

Prediction and Comparative Analysis of MHC Binding Peptides and Epitopes in *Nanoviridae* Nano-organisms

Dangre DM^{1*}, Deshmukh SR², Rathod DP², Umare VD³ and Ullah I⁴

¹Department of Public Health and Geriatric Medicine, Maharashtra University of Health Sciences, Regional Centre, Aurangabad, India

²Department of Biotechnology, SGB Amravati University, Amravati, Maharashtra, India

³Department of Biotechnology, Centre for Advanced Life Sciences, Deogiri College, Aurangabad, India

⁴Department of Biology, Gyeongsang National University, Jinju, South Korea

Abstract

Nanoviridae is a family of single stranded DNA viruses which infect the plants through their phloem tissues. The few members of these viruses are now turning towards animals. It includes two genera, *Nanovirus* and *Babuvirus*. *Nanovirus* includes three species namely *Faba beans necrotic yellows virus* (FBNYV), *Milk vetch dwarf virus* (MVDV) and *Subterranean clover stunt virus* (SCSV) while *Babuvirus* have two species accounted yet, namely *Abaca bunchy top virus* (ABTV) and *Banana bunchy top virus* (BBTV). The viral coat proteins are likely to be responsible for many diseases in plants as well as animals. The intra-genomic changes or post translational modifications within SCSV have decreased the length of the sequence, but increased the antigenic potential and numbers of antigenic binding regions on the surface of the protein. We have predicted the most probable responsible antigenic determinants and MHC binders within the viruses in *Nanoviridae* family. MHC molecules play a crucial role in host immune response. The nonamers of antigenic determinants in *Nanoviridae* family viral coat proteins are highly sensitive to H-2Kd and I-Ag7 molecules. The species within particular genera shows their common epitopes along with diverse colleague epitopes. These common antigenic peptides can be used as their identifiers. These identifiers along with their diverse colleagues could be most informative for antidotes production against themselves. Also, these are important for synthetic peptide vaccine production against the relevant viruses.

Keywords: *Nanoviridae* family; Antigenicity; Epitopes; MHC binders; Antidotes

Introduction

Nanoviridae is a family of single stranded DNA viruses which infect the plants through their phloem tissues. They are characterized by non enveloped, spherical virions of 17-20 nm in diameter with icosahedral symmetry and are transmitted by aphids in a persistent but not in a propagative manner. These viruses retained when the vector moults, don't multiply in a vector. The genome of *Nanoviridae* viruses consists of 6 to 8 circular (+) ssDNA about 1kb in size. Each ssDNA have a common stem-loop region and are encapsidated in a separate particle. In addition to genomic DNA, satellite-lake DNA are commonly found, usually encoding for Rep proteins. These satellite-rep proteins are only able to initiate replication of their genomic DNA, unlike genomic rep which promotes replication of all 6-8 viral genomic ssDNAs.

Recently, the many plant pathogenic viruses have been found in animal feces (Zhang et al., 2006; Li et al., 2010) which indicate the prevalence of these plant viruses in animals. Viruses with small (10 kb) circular DNA genomes, either single- or double-stranded, have been found to infect vertebrates as well as plants (Fauquet et al., 2005). Blinkova and his group performed the alignment of loop nonanucleotide sequences from ChiSCVs, circoviruses, nanoviruses and a geminivirus. They found conserved regions between ChiSCV and other circular ssDNA viruses. ChiSCV ORF 1 was 15–24% similar in amino acid sequence to the replicase of the multi-segmented plant viruses of the *Nanoviridae* family but only 12–17% similar to the replicases of circoviruses and geminiviruses. Thus, it can be assume that ChiSCV genomes share features with plant nanoviruses. The replicase genes of these viruses were most closely related to those of the much smaller (~1 kb) plant nanovirus circular DNA chromosomes (Blinkova et al., 2010). This is the main reason to make a keen interest in finding whether these plant viruses can be harmful to human being in future or not.

Nanoviridae family of viruses includes two genera viz. *Nanovirus* and *Babuvirus*. The genus *Nanovirus* which is closely related to the genus *Babuvirus*, accommodates *Subterranean clover stunt virus* (SCSV) as type species and other two species: *Faba bean necrotic yellows virus* (FBNYV) and *Milk vetch dwarf virus* (MVDV) whereas the genus *Babuvirus* accommodates *Banana bunchy top virus* (BBTV) as type species and one new species: *Abaca bunchy top virus* (ABTV).

The three species of *Nanovirus* genus mainly infect a wide range of leguminous plants along with some other crops. The brief description of each species is given as follows:

Faba beans necrotic yellows virus (FBNYV)

It is the causal agent of economically important disease of faba beans (*Vicia faba*), lentil (*Lens esculanta*) and pasture legumes in West Asia, North Africa, Sudan and Ethiopia. FBNYV infected crops show stunting, leaf rolling, yellowing and systemic reddening symptoms. The leaves become thick, brittle and have interveinal chlorotic blotches and nodulation and yield markedly reduced (Makkouk et al., 1998, Katul et al., 1995). FBNYV particles are efficiently and persistently transmitted by the green pea aphids and *Acyrtosiphon pisum*.

***Corresponding author:** Devanand M. Dangre, Department of Public Health and Geriatric Medicine, MUHS, Regional Centre, Shivaji (Amkhas) ground, Civil Hospital Campus, Aurangabad – 431001, Tel: 0240-2336181, +919372757601; E-mail: dangre_anand@yahoo.com

Received April 01, 2010; **Accepted** May 03, 2010; **Published** May 03, 2010

Citation: Dangre DM, Deshmukh SR, Rathod DP, Umare VD, Ullah I (2010) Prediction and Comparative Analysis of MHC Binding Peptides and Epitopes in *Nanoviridae* Nano-organisms. J Proteomics Bioinform 3: 155-172. doi:10.4172/jpb.1000135

Copyright: © 2010 Dangre DM, 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.



Milk vetch dwarf virus (MVDV)

It was first reported in *Astragalus sinicus* from Japan by Ohki in 1975 (Ohki et al., 1975). It is a causal agent of *Spinacia oleracea*, *Vigna unguiculata ssp. sesquipedalis*, *Astragalus sinicus*, *Vicia sativa*, *Phaseolus vulgaris*, *Pisum sativum*, *Vicia faba*, *Datura stramonium*, *Nicotiana rustica*, etc. and it gets transmitted by vectors like *Aphis craccivora*, *A. gossypii*, *Acyrtosiphon pisum*, *Acyrtosiphon (Aulocorthum) solani*, etc. The MVDV infected plants show yellow dwarf, stunting and leaf rolling symptoms.

Subterranean clover stunt virus (SCSV)

It was first reported in *Trifolium subterraneum* from New South Wales, Australia by Grylls and Butler in 1956 (Grylls and Butler, 1956). It is a causal agent of diseases of *Trifolium subterraneum*, *T. cernuum*, *Medicago lupulina*, *M. hispida* var. *denticulata*, *M. minima*, *Trifolium repens*, *T. glomeratum*, *Wisteria sinensis*, *Phaseolus vulgaris*, *Pisum sativum*, *Vicia faba*, *Trifolium dubium*, *T. pratense*, *Medicago arabica*, etc. and it gets transmitted by *Aphis craccivora*, *A. gossypii*, *Myzus persicae*, *Macrosiphum euphorbiae*, etc. SCSV infected crops show marginal chlorosis and puckering or cupping of leaflets, epinasty of leaves and the whole plant being markedly stunted with a reduction in the length of internodes.

The genus *Babuvirus* includes two species, *Banana bunchy top virus* and *Abaca bunchy top virus*, which differ from *Nanoviruses* slightly in number of DNA components, the host range and aphid vectors. These two species are described in brief as follows:

Banana bunchy top virus (BBTV): The BBTV disease was first reported in *Musa* species from Fiji by Magee in 1953 (Magee, 1953). BBTV is considered to be the most economically destructive disease of banana. The disease is widespread in Asia including China and Japan, Africa and Oceania but not in Central and South America (Dale, 1987). It is transmitted in a persistent manner only by banana aphid (*Pentalonia nigronervosa*). BBTV infection results in narrow, bunched leaves and stunted, fruitless plants, which eventually die. BBTV is regarded as one of the 100 most important emerging pathogens afflicting humankind.

Abaca bunchy top virus (ABTV): ABTV disease was first recognised in 1910 in Albay Province, The Philippines (Ocfemia, 1926) and has long been regarded as the most important biological constraint to abaca production. The disease occurs in all major production areas of abaca plants (Raymundo and Bajet, 2000). The ABTV infected plants show stunting, bunched and rosetted leaves, dark green flecks or veins clearing of leaves, upcurling and chlorosis of leaf margins etc. The symptoms and transmitting vector of ABTV disease are reminiscent of those of BBTV disease which led to the conclusion that both diseases are caused by the same virus. However, conflicting evidence was presented suggesting that two separate viruses were involved. In The Philippines, ABTV was transmitted to abaca but not to banana by *P. nigronervosa* (Ocfemia, 1930). BBTV was not observed in banana even when growing adjacent to abaca plantations seriously affected with ABTV (Ocfemia and Buhay, 1934). Therefore, ABTV is included as a new member of the genus *Babuvirus*.

Nanoviridae viruses possess small single-stranded DNA (ssDNA) genomes, as opposed to the double-stranded genomes of the mammalian tumor viruses such as *simian virus 40* (SV40) or the papillomaviruses (Lageix et al., 2007) but still show striking similarities with them in the way they induce host cells to enter

S phase or trigger progress beyond the G₁/S checkpoint (Hanley-Bowdoin et al., 2004). In *Nanoviruses*, Rep (M-Rep) plays a master role in initiation of replication (Timchenko et al., 2006). This would support the future host of the modified nanoviruses could be animals or human. Nevertheless, the stability of these plant viruses in the human gastro-intestinal tract may allow them to be used as a platform for oral vaccine development (Yusibov et al., 2002). Nanovirus proteins have found to be harmful for cell cycle-regulatory proteins in animals and plants. For its potential to link viral DNA replication with key regulatory pathways of the cell cycle, Aronson and his coworkers (Aronson et al., 2000) named the FBNYV C10 protein Clink, for "cell cycle link." This may be a clue to be alert from these plant viruses which are altering themselves for a new habitat.

MHC binding peptides

Innovations and discoveries of archetype in immunology are achieving impressive task in vaccine and drug design, anti-dotes and their development. Development of new MHC class-I and II binding peptides prediction tools are also supporting to this task.

The major histocompatibility complex (MHC) molecules are cell surface glycoproteins, which plays an important role in the host immune system, autoimmunity and reproductive success. MHC class-I encodes heterodimeric peptide-binding proteins as well as antigen-processing molecules such as TAP and Tapasin. MHC class-II encodes heterodimeric peptide-binding proteins and proteins that modulate antigen loading onto MHC class-II proteins in the lysosomal compartment such as MHC-II DM, MHC-II DQ, MHC-II DR, and MHC-II DP. The MHC Class-I molecules present the parts of almost all antigens to the T-cells with certain specificity. The binding mechanism appears to be the most selective step in the recognition of T-cell epitopes. The molecular mechanisms underlying this selectivity are still debated (Rhodes and Trowsdale, 1999), but a crucial factor is the complementarity between amino acids in the antigen peptide and the MHC binding pocket (Yewdell and Bennink, 1999). Successfully modeling the behavior exhibited by MHCs can be used to pre-select candidate peptides.

MHC alleles are grouped according to their structures. For class I MHC alleles, the close binding groove at both ends in MHC Class-I makes it possible to predict exactly which residues are positioned in the binding groove. For Class II MHC molecules, the binding groove is open at both ends and peptides which bind class II alleles are generally longer than those which bind class I MHCs, typically 9 to 25 residues. Moreover, the grooves of MHC Class-II alleles will only accommodate 9 to 11 residues of the target peptide (Kropshofer et al., 1993). Thus class II peptides have the potential to bind to the MHC groove in one of several registers (potential alignments between groove and antigenic peptide). A peptide binds through a network of hydrogen bonds between its backbone and the binding groove, and through interactions between the peptide side chains and pockets inside the binding groove (Madden et al., 1993; Stern et al., 1994). Interaction, within the groove, between MHC and peptide side chains is generally considered the principal determinant of binding affinity (Brusic et al., 1998). However, for MHC Class-II type alleles, a recent study speculates that binding may not be completely deterministic, and that the same peptide can have multiple possible binding cores (Tong et al., 2006). The binders and their subsets are most important for vaccine designers.

As viruses infect a cell by entering its cytoplasm, this cytosolic, MHC class-I dependent pathway of antigen presentation is the

primary way for a virus-infected cell to signal immune cells in plants. Parameters such as hydrophilicity, flexibility, accessibility, turns, exposed surface, and antigenic propensity of polypeptides chains have been correlated with the location of continuous epitopes.

Determination and identification of epitopes, their efficiency and the MHC-I and II binders are technically skillful and time consuming tasks. Thus, our predictions would be a milestone for the antidiotes designers and researchers who are working continuously within the study area of *Nanoviridae* family viruses.

Materials and Methods

Protein sequence analysis

We retrieved the protein sequences of interest of *Nanovirus* species and *Babuvirus* species and analyzed. Simultaneously, we extracted the protein and genomic sequences of SCSV, MVDV, FBNYV, BBTB and ABTV from NCBI (<http://www.ncbi.nlm.nih.gov/Protein/>) to perform the entropy measurement and phylogenetic reconstruction among the respective species.

Assessment of solvent accessibility regions

The calculation of surface accessible regions was based on surface accessibility scale on a product instead of an addition within the window. The accessibility profile was obtained using the formulae mentioned by Emani (Emani et al., 1985). A hexapeptide sequence with surface probability greater than 1.0 indicates an increased probability for being found on the surface. We also concentrated on the flexibility data of all the peptides to increase the prediction accuracy.

This has been done with Karplus and Schulz flexibility prediction. In this method, flexibility scale based on mobility of protein segments on the basis of the known temperature B factors of the α -carbons of 31 proteins of known structure was constructed (Karplus and Schulz, 1985). The calculation based on a flexibility scale is similar to classical calculation, except that the center is the first amino acid of the six amino acids window length and there are three scales for describing flexibility instead of a single one. For the assessment of solvent accessible regions in proteins, assay of different measurement was performed. This data and assessment may be useful for the prediction of the participant peptides in antigenic activity, surface region peptides and useful domain(s) in the sequence (Kyte and Doolittle, 1982; Abraham and Leo, 1987; Bull and Breese, 1974; Guy, 1985; Roseman, 1988; Wilson et al., 1981; Aboderin, 1971; Chothia, 1976; Janin, 1979; Cowan and Whittaker, 1990; Rose et al., 1985; Hopp and Woods, 1981; Eisenberg et al., 1984a; Eisenberg et al., 1984b; Gomase et al., 2007).

The comparative analysis has been done to predict the similarity and difference in hydrophobicity and hydrophilicity among the coat proteins of the species (see supplementary data).

Protein secondary structure prediction

The keys for secondary structure prediction were the residue conformational propensities, sequence edge effects, moments of hydrophobicity, position of insertions and deletions in aligned homologous sequences, moments of conservation, auto-correlation, residue ratios and filtering (Garnier et al., 1996; Gomase et al., 2008). The comparison among the secondary structures of respective protein sequences has been performed to predict the most probable regions and structures involved in antigenicity.

Prediction of antigenicity

The line up predicts the peptides that are likely to be antibody responsive. The antigenicity and antigenic epitopes for MHC Class-I and II have been predicted by using BepiPred server, MHC2pred server, TmhcPred server, Hopp and Woods, Kolaskar and Tongaonkar antigenicity methods.

The Parker hydrophilicity scales were also used within the assay to predict the hydrophilic peptides. In this method, hydrophilic scale based on peptide retention times during high-performance liquid chromatography (HPLC) on a reversed-phase column was constructed (Parker et al., 2003). A window of seven residues was used for analyzing epitope region. The corresponding value of the scale was introduced for each of the seven residues and the arithmetical mean of the seven residue value was assigned to the fourth, (i+3), residue in the segment. The comparative assessment has been calculated for the accuracy of the further prediction. Predictions are based on the comparative analysis, graphical representations and the tables which reflect the occurrence of amino acids at particular positions of experimentally known epitopes.

MHC binding peptide prediction

MHC-I and MHC-II molecules have their own specificity for the binding with their respective epitopes. We used TmhcPred server and MHC2pred server (Bhasin et al., 2003; Bhasin and Raghava, 2005) to predict the MHC Class-I and MHC Class-II epitopes. The servers allow the prediction of potential MHC class-I and class-II binding regions from antigenic sequences. The server can predict MHC binding regions or peptides for 97 MHC alleles. The server uses the matrix data in linear fashion for prediction. In MHC2Pred, the molecules from protein sequences or sequence alignments use Position Specific Scoring Matrices (PSSMs). This server uses Support Vector Machine (SVM) for the predictions.

The resultant epitopes were the output after proteosomal cleavage.

Comparative analysis

Finally we compared the proteomic analysis and genomic analysis mentioned above for the interpretation of resultant antigenic efficiency and rate of possible mutational changes or post translational modifications.

Results and Discussion

The randomly selected coat proteins sequences of FBNYV (NP_619570), MVDV (BAB78734), SCSV (AAA68021), BBTB (AAQ01659) and ABTV (ACN79533) are 172, 172, 169, 170 and 170 residues longer respectively.

Prediction of antigenic peptides

The epitopes of *Nanoviridae* family have been predicted by comparative analysis of greatest hydrophilic regions. Kyte-Doolittle is a widely applied scale for delineating hydrophobic character of a protein (Kyte and Doolittle, 1982). Hydrophobic regions achieve a positive value. Setting window size to 5-7 is suggested to be a good value for finding putative surface-exposed regions. Regions with values above 0 are hydrophobic in character. Hopp-Woods scale was designed for predicting potentially antigenic regions of polypeptides (Figure1) on the basis of the assumption that antigenic determinants would be exposed on the surface of the protein and thus would be

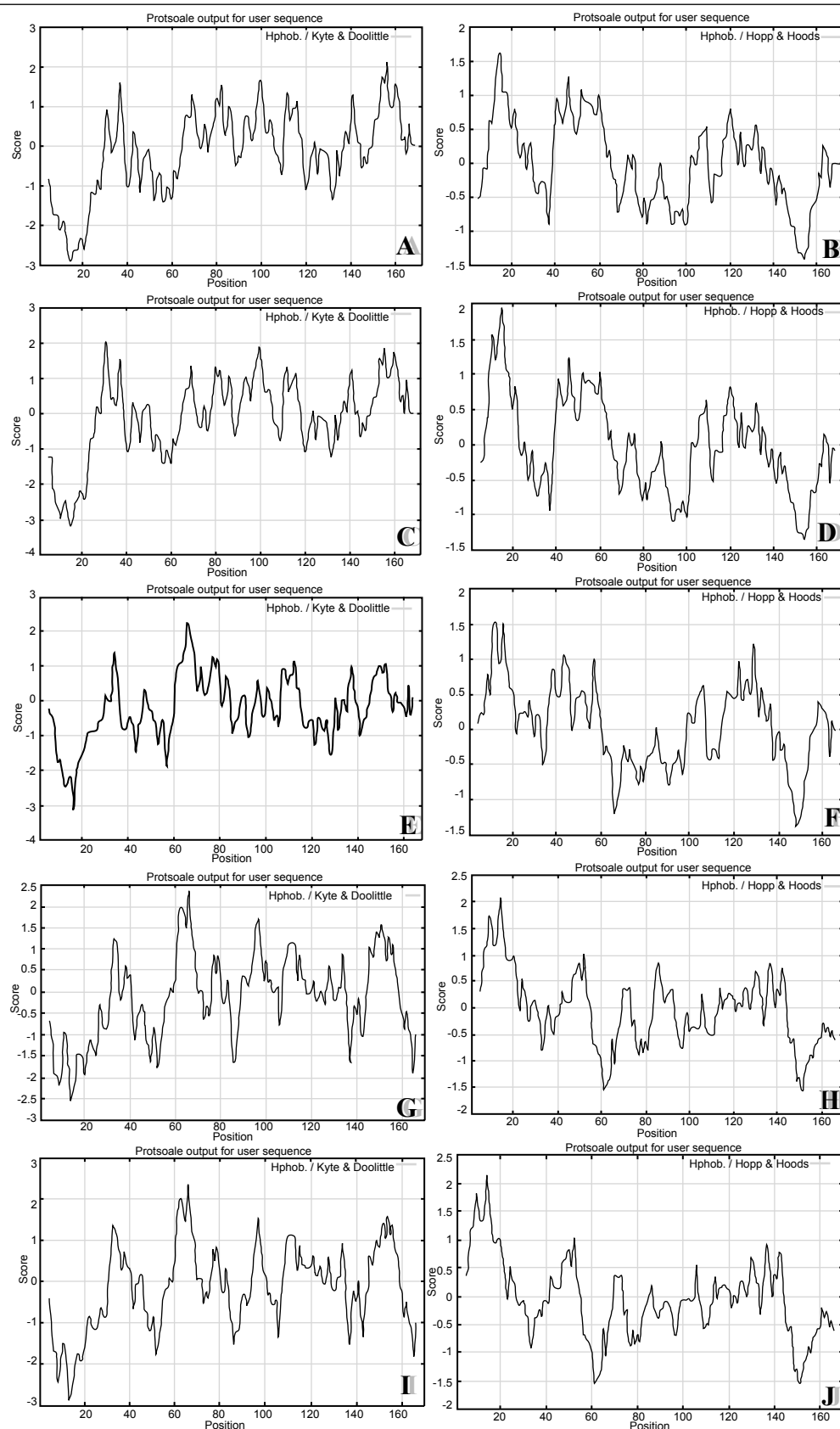


Figure 1: Kyte-Doolittle hydrophobicity plots (A, C, E, G and I) and Hopp and Woods hydrophilicity plots (B, D, F, H and J) of coat proteins in FBVNV, MVDV, SCSV, ABTV and BBTV respectively. In Kyte-Doolittle hydrophobicity plots (A, C, E, G and I), the regions with values above 0 are *hydrophobic* in character. In Hopp and Woods hydrophilicity plots (B, D, F, H and J), the regions with values above 0 are *hydrophilic* in character.

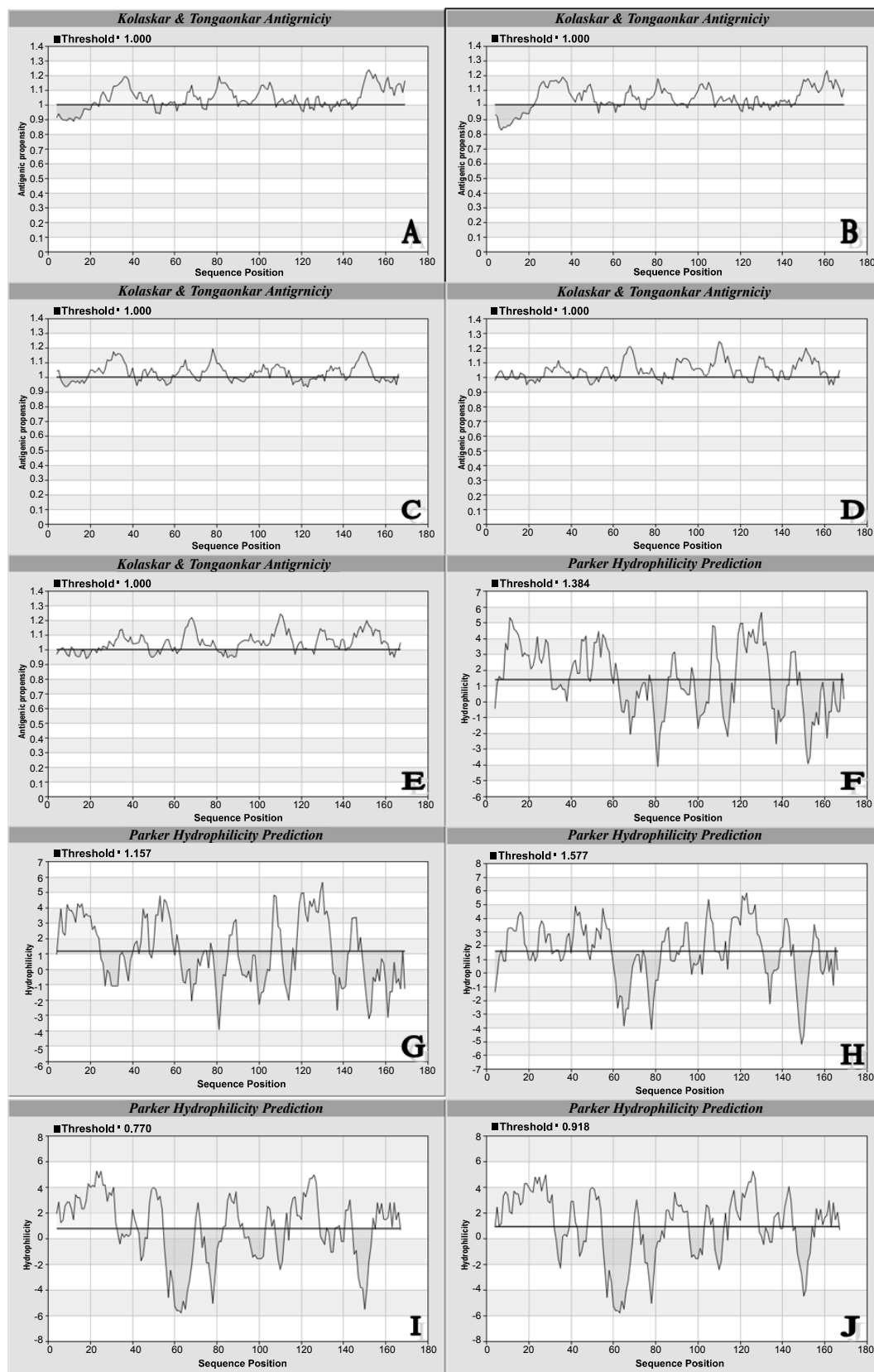


Figure 2: Kolaskar and Tongaonkar antigenicity prediction plots and Parker hydrophobicity plots of FBNYV (A and F), MVDV (B and G), SCSV (C and H), ABTV (D and I) and BBTV (E and J). Threshold = 1.000 for Kolaskar and Tongaonkar antigenicity prediction plots (A - E). Threshold = 1.384; 1.157; 1.577; 0.770 and 0.918 for Parker hydrophobicity plots (F - J) respectively.

located on the hydrophilic regions (Hopp and Woods, 1981). This scale was developed for predicting potential antigenic sites of globular proteins, which are likely to be rich in charged and polar residues. This scale is essentially a hydrophilic index with apolar residues assigned negative values. Moreover, using a window size of 6, the region of maximal hydrophilicity is likely to be an antigenic site. The values, greater than 0, are hydrophilic and thus likely to be exposed on the surface of folded proteins.

The prediction of antigenicity and epitopes has been done with BepiPred server and Kolaskar and Tongaonkar antigenicity (Figure 2A-2E). The antigenicity predictions for all the coat proteins in *Nanoviridae* family viruses are given in Table 1. The species within particu-

lar genera shows their common epitopes along with some positive modifications and diverse colleague epitopes as well. The common and/or similar antigenic peptide regions in *Nanoviruses* are 76-GELV-NYLIVKCNSP-89, 109-QDMITIIAKGK-119; 76-GELVNYIIVKSSSP-89, 109-QDMISIIAKGK-119 and 73-GELVNYLIVKSNS-85, 106-QDTV-TIV-112 in FBNVY, MVDV and SCSV respectively. Whereas, the common or similar antigenic regions in *Babuviruses* are 63-FMLLVCKVR-PGRILHWA-79 and 63-FMLLVCKVKPGRILHWA-79 in ABTV and BBTV respectively.

These epitopes can be a milestone for vaccine and antidotes design against their respective viruses.

Viruses	Start Position	End Position	Peptide	Peptide Length
FBNVY	25	50	KSSVPTTRVVVHQSAVLKKDEVVGTE	26
	64	72	KVMLTCTLR	9
	76	89	GELVNYLIVKCNSP	14
	91	107	SSWSAAFTSPALLVKES	17
	109	119	QDMITIIAKGK	11
MVDV	23	52	PYKPVVPIITRVVHQSAVLKKDEVVGCEIK	30
	66	72	MLTCTLR	7
	76	89	GELVNYIIVKSSSP	14
	96	107	AFTAPALLVKES	12
	109	119	QDMISIIAKGK	11
SCSV	19	40	SRIAYKPPSSKVVSHVESVLNK	22
	45	51	GAEVKPF	7
	59	69	MKKVMLIATLT	11
	73	85	GELVNYLIVKSNS	13
	96	104	NPSLMVKES	9
	106	112	QDTVIV	7
	132	140	RKFVKLGSG	9
ABTV	143	154	QTQHLIYSS	12
	26	40	SHDYAVDTSFIVPEN	15
	42	47	IKLYRI	6
	63	79	FMLLVCKVRPGRILHWA	17
	89	116	DPTVVLEAPGLFIKPAHSHLVKLVCSGE	28
BBTV	127	135	EVECLLRKT	9
	144	161	ELDFLYLAFYCSSGVTIN	18
	26	47	SHDYSSLGSLVPENTVKVFRI	22
	63	79	FMLLVCKVKPGRILHWA	17
	90	116	PTTCLEAPGLFIKPEHSHLVKLVCSGE	27
	118	125	EAGVATGT	8
	127	140	DVECLLRKTTVLRK	14
	142	161	VTEVDYLYLAFYCAGVSIN	20

Table 1: Predicted antigenic peptides of *Nanoviridae* family viruses.

Viruses	Start Position	End Position	Peptide	Peptide Length
FBNVY	10	27	TKGRRTPRRPYGRPYKSS	18
	58	63	ARYKMK	6
MVDV	4	24	NWNRNGMKRRRTPRRGYGRPY	21
	58	63	ARYKMR	6
SCSV	9	19	KGLRSQRRKYS	11
	21	27	IAYKPPS	7
	54	59	GSRYSM	6
ABTV	3	27	RYPKKALKKRKAVRRKYGSKATTSH	25
	45	55	YRIPTDKTLP	11
	133	138	RKTTLL	6
	161	166	NYQNRI	6
BBTV	3	20	RFPKKSIIKRRVGRKYG	18
	23	29	AATSHDY	7
	47	55	IEPTDKTLP	9
	133	138	RKTTVL	6
	161	166	NYQNRI	6

Table 2: Predicted surface accessible peptides of *Nanoviridae* family viruses.

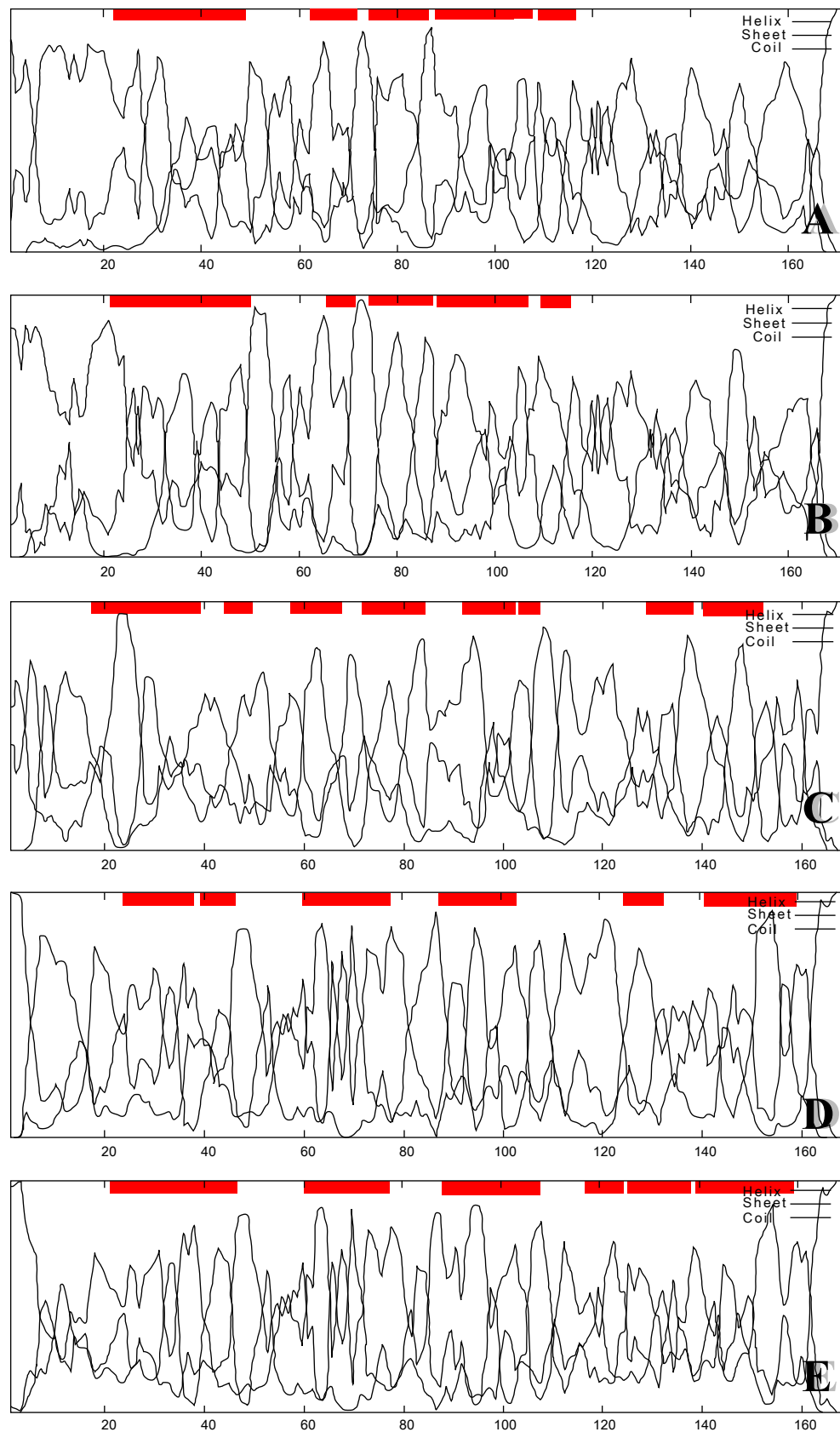


Figure 3: Secondary structures prediction plots of coat proteins in FBVYV (A), MVDV (B), SCSV (C), ABTV (D) and BBTV (E) using GOR method. Red bars indicate the probable antigenic peptide regions.

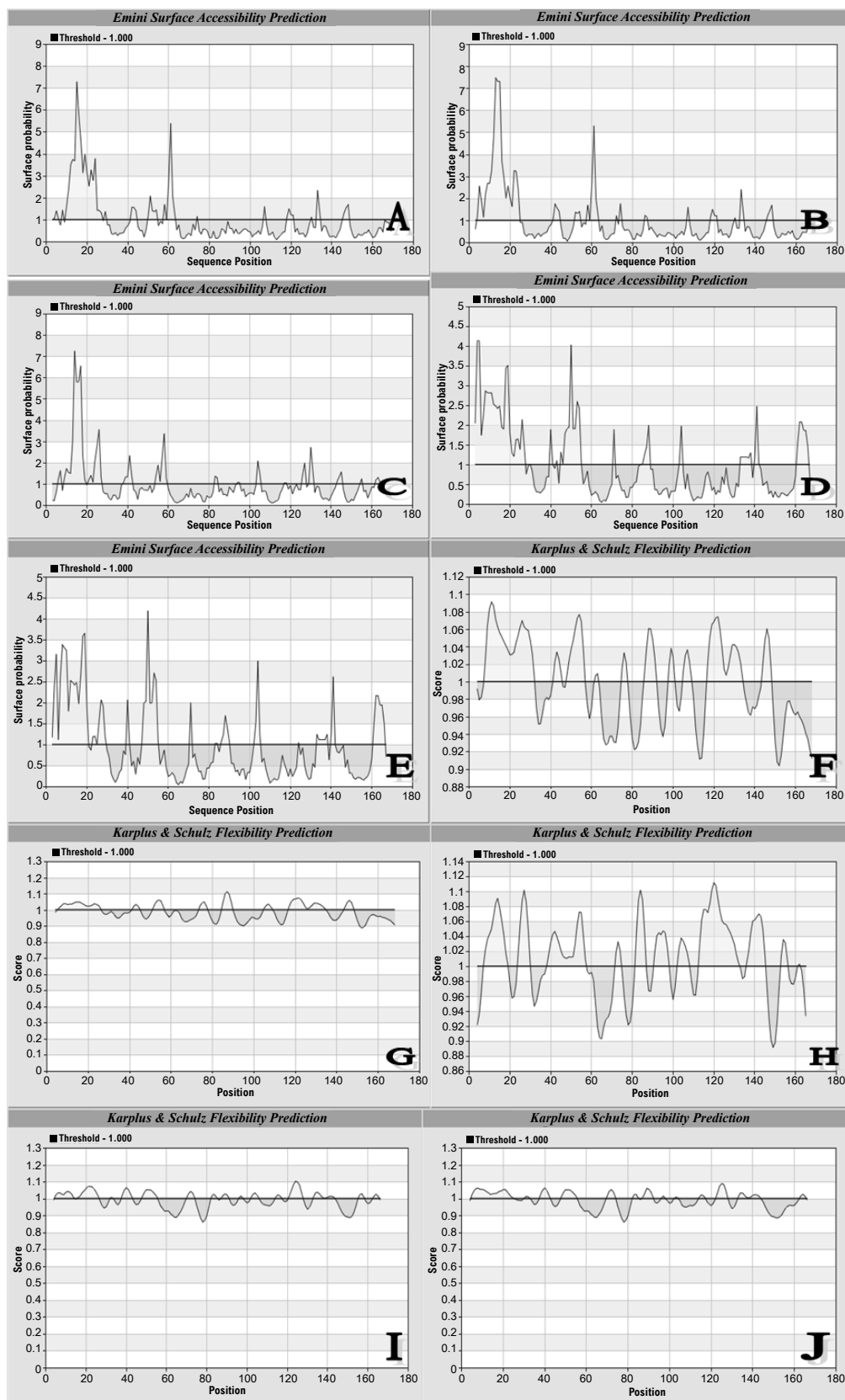


Figure 4: Emani surface accessibility prediction plots (A-E) and Karplus and Schulz flexibility plots (F-J) of FBNYV, MVDV, SCSV, ABTV and BBTV respectively. Threshold = 1.000.

Secondary structure comparison

The Garnier and Robson method (GOR) (Garnier et al., 1978) predicted the secondary structure of antigenic proteins of 5 coat proteins of *Nanoviridae* family species. Each residue has been assigned for the probability of alpha helix, beta strand or random coils by using a 7 residue window model. The comparison among the secondary structures of these species reveals that random coil residues and extended strands are more involved in antigenic propensity (Figure 3).

Accessible surface area

The protein folded the hydrophobic side chains were preferentially buried away from the external solvent. For this property measurement and analysis, the scales for solvent accessibility have been improved. These scales have proved a beneficial role in prediction of antigenic potential sites in globular proteins (Linding et al., 2003). We

predicted that the coat proteins of *Nanoviridae* family species are globular, hydrophobic and highly flexible. GlobPlot version 2.3 was used to predict the globularity and disorder in protein sequences by Russell/Linding definition (Linding et al., 2003). The Karplus and Schulz flexibility prediction (Karplus and Schulz, 1985) reflects that the coat proteins are highly flexible and somehow related to their solvent accessibility properties.

Solvent accessible regions segregate the molecules on the basis of their hydrophobic and hydrophilic properties. It has been assumed that often the hydrophilic molecules of peptides are active in their antigenicity. The scales are developed for predicting the potential antigenic sites of globular proteins. It is shown that viral capsids are highly flexible (Figure 4F-4J) and thus have highly active antigenic peptides.

Also we predicted the surface accessibility by Emani method (Emani et al., 1985). The Emani surface accessibility data reveals the

Viruses	Position	Coat protein disorder	No. of disorders
FBNVY	2-26; 123-129	ASKWNWSGTKGRRTPRRYPYGRPYKS; NGVAGTD	2
MVDV	2-24; 123-129	VSNWNRNGMKRRRTPRRGYGRPY; NGVAGTD	2
SCSV	6-10; 87-95; 114-124	WGRKG; IANWSSSFS; GKGLESSGTAG	3
ABTV	118-123	EAPVGG	1
BBTV	---	---	0

Parameters: Propensities=Russell/Linding smooth=10 dy/dx_smooth=10; Disorder frames: peak-frame=5 join-frame=4; Globularity frames: peak-frame=74 join-frame=15.

Table 3: Predicted coat proteins disordered regions in *Nanoviridae* family viruses.

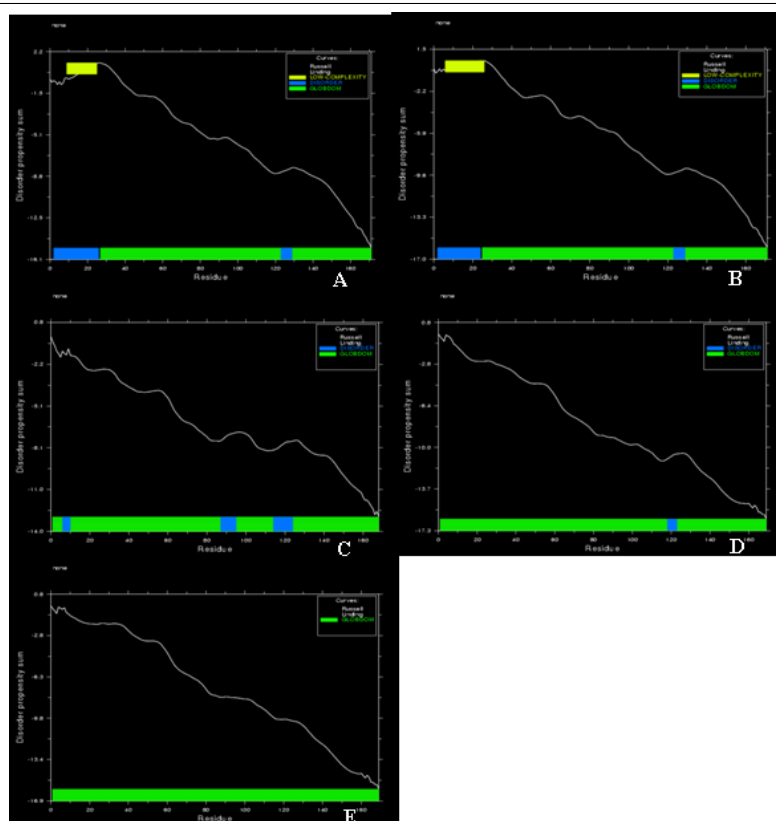


Figure 5: Protein globular domain and disorder prediction of (A) FBNVY (B) MVDV (C) SCSV (D) BBTV and (E) ABTV. Residue number on X-axis and Disorder propensity on Y-axis. Green = globular domains; Blue = Disorder and Yellow = low complexity region.

most probable residues and regions of the proteins being found on the surface (Figure 4A- 4E).

The predictions when compared with antigenicity data revealed that all the surface accessible peptides or residues are not mandatory to be involved in antigenicity.

Prediction of disordered regions

It has been postulated that disorder-to-order transitions provide a mechanism for uncoupling binding affinity and specificity. Disordered

protein sequences function in some cases to mechanically uncouple structured domains, making their dynamics less constrained. Linkers of this type are important in a diverse collection of proteins, from viral attachment proteins to transcription factors (Kissinger et al., 1999). The globular domains of protein are likely to be rich in charged and polar residues. Our disordered region prediction using GlobPlot version 2.3 (Linding et al., 2003) reveals four disordered regions containing proteins among five (Table 3). *Nanoviruses* (FBNYV, MVDV and SCSV) shows more disordered regions than *Babuviruses* (ABTV and BBTV). In fact, BBTV do not have any disordered peptide region.

Allele	Rank	Sequence	Residue No.	Score	Predicted affinity
H-2Db	1	GELVNYLIV	76	3.583	Low
	2	TSVALKAVL	156	2.149	Low
	3	AGITQTQHL	142	2.149	Low
	4	VMLTCTLRM	65	2.149	Low
H-2Dd	1	AGITQTQHL	142	2.995	Low
	2	AGTDCTKSF	126	1.974	Low
	3	KGKVESNGV	117	1.974	Low
	4	KSFNRFIKL	132	1.791	Low
H-2Ld	1	APGELVNYL	74	5.010	High
	2	SPALLVKES	99	3.401	Low
	3	SPISSWSAA	88	3.401	Low
	4	RPYKSSVPT	22	3.401	Low
H-2Kb	1	QTQHLYVVL	146	4.094	Moderate
	2	KSFNRFIKL	132	2.995	Low
	3	AGITQTQHL	142	1.335	Low
	4	LEHRVYVEV	164	0.970	Low
H-2Kd	1	RYKMKVVML	59	7.783	High
	2	RFIKLGAGI	136	7.560	High
	3	LYVLYTSV	150	7.272	High
	4	AFTSPALLV	96	5.662	High
H-2Kk	1	LEHRVYVEV	164	4.605	Moderate
	2	TEIKPEGDV	49	4.605	Moderate
	3	KDEVVGTEI	43	4.605	Moderate
	4	GELVNYLIV	76	3.912	Moderate

Table 4: Peptide binders of FBNYV coat protein to MHC-I molecules.

Allele	Rank	Sequence	Residue No.	Score	Predicted affinity
H-2Db	1	VSNWNRNGM	2	5.655	High
	2	GELVNYIIV	76	3.583	Low
	3	SPIANWAAA	88	3.496	Low
	4	TSVALKVVL	156	2.149	Low
H-2Dd	1	AGISQTQHL	142	2.995	Low
	2	AGTDCTKSF	126	1.974	Low
	3	KGKVESNGV	117	1.974	Low
	4	KSFNKFIRL	132	1.791	Low
H-2Ld	1	MPPGELVNY	73	4.682	Moderate
	2	KPVVPITRV	25	4.094	Moderate
	3	APALLVKES	99	3.401	Low
	4	SPIANWAAA	88	3.401	Low
H-2Kb	1	QTQHLYVVM	146	4.094	Moderate
	2	LEHRVYIEL	164	3.273	Low
	3	KSFNKFIRL	132	2.995	Low
	4	AGISQTQHL	142	1.153	Low
H-2Kd	1	RYKMRKVML	59	7.783	High
	2	KFIRLGAGI	136	7.742	High
	3	LYVVMYTSV	150	7.272	High
	4	MYTSVALKV	154	6.579	High
H-2Kk	1	KDEVVGCEI	43	4.605	Moderate
	2	GELVNYIIV	76	3.912	Moderate
	3	CEIKPDGDV	49	3.912	Moderate
	4	LEHRVYIEL	164	3.688	Low

Table 5: Peptide binders of MVDV coat protein to MHC-I molecules.

Allele	Rank	Sequence	Residue No.	Score	Predicted affinity
H-2Db	1	SSFSNPGLM	92	5.655	High
	2	ESVLNKRDV	35	3.663	Low
	3	GELVNYLIV	73	3.583	Low
	4	ESVQDTVIT	103	2.236	Low
H-2Dd	1	SGISQTQHL	139	3.178	Low
	2	VGGGKLESS	112	2.484	Low
	3	AGKDVTKSF	123	1.974	Low
	4	KSFRKFVKL	129	1.791	Low
H-2Ld	1	APGELVNYL	71	5.010	High
	2	KPFADGSRV	49	4.094	Moderate
	3	KPPSSKVVV	24	3.806	Moderate
	4	ISQTQHLYL	141	3.624	Low
H-2Kb	1	KSFRKFVKL	129	3.178	Low
	2	QTQHLYLII	143	2.890	Low
	3	ISQTQHLYL	141	1.376	Low
	4	SGISQTQHL	139	1.153	Low
H-2Kd	1	IYSSDAMKI	151	7.965	High
	2	RYSMKKVML	56	7.783	High
	3	KFVKLGSGI	133	7.742	High
	4	AYKPPSSKV	22	6.579	High
H-2K	1	LETRMYIDV	161	4.605	Moderate
	2	GELVNYLIV	73	3.912	Moderate
	3	QTQHLYLII	143	2.302	Low
	4	ESVQDTVIT	103	2.302	Low

Table 6: Peptide binders of SCSV coat protein to MHC-I molecules.

Allele	Rank	Sequence	Residue No.	Score	Predicted affinity
H-2Db	1	IVPENTIKL	36	6.674	High
	2	LLRRNVTEL	137	4.094	Moderate
	3	IKPANSHLV	101	3.496	Low
	4	VKLVCSEGL	109	2.069	Low
H-2Dd	1	GGTSEVECL	123	3.178	Low
	2	IVPENTIKL	36	3.178	Low
	3	VRPGRILHW	70	2.302	Low
	4	IKPANSHLV	101	1.791	Low
H-2Ld	1	LPRYFIWKM	54	3.806	Moderate
	2	APGLFIKPA	96	3.401	Low
	3	RPGRILHWA	71	3.401	Low
	4	VPENTIKLY	37	2.890	Low
H-2Kb	1	FIWKMFMLL	58	2.667	Low
	2	IVPENTIKL	36	2.069	Low
	3	FLYLAFYCS	147	1.609	Low
	4	CLLRKTTLL	130	0.875	Low
H-2Kd	1	FYCSEGVIT	152	7.965	High
	2	DYAVDTSEI	28	7.783	High
	3	SFVPENTI	34	7.560	High
	4	YFIWKMFML	57	7.377	High
H-2Kk	1	VECLLRKTT	128	3.401	Low
	2	LEAPVGGGT	117	2.995	Low
	3	TELDLYLYL	143	2.302	Low
	4	WDVKDPTVV	85	2.302	Low

Table 7: Peptide binders of ABTV coat protein to MHC-I molecules.

Allele	Rank	Sequence	Residue No.	Score	Predicted affinity
H-2Db	1	LVPENTVKV	36	3.583	Low
	2	AMIKSSWEI	79	2.331	Low
	3	VKLVCSEGL	109	2.069	Low
	4	CKVKPGRIL	68	1.974	Low
H-2Dd	1	TGTSDEVECL	123	3.178	Low
	2	VKPGRILHW	70	2.302	Low
	3	IKPEHSHLV	101	1.974	Low
	4	LVPENTVKV	36	1.974	Low
H-2Ld	1	VPENTVKVF	37	4.499	Moderate
	2	LPRYFIWKM	54	3.806	Moderate
	3	KPGRILHWA	71	3.401	Low
	4	YSSLSGILV	29	2.564	Low
H-2Lb	1	ATSHDYSSL	24	3.091	Low
	2	FIWKMFMLL	58	2.667	Low
	3	LYLAFYCS	147	1.609	Low
	4	VTEVDLYLYL	142	1.386	Low

H-2Kd	1	DYSSLGSIL	28	7.965	High
	2	FYCSAGVSI	152	7.783	High
	3	YFIWKMFML	57	7.377	High
	4	MFMLLVCKV	62	6.173	High
H-2Kk	1	HDYSSLGSI	27	4.605	Moderate
	2	VECLLRKTT	128	3.401	Low
	3	LEAGVATGT	117	2.995	Low
	4	TEVDYLYLA	143	2.302	Low

Table 8: Peptide binders of BBTv coat protein to MHC-I molecules.

Allele	Rank	Sequence	Residue No.	Score	Predicted affinity
I-Ab	1	YKSSVPTTR	24	1.154	High
	2	PTTRVVVHQ	29	1.056	Moderate
	3	QHLYVVLYT	148	1.041	Moderate
	4	ITQTQHLYV	144	1.000	Moderate
I-Ad	1	PISSWSAAF	89	0.575	Moderate
	2	FIKLGAGIT	137	0.505	Moderate
	3	GTKGRRTPR	9	0.504	Moderate
	4	TSVALKAVL	156	0.464	Low
I-Ag7	1	NGVAGTDCT	123	1.433	High
	2	GKVESNGVA	118	1.318	High
	3	EIKPEGDVA	50	1.316	High
	4	SWSAAFSTP	92	1.310	High
RT1.B	1	ITQTQHLYV	144	0.606	Moderate
	2	AFTSPALLV	96	0.602	Moderate
	3	TQTQHLYVV	145	0.590	Moderate
	4	AAFTSPALL	95	0.586	Moderate

Table 9: Peptide binders of FBNV coat protein to MHC-II molecules.

Allele	Rank	Sequence	Residue No.	Score	Predicted affinity
I-Ab	1	VALKVLEH	158	1.237	High
	2	PITRVVVHQ	29	0.822	Moderate
	3	ISQTQHLYV	144	0.808	Moderate
	4	QHLYVVMYT	148	0.772	Moderate
I-Ad	1	GAGISQTQH	141	0.578	Moderate
	2	VALKVLEH	158	0.497	Low
	3	TSVALKVVL	156	0.466	Low
	4	TAPALLVKE	98	0.421	Low
I-Ag7	1	SPIANWAAA	88	2.138	High
	2	WAAAFAPTA	93	1.698	High
	3	PIANWAAAF	89	1.642	High
	4	NWAAAFAP	92	1.439	High
RT1.B	1	AAAFAPAL	94	1.010	Moderate
	2	AFTAPALLV	96	0.716	Moderate
	3	FTAPALLVK	97	0.636	Moderate
	4	ANWAAAFAT	91	0.611	Moderate

Table 10: Peptide binders of MVDV coat protein to MHC-II molecules.

Allele	Rank	Sequence	Residue No.	Score	Predicted affinity
I-Ab	1	PSLMVKESV	97	0.990	Moderate
	2	RYSMKKVML	56	0.785	Moderate
	3	IAYKPPSSK	21	0.776	Moderate
	4	DAMKITLET	155	0.755	Moderate
I-Ad	1	KLESSGTAG	116	0.579	Moderate
	2	GGGKLESSG	113	0.571	Moderate
	3	GSGISTQH	138	0.541	Moderate
	4	APGELVNYL	71	0.447	Low
I-Ag7	1	SGTAGKDVT	120	1.631	High
	2	SRIAYKPPS	19	1.379	High
	3	SPIANWSSS	85	1.360	High
	4	TLTMAPGEL	67	1.309	High
RT1.B	1	TLTMAPGEL	67	0.624	Moderate
	2	AMKITLETTR	156	0.426	Low
	3	QTQHLYLII	143	0.392	Low
	4	ATLTMAPGE	66	0.379	Low

Table 11: Peptide binders of SCSV coat protein to MHC-II molecules.

Allele	Rank	Sequence	Residue No.	Score	Predicted affinity
I-Ab	1	IKLYRIEPT	42	1.076	Moderate
	2	KALKKRKAV	7	1.001	Moderate
	3	IWKMFMLLV	59	0.857	Moderate
	4	TTLRRNVT	135	0.820	Moderate
I-Ad	1	LEAPVGGGT	117	0.561	Moderate
	2	LDLYLAFY	145	0.547	Moderate
	3	FIWKMFMLL	58	0.496	Low
	4	GTSEVECLL	124	0.491	Low
I-Ag7	1	CKVRPGRIL	68	1.439	High
	2	TVVLEAPGL	91	1.438	High
	3	SGEAPVPG	114	1.421	High
	4	SHDYAVDTS	26	1.296	High
RT1.B	1	TTSHDYAVD	24	0.773	Moderate
	2	TSHDYAVDT	25	0.443	Low
	3	SKATTSHDY	21	0.434	Low
	4	TELDLYLA	143	0.426	Low

Table 12: Peptide binders of ABTV coat protein to MHC-II molecules.

Allele	Rank	Sequence	Residue No.	Score	Predicted affinity
I-Ab	1	RKTTVLRKN	133	1.088	Moderate
	2	IWKMFMLLV	59	0.857	Moderate
	3	GSILVPENT	33	0.732	Moderate
	4	PGLFIKPEH	97	0.630	Moderate
I-Ad	1	DYSSLGSIL	28	0.720	Moderate
	2	LEAGVATGT	117	0.661	Moderate
	3	TTCLEAPGL	91	0.528	Moderate
	4	FIWKMFMLL	58	0.496	Low
I-Ag7	1	SINYQNRIT	159	1.496	High
	2	YLYLAFYCS	147	1.360	High
	3	TEVDYLYLA	143	1.321	High
	4	SGELEGVA	114	1.318	High
RT1.B	1	ATSHDYSSL	24	0.757	Moderate
	2	INQPTTCLE	87	0.537	Moderate
	3	TEVDYLYLA	143	0.485	Low
	4	TTCLEAPGL	91	0.473	Low

*Optimal score for MHC peptide binder in mouse

Table 13: Peptide binders of BBTB coat protein to MHC-II molecules.

The whole peptide regions of BBTB confer the globular protein domains and absence of any disorder (Figure 5). The globularity and disorder of proteins were analysed using following parameters: Propensities=Russell/Linding smooth=10 dy/dx_smooth=10; Disorder frames: peak-frame=5 join-frame=4; Globularity frames: peak-frame=74 join-frame=15.

Prediction of MHC binding peptides

The MHC binding peptides plays an important role in immune response. Our prediction was based on cascade support vector machine, using properties of all the sequences and their residual characteristics. The correlation coefficient of 0.86 was obtained by Jack-knife validation test. This test resulted in the MHC Class I and II binding regions (Table 4 - 13).

MHCs are cell surface glycoproteins important for immunogenic reactions within the host organism. The MHC molecules show response to almost all antigens from the parasitic proteins. In our assay system, we predicted the binding affinity of coat proteins from five species in *Nanoviridae* family. The two *Nanovirus* species among three are having same residual length of 172 amino acids, while rest one having 169 amino acid residues. *Babuvirus* species show the

same amino acid length of 170 residues. All these proteins resulted in different nonamers (Table 4-13).

FBNYV and MVDV have 172 amino acids each and show 164 nonamers. However, SCSV has 169 amino acid residues in its coat protein which shows 161 nonamers. Moreover, the *Babuvirus* species i.e. ABTV and BBTB are having 170 residues longer and shows 162 nonamers. The small MHC-I and MHC-II binders found are given in the tables 4-13. The binding of MHC to the respective peptide were analyzed as a log-transformed value related to the IC50 values in nM units. These epitopes are enough to elicit the desired immune response. These are the regions responsible for antigenic binding to the MHC associated with immune response of the host.

Genomic analysis

We retrieved genomic sequences of involved viruses in our research, from NCBI (<http://www.ncbi.nlm.nih.gov/Genomes/>) ((ABTV NC_010319, (BBTB) NC_003479, (FBNYV) NC_003567, (MVDV) NC_003646 and (SCSV) NC_003817). The alignment of the sequences revealed no conserved sequences (see the multiple sequence alignment below).

CLUSTAL 2.0.12 multiple sequence alignment

```

NC_010319      -----GGCAGGGGGGCTTATTATTACCCCCCTGCCCG
NC_003479      AGATGTCCCGAGTTAGTGCGCCACGTAAGCGCTGGGGCTTATTATTACCCCCAGCGCTCG
NC_003646      -----CTGGGGCGGGGCTTAGTATTACCCCC-GCCCCAG
NC_003817      -----TAGTATTACCCC--GTGCCCG
NC_003567      -----GGCTCCAAGGTGGTTTGTAGTATTACCAC----CTTG
                                     *  * * * * * * *

```

NC_010319	GGACGGGACATTTGCATCTATAAATAGAAGCGCCCTCGCTCAACCA---GATCA-----G
NC_003479	GGACGGGACATTTGCATCTATAAATAGACCTCCCCCTCTCCATTACAAGATCATCATCG
NC_003646	GATCAGCGGAGTC-----ATTTAGA-----CTCGCTATAA-----
NC_003817	GGTCAGAGACATTTGA-CTAAATATTGA-----CTTGGAAATAA-----
NC_003567	GAGCCCCCACTTGC-----TAGAGAGAGAAAGAAGAGAGAAATG-----TCT
	* * * * *
NC_010319	GCGCTGCAATGGCTAGATATGTCGTATGTTGGATGTTCAACATCAACAATCC---CGAAG
NC_003479	ACGACAGAATGGCGCGATATGTGGTATGCTGGATGTTCAACATCAACAATCC---CACAA
NC_003646	GCCGTTAGATGTGTAGACACGT-----GGACGATCAGGATCTGTGATTC--GTGAAG
NC_003817	GCCCTTGGATTAGATGACACGT-----GGACGCTCAGGATCTGTGATGCTAGTGAAG
NC_003567	GCAGTGAAGTGGGTTTTACGT-----TGAAGTTCGCCGGCGAAGTTCC-----TG
	* * * * *
NC_010319	CTCTTCCAGAGATGAGGGAAGAATACAAATACCTGGTTTACCAGGTGGAGCGAGGCGAAA
NC_003479	CACTACCAAGTATGAGGGATGAGATAAAATATATGGTATATCAAGTGGAGAGGGGACAGG
NC_003646	CG-----AATCTGACGGAAGAT-CGTCCGAGCTTCGTG-----GTAGG
NC_003817	CGCTTAAGCTGAACGAATCTGACGGAAGAG-CGTCTGGTCCACATGTCTAAAGAATAAT
NC_003567	TTCTCTCGTT--CGACGAGAGAGTTCA-ATACGCCGTCTGGCAACACGAAAGAGTAAATC
	* * * *
NC_010319	GCGGTACACGACATGTGCAGGGCTATGTTGAAATGAAGAGACGAAGTTCTCTGAAACAAA
NC_003479	AGGGTACTCGTCATGTGCAAGGTTATGTCGAGATGAAGAGACGAAGCTCTCTGAAGCAGA
NC_003646	GCCC-----CTATGTTG-----CTTTAT--CTTTACTTTAATA-----AAGTAAA
NC_003817	GCTTTACA---GCTGTATTGATTTGACTTTACGCGCTTTACTTTAATTGCTTTAAGTAAA
NC_003567	ACGAC-----CATATTCAGGGAGTGATTCAATTAAAGAAGAAGGCAAAGATGAACACAG
	* * * *
NC_010319	TGAGGGCTTTAATTCCTGG--TGCCCATCTCGAAAAGAGAAGGGGCACACAGGAAGAAG
NC_003479	TGAGAGGCTTCTTCCCAGG---CGCACACCTTGAGAAACGAAAGGGAAGCCAAGAAGAAG
NC_003646	GTAAGATGCTGTCCCTTAC---TTTATTCGTTTGTCACTGG-----GTTACA----GCTG
NC_003817	GTAAGATGCTTTACTTTGC---TCGCACGAAGCAAAGTGATTGTAGCTGCA---GAAA
NC_003567	TGAAGAACATCATTGGTGAAATCCTCATCTGGAGAAGATGAAGGGTTCGATAGAAGAAG
	* * *
NC_010319	CTAGAGCTTATTGTATGAAGGCAGATACGAGAGTCGAAGGTCCCTTCGAGTTTGGTCTTT
NC_003479	CGCGGTCATACTGTATGAAGGAAGATACAAGAATCGAAGGTCCCTTCGAGTTTGGTTCAT
NC_003646	TCTTTGCTTCGTCT-CCAAGCAAAGCATAATTTCTCTCTCTATAAAAAG---CTGTTAAA
NC_003817	TTGATGCTTTAATTACCGGGTAACACGGTTTGATTGTGGGTATAAATATGTTCTGTTCTGT
NC_003567	CTTCTGCGTATGCCCAGAAAGAAGAATCAAGAGTCGCCCGGACCCTGGA-----GTTACG
	* * * *
NC_010319	TCAAAGTATCATGTAATGATAATTTGTTTGAT-GTCATACAGGATATGAGAGAAACGCAC
NC_003479	TTAAATTGTCATGTAATGATAATTTATTTGAT-GTCATACAGGATATGCGTGAAACGCAC
NC_003646	TCTATTCGTTGTG--TTCACAACGAAAATGGTGAGCAATTGGAATTGGAATGGGATGAAG
NC_003817	TTTCTTCGTTGTCATTTTACAACGAAGATGGTTGCTGTTTCG--ATGGGGAGAAAGGGTC
NC_003567	GTGAATTATTGAAGAAAGGTAGTC-----AT----AAACGGAAGATTATGAGAGTTAATT
	* * * *
NC_010319	AAACGGC---CGATTGAGTATTTATACGACTGTCCTAATACCTTCG-----ATAGAAGT
NC_003479	AAAAGGC---CTTTGGAGTATTTATATGATTGTCCTAACACCTTCG-----ATAGAAGT
NC_003646	AGACGACGAACCTCCTCGTCGTGGTTACGGCAGGCCTTACAAGCCTGTTGTTCCTATAACC
NC_003817	TGAGGTCTCAAAGGAGAAAATATTCGCGAATTGC-TTACAAACCT-----CCTTCGTCT
NC_003567	AAAGATC-----CGGAGAACGAATTGGAAGAACCCAGAAAATAC-----AGAAGAGC
	* * * *
NC_010319	AAGGATACATTATACAGGGTACAAGCGGAAATGAATAAAATGCAAGCTATGATGTCGTGG
NC_003479	AAGGATACATTATACAGAGTACAAGCAGAGATGAATAAAACGAAGGCGATGAATAGCTGG
NC_003646	AGGGTCGTCGTCCAT---CAATCAGCTCTGCTGAAGAAAGATGAAGTTGTTGGGTGTGAG
NC_003817	AAGGTTGTAAGTCAT---GTGGAGTCTGTTCTGAATAAGAGAGATGTTACTGGAGCGGAG
NC_003567	GATGGCTTGGTCCGC-----CATGGACGAATCTCGGAAGCTTGCTGAAGAAGGAGGCTTTC
	* * *
NC_010319	TCGGAACCTATGGTTGC--TGGACGAAGGAAG-TGGAGGAACTAATGGCG-GAGCCATG
NC_003479	AGAACTTCTTTCAGTGCT--TGGACATCAGAGG-TGGAGAATATCATGGCG-CAGCCATG
NC_003646	ATAAAACCAGATGGTGATGTTGCTCGTTATAAGATGATGAAGGTTATGTTAACTGTACT
NC_003817	GTTAAGCCATTTCGCTGATGGTTCAAGGTATAGTATGAAGAAGGTAATGTTGATTGCAACA
NC_003567	CCTATACGCTTTACAGCT--GGCAAGAAACAGTGTGGGCTATTAGAAGAAGAGCCCAA
	* * *

NC_010319	TCACCGACGGATTATTTGGGTCTA-TGGCCCAAATGGTGGTGAAGGTAAAACAACCTATG
NC_003479	TCATCGGAGAATAATTTGGGTCTA-TGGCCCAAATGGAGGAGAAGGAAAGACAACGTATG
NC_003646	TTGAGGATGCCCTCTGGAGAATTAGTCAATTATATCATCGTTAAGAGCAGTTCTCCCAT
NC_003817	TTAACTATGGCTCCTGGAGAATTAGTTAATTATCTTATTGTGAAGAGTAATTCGCCTATT
NC_003567	TGACCGTATTATTATTTGGGTCTA-CGGCCCAAATGGTAACGAAGGAAATCACAGTTTG
	* * * * *
NC_010319	CGAAGCATCTAATCAAGACCAGAAATGCATTTTATACACCTGGCGGAAAGACACTGGATA
NC_003479	CAAAACATCTAATGAAGACGAGAAATGCGTTTTATTCTCCAGGAGGAAATCATTGGATA
NC_003646	GCTA--ATTGGTCTGCAGCTTTTACTAC-TCCTGCTCTGTTAGTGAAGGAAAGTTGTCAA
NC_003817	GCGA--ATTGGAGTTCGTCTTTCAGTAA-TCCTTCGTTGATGGTGAAGAGTCTGTTCAA
NC_003567	GTAAATTCCTGGGATTAAAAAAGATTACCTTTATTACCTGGAGGTAAAACCCAAGATA
	* * * *
NC_010319	T---ATGTAGGCTG--TATAATTAT-GAGGGAATTGTAATATTTGATATTTCCAGATGCA
NC_003479	T---ATGTAGACTG--TATAATTAC-GAGGATATTGTTATATTTGATATTTCCAGATGCA
NC_003646	G---ACATGATATC--CATTATTGCTAAGGGCAAGGTTGAGTCTAACGGAGTTGCAGGTA
NC_003817	G---ATACAGTTAC--GATTGTTGGAGGAGGAAAGCTTGAGTCTTCTGGTACTGCTGGTA
NC_003567	TGACATATATGTTAATGAAAAATCCAAAGGCAAATGTTGTGATGGATATTCCTCGTTGTA
	* * * *
NC_010319	AAGAGGATTACTTGAATTACGGAATTCCTTGAGGAATTCAGAATGGCATCATTCAGAGCG
NC_003479	AAGAGGATTATTTAAATTATGGGTTATTAGAGGAATTTAAGAATGGAATAATTCAAAGCG
NC_003646	CTGACTGTACAAAGTCATTTAATAAATTTATTAGATTGGGAGCCGGTATCAGCCAAACCC
NC_003817	AAGATGTAACATAAGTCTTTTAGGAAGTTTGTAAAGCTGGGTTTCAGGTATTAGTCAGACCC
NC_003567	ATTCTGAATATTTAAATTATCAATTTATGGAATTAATAAAAATAGAACCATATTTAGTT
	* * * *
NC_010319	GGAAATATGAACCGTTTTAAAAATGTAGAGTATG---TGGAGGTCATTGTCATGGCTA
NC_003479	GGAAATATGAACCCGTTTTGAAGATAGTAGAATATG---TCGAAGTCATTGTAATGGCTA
NC_003646	AGCATTTGTATGTTGTATGTACACTAGTGTGCTT---TGAAGCTGTGTTAGAACATA
NC_003817	AGCATTTGTATTTAATTATTTATTCAGTGATGCGA---TGAAGATCACACTGGAGACGA
NC_003567	ATAAATATGAACCGTTGGATGTATTATAAATAATAAAATACATGTAATTGTATTAGCTA
	* * * *
NC_010319	-ACTTCCTGCCGAAGGAAGGAATATCTCGGAAGACCGAATAAAGCTTGTAACCTGTTGA
NC_003479	-ACTTCCTTCCGAAGGAAGGAATCTTTTCTGAAGATCGAATAAAGTTGGTTTTCTTGCTGA
NC_003646	GAGTGTATATTGAAC--TGTAAT-GTAATGAAGAACACTATGAAATAATGAAATCAACAA
NC_003817	GAAATGTATATTGATG--TATAATTGTGATGATTAAATGA-ATAAAGAGTTG---TTTTTAT
NC_003567	-ATGTATTGCCTGATTATGAAAAAATTAGTCAGGACAGAATTAAATAATTTATTGTTAA
	* * * *
NC_010319	ACACGCTATGCAATAAAGGGGAAAAATGCAATTATGACCTGTCACGTTTACACTTTTCGT
NC_003479	ACAAGTAATGACTTTACAGCGCACGCTCCGACAAAAGCACACTATGACAAAA-----GT
NC_003646	TCATTGAATCTTATTACTCCGCGTAGCGGTATGTTTCCGTGTTTTTGTGGCCAATAA-TG
NC_003817	TCTTTGAA-----TTACTCCGCGAAGCGGTGTGTT--ATGTTTTTGTGGAGACAT-AT
NC_003567	GTATTCGGCGAAGCCATA-TATATAAAAAAAAAAATTTGCGTTTTTGGTATCAAAACG---
	*
NC_010319	AAAGATGTAGGGCCGAAG-GCCCTAATGACGC-GTGCATATTCTCTATAGTGGTGGGTC
NC_003479	ACGGGTATCTGATTGGGTTATCTTAAACGATCTAGGGCCGTAGGCCCGTGAGCAATGAACG
NC_003646	CCCTTCATTAATGAAGGAGAATTTACAAATATGACCTGTGACGTCATTTGATCCCGTGC
NC_003817	GACGTCATATGTCCTCCGCGA-CAGGCTGGCACGGGGCT-----
NC_003567	ACGTCGTTTTTACCTCGGCGCCCTATAAATAGA-----
	*
NC_010319	ATATGTCCCGAGTTAGTGCGCCACGTG
NC_003479	GCGAGATC-----
NC_003646	TGAG-----

This confers the high flexibility within the sequences. The entropy analysis proves the high flexibility and variability in the sequences (Figure 6). This might be cause of mutational changes or post translational modifications within the genome.

A maximum parsimonious tree has been generated to check the evolutionary changes, mutational changes and/or post translational modifications and comparison with the predictions of our proteomic analyses. The evolutionary history was inferred using the Maximum Parsimony method. The most parsimonious tree with length =

1211 is shown. The consistency index is (0.853035), the retention index is (0.752022) and the composite index is 0.694890 (0.641501) for all sites and parsimony-informative sites (in parentheses). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, 1985). The MP tree was obtained using the Close-Neighbor-Interchange algorithm (Nei and Kumar, 2000) with search level 3 (Felsenstein, 1985; Nei and Kumar, 2000) in which the initial trees were obtained with the random addition of sequences (10 replicates). The tree is drawn to scale with branch lengths calculated

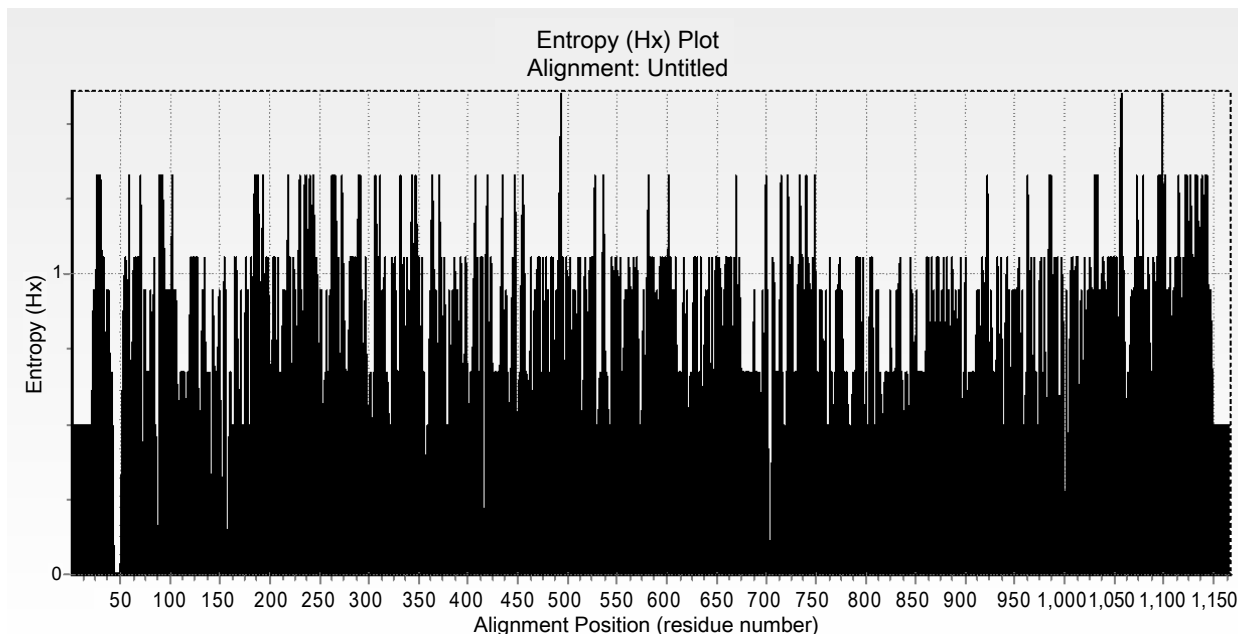


Figure 6: Graphical representation of entropy analysis of *Nanoviridae* family viruses. Residue numbers or alignment positions on X-axis and entropy (Hx) on Y – axis.

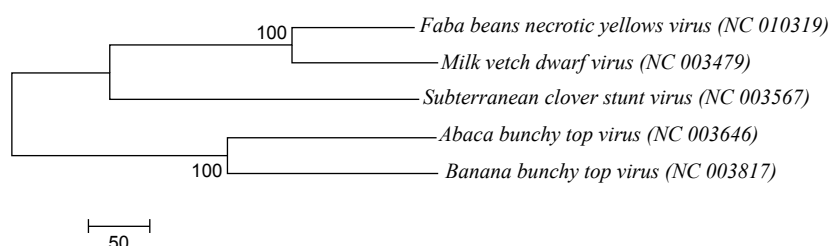


Figure 7: Evolutionary relationships of *Nanoviridae* family viruses on the basis of genomic sequences. The most parsimonious tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches.

using the average pathway method (Nei and Kumar, 2000) and are in the units of the number of changes over the whole sequence. All positions containing gaps and missing data were eliminated from the dataset (Complete Deletion option). The reconstructed tree presumes that SCSV is distinct from rest of the *Nanoviruses* (Figure 7).

There were a total of 884 positions in the final dataset, out of which 371 were parsimony informative. Phylogenetic analyses were conducted in MEGA4 (Tamura et al., 2007).

In all phylogenetic analyses of *Nanoviridae* isolates, ABTV and BBTv fall in the same clade, but on separate branches. From the tree, sequence length comparison and flexibility data; we can infer that the rate of mutation is high among the *Nanoviridae* family genera. Our genomic analysis supports the proteomic analysis of predictions.

The comparative binding affinity analysis shows the mutational changes within SCSV made it highly potential virus among all *Nanoviridae* viruses against MHC Class-I molecules (Figure 8). However, MVDV shows greater affinity towards the MHC Class-II molecules as compare to other viruses within the analysis even though the SVSC has more numbers of antigenic determinants than MVDV.

Conclusion

Despite the fact that, the nanoviruses attack plants mostly, the MHC binding peptide and their increasing competence forecast be a sign of future host could be a man. We should consider the severity of this issue. We confirm that the plant viruses can be harmful to humans and animals. Our finding would help in future pharmacologists as well as agriculture scientists.

The coat proteins of *Nanoviridae* family viruses are highly active for mutations. The possible mutational changes in the genome or post translational modifications within SCSV have decreased the length of the sequence but have increased the antigenic potential and numbers of antigenic binding regions on the surface of the protein. SCSV has the high antigenicity among all *Nanoviridae* family viruses. The probable mutational changes or post translational modifications within the epitopes may cause the selectivity and specificity of the virus to the host. Nanoviruses (FBNYV, MVDV and SCSV) contain few similar epitopes with either positive mutational changes or post translational modifications. However, *Babuviruses* also shows some antigenic similarity. Indeed, *Nanoviruses* have different epitopes as compare to *Babuviruses*. The MHC binding regions involve the random coils and extended strands with their flexible property. The dis-

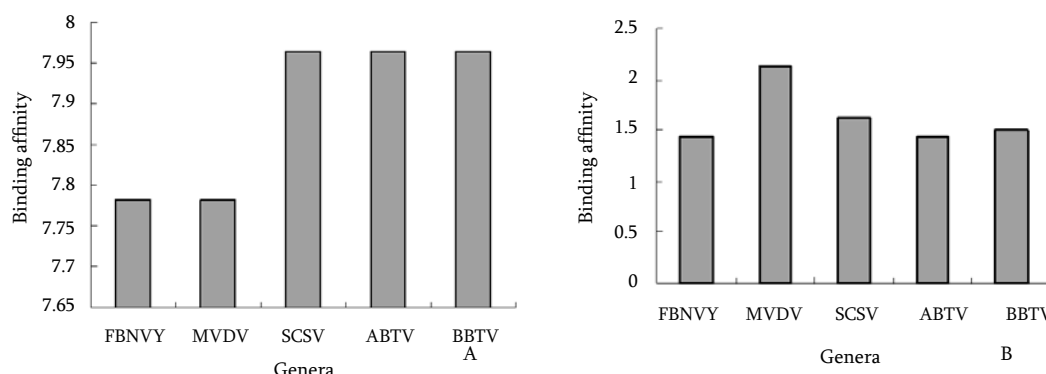


Figure 8: Binding affinity potential analysis of *Nanoviridae* family viruses against MHC-I (A) and MHC-II (B). Species of the genera are on X – axis and binding affinity potential on Y – axis.

orders within the surface protein sequences may play a crucial role in the generation of new antigenic binding regions. Apparently, the disordered regions do not take part in antigenicity of the protein, but can be found at the initial or terminal ends of the antigenic region on the peptides. All the surface accessible peptides or residues are not mandatory to being involved in antigenicity. The nonamers antigenic determinants in *Nanoviridae* family viral coat proteins are highly sensitive to H-2Kd and I-Ag7. These epitopes may play a highly informative and crucial role in antidotes production against FBNVY, MVDV, SCSV, ABTV and BBTV. One can apply one antidote for more than one species in its respective genera. The predicted MHC binding regions acts as barriers for antigens and also are responsible to generate immune response within the host. Thus, the antigenic epitopes are vital for activation of defense mechanism within the host. The species within particular genera shows their common epitopes along with some positive modifications and diverse colleague epitopes as well. The common antigenic peptide regions in *Nanoviruses* are 76-GELVNYLIVKCN89, 109-QDMITIIAKGK-119, 76-GELVNYIIVKSSSP-89, 109-QDMISIIAKGK-119 and 73-GELVNYLIVK-SNS-85, 106-QDTVIV-112 in FBNVY, MVDV and SCSV respectively. Whereas, the common antigenic regions in *Babuviruses* are 63-FMLLVCKVRPGRILHWA-79, 89-DPTVVLEAPGLFIKPANSHLVKLVCSGE-116 and 63-FMLLVCKVRPGRILHWA-79, 90-PTTCLLEAPGLFIKPEHSHLVKLVCSGE-116 in ABTV and BBTV respectively. These common antigenic peptides can be used as their identifiers. These are also important for synthetic peptide vaccine production or antidotes production against the relevant viruses.

Acknowledgement

We are grateful to Department of Biotechnology (DBT) – Ministry of Science and Technology, Govt. of India, for providing us the Bioinformatics Infrastructure Facility (BIF) for our research work. We are also thankful to Dr. Mrs. Nilima Kshirsagar, Hon'ble Vice Chancellor, Maharashtra University of Health Sciences, Nashik, India for the kind support.

References

- Aboderin AA (1971) An empirical hydrophobicity scale for alpha-amino-acids and some of its applications. *Int J Biochem* 2: 537-544. » CrossRef » PubMed » Google Scholar
- Abraham DJ, Leo AJ (1987) Extension of the fragment method to calculate amino acid zwitterions and side chain partition coefficients. *Proteins* 2: 130-152. » CrossRef » PubMed » Google Scholar
- Aronson MN, Meyer AD, Györgyey J, Katul L, Vetten HJ, et al. (2000) Clink, a Nanovirus-Encoded Protein, Binds both pRB and SKP1. *J Virology* 74: 2967-2972. » CrossRef » PubMed » Google Scholar
- Bhasin M, Raghava GP (2005) Pcleavage: an SVM based method for prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences. *Nucleic Acids Res* 33: W202-207. » CrossRef » PubMed » Google Scholar
- Bhasin M, Singh H, Raghava GPS (2003) MHCBN: a comprehensive database of MHC binding and non-binding peptides. *Bioinformatics* 19: 665-666. » CrossRef » PubMed » Google Scholar
- Brusic V, Rudy G, Honeyman G, Hammer J, Harrison L (1998) Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network. *Bioinformatics* 14: 121-130. » CrossRef » PubMed » Google Scholar
- Bull HB, Breese K (1974) Surface tension of amino acid solutions: A hydrophobicity scale of the amino acid residues. *Arch Biochem Biophys* 161: 665-670. » CrossRef » PubMed » Google Scholar
- Chothia C (1976) The nature of accessible and buried surfaces in proteins. *J Mol Biol* 105: 1-12. » CrossRef » PubMed » Google Scholar
- Cowan R, Whittaker RG (1990) Hydrophobicity indices at pH 3.4 determined by HPLC. *Pept Res* 3: 75-80. » CrossRef » PubMed » Google Scholar
- Dale JL (1987) Banana bunchy top: An economically important tropical plant virus disease. *Adv Virus Res* 33: 301-325. » CrossRef » PubMed » Google Scholar
- Eisenberg D, Schwarz E, Komaromy M, Wall R (1984a) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J Mol Biol* 179: 125-142. » CrossRef » PubMed » Google Scholar
- Eisenberg D, Weiss RM, Terwilliger TC (1984b) The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc Natl Acad Sci USA* 81: 140-144. » CrossRef » PubMed » Google Scholar
- Emani EA, Hughes JV, Perlow DS, Boger J (1985) Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol* 55: 836-839. » CrossRef » PubMed » Google Scholar
- Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA (2005) *Virus Taxonomy. Eighth Report of the International Committee on Taxonomy of Viruses*. Amsterdam: Academic Press. » CrossRef » PubMed » Google Scholar
- Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783-791. » CrossRef » PubMed » Google Scholar
- Garnier J, Osguthorpe DJ, Robson B (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J Mol Biol* 120: 97-120. » CrossRef » PubMed » Google Scholar
- Garnier J, Gibrat JF, Robson B (1996) GOR: secondary structure prediction method version IV. *Methods Enzymol* 266: 540-553. » CrossRef » PubMed » Google Scholar
- 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: 117-121. » CrossRef » PubMed » Google Scholar
- Gomase VS, Changbhale SS, Kale KV (2008) Insilico analysis of nucleocapsid protein from groundnut ringspot virus. *Advancements in Information Technology and Internet Security* 370-378. » CrossRef » PubMed » Google Scholar



20. Grylls NE, Butler FC (1956) An aphid-transmitted virus affecting subterranean clover. *J Aust Znst agric Sci* 22: 73-74. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
21. Guy HR (1985) Amino acid side chain partition energies and distributions of residues in soluble proteins. *Biophys J* 47: 61-70. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
22. Hanley-Bowdoin L, Settlage SB, Robertson D (2004) Reprogramming plant gene expression: a prerequisite to geminivirus DNA replication. *Mol Plant Pathol* 5: 149-156. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
23. Hopp TP, Woods KR (1981) Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci USA* 78: 3824-3828. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
24. Janin J (1979) Surface and inside volumes in globular proteins. *Nature* 277: 491-492. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
25. Karplus PA, Schulz GE (1985) Prediction of Chain Flexibility in Proteins - A tool for the Selection of Peptide Antigens. *Naturwissenschaften* 72: 212-213. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
26. Katul L, Maiss E, Vetten HJ (1995) Sequence-Analysis of A Faba Bean Necrotic Yellows Virus-Dna Component Containing A Putative Replicase Gene. *J Gen Virol* 76: 475-479. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
27. Kissinger P, Rice J, Farley T, Trim S, Jewitt K, et al. (1999) Application of Computer-assisted Interviews to Sexual Behavior Research. *Am J Epidemiol* 149: 950-954. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
28. Kropshofer H, Max H, Halder T, Kalbus M, Muller CA, et al. (1993) Self-peptides from four HLA-DR alleles share hydrophobic anchor residues near the NH2-terminal including proline as a stop signal for trimming. *J Immunol* 151: 4732-4742. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
29. Kyte J, Doolittle RF (1982) A Simple Method for Displaying the Hydropathic Character of a Protein. *J Mol Biol* 157: 105-132. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
30. Lageix S, Catrice O, Deragon JM, Gronenborn B, Pelissier T, et al. (2007) The nanovirus-encoded Clink protein affects plant cell cycle regulation through interaction with the retinoblastoma-related protein. *J Virol* 81: 4177-4185. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
31. Li L, Kapoor A, Slikas B, Bamidele OS, Wang C, et al. (2010) Multiple diverse circoviruses infect farm animals and are commonly found in human and chimpanzee feces. *J Virol* 84: 1674-1682. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
32. Linding R, Russell RB, Neduva V, Gibson TJ (2003) GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Res* 31: 3701-3708. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
33. Madden DR, Garboczi DN, Wiley DC (1993) The antigenic identity of peptide-MHC complexes: a comparison of the conformations of five viral peptides presented by HLA-A2. *Cell* 75: 693-708. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
34. Magee CJP (1953) Some aspects of the bunchy top disease of banana and other Musa spp. *Proc Royal Society of New South Wales* 87: 3-18. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
35. Makkouk KM, Vetten HJ, Katul L, Franz A, Madkour MA (1998) Epidemiology and control of faba bean necrotic yellows virus. In: Plant Virus Disease Control. A. Hadidi, R.K. Khetarpal and H. Koganezawa (eds.). APS Press, The American Phytopathological Society, St. Paul, Minnesota, USA pp534-540. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
36. Nei M, Kumar S (2000) Molecular Evolution and Phylogenetics. Oxford University Press, New York. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
37. Ocfemia GO (1926) Progress report on bunchy-top of abaca or manila hemp. *Phytopathology* 16: 894. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
38. Ocfemia GO (1930) Bunchy-top of abaca or manila hemp I. A study of the cause of the disease and its method of transmission. *Am J Bot* 17: 1-18. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
39. Ocfemia GO, Buhay GG (1934) Bunchy top of abaca or Manila hemp-2. Further studies on the transmission of the disease and a trial planting of abaca seedlings in a bunchy top devastated field. *Philippine Agriculturists* 22: 267-80. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
40. Ohki ST, Doi Y, Yora K (1975) Small spherical partials found in broad been plants infected with Milk- Vetch Dwarf Virus. *Phytopathological note* 41: 508-510. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
41. Parker JMR, Guo D, Bhasin M, Singh H, Raghava GPS (2003) A comprehensive database of MHC binding and non-binding peptides. *Bioinformatics* 19: 665-666. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
42. Raymundo AD, Bajet NB (2000) Epidemiology and integrated management of Abaca Bunchy Top in Philippines, In: Molina A, Roa V, Bay-Petersen J, Carpio A, Joven J (Eds) Proceedings in Davao City. Philippines 89-97. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
43. Rhodes DA, Trowsdale J (1999) Genetics and molecular genetics of the MHC. *Rev in immunogenetics* 1: 21-31. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
44. Rose GD, Geselowitz AR, Lesser GJ, Lee RH, Zehfus MH (1985) Hydrophobicity of amino acid residues in globular proteins. *Science* 229: 834-838. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
45. Roseman MA (1988) Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. *J Mol Biol* 200: 513-522. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
46. Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, et al. (1994) Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* 368: 215-221. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
47. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596-1599. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
48. Timchenko T, Katul L, Aronson M, Vega-Arreguin JC, Ramirez BC, et al. (2006) Infectivity of Nanovirus DNAs: Induction of disease by cloned genome components of Faba bean necrotic yellows virus. *J Gen Virol* 87: 1735-1743. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
49. Tong JC, Zhang GL, Tan TW, August JT, Brusica V, et al. (2006) Prediction of HLA-DQ3.2beta ligands: evidence of multiple registers in class II binding peptides. *Bioinformatics* 22: 1232-1238. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
50. Wilson KJ, Honegger A, Stotzel RP, Hughes GJ (1981) The behaviour of peptides on reverse-phase supports during high-pressure liquid chromatography. *Biochem J* 199: 31-41. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
51. Yewdell JW, Bennink JR (1999) Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. *Annu Rev Immunol* 17: 51-88. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
52. Yusibov V, Hooper DC, Spitsin SV, Fleish N, Kean RB (2002) Expression in plants and immunogenicity of plant virus-based experimental rabies vaccine. *Vaccine* 20: 3155-3164. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
53. Zhang T, Breitbart M, Lee WH, Run J, Wei CL (2006) RNA viral community in human feces: Prevalence of plant pathogenic viruses. *PLoS Biol* 4: e3. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)