International Journal of Advanced Biotechnology and Research (IJBR)
ISSN 0976-2612, Online ISSN 2278–599X,
Vol-8, Issue-1, 2017, 321-333
http://www.bipublication.com

Research Article

Detection of b-cell & t-cell epitopes in major proteins of the order *Mononegavirales* of ssRNA negative strand viruses

Jaspreet Kaur*, Mansi Arora, Sanjot Kaur,

Aishwarya Ahuja, Swati Sharma and Shruti Jaluthria

Department of Zoology, Maitreyi College, Bapu Dham,
Chanakyapuri, New Delhi-110021,
University of Delhi, India
*kaurpreet21@gmail.com, Tel: 09899660599

ABSTRACT:

The maximum number of ssRNA viruses, which infect both plants and animals, belong to the category of non-segmented negative strand RNA viruses (NS-NSVs). This group belongs to the order *Mononegavirales* and it consists of five families: *Bornaviridae*, *Filoviridae*, *Nyamviridae*, *Paramoxyviridae* and *Rhabdoviridae* as well as some unclassified viral isolates. The genetic structure of these viruses is extremely similar, comprising of only one molecule of negative sense single stranded RNA, which codes for 5-10 genes. In this study, the major proteins found in the members of the order *Mononegavirales* have been screened for the detection of B-cell and T-cell epitopes that bind with Class I MHC CD8+ and Class II MHC CD4+ molecules. These epitopes may further help in developing immunodiagnostic methods or antibody based therapeutics against these viruses. Also, the identity of hypothetical proteins and proteins with unassigned function found in some viral isolates was also investigated.

Keywords: *Mononegavirales, B-cell epitope, T-cell epitope, Class I MHC, Class II MHC*

1. INTRODUCTION

Non-segmented negative strand RNA viruses (NS-NSVs) comprise a group of viruses infecting plants and animals, including mammals and they belong to a single large order named *Mononegavirales*. At the time of writing, a total of 5444 viral genome sequences, including the genomes of unclassified and unassigned viruses and virophages were available (http://www.ncbi.nlm.nih.gov/genome/viruses/, Figure-1).

The order *Mononegavirales* comprises of five distinct families, namely, *Bornaviridae*, *Filoviridae*, *Nyamviridae*, *Rhabdoviridae* and *Paramoxyviridae*. It includes some unclassified members also. These viruses have a simple genomic structure, comprising of a 5′ leader sequence or UTR, nucleoprotein or nucleocapsid (N), phosphoprotein (P), matrix protein (M), glycoprotein (G), RNA dependent RNA polymerase (RdRp) and a trailer sequence or 3′ UTR. The RNA is not infectious by itself until packaged by the nucleocapsid protein and transcribed by the RdRp and co-factors [1].

Members of *Nyamviridae* and unclassified viral isolates do not infect humans, while maximum number of viral strains that infect vertebrates and humans belong to the family *Paramoxyviridae*,

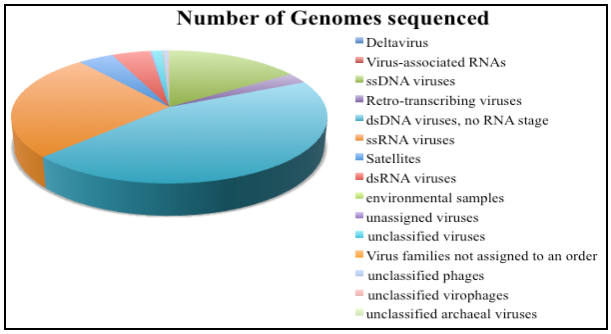
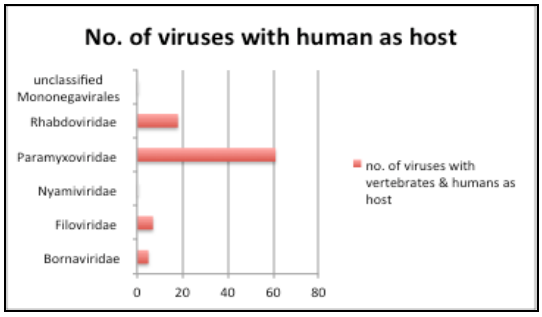
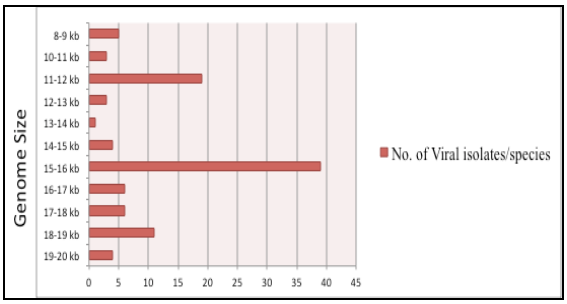


Figure 1: Distribution of complete sequenced genomes of different taxonomical categories of viruses. The maximum of genomes were available from dsDNA viruses with no RNA stage (2421), followed by ssRNA viruses (1422) and ssDNA viruses (859).

followed by *Rhabdoviridae* (Figure-2a). The members of *Paramoxyviridae* have the largest genome size of ~15-16 kb, while in the members of *Bornaviridae*, it is ~8-9 kb (Figure-2b). *Bornaviridae* comprises of only one genus *Bornavirus*, which causes severe neurobehavioral changes in horses and sheep [1].



(a)



(b)

Figure 2: Distribution plots*:* (a) Number of ssRNA viruses in different families of the order *Mononegavirales* with vertebrates and humans as their host; (b) Variability in genome size of members of order *Mononegavirales*.

Filoviridae consists of three genera, *Cuevavirus*, *Ebolavirus* and *Marburg*, which are long 'thread-like' (Latin; filum, meaning thread) in structure. The former two viruses cause severe hemorrhagic fevers in humans and non-human primates, leading to high mortality rate. *Ebolavirus* (EBOV) is highly virulent with mortality rate of up to 90% in African epidemics [2]. The *Nyamviridae* is a small family like *Filoviridae*, with a single genus, named Nyavirus and two species *Nyamanini* virus (NYMV) and Midway virus, which were isolated from insects and birds [3]. *Paramyxoviridae* is the largest family and is further divided in two subfamilies: *Paramoxyvirinae* and *Pneumovirinae*. *Paramoxyviridae* is different from other families as it has additional proteins like fusion protein (F), hemagglutinin-neuraminidase glycoprotein (HN), attachment glycoprotein, transmembrane protein, and small hydrophobic protein (SH). *Paramoxyvirinae* includes species of *Avian paramoxyvirus*, which infect a broad range of avian species & is the causative agent of Newcastle disease. Members of the other subfamily, *Pneumovirinae* consist of two nonstructural proteins (NS1 and NS2), which are absent in other NS-NSVs. These genes are located upstream of the nucleoprotein and interfere with the host innate immunity [1]. Lastly, *Rhabdoviridae* includes rabies virus, (RABV) which causes more than 25,000 human deaths per year [4].

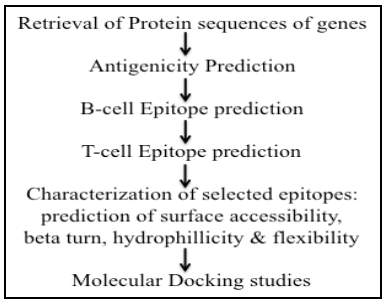
Thus, with such great impact on human and animal health, it becomes imperative to study the antigenicity of these NS-NSVs, with the aim to develop potent vaccines or immunodiagnostic methods against them. With this target, we screened the important genes present in all the ssRNA viral strains of the order *Mononegavirales* in order to detect potent B-cell and T-cell epitopes in them.

2. MATERIALS AND METHODS

2.1. Retrieval of Sequences: The protein sequences of RNA dependent RNA polymerase (RdRp), nucleoprotein, phosphoprotein, glycoprotein and hemagglutinin-neuraminidase genes were retrieved from NCBI Viral Genome

Browser (http://www.ncbi.nlm.nih.gov/genome/viruses/). The sequences were filtered for viral strains with vertebrates and humans as their host.

2.2. Immunoinformatic Predictions: The following methodology was adopted for immunoinformatic predictions:



2.3. Determination of Antigenic Proteins: The antigenicity of all the protein sequences was determined using VaxiJen [5] (v2.0) with threshold value of 0.5.

2.4. Prediction of B-cell and T- Cell epitopes: Each antigenic sequence was further investigated for B-cell epitopes using BCPred [6] using default parameter of 75 % specificity and length of 20 amino acids. The epitopes were then subjected to ProPred-I [7] and ProPred [8] with 1% threshold of highest scoring peptides. ProPred-I allows predicting Class-I MHC binding peptide and ProPred predicts Class II MHC binding peptide. Also, the peptides were predicted using proteasome & immunoproteasome filters at 1% threshold. The common epitopes that bind both the MHC classes were selected for further analysis.

2.5. Prediction of surface accessibility, hydrophillicity, flexibility and beta-turn: was done with the tools provided by Immune Epitope Database –Analysis Resource (IEDB-AR, *tools.iedb.org/main*). The tools included Emini surface accessibility prediction [9], Parker hydrophilicity scale [10], Karplus and Schulz flexibility scale [11] and Chou and Fasman beta-turn prediction tool [12]. All the analyses were done with default parameters using threshold value of 1.0.

2.6. Allergenicity assessment: Allerdictor server [13] was used to predict the allergenicity of the proposed epitopes for vaccine development.

2.7. Protein Sequence Analysis: ProtParam tool [14] was used to compute various physico-chemical properties of hypothetical protein sequence as well as proteins with no putative function.

2.8. Secondary Structure Prediction: The X protein sequences (proteins with no putative function) were subjected to prediction of secondary structures using PSIPRED [15].

2.9. Epitope Modelling: The 3D structures of epitope "FAGDQAAPQ" in nucleoprotein sequence of *Avian paramyxovirus* 7 (NC_025347) and "QPTSPLSTI" in nucleoprotein sequence of *Lloviu cuevavirus* (NC_016144) were predicted by using PEPstr peptide tertiary structure prediction server [16,17].

2.10. Docking: The protein sequence of HLA molecules: HLA-B*5401 and HLA-B*5102 were retrieved from NCBI (Accession No. AAA03686 & AAB07699 respectively) and Phyre2 [18] was used to predict their 3D structure. The epitope receptor docking studies were carried out using the docking server PatchDOCK [19,20] and the results were further refined with FireDock [21, 22]. The results of docking were viewed in NGL Viewer [23] and iMOL (https://www.macupdate.com/app/mac/8474/imol) .

3. RESULTS & DISCUSSION

3.1. Antigenicity and Epitope prediction: Immunoinformatics, as it is termed, aids in prediction of potential B-cell and T-cell epitopes with Class-I and Class-II MHC binding ligands. This strategy has provided useful information in designing novel drugs for multiple sclerosis [24], malaria [25], tuberculosis [26] and Human Immunodeficiency Virus [27].

In our study, we have predicted potential B-cell and T-cell epitopes, from different viral strains/isolates of the order *Mononegavirales*, with the expectation that they will help in designing potent and efficient vaccines or diagnostic tools

against them. In the study, the viral strains with humans & vertebrates as their hosts were selected. Hence, the members of *Nyamviridae* and the unclassified viral isolates, which do not infect humans, were not included in the study.

In order to determine the antigenicity, a total of 95 RdRp, 95 nucleoprotein, 148 glycoprotein, 87 phosphoprotein and 36 hemagglutinin-neuraminidase protein sequences were retrieved from different members of the order *Mononegavirales* and were subjected to VaxiJen (v2.0) server. The antigenic sequences were selected using a threshold value of 0.5. Predictions done at this threshold have higher accuracy, sensitivity and specificity [5]. Out of 95 RdRp sequences, only 3 sequences, which belonged to family *Paramoxyviridae* were found antigenic: *Avian paramyxovirus* 5 (NC_025361), *Fer-de-Lance paramyxovirus* (NC_005084) and *Human parainfluenza virus 4a* (NC_021928). Out of these 3 viral strains, only a single effective epitopic peptide (LRGTRIENQ) that showed binding affinity with Class-I MHC alleles (HLA-B14 & HLA-B*2705) and Class-II MHC allele (HLA-DRB1*0401) was found in *Avian paramyxovirus* 5. The peptide scores of Class-I MHC binding were less than 50.0, but the predicted epitope was found to be non-allergenic in nature. Thus, stable antiviral therapies against *Avian paramyxovirus* 5 can be made, since RdRp is an important cellular component for the transcription and replication of the viral genome. On the other hand, 28 proteins out of 95 nucleoprotein sequences were predicted as probable antigens and 13 proteins were less antigenic, with values slightly less than 0.5. The antigenic nucleoproteins were further predicted for B-cell epitopes. The detected peptides with score >50 have been listed in Table 1. All the predicted epitopes had binding affinity with different Class-I MHC alleles, but only one 9-mer sequence (FAGDQAAPQ) from *Avian paramyxovirus* 7 (NC_025347) had binding affinity with HLA-DRB1*0305 Class-II MHC binding allele.

Table 1: Predicted binders of nucleoproteins for Class-I MHC alleles, with 1% threshold (the high scoring peptides depicting maximum probability of binding have been shown only).

Genome Acc. No.	B cell epitopes	Position	Peptide	Interacting MHC-I allele (peptide score)
NC_025347	FAGDQAAPQANPPAQPRQND	416	FAGDQAAPQ	HLA-B*5401 (153.29)
NC_001906	KESTAQSSSERNPPNNRPQA	489	STAQSSSER	HLA-A68.1 (100.0)
NC_025344	EMEQVVEVYEYAQKQGGEAG	280	EQVVEVYEY	HLA-B62 (80.0)
NC_006428	DPFAGAAGDYQGAAAGGAQG	415	FAGAAGDYQ	HLA-B*5401 (149.23)
NC_005339	APRNNRNRRRRGQPDQGDDD	436	RNNRNRRRR	HLA-A20 Cattle (80.0)
NC_016144	KQPTSPLSTIPEEEGGHEAN	632	KQPTSPLST	HLA-B*2705 (60.0)
			QPTSPLSTI	HLA-B*5101 (484.0), HLA-B*5102 (440.0)
	DHRRNQVKPGRRGNDPRTLP	522	RRNQVKPGR	HLA-A20 Cattle (80.0), HLA-B*2705 (3000.0)
			RNQVKPGRR	HLA-A20 Cattle (80.0)
NC_016144	SRRPAYDWPPGDRPHTTQAT	523	RRPAYDWPP	HLA-B*2705 (60.0)
			RPAYDWPPG	HLA-B*0702 (134.58)

Further, 20 phopsphoprotein, 25 hemagglutinin-neuraminidase and 67 glycoprotein sequences were found to be antigenic. The detected epitopic peptides for antigenic phosphoprotein and hemagglutinin-neuraminidase proteins with score >50 have been listed in Table 2 and 3. Only one B-cell epitopic peptide (YQINNAANN) predicted in hemagglutinin-neuraminidase protein of *Avian paramyxovirus* 9 strain duck/NewYork/22/1978 (NC_025390), was found to bind with HLA-DRB1*0101, HLA-DRB1*0801, HLA-DRB1*1321 Class-II MHC alleles. The epitope was found to be non-allergenic in nature. B-cell epitopes were also detected in antigenic glycoproteins, but they were not found to bind with either Class-I or Class-II MHC alleles.

Table 2: Predicted binders of phosphoproteins for Class-I MHC alleles, with 1% threshold (the high scoring peptides depicting maximum probability of binding have been shown only).

Genome Acc. no.	B cell epitopes	Position	Peptide with Class-I MHC binding	Class-I MHC allele
NC_028362	ERREGQQRREEGEKREREEK	252	ERREGQQRR	HLA-A20 Cattle (80.0), HLA-B*2705 (100.0)
NC_025360	PTIGKTPPATTTAEKVRPNN	505	GKTPPATTT	HLA-A20 Cattle (100.0)
NC_025256	TIEEEGGTAVDIMVNETREG	514	EEGGTAVDI	MHC-Kk (1000.0)
			IEEEGGTAV	MHC-Kk (100.0)
	NIKDPTTAPNKTSIQTKPAR	380	IKDPTTAPN	HLA-A20 Cattle (100.0)
NC_017937	EPKPGVFPEPIDPDKTNAAG	438	KPGVFPEPI	HLA-B*5101 (242.0), HLA-B*5102 (440), HLA-B*0702 (152.23)
NC_007620	GGKDLRHGPSDIGPGAIGGR	151	GKDLRHGPS	HLA-A20 Cattle (100.0)
			DLRHGPSDI	MHC-Kd (57.6)
	GAVVQGGEKGDRDNAKKEVT	57	VQGGEKGDR	HLA-A20 Cattle (80.0), HLA-B*2705 (100.0)
	NSKLIPPVTPKKPPRHKDRI	108	KLIPPVTPK	HLA-A3 (607.5), HLA-B*2705 (90.0)
NC_004148	KQTIKVMDPIEEEEFTEKRV	83	QTIKVMDPI	MHC-Kd (115.2)
NC_001921	VEAERPPREDIQPGPGIRCD	91	AERPPREDI	MHC-Kk (500.0)
NC_009527	SKSTQTTPKKVKPEPPSAPT	143	SKSTQTTPK	HLA-A20 Cattle (1000.0)
			TQTTPKKVK	HLA-B*2705 (60.0)
NC_025344	KLTKPSTPPSTPPPTPPQNK	5	TKPSTPPST	HLA-A20 Cattle (100.0)

Table 3: Predicted binders of hemagglutinin-neuraminidase proteins for Class-I MHC alleles, with 1% threshold (the high scoring peptides depicting maximum probability of binding have been shown only).

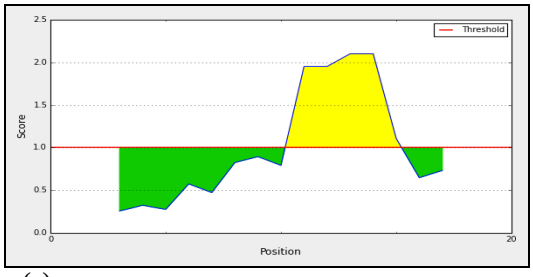
Genome Acc. No.	B cell epitopes	Position	Peptide	Class-I MHC
NC_028362	IPTISASPKIRLIPGPGLLA	165	ISASPKIRL	MHC-Kd (57.6)
NC_025390	SYQINNAANNSGCGAPVHDK	111	YQINNAANN	HLA-B*5401 (154.72)
NC_025349	EHPSFIPGPTTIGGCTRIPT	155	HPSFIPGPT	HLA-B*0702 (133.86)
			SFIPGPTTI	MHC-Kd (2764.8)
NC_001498	AKWAVPTTRTDDKLRMETCF	363	AKWAVPTTR	HLA-A20 Cattle (4000.0)
			WAVPTTRTD	HLA-B*5401 (154.86)
			SKGNCSGPT	HLA-A20 Cattle (100.0)
			FYKDNPHPK	HLA-B*5401 (152.45)
NC_005036	DWVANYPGAGGGSFIDDRVW	294	WVANYPGAG	HLA-B*5401 (151.15)
NC_002161	YWKHTNSINNTNNETETARG	3	WKHTNSINN	HLA-A20 Cattle (100.0), HLA-B*5401 (153.91)
			YWKHTNSIN	HLA-B*5401 (147.52)

A total of 47 MHC Class-I alleles and 51 MHC Class-II alleles were screened using ProPed-I and Propred respectively. Since, both servers cover a maximum number of HLA (Human Leukocyte antigens), therefore, were considered suitable for predicting epitopes. Also, length is an important factor to consider for peptide antigen binding with MHC or TCR or both, since most MHC class I molecules prefer binding to peptides of 9 amino acids in length [28]. The predicted epitopic

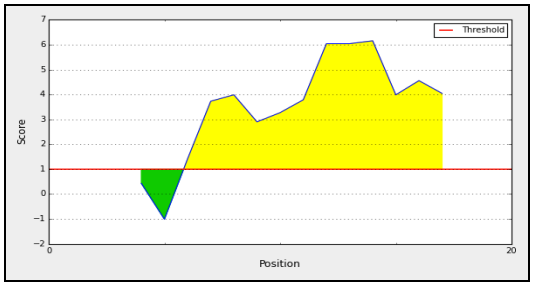
peptides in this study are also of 9 amino acids in length. The prediction of epitopes in all the major proteins was done at the lowest threshold value of 1.0 % (in Proped-I server) as the threshold plays a vital role in determining the stringency of prediction. A lower threshold value increases the stringency of prediction with lower rate of false positives [7].

A B-cell epitope is usually characterized by its antigenicity, hydrophillicity and accessibility in a flexible region of an immunogen [29,30]. The epitope of RdRp sequence of *Avian paramyxovirus* 5 (NC_025361) was found to be hydrophilic with maximum propensity score of 6.157 and minimum score of 0.429 and surface accessible with maximum propensity score of 2.104 and minimum score of 0.254 (Figure-3). On the graphs, the X-axes depicts the residue positions in the sequence while the Y-axes depicts for each residue the correspondent calculated score. The residues that are colored in yellow on the graphs are the residues that have a higher probability to be part of the detected epitope at a threshold of 1.0. For example, a peptide sequence with surface probability greater than 1.0 indicates an increased probability for being found on the surface in the Emini surface accessibility scale.

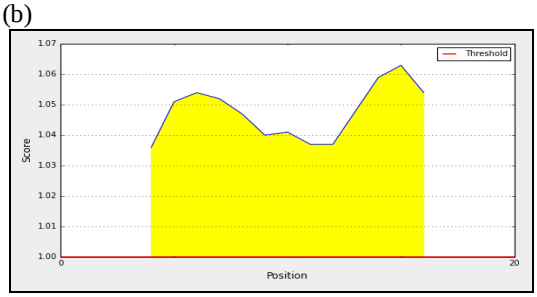
The B-cell epitope predicted for nucleoprotein of *Avian paramyxovirus* 7 (NC_025347) was also examined for its hydrophillicity, surface accessibility, beta turn and flexibility. The epitope qualified all the parameters (Figure-4). The B-cell epitope predicted in hemagglutinin-neuraminidase protein of *Avian paramyxovirus* 9 strain duck/New York/22/1978 (NC_025390) was also found to qualify all the four parameters (data not shown).



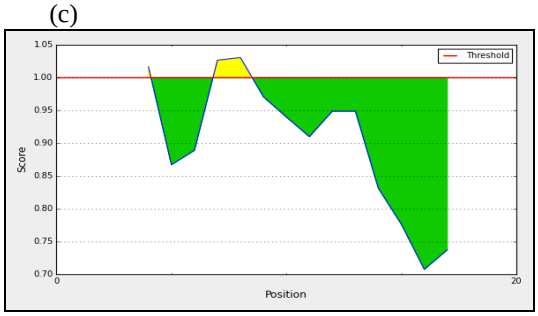
(a)



(b)

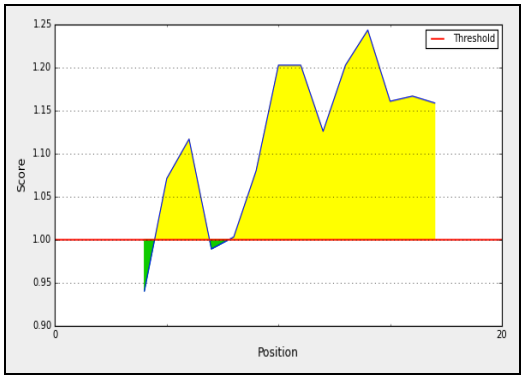


(c)



(d)

Figure 3: Plots for hydrophillicity, surface accessibility, flexibility and beta turn prediction. (a) Hydrophilicity and (b) Surface accessibility of the predicted epitope of RdRp sequence of *Avian paramyxovirus* 5 (NC_025361); (c) Flexibility and (d) Beta turn prediction of the predicted epitope of RdRp sequence of *Avian paramyxovirus* 5 (NC_025361).The hydrophilic residues are in the yellow colored region. The residues below the cut off (red line, 1.00) are in the green region.



(a)

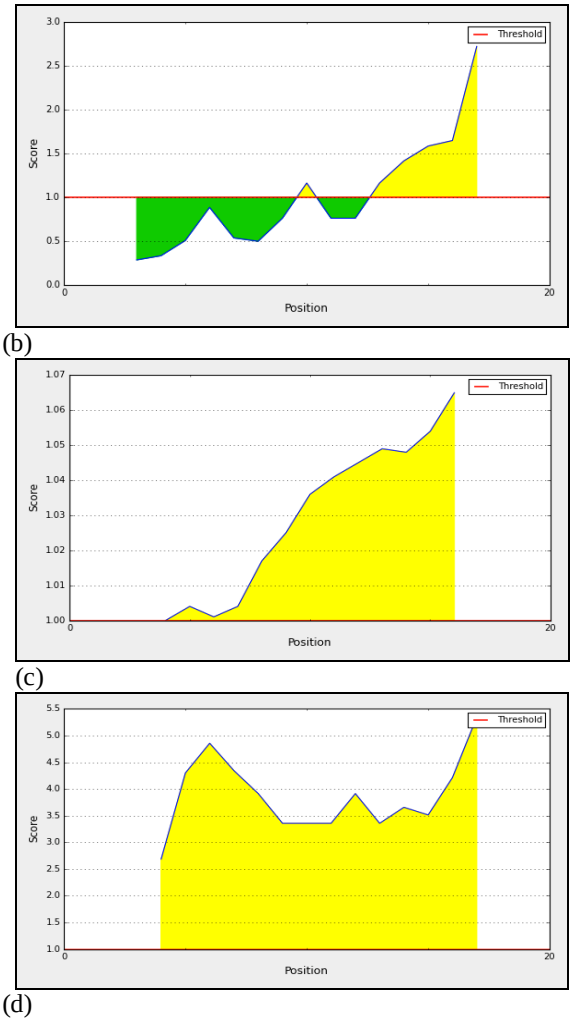


Figure 4: Plots for hydrophillicity, surface accessibility, flexibility and beta turn prediction. (a) Beta turn prediction and (b) Surface accessibility of predicted B-cell epitope of nucleoprotein of *Avian paramyxovirus* 7 (NC_025347). (c) Flexibility prediction and (d) Hydrophillicity prediction of the B-cell epitope of nucleoprotein of *Avian paramyxovirus* 7 (NC_025347).

Assessment of HLA-epitope interaction: To further investigate the binding of peptides predicted from B-cell epitopes of antigenic proteins with their corresponding Class-I MHC alleles, molecular docking study was done. The docking complex structures were selected on the basis of energy functions like global energy and the contribution of atomic contact energy (ACE), hydrogen bonds (HB) and van der Waals interactions (VdW) to the global binding energy (Table 4). Lower global energy specifies higher binding affinities with HLA receptors [23].

The epitope "FAGDQAAPQ" predicted in the nucleoprotein sequence of *Avian paramyxovirus* 7 (NC_025347) can be seen in close association with its corresponding HLA allele (HLA-B*5401; Figure-5). Likewise, the epitopic peptide "QPTSPLSTI" predicted in nucleoprotein sequence of *Lloviu cuevavirus* (NC_016144) can be seen in close association with its corresponding HLA allele (HLA-B*5102; Figure-6).

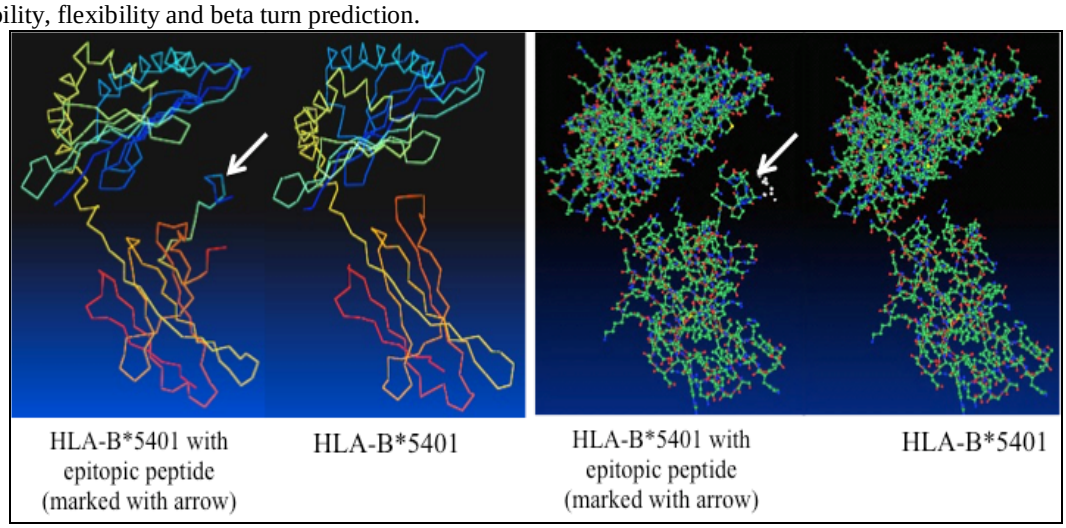
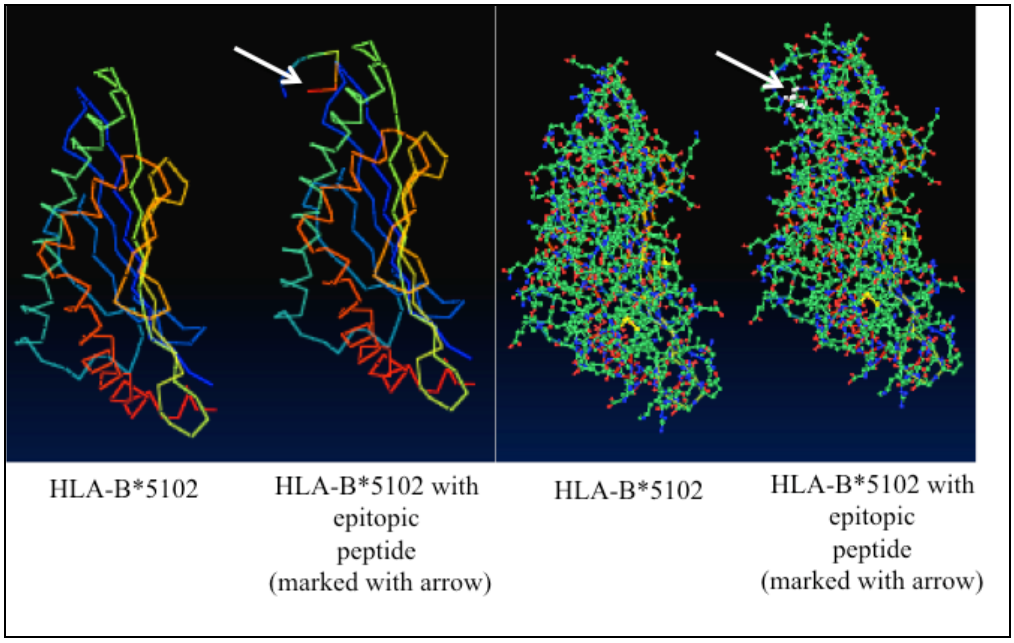


Figure 5: The docked complex of FAGDQAAPQ epitope with HLA-B*5401 (a) wire illustration; (b) ball and stick illustration (as viewed in iMOL)



(a) (b)

Figure 6: The docked complex of QPTSPLSTI epitope with HLA-B*5102, (a) wire illustration; (b) ball and stick illustration (as viewed in iMOL)

Table 4: The details of docking studies

Epitope	Allele	Global energy	Attractive VdW	Repulsive VdW	ACE	HB
FAGDQAAPQ	HLA-B*5401	-30.98	-23.17	7.06	-2.88	-1.87
QPTSPLSTI	HLA-B*5102	-40.29	-22.71	8.63	-7.09	-2.34

X protein analysis: The genome of the members of family *Bornaviridae* consists of a small protein, annotated as 'X protein' (~86-89 amino acids in length) in the NCBI database. It is located downstream to gene N (nucleoprotein) and its function is not known. So, it was subjected to structural and functional investigation about its plausible role in the organism. Primary structure analysis of the X proteins was done by ProtParam tool and molecular weight and theoretical pI were obtained (Table 5). The X protein in *Canary bornavirus 2 and Parrot bornavirus* 2 had the highest molecular weight of 9501.5 Da. Isoelectric points of these proteins ranged 5.38 to 9.18. The X proteins of *Canary bornavirus* 2, *Canary bornavirus* 3 and *Reptile bornavirus* 1 were negatively charged since their isoelectric point was below 7. The instability index classified the proteins to be unstable. The negative Grand average of hydropathicity (GRAVY) of X protein in all the members of *Bornaviridae* was negative, indicating that these proteins were hydrophilic in nature.

A very low degree of sequence similarity was found between X protein sequences of different members of *Bornaviridae* (Table 6). Since the protein might be responsible for antigenicity, hence, it was subjected to VaxiJen (v2.0) analysis-using threshold of 0.5. It was found that all the X proteins were probable antigens. The B-cell epitopes were also detected in the X protein of *Borna disease virus* 1, *Parrot bornavirus* 2 and *Reptile bornavirus* 1. The peptides binding with Class-I MHC alleles were also predicted (Table 7).

The secondary structures for all the X-proteins were predicted using PsiPred (Figure-6). It was found that these proteins shared similar structure, comprising of a helix, strand and coil. The structures of X protein in *Canary bornavirus* 2 and *Reptile bornavirus* 1 consisted of two and three helices respectively, while others had only 1 helix. The detected epitopes were also checked for their surface accessibility, hydrophillicity, flexibility and beta-turn and were found to qualify all the features (data not shown).

The X proteins had low sequence similarity but shared similar structures due to the presence of large number of residues with strongly similar properties, as indicated by ':' (colon) in the sequence alignment (Figure-7). By simply looking at the sequences, the homology among these proteins could not be discerned as the sequence changes more rapidly than three-dimensional structures.

Table 5: Physicochemical properties of X protein in members of *Bornaviridae*

Name	Accession No.	No. of amino acids	M.W	pI	GRAVY	Instability index
Borna disease virus 1	NC_001607	87	9385.5	9.18	-0.948	85.45
Canary bornavirus 2	NC_027892	87	9501.5	5.38	-0.754	69.41
Canary bornavirus 3	NC_024296	87	9498.5	5.81	-0.754	80.73
Parrot bornavirus 2	NC_028106	86	9501.5	9.09	-1.002	60.82
Reptile bornavirus 1	NC_024778	89	9341.1	6.26	-0.807	64.84

Table 6: Results of pairwise sequence alignment using BLAST-P

Sequence similarity between X protein sequences		Alignment score	Query coverage	E value	Identity
Born disease virus 1	*Canary bornavirus2*	98.6	100%	9.00E-33	55%
Canary bornavirus 2	*Canary bornavirus3*	150	100%	4.00E-53	85%
Parrot bornavirus 2	*Reptile bornavirus 1*	35.8	100%	2.00E-09	35%
Parrot bornavirus 2	*Born disease virus 1*	51.6	100%	6.00E-15	41%
Reptile bornavirus 1	*Born disease virus 1*	28.1	100%	2.00E-06	33%
Canary bornavirus 2	*Parrot bornavirus 2*	76.6	100%	2.00E-24	51%
Canary bornavirus 2	*Reptile bornavirus 1*	36.6	100%	1.00E-09	36%

Table 7: Predicted binders of X proteins for Class-I MHC alleles, with 1% threshold (the high scoring peptides depicting maximum probability of binding have been shown only).

Name	Position	B-cell Epitope	Peptide	Class-I MHC allele
Born disease virus 1	25	GRLPGGRRRSPDTTTGTTGV	GRLPGGRRR	HLA-A20 Cattle (80.0), HLA-B*2705 (300)
Parrot bornavirus 2	50	TQGGPNCPRGETSRASQEES	TQGGPNCPR	HLA-A20 Cattle (80.0), HLA-B*2705 (100.0)

Figure 6: Secondary structure of X protein in *Borna disease virus* 1

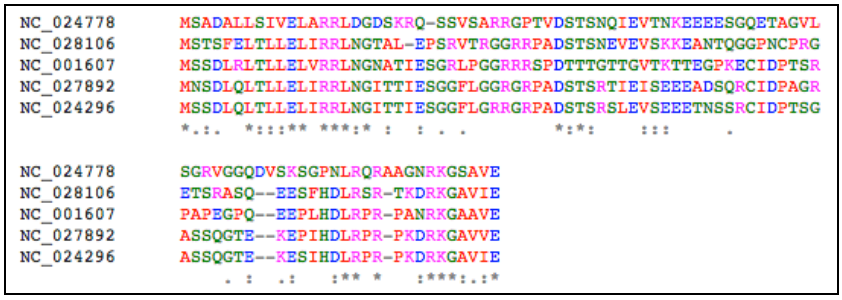


Figure 7: Multiple sequence alignment of X protein sequences of members of *Bornaviridae* done by Clustal omega (http://www.ebi.ac.uk/Tools/msa/clustalo/).

Hypothetical protein analysis: Two hypothetical proteins were present in *Soybean cyst nematode virus* (NC_024702), a member of *Nyamviridae* and three in *Sclerotinia sclerotiorum* negative-stranded RNA virus 1 strain AH98, an unclassified member of the order *Mononegavirales*. So, like the X proteins, the role of these hypothetical proteins was also investigated. The physico-chemical properties of these proteins were studied (Table 8). Upon inspection of their antigenicity, it was found that 4 out of these proteins were antigenic. But, B-cell epitopes were detected only in one hypothetical protein of *Soybean cyst nematode virus* (YP_009052464.1) (Table 9). The predicted epitope qualified the parameters of a potent epitope (data not shown). Thus, even though the exact role of these proteins cannot be deciphered, but they can be easily used in development of diagnostic methods or vaccine against these viruses.

Table 8: Physicochemical properties of hypothetical proteins

Name	Protein id	No. of amino acids	M.W (Da)	pI	GRAVY	Instability index
Soybean cyst nematode virus	YP_009052464.1	281	29413.3	4.72	-0.689	50.26
Sclerotinia sclerotiorum negative-stranded RNA virus 1	YP_009052465.1	80	8998.3	9.64	-0.526	54.33
	YP_009094313.1	242	26432.1	5.58	-0.321	49.99
	YP_009094315.1	269	30559.7	6.21	-0.312	49.59
	YP_009094316.1	62	7037.7	3.64	-0.061	51.94

Table 9: Potential B- cell epitopes in hypothetical proteins and their HLA class allele coverage (the high scoring peptides depicting maximum probability of binding have been shown only).

B cell epitopes	Position	Peptide	Class-I MHC alleles
AYRQANPPPPPPAPSAIDPD	162	AYRQANPPP	HLA B*5401 (152.86)
TTARPPPNPNSLAPPPVVVE	49	TTARPPPNP	HLA-B*0702 (143.79)

CONCLUSION:

In this study, an attempt has been made to predict the antigenicity of the important proteins (RNA-dependent RNA Polymerase, nucleoprotein, hemagglutinin-neuraminidase, phosphoprotein) present in the non-segmented negative strand RNA viruses (NS-NSVs) of the order *Mononegavirales*. A large number of B-cell and T-cell epitopes have been detected in these proteins. The prediction of effective binding of the epitopes with Class-I and/or Class-II MHC alleles was also done as peptide binding to MHC molecules is the key feature in cell-mediated immunity, because it is the peptide–MHC class I complex that can be recognized by the T-cell receptor (TCR) and thereby initiate the immune response. Also, importantly, the predicted MHC binders were filtered based on prediction of proteasome cleavage sites. This is based on observations of previous studies where it has been demonstrated that most of MHC binders having proteasome cleavage site at their C terminus have high potency to become T cell epitopes [31,32]. Moreover, the residues of the selected epitopes were above the threshold value of 1.0 (those residues are colored in yellow on the graphs) for all the tested parameters (hydrophillicity, surface accessibility, beta turn and flexibility) required for a potent epitope. Thus, it can be concluded that the efficacy of the predicted epitopes can observed through substantial data obtained in the present study. Though, wet lab experiments would be required to substantiate our results but it is definitely a step towards focusing the experimental efforts on the subset of potential immunogenic peptides that will increase the probability of success.

ACKNOWLEDGEMENT

The authors acknowledge the financial assistance from Department of Biotechnology (DBT) under Star College Scheme.

REFERENCES

1. Pfaller, C.K, Cattaneo, R, Schnell, M.J, (2015) Reverse genetics of *Mononegavirales*: How they work, new vaccines, and new cancer therapeutics. Virology. Vol- 479-480, pg 331–344.
2. Nakayama, E, Saijo, M, (2013) Animal models for Ebola and Marburg virus infections. Fronters in Microbiology. Vol 4, issue- 267, doi:10.3389/fmicb.2013.00267.
3. Mihindukulasuriya, K.A, Nguyen, N.L, Wu, G, Huang, H.V, da Rosa, A.P, Popov, V.L, et al., (2009) Nyamanini and midway viruses define a novel taxon of RNA viruses in the order *Mononegavirales*. Journal of Virology. Vol-83, issue-10, pg 5109–5116.
4. Lozano, R, Naghavi, M, Foreman, K, Lim, S, Shibuya, K, et al., (2012) Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study. Lancet. Vol- 380, issue-9859, pg 2095–2128.
5. Doytchinova, I.A, Flower, D.R, (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. Vol- 8, issue-4, doi: 10.1186/1471-2105-8-4.
6. EL-Manzalawy, Y, Dobbs, D, Honavar, V, (2008) Predicting flexible length linear B-cell epitopes. In: 7th International Conference on Computational Systems Bioinformatics, 26–29 Stanford, California, United States of America.

7. Singh, H, Raghava, G.P.S, (2003) ProPred1: Prediction of promiscuous MHC class-I binding sites. Bioinformatics. Vol- 19, no.-8, pg 1009-14.
8. Singh, H, Raghava, G.P, (2001) ProPred: prediction of HLA-DR binding sites. Bioinformatics. Vol-17, no.-12, pg 1236-7.
9. Emini, E.A, Hughes, J.V, Perlow, D.S, Boger, J, (1985) Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. Journal of Virology. Vol- 55, issue-3, pg 836-839.
10. Parker, J.M, Guo, D, Hodges, R.S, (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. Vol- 25, issue-19 pg 5425-32.
11. Karplus, P.A, Schulz, G.E, (1985) Prediction of Chain Flexibility in Proteins. A tool for the selection of Peptide Antigens. Naturwissenschafren. Vol- 72, pg 212-3.
12. Chou, P.Y, Fasman, G.D, (1978) Prediction of the secondary structure of proteins from their amino acid sequence. Advances in Enzymology and Related Areas of Molecular Biology. Vol- 47, pg 45–148.
13. Dang, H.X, Lawrence, C.B, (2014) Allerdictor: fast allergen prediction using text classification techniques. Bioinformatics. Vol- 30, issue-8, pg 1120-1128.
14. Gasteiger, E, Hoogland, C, Gattiker, A, Duvaud, S, Wilkins, M.R, et al, (2005) Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press, pg 571-607.
15. Buchan, D.W.A, Minneci, F, Nugent, T.C.O, Bryson K, Jones DT, (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acid Research. Vol-41, (Web Server issue), W349–W357.
16. Kaur, H, Garg, A, Raghava, G.P.S, (2007) PEPstr: A de novo method for tertiary structure prediction of small bioactive peptides. Protein Peptide Letters. Vol-14, issue-7, pg 626-31.
17. Singh, S, Singh, H, Tuknait, A, Chaudhary, K, Singh, B, Kumaran, S, Raghava, G.P.S, (2015) PEPstrMOD: structure prediction of peptides containing natural, non-natural and modified residues. Biology Direct. Vol-10, issue-73, doi: 10.1186/s13062-015-0103-4.
18. Kelley, L.A, Mezulis, S, Yates, C.M, Wass, M.J.E, (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols. Vol- 10, issue-6, pg 845-858.
19. Duhovny, D, Nussinov, R, Wolfson, H.J, (2002) Efficient Unbound Docking of Rigid Molecules. In Gusfield et al., Ed.", Proceedings of the 2'nd Workshop on Algorithms in Bioinformatics (WABI) Rome, Italy, Lecture Notes in Computer Science 2452, 185-200.
20. Schneidman-Duhovny, D, Inbar, Y, Polak, V, Shatsky, M, Halperin, I, Benyamini, H, et al, (2003) Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking. Proteins. Vol-52, issue-1, pg 107-12.
21. Andrusier, N, Nussinov, R, Wolfson, H.J, (2007) FireDock: Fast Interaction Refinement in Molecular Docking. Proteins. Vol- 69, issue-1, pg 139-159.
22. Mashiach, E, Schneidman-Duhovny, D, Andrusier, N, Nussinov, R, Wolfson, H.J, (2008) FireDock: a web server for fast interaction refinement in molecular docking. Nucleic Acids Research. Vol-36 (Web Server issue): W229-32.
23. Rose, A.S, Hildebrand, P.W, (2015) NGL Viewer: a web application for molecular visualization. Nucleic Acids Research. Vol-43, doi: 10.1093/nar/gkv402.
24. Bourdette, D.N, Edmonds, E, Smith, C. et al, (2015) A highly immunogenic trivalent T cell receptor peptide vaccine for multiple sclerosis. Multiple Sclerosis. Vol-11, issue-5, pg 552–561.
25. López, J.A, Weilenman, C, Audran, R. et al, (2001) A synthetic malaria vaccine elicits a

potent CD8+ and CD4+ T lymphocyte immune response in humans. Implications for vaccination strategies. European Journal of Immunology. Vol- 31, issue-7, pg 1989–1998.

26. Robinson, H.L, Amara, R.R, (2005) T cell vaccines for microbial infections. Nature Medicine. Vol-11, suppl-4, pg 25–32.
27. Wilson, C.C, McKinney, D, Anders, M. et al, (2003) Development of a DNA vaccine designed to induce cytotoxic T lymphocyte responses to multiple conserved epitopes in HIV-1. Journal of Immunology. Vol-171, pg 5611–5623.
28. Lundegaard, C, Lund, O and Nielsen, M, (2012), Predictions versus high-throughput experiments in T-cell epitope discovery: competition or synergy?. Expert Review of Vaccines. Vol-11, issue-1, pg 43–54.
29. Yasmin, T, (2015) Prediction of B and T cell epitope-based peptide vaccines from highly conserved regions in *Enterovirus* D68 capsid protein VP1: a computational approach. Journal of Innovations in Pharmaceuticals and Biological Sciences. Vol-2, issue-4, pg 426-439.
30. Fieser, T.M, John, A, Tainer, H, et al, (1987) Influence of protein flexibility and peptide conformation on reactivity of monoclonal anti-peptide antibodies with a protein a-helix. PNAS. Vol-84, issue-23, pg 8568–72.
31. Ayyoub, M, Stevanovic, S, Sahin, U, Guillaume, P, Servis, C, Rimoldi, D, Valmori, D, Romero, P, Cerottini, J.C, Rammensee, H.G, et al, (2002) Proteasome-assisted identification of a SSX-2-derived epitope recognized by tumorreactive CTL infiltrating metastatic melanoma. Journal of Immunology. Vol-168, issue-4, pg 1717–1722.
32. Kessler, J.H, Beekman, N.J, Bres-Vloemans, S.A, Verdijk, P, vanVeelen, P.A, Kloosterman-Joosten A.M, Vissers, D.C.J, ten Bosch, G.J.A, Kester, M.G.D, Sijts, A, Drijfhout, J.W, et al, (2001) Efficient identification of novel HLA-A*0201- presented cytotoxic T lymphocyte epitopes in the widely expressed tumor antigen PRAME by proteasome-mediated digestion analysis. Journal of Experimental Medicine Vol-193, pg 73–88.