

## IN SILICO PREDICTION OF EPITOPE-BASED PEPTIDES FROM PROTEOME OF NIPAH VIRUS

\*

B. R. PANDEY<sup>1</sup>, SUGANDHA SINGH<sup>2</sup>, SHIPRA SRIVASTAVA<sup>3</sup>, NIDHI SHARMA

<sup>1</sup> Sky Institute, Lucknow, India, <sup>2</sup> Research scholar, Sai Nath University, Ranchi, India, <sup>3</sup> Dr. Rawat Institute of Bioinformatics, Lucknow, India

\* Address for Correspondence : Dr. B. R. Pandey, Director ( Research ), Sky Institute, Lucknow, India ,  
Email ID : drbrpandey@gmail.com

### ABSTRACT

*Prediction and modeling of T-cell epitopes of Nipah virus antigenic proteins nucleocapsid, phosphoprotein, matrix, fusion, glycoprotein, L protein, W protein, V protein and C protein followed by the binding simulation studies of predicted highest binding scorers with their corresponding MHC class I alleles were done in this study. ProPred1 tool was used to predict the promiscuous MHC class I epitopes of viral proteins. 3D structures of epitopes were built with the help of PEPstr server. Molecular dynamics simulation studies were performed through the NAMD graphical user interface embedded in visual molecular dynamics. Epitopes IRTIAAYPL and NPTAVPFTL of Matrix Protein and W-protein have lowest binding energy and highest score with HLA-B\*2705 and HLA-B\*5101 MHC class I allele, respectively. These predicted peptides are highly potential to induce T-cell-mediated immune response and are expected to be useful in designing epitope-based vaccines against Nipah virus.*

**Keywords:** Nipah virus, Encephalitis, ribavirin, Immunoinformatics, Nucleocapsid, Phosphoprotein, Matrix, Fusion, Glycoprotein, L protein, W protein, V protein and C protein, T-cell epitopes, Peptide- or Epitope-based Vaccine, NCBI, ProPred I, Histocompatibility Complex (MHC) Molecules, Molecular docking.

### INTRODUCTION

Nipah virus (NiV), of the family Paramyxoviridae<sup>[1,2]</sup> and the genus Henipavirus, is a zoonotic virus that causes outbreaks of fatal encephalitis in humans<sup>[3]</sup>. The human Nipah virus (NiV) infection was first recognized in a large outbreak of 276 reported cases in peninsular Malaysia and Singapore from September 1998 through May 1999<sup>[4, 5, 6, 7, 8]</sup>. The virus was first isolated from a patient from Sungai Nipah village in Malaysia and the name 'Nipah' was first introduced according to the name of that village. It was also identified in India for the first time in 2001 and second time in 2007. Unfortunately, eleven outbreaks have already occurred in Bangladesh since the first detection of NiV in 2001, with high mortality rate an estimated 80% in an average and 100% in some cases<sup>[9]</sup>. The most alarming fact is that almost every year in winter (December to March), the deadly NiV strikes in the north and western regions of Bangladesh. Infection with Nipah virus is associated with encephalitis (inflammation of the brain). After exposure and an incubation period

of 5 to 14 days, illness presents with 3-14 days of fever and headache, followed by drowsiness, disorientation and mental confusion<sup>[10]</sup>. These signs and symptoms can progress to coma within 24-48 hours. Some patients have a respiratory illness during the early part of their infections, and half of the patients showing severe neurological signs showed also pulmonary signs.

During the first NiV outbreak, the virus infected both pigs and humans, in addition to a small number of cats, dogs and horses<sup>[5, 11]</sup>. NiV possesses a negative-sense, non-segmented RNA genome that is 18246 nt (Malaysian isolate) or 18252 nt (Bangladesh isolate) in length<sup>[12]</sup>. In Bangladesh, 135 probable or confirmed cases of NiV infection in humans were identified from 2001 to 2008; 98 (73%) were fatal<sup>[7]</sup>. There is no effective treatment and vaccine for Nipah virus disease, but ribavirin may mitigate the symptoms of nausea, vomiting, and convulsions.

Vaccination is the most effective of all the medical interventions to save human and animal lives and to increase production<sup>[13,14]</sup>. Compared

to the conventional vaccines, peptide- or epitope-based vaccines are easy to produce, more specific, cost effective, less time consuming and also safe<sup>[15]</sup>. It is well established that T cells play a critical role in inducing cellular immune response against foreign antigens but they recognize antigenic fragments only when they are associated with major histocompatibility complex (MHC) molecules exposed on surface of all vertebrate cells<sup>[16,17]</sup>. Immunoinformatics approach uses computational algorithms to predict potential vaccine candidates or T-cell epitopes. The advantage of a peptide- or epitope-based vaccine is the ability to deliver high doses of the potential immunogen at a low cost<sup>[18, 19]</sup>. Viral protein which could act as a vaccine candidate must be surface-exposed, antigenic and responsible for pathogenicity<sup>[20,21]</sup>.

## MATERIALS AND METHODS

The amino acid sequence of Nucleocapsid, phosphoprotein, matrix, fusion, glycoprotein, L protein, W protein, V protein and C protein was retrieved from the protein sequence database of NCBI (<http://www.ncbi.nlm.nih.gov/protein>) and their accession number is shown in Table 1.

### Prediction of MHC class I binding peptides

An online web tool ProPred I<sup>[22]</sup> has been

used for the prediction of promiscuous MHC class I binding peptides. It uses matrix-based method that allows the prediction of MHC-binding sites in an antigenic sequence for MHC class I alleles, and also allows the prediction of the standard proteasome and immunoproteasome cleavage sites in an antigenic sequence. The simultaneous prediction of MHC binders and proteasome cleavage sites in an antigenic sequence leads to the identification of potential T-cell epitopes.

### Structure-based modeling of T-cell epitopes

3D structures of identified potential T-cell epitopes were predicted by PEPstr (peptide tertiary structure prediction server) server<sup>[23]</sup>. The prediction strategy is based on the realization that b-turn is an important and consistent feature of small peptides in addition to regular structures. Thus, the methods use both the b-turns information predicted from Beta Turns and regular secondary structure information predicted from PSIPRED. The side-chain angles are placed using standard backbone-dependent rotamer library. The structure is further refined with energy minimization and molecular dynamic simulations using Amber version 6.

**Table 1: Scores generated by ProPred for MHC class I alleles.**

Protein	Accession No	Length of amino acids	Start position	Epitope	Allele	ProPred score
Nucleocapsid	ACT32611	532	474	SLLNLR SRL	HLA-A*0201	3729.239
			446	KREMSIS SL	HLA-B*2705	6000.000
			474	SLLNLR SRL	HLA-A2	5968.882
Phosphoprotein	ACT32612	709	168	DRETDLVHL	HLA-B*2705	2000.000
			624	EPYGA AVQL	HLA-B*5101	572.000
Matrix	ACT32613	352	85	IRTIAAYPL	HLA-B*2705	6000.000
			201	IAFNLLVYL	HLA-A*0201	4702.218
			293	FQKNLCFSL	HLA-B*2705	3000
Fusion	ACT32614	546	192	KQTELSLDL	HLA-B*2705	9000
			125	AQITAGVAL	HLA-B*2705	2000
Glycoprotein	ACT32615	602	45	ILSAFNTVI	HLA-A*0201	2489.047
			247	RIIGVGEVL	HLA-B*2705	2000
			45	ILSAFNTVI	HLA-A2	3901.211
L-Protein	ACT32616	2244	2087	VLLQAGLKL	HLA-A2	3475.964
			1567	TMVDLLSDL	HLA-A*0201	3804.077
			1564	GRHTMVDLL	HLA-B*2705	30000.000
			1537	DPELFALYL	HLA-B*5101	1522.664
W-protein	YP_007188592	449	316	KEEPPQKRL	HLA-B*2705	3000.000
			186	NPTAVPFTL	HLA-B*5101	880.000
V-protein	NP_112023	456	186	NPTAVPFTL	HLA-B*2705	2000
			66	DGDVERRNL	HLA-B*5101	520
			316	KEEPPQKRL	HLA-B*2705	2000
C-protein	NP_112024	166	40	FCSAPVENL	HLA-B*2705	3000.000
			116	PDMDLLQAL	HLA-B*2705	2000
			0	MMASILLTL	HLA-B*2705	2000

### 3D structure of MHC I alleles

Information and 3D structure of selected HLA alleles were retrieved from IMGT/HLA database (<http://www.ebi.ac.uk/ipd/imgt/hla/intro.html>)<sup>[24]</sup>. The Database currently contains 10,103 allele sequences. In addition to the physical sequences, the database contains detailed information concerning the material from which the sequence was derived and data on the validation of the sequences. PDB codes of relevant MHC class I alleles were shown in table 2.

**Table 2: List of class I MHC alleles considered in this study for prediction of binding peptides.**

S. No.	Allele	Crystal Structure (PDB ID)
1	HLA-A2	1AKJ
2	HLA-A*0201	1AKJ
3	HLA-B*2705	1HSA
4	HLA-B*5101	1E27

### Molecular docking

AutoDock 4.2<sup>[25, 26]</sup> has been used for In-silico docking of peptides and alleles structure. Gasteiger charges were added to the ligand and maximum six numbers of active torsion are given to the lead compound using AutoDock tool (<http://autodock.scripps.edu/resources/adt>). Kollman charges and salvation term were added to the protein structure using AutoDock tool. The Grid for docking calculation was centered to cover the protein-binding site residues and accommodate ligand to move freely. During the docking procedure, a Lamarckian genetic algorithm (LGA) was used for flexible peptide and rigid protein docking calculation. Docking parameters were as follows: 30 docking trials, population size of 150, maximum number of energy evaluation ranges of 250,000, maximum number of generations is 27,000, mutation rate of 0.02, cross-over rate of 0.8, other docking parameters were set to the software's default values.

### Molecular dynamics simulation of epitope and HLA allele complex

Molecular dynamics simulation was done using the NAMD graphical interface module<sup>[27]</sup> incorporated visual molecular dynamics (VMD 1.9.2)<sup>[28]</sup>. A protein structure file (psf) was created from the initial pdb and topology files. The psfgen package of VMD is used to create this. To create a psf file a pgn file is made, which is the target of psfgen. After running psfgen, two new files were generated protein pdb and protein psf and by accessing PSF and PDB files; NAMD generated the trajectory DCD file. Root mean square deviation (RMSD) of the complex was completed using rmsd tcl source file from the Tk console and finally rmsd .dat was saved and accessed in Microsoft office excel.

### RESULTS AND DISCUSSION

#### Prediction and analysis of MHC class I binding peptides

The Matrix protein peptide IRTIAAYPL at position 85–93 showed ProPred score of 6000 with HLA-B\*2705 MHC I allele. The W-protein peptide NPTAVPFTL at position 186–194 showed ProPred score of 880 with the HLA-B\*5101 allele. ProPred scores of peptides with MHC I alleles are shown in Table 1.

#### Docking energy determination by AutoDock

AutoDock binding simulation studies showed that IRTIAAYPL epitope of Matrix protein with HLA-B\*2705 allele as well as NPTAVPFTL epitope of W-protein with HLA-B\*5101 allele formed stable HLA-peptide complexes with the energy minimization values of -4.03, and -4.14 kcal/mol, respectively (Table 3). After docking studies, we determined the number of H bonds present in the stable complex formed. Using AutoDock, it was found that two H-bond were present in Matrix protein IRTIAAYPL-HLA-B\*2705 complex (fig. 1) via residue GLY1 and LYS243 of allele HLA-B\*2705 with epitope IRTIAAYPL. No hydrogen bond was present in W protein peptide NPTAVPFTL- HLA-B\*5101 complex (fig. 2).

**Table 3: Docking result of epitopes with Allele structures.**

Protein	Epitope	Allele	BE	IME	IE	TorE	VdwE	EE
Nucleocapsid	SLLNLR SRL	HLA*0201	-0.57	-11.61	-6.7	11.04	-10.74	-0.87
	KREMSISSL	HLA-B*2705	-0.72	-12.05	-5.25	11.34	-9.77	-2.29
	SLLNLR SRL	HLA-A2	-1.45	-12.49	-4.94	11.04	-11.75	-0.74
Phosphoprotein	DRETDLVHL	HLA-B*2705	0.01	-10.73	-6.03	10.74	-10.69	-0.04
	EPYGAAVQL	HLA-B*5101	-2.96	-11.02	-6.84	8.05	-10.79	-2.23
Matrix	IRTIAAYPL	HLA-B*2705	-4.03	-12.68	-6.47	8.65	10.74	-1.94
	IAFNLLVYL	HLA-A*0201	-1.88	-11.42	-4.85	9.55	-10.82	-0.6
	FQKNLCFSL	HLA-B*2705	0.36	-10.38	-4.64	10.74	-9.12	-1.26
Fusion	KQTELSLDL	HLA-B*2705	2.99	-8.04	-4.27	11.04	-7.74	-0.3
	AQITAGVAL	HLA-B*2705	-0.61	-8.37	-3.82	7.76	-8.32	-0.04
Glycoprotein	ILSAFNTVI	HLA-A*0201	-2.28	-11.22	-6.9	8.95	-11.71	0.49
	RIIGVGEVL	HLA-B*2705	1.1	-8.44	-3.43	9.55	-7.99	-0.45
	ILSAFNTVI	HLA-A2	-0.86	-9.81	-4.73	8.95	-9.23	-0.58
L protein	VLLQAGLKL	HLA-A2	-1.41	-11.26	-6.49	9.84	-9.53	-1.73
	TMVDLLSDL	HLA-A*0201	-0.23	-10.08	-6.47	9.84	-10.65	0.57
	GRHTMVDLL	HLA-B*2705	0.65	-9.49	-6.91	10.14	-8.65	0.84
	DPELFALYL	HLA-B*5101	-2.92	-12.17	-8.04	9.25	-10.74	-1.43
W protein	KEEPPQKRL	HLA-B*2705	1.7	-9.64	-6.1	11.34	-7.44	-2.2
	NPTAVPFTL	HLA-B*5101	-4.14	-11.3	-7.46	7.16	-10.76	-0.54
V protein	NPTAVPFTL	HLA-B*2705	-2.53	-9.69	-4.23	7.16	-8.89	-0.8
	DGDVERRNL	HLA-B*5101	3.05	-7.99	-4.52	11.04	-7.44	-0.55
	KEEPPQKRL	HLA-B*2705	-0.11	-11.45	-3.99	11.34	-9.43	-2.02
C protein	FCSAPVENL	HLA-B*2705	-2.51	-10.87	-7.44	8.35	-6.23	0.45
	PDMDLLQAL	HLA-B*2705	1.86	-7.98	-5.07	9.84	-6.28	-1.7
	MMA SILLTL	HLA-B*2705	-1.54	-11.38	-3.75	9.84	-11.43	0.15

BE: Binding Energy; IME: Intermolecular Energy; IE: Internal Energy; TorE: Torsional Energy; VdwE: Vdw-lbDesolv Energy; EE: Electrostatic Energy.



**Fig. 1: Docked Matrix protein peptide IRTIAAYPL - HLA-B\*2705 allele complex depicting position of amino acids along with the formation of two Hydrogen bonds (shows as dotted line) with GLY1 and LYS243 residues of Protein 1HSA**

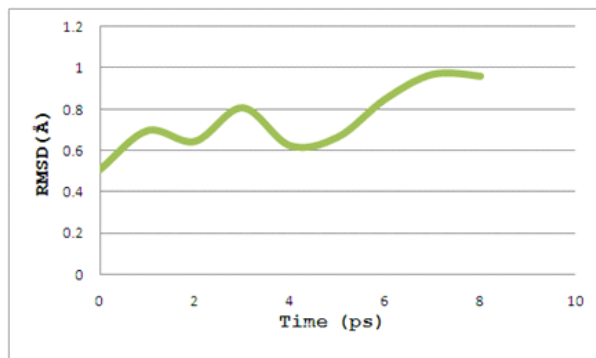


**Fig. 2: Docked W protein peptide NPTAVPFTL-HLA-B\*5101 allele (PDB Id: 1E27) complex**

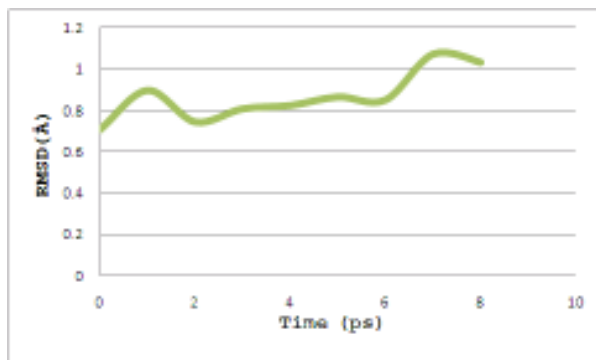
### Molecular dynamics simulation of peptide-allele Complex through NAMD

The peptide-allele complexes formed by AutoDock were subjected to molecular dynamics

simulation and RMSD. Matrix protein epitope IRTIAAYPL - HLA-B\*2705 allele complex displayed the highest peak at RMSD value of 0.96 Å (Fig. 3). W protein peptide NPTAVPFTL - HLA-B\*5101 allele complex resulted in highest peak at RMSD value of 1.03 Å (Fig. 4). The current study incorporates immunoinformatics approach for reducing the time consumed in the long array of experiments to avoid hit and trial sets.



**Fig. 3: Graph displaying molecular dynamic simulation of Matrix protein peptide–allele complex, resulted in highest peak at 0.96 Å**



**Fig. 4: Graph displaying molecular dynamic simulation of W protein peptide–allele complex, resulted in highest peak at 1.03 Å**

The molecular dynamics simulation showed that complex formed between a peptide and allele was attaining proper stability by creating a parallelism in RMSD over a time window. The mentioned peptides can be either isolated or formulated for further *in vitro* and *in vivo* testing.

## CONCLUSION

From the present study it is concluded that the epitopes IRTIAAYPL and NPTAVPFTL of Matrix and W protein, respectively, have considerable binding with HLA-B\*2705 and HLA-

B\*5101 class I allele and low-energy minimization values providing stability to the peptide–MHC complex. These peptide constructs may further be undergone wet laboratory studies for the development of targeted vaccine against Nipah virus.

## REFERENCES

1. Aljofan M. Hendra and Nipah infection: emerging paramyxoviruses. *Virus Res.* 2013; 177:119–26.
2. Field H.E., Mackenzie J.S., Daszak P. Henipaviruses: emerging paramyxoviruses associated with fruit bats. *Curr Top Microbiol Immunol.* 2007; 315:133–59.
3. Chong H.T., Hossain M.J., Tan C.T. Differences in epidemiologic and clinical features of Nipah virus encephalitis between the Malaysian and Bangladesh outbreaks. *Neurology Asia.* 2008; 13:23-26.
4. Chong H.T., Bakar S.A., Tan (Edts.) C.T. Proceedings of the Nipah virus colloquium 2008, University of Malaya and related articles, *Neurology Asia.* 2009; 14:45-47.
5. Chua K.B., Bellini W.J., Rota P.A. et al. Nipah virus: A recently emergent deadly paramyxovirus. *Science.* 2000; 288(5470):1432-1435.
6. Chua K.B. Nipah virus outbreak in Malaysia. *J. Clin. Virol.* 2003; 26(3):265-275.
7. Luby S.P., Gurley E.S., Hossain M.J. Transmission of human infection with Nipah virus, *Clin. Infect. Dis.* 2009; 49(11):1743-1748.
8. Paton N.I., Leo Y.S., Zaki S.R. et al., Outbreak of Nipah-virus infection among abattoir workers in Singapore. *Lancet.* 1999; 354(9186):1253-1256.
9. Biswas M.H.A. Optimal control of Nipah virus (NiV) infections: a bangladesh scenario. *Journal of Pure and Applied Mathematics: Advances and Applications.* 2014; 12(1):77-104.
10. Centers for Disease Control and Prevention (CDC), March 20, 2014.
11. Epstein J.H., Abdul Rahman S., Zambriski J.A., Halpin K., Meehan G. et al. Feral cats and risk for Nipah virus transmission. *Emerg Infect Dis.* 2006; 12:1178–1179.
12. Harcourt B.H., Lowe L., Tamin A., Liu X., Bankamp B. et al. Genetic characterization of Nipah virus, Bangladesh, 2004. *Emerg Infect Dis.* 2005; 11:1594–1597.
13. Horzinek M.C. Vaccination: A philosophical view. *Adv. Vet. Med.* 1999; 41: 1-6.
14. Tang H., Liu X.S., Fang Y.Z., Pan L., Zhang Z.W. et al. The epitopes of foot and mouth disease. *Asian J. Anim. Vet. Adv.* 2012a; 7: 1261-1265.

15. Kumar A., Jain A., Verma S.K. Screening and Structure-based Modeling of T-Cell Epitopes of Marburg Virus NP, GP and VP40: An Immunoinformatic Approach for Designing Peptide-based Vaccine. Trends in Bioinformatics. 2013; 6: 10-16.
16. Shekhar C., Dev K., Verma S.K., Kumar A. In-silico: Screening and modeling of CTL binding epitopes of crimeancongo hemorrhagic fever virus. Trends Bioinformatics. 2012; 5:14-24.
17. Mohabatkar H., Mohammadzadegan R. Computational comparison of T-cell epitopes of gp120 of Iranian HIV-1 with different subtypes of the virus. Pak. J. Biol. Sci. 2007; 10: 4295-4298.
18. Von Hoff D.D., Evans D.B., Hruban R.H. (eds). Pancreatic cancer. Jones and Bartlett, Sudbury, MA. 2005.
19. Tang H., Liu X.S., Fang Y.Z., Pan L., Zhang Z.W. et al. Advances in studies on vaccines of foot-and-mouth disease. Asian J. Anim. Vet. Adv. 2012; 7: 1245-1254.
20. Cerdino-Tarraga A.M., Efstratiou A., Dover L.G., Holden M.T.G., Pallen M. et al. The complete genome sequence and analysis of *Corynebacterium diphtheria* NCTC13129. Nucl. Acids Res. 2003; 31: 6516-6523.
21. Verma S.K., Yadav S.S.P., Kumar A. In silico T cell antigenic determinants from proteome of H1N2 swine influenza A virus. Online J. Bioinformatics. 2011; 12: 371-378.
22. Singh H., Raghava G.P.S. ProPred: Prediction of HLA-DR binding sites. Bioinformatics. 2001; 17(12), 1236-37.
23. Kaur H., Garg A., Raghava G.P.S. PEPstr: A de novo method for tertiary structure prediction of small bioactive peptides. Protein Pept Lett. 2007; 14:626-30.
24. Robinson J., Halliwell J.A., McWilliam H., Lopez R., Parham P., Marsh S.G.E. The IMGT/HLA Database. Nucleic Acids Research. 2013; 41:D1222-7.
25. Goodsell D.S., Olson A.J. Automated docking of substrates to proteins by simulated annealing. Proteins. 1990; 8(3):195-202.
26. Morris G.M., Goodsell D.S., Halliday R.S., Huey R., Hart W.E., Belew R.K, Olson A.J. Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function. J. Comput. Chem. 1998; 19:1639-1662.
27. James C.P., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R.D., Kale L., Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem. 2005; 26:1781–1802.
28. Humphrey W., Dalke A., Schulten K. VMD – Visual Molecular Dynamics. J. Molec. Graphics. 1996; 14.1, 33-38.