immune response [13]. 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. Several methods have been used to predict MHC binding peptides, including those based on binding motifs, quantitative matrices, artificial neural networks (ANNs) and support vector machine (SVM). Binding motifs specify which residues at given positions within the peptide are necessary or favourable for binding to a specific MHC molecule [14].

In the present study, an important protein i.e. non-structural movement protein of GBNV has been used to identify highly suitable MHC binding peptides and epitopes that in turn may be used for the development of serodiagnostics. It is also possible that the nucleotide template of predicted peptide might be used as trasngene to develop the genetically modified plants for resistance to GBNV. In such transgenic plants the expression of trasngene would hinder the multiplication of challenged virus (GBNV), as happens in the mechanism of cross protection.

2. MATERIALS AND METHODS

2.1. Protein Sequence Analysis

Since most biologically important antigens are proteins, we consider here mainly the antigenecity, solvent accessible regions and MHC class peptide binding of the non-structural movement protein sequence of *Groundnut bud necrosis virus* (AY259522) [15] for identifying active sites.

2.2. Antigenic Determinants in the Used Protein

The antigenic specificity of a protein resides in restricted areas of the molecule, known as antigenic determinants or epitopes, which are recognised by the combining sites of paratopes of certain immunoglobulin molecules. Once an immunoglobulin has been shown to bind to an antigen, it becomes an antibody specific for that antigen. Antigenicity prediction tools adopted in this study predict those segments from NSm protein that are likely to be antigenic by raising an antibody response. Here, antigenic epitopes are determined using Hopp and Woods, Welling, B-EpiPred Server and Kolaskar and Tongaonkar antigenicity methods [16-19].

2.3. Identification of Solvent Accessible Regions

There exist several rules to determine the peptide fragments from protein which are likely to be antigenic. Accordingly, antigenic peptides need to be located in solvent accessible regions and contain both hydrophilic and hydrophobic residues. In the study, measure of the distribution of polar and apolar amino acid residues within the protein sequence is provided by various plots [20-24].

2.4. 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. Several methods have been used to predict MHC binding peptides, including those based on binding motifs, quantitative matrices, artificial neural networks (ANNs) and support vector machine (SVM). Binding motifs specify which residues at given positions within the peptide are necessary or favorable for binding to a specific MHC molecule. 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 in comparison to other methods, because SVM can be trained on the binary input of single amino acid sequence [9, 13, 25, 26].

3. RESULT AND INTERPRETATION

3.1. 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 finding the area of greatest local hydrophilicity. Hopp-Woods scale was designed for predicting potential antigenic sites of protein which is essentially a hydrophilic index, with apolar residues assigned negative values (Fig. **1**). 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 (Fig. **2**). Kolaskar and Tongaonkar antigenicity methods and B-EpiPred Server were also studied (Figs. **3**, **4**).

3.2. 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 trace hydrophobic and hydrophilic characteristics of amino acids. From the analysis, it was interpreted that the non-structural movement protein under study has high-prediction flexibility (Figs. **5-7**).

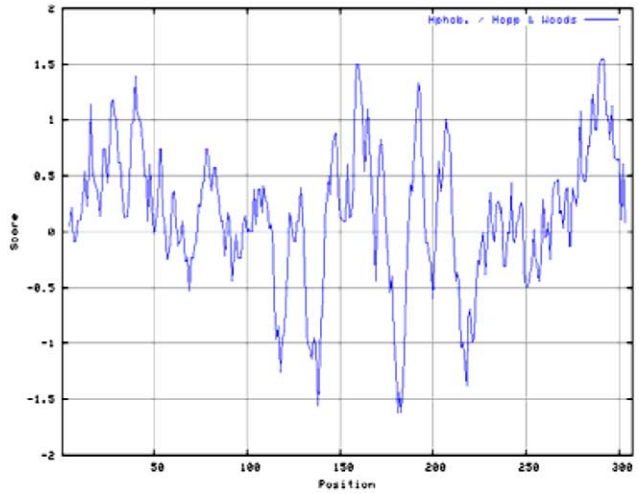

Fig. (1). Hydrophobicity plot of Hopp & Woods of non-structural movement protein.

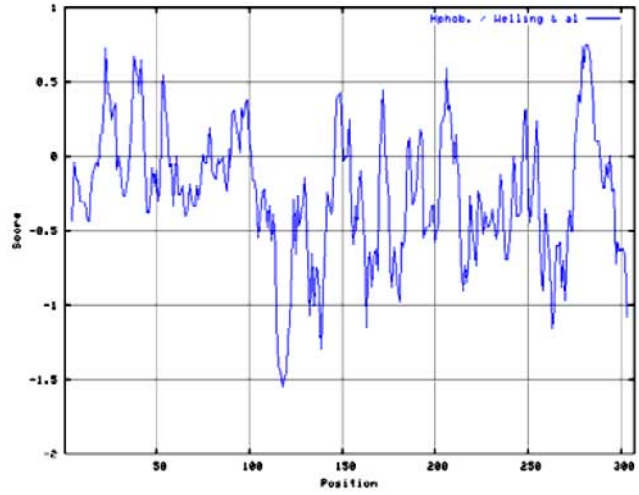

Fig. (2). Hydrophobicity plot of Welling & al of non-structural movement protein.

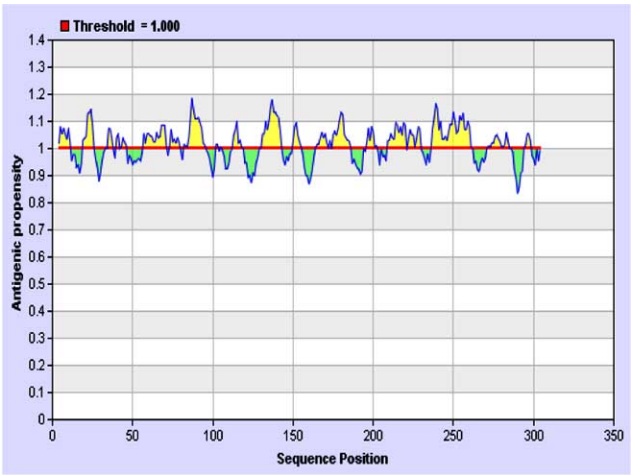

Fig. (3). Kolaskar and Tongaonkar antigenicity sites of the non-structural movement protein.

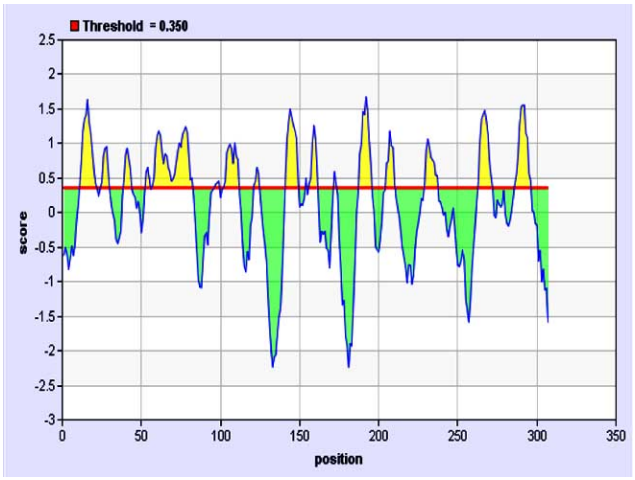

Fig. (4). B-cell epitopes sites of the non-structural movement protein.

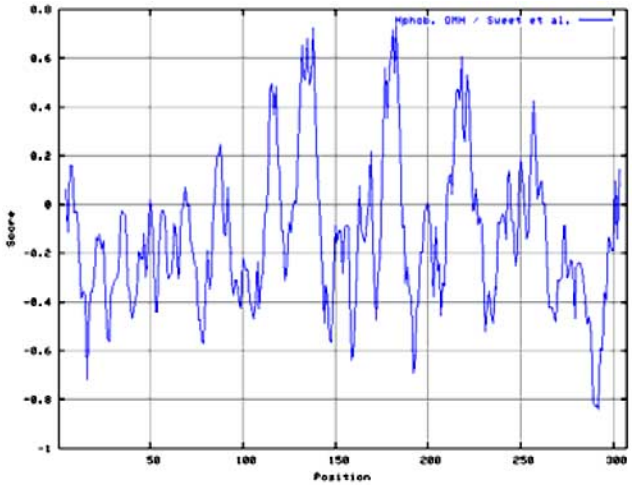

Fig. (5). Hydrophobicity Sweet plot of OMH for the non-structural movement protein.

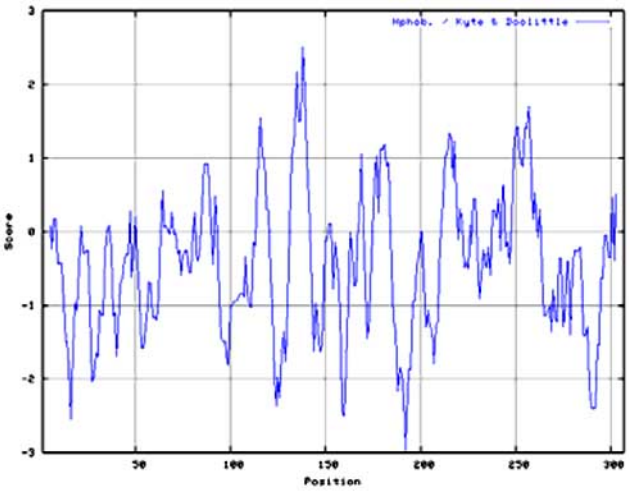

Fig. (6). Hydrophobicity plot of Kyte & Doolittle for the non-structural movement protein.

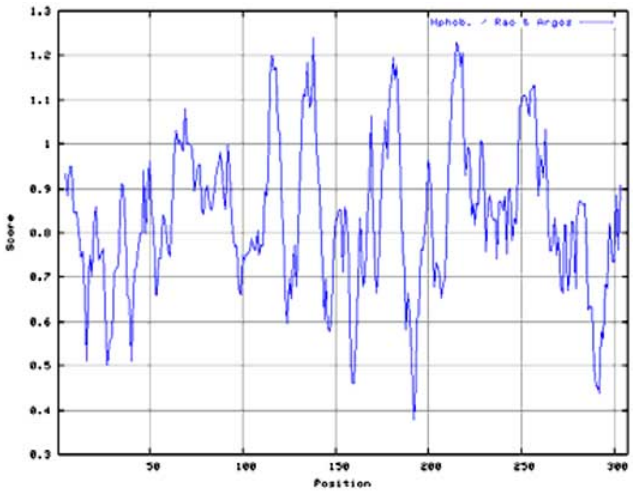

Fig. (7). Hydrophobicity plot of Rao & Argos for the non-structural movement protein.

3.3. Determination of MHC Binding Peptides

The binding between peptide epitopes and MHC protein(s) is an important event in the cellular immune response. MHC binding peptides are sufficient to elicit the desired immune response. The cascade support vector machine approach based on amino acid sequence and properties was used to predict MHCI and MHCII binding regions. In this study, prediction of the binding affinity of NSm protein having 257 amino acids, showing 249 nonamers was performed. The binding regions obtained are reported in Tables **1** and **2**.

Eighteen peptide regions were found to have high affinity. Table **1** reports top fourteen high affinity peptide regions. Some of these are 126- RRYMHISRL with score 11.638, 125- NRRYMHISR with score 10.280, 46- AIMNKAKTL with score 7.762, 120- PTWNSNRRY with score 7.632 and 171- ASLKDPMCF with score 7.277, which are known as non-structural movement protein TAP transporter. The SVM based method for prediction of promiscuous MHC Class II binders are reported in Table **2**. MHCII-IAb peptide regions, 45- SAIMNKAKT, 151- ASLIDPNKM, 23- PAVKKENNR, 229- PIAAENNTC, (optimal score is 0.938); MHCII-IAd peptide regions, 208- YAKGVGFAS, 101- NDSLVGNGN, 55- NGKQYVSSG, 63- GDSSVLGTY, (optimal score is 0.852); MHCII-IAg7 peptide regions, 277- LQKAAERLA, 145- SKNNVKASL, 228- TPIAAENNT, 276- SLQKAAERL, (optimal score is 1.640) and MHCII- RT1.B peptide regions, 193- TPKQCMQLN, 195- KQCMQLNLT, 246- KVIQSAALI, 166- IISRQASLK, (optimal score is 0.800) represent predicted binders from non-structural movement protein under study. Table **3** reports the predicted antigenic epitopes from non-structural movement protein.

4. CONCLUSION

GBNV being one of the most economically important viruses that affects several crops including peanut, potato, tomato and soybean in parts of China, India, Iran, Nepal, Sri Lanka and Thailand, needs much attention.

The scales of Sweet hydrophobicity, Kyte & Doolittle hydrophobicity and Rao & Argos hydrophobicity depict hydrophilic index, with polar residues assigned negative values. Small peptide regions, 126- RRYMHISRL (score-11.638), 125- NRRYMHISR (Score- 10.280), 46- AIMNKAKTL (Score- 7.762) and 120- PTWNSNRRY (Score- 7.632) were few of the non-structural movement protein TAP transporters. The SVM based MHCII-IAb, MHCII-IAd, MHCII-IAg7 and MHCII- RT1.B peptide regions with highest rank were found to be 45- SAIMNKAKT, (optimal score is 0.938), 208-YAKGVGFAS (optimal score is 0.852), 277- LQKAAERLA, (optimal score is 1.640) and 193- TPKQCMQLN (optimal score is 0.800) respectively, which represented predicted binders from non-structural movement protein.

Kolaskar and Tongaonkar antigenicity sites of molecules are recognized by antibodies of the immune system for the non-structural movement protein. In general, the maximal

Table 1. TAP Peptide Binders of Non-Structural Movement Protein

Peptide Rank	Start Position	Sequence	Score	Predicted Affinity
1	126	RRYMHISRL	11.638	High
2	125	NRRYMHISR	10.280	High
3	46	AIMNKAKTL	7.762	High
4	120	PTWNSNRRY	7.632	High
5	171	ASLKDPMCF	7.277	High
6	173	LKDPMCFIF	7.255	High
7	215	ASVMYSWVK	6.883	High
8	245	AKVIQSAAL	6.759	High
9	51	AKTLNGKQY	6.656	High
10	132	SRLIIWVVP	6.580	High
11	217	VMYSWVKNF	6.219	High
12	79	ATSDDILSR	6.200	High
13	178	CFIFHLNWS	6.197	High
14	49	NKAKTLNGK	6.144	High

Table 2. Peptide Binders to MHCII Molecules of Non-Structural Movement Protein

Prediction Method	Allele	Rank	Sequence	Residue No.	Peptide Score
SVM	I-Ab	1	SAIMNKAKT	45	0.938
SVM	I-Ab	2	ASLIDPNKM	151	0.817
SVM	I-Ab	3	PAVKKENNR	23	0.774
SVM	I-Ab	4	PIAAENNTC	229	0.733
SVM	I-Ad	1	YAKGVGFAS	208	0.852
SVM	I-Ad	2	NDSLVGNGN	101	0.547
SVM	I-Ad	3	NGKQYVSSG	55	0.528
SVM	I-Ad	4	GDSSVLGTY	63	0.495
SVM	I-Ag7	1	LQKAAERLA	277	1.640
SVM	I-Ag7	2	SKNNVKASL	145	1.378
SVM	I-Ag7	3	TPIAAENNT	228	1.350
SVM	I-Ag7	4	SLQKAAERL	276	1.348
SVM	RT1.B	1	TPKQCMQLN	193	0.800
SVM	RT1.B	2	KQCMQLNLT	195	0.624
SVM	RT1.B	3	KVIQSAALI	246	0.565
SVM	RT1.B	4	IISRQASLK	166	0.538

Table 3. Predicted Antigenic Epitopes from Non-Structural Movement Protein

No.	Start Position	Peptide	End Position	Peptide Length
1	4	LSNVLESF	11	8
2	19	KELVPAV	25	7
3	57	KQYVSSGDSSVLGTY	71	15
4	73	SESAVEA	79	7
5	82	DDILSRLVVEQSTH	95	14
6	113	SFTISIM	119	7
7	129	MHISRLIIWVVPTI	142	14
8	150	KASLID	155	6
9	164	KIIISRQASLKDPMCFIFHLNWS	186	23
10	196	QCMQLNL	202	7
11	209	AKGVGFASVMYS	220	12
12	222	VKNFCDTP	229	8
13	237	CDVVPINRAKVIQSAALIEACKLMIP	262	26
14	271	SNQIKSLQKAAERLAL	286	16

hydrophilicity region is likely to be considered as an antigenic site because C- terminal regions of non-structural movement protein is solvent accessible and unstructured. Antibodies against those regions are also likely to recognize the native protein. Fourteen antigenic determinant sites in the NSm protein sequence were predicted. The highest pick is recorded between sequence of amino acid in the regions 164-KIIISRQASLKDPMCFIFHLNWS-186 and 237-CDVVPINRAKVIQSAALIEACKLMIP-262 (Table **3**). The average propensity for the non-structural movement protein is found to be 1.016 and residues having propensity greater than 1.0 are considered potentially antigenic.

The fragments thus 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 for

the development of serodiagnostics for GBNV. Further, it is also possible that the nucleotide template of predicted peptide might be used as trasngene to develop the genetically modified plants for resistance to GBNV. In such transgenic plants the expression of trasngene would hinder the multiplication of challenged virus (GBNV), as happens in the mechanism of cross protection.

REFERENCES

[1] Pappu HR, Jones RAC, Jain RK. Global status of tospovirus epidemics in diverse cropping systems: Successes achieved and challenges ahead. Virus Res 2009; 141: 219-36.

[2] Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA. Virus Taxonomy-8th Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, CA, 2005.

[3] Whitfield AE, Ullman DE, German TL. Tospovirus-thrips interactions. Annu Rev Phytopathol 2005; 43: 459-89.

[4] Mound LA. So many thrips-so few tospoviruses? In proceedings of the 7th International Symposium on Thysanoptera- Thrips, Plants, Tospoviruses: The Millennial Review; Bari, Italy. Marullo R, Mound L, Eds. 2001; pp. 15-18.

[5] Anfoka GH, Abhary M, Stevens MR. Occurrence of tomato spotted wilt virus (TSWV) in Jordan. EPPO Bull 2006; 36: 517-22.

[6] Reddy DVR, Buiel AAM, Satyanarayana T, *et al.* Peanut bud necrosis virus disease: an overview. In proceedings of Recent Studies on Peanut Bud Necrosis Disease; ICRISAT, India. Buiel AAM, Parlevliet JE, Lenne JM, Eds. March 20, 1995; pp. 3-7.

[7] Singh AB, Srivatava SK. Status and control strategy of peanut bud necrosis disease in Uttar Pradesh. In proceedings of Recent Studies on Peanut Bud Necrosis Disease; ICRISAT, India. Buiel AAM, Parlevliet JE, Lenne JM, Eds. March 20, 1995; pp. 65-8.

[8] Singh RB, Srivastava KK, Pandey SK, Paul Khurana SM. Assessment of yield losses due to potato stem necrosis disease. Ind J Virol 1997; 13: 135-7.

[9] Bhasin M, Raghava GPS. Analysis and prediction of affinity of TAP binding peptides using cascade SVM. Protein Sci 2004; 13(3): 596-607.

[10] Ding CHQ, Dubchak I. Multi-class protein fold recognition using support vector machines and neural networks. Bioinformatics 2001; 17(4): 349-58

[11] Brown M, Grundy W, Lin D, Cristianini N, Sugnet C, Furey T, Ares M, Haussler D. Knowledge-based analysis of microarray gene expression data using support vector machines. Proc Natl Acad Sci USA 2000; 97: 262-7.

[12] Larsen JEP, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immun Res 2006; 2: 2.

[13] Gomase VS, Kale KV, Shyamkumar K. Prediction of MHC Binding Peptides and Epitopes from Groundnut Bud Necrosis Virus (GBNV). J Proteomics Bioinform 2008; 1: 188-205.

[14] Rotzschke O, Falk K, Stevanovic S, Jung G, Rammensee H. Peptide motifs of closely related HLA class I molecules encompass substantial differences. Eur J Immun 1992; 22: 2453-6.

[15] Akram M, Jain RK, Chaudhary V, Ahlawat YS, Khurana SMP. Comparision of groundnut bud necrosis isolates based on movement protein (NSm) gene sequences. Ann appl Biol 2004; 145: 285-9.

[16] Hopp TP, Woods KR. Prediction of Protein Antigenic Determinants from Amino Acid Sequences. Proc Natl Acad Sci USA 1981; 78: 3824-8.

[17] Welling GW, Weijer WJ, Van Der ZR, Welling WS. Prediction of sequential antigenic regions in proteins. FEBS Lett 1985; 188: 215-8.

[18] Jens EPL, Ole L, Morten N. Improved method for predicting linear B-cell epitopes. Immun Res 2006; 2: 2.

[19] Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett 1990; 276: 172-4.

[20] Sweet RM, Eisenberg D. Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure. J Mol Biol 1983; 171: 479-88.

[21] Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982; 157: 105-132.

[22] Rao MJK, Argos P. A conformational preference parameter to predict helices in integral membrane proteins. Biochim Biophys Acta 1986; 869: 197-214.

[23] Wilkins MR, Gasteiger E, Bairoch A, *et al.* Protein identification and analysis tools in the ExPASy server. Methods Mol Biol 1999; 112: 531-52.

[24] Eisenberg D, Schwarz E, Komaromy M, Wall R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J Mol Biol 1984; 179: 125- 42.

[25] Brusic V, Rudy G, Honeyman G, Hammer J, Harrison L. Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network. Bioinformatics 1998; 14: 121-30.

[26] Bhasin M, Raghava GPS. P cleavage: an SVM based method for prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences. Nucleic Acids Res 2005; 33: W202-7.

Received: October 21, 2009 Revised: April 19, 2010 Accepted: June 01, 2010