

A Computational Approach for Designing a Universal Epitope-Based Peptide Vaccine Against Nipah Virus

Mohammad Tuhin Ali¹ · Mohammed Monzur Morshed¹ · Faizule Hassan²

Received: 14 March 2014 / Revised: 27 January 2015 / Accepted: 9 April 2015 / Published online: 9 July 2015
© International Association of Scientists in the Interdisciplinary Areas and Springer-Verlag Berlin Heidelberg 2015

Abstract Nipah virus (NiV) is highly pathogenic single-stranded negative sense RNA virus. It can cause severe encephalitis and respiratory disease in humans. In addition, NiV infects a large range of host including mammals. As a result of its higher zoonotic potential and pathogenicity for human, it has been rated as an alert in recent days. A therapeutic treatment or vaccines has become elusive to fight against this virus. In this study, the attachment (G) and fusion (F) glycoproteins of NiV, responsible for the viral attachment and entry to the host cell, were selected to develop epitope-based vaccine against Nipah virus. Epitopes were identified from the conserved region of G and F protein of NiV. Both B-cell and T-cell immunity were checked to affirm it that these epitopes will be able to induce humoral and cellular immunity. A total of 6 T-cell epitopes and 19 significant HLA–epitope interactions were identified. Eventually it has shown an acceptable percentage in population coverage (46.45 %) and efficient binding with HLA molecule by molecular docking study.

Keywords Epitope-based vaccine · HLA ligand · Population coverage · Zoonotic virus

Electronic supplementary material The online version of this article (doi:10.1007/s12539-015-0023-0) contains supplementary material, which is available to authorized users.

✉ Faizule Hassan
hassanf@miamioh.edu

¹ Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka 1000, Bangladesh

² Department of Chemistry and Biochemistry, Miami University, 107 Hughes Laboratory, Oxford, OH 45056, USA

1 Introduction

The initial outbreak of human Nipah virus (NiV) infection caused 276 patients in Malaysia and Singapore in 1998–1999 [1, 2]. Recurrent Nipah outbreaks have been also reported since 2001 in India and Bangladesh [3–5]. The febrile encephalitis with about 40 % mortality rate was displayed in the large Malaysian outbreaks. In the more recent Bangladeshi and Indian outbreaks, an increased rate of respiratory disease and a potential human to human transmission were prominent. The latter one had the mortality rate about 75–92 % [3, 6–8]. Fruit bats of the genus *Pteropus* were recognized as natural reservoir of NiV [9]. NiV is closely related to Hendra virus and represents Henipavirus in the paramyxovirus family [10]. Unlike most of the paramyxoviruses, henipaviruses cause diseases in mammals, e.g. pigs, cats, horses, humans [11, 12] and are considered as biosafety level 4 (BSL-4) organisms. Its potential to cross the species boundaries like from bats to domestic animals and humans causing fatal infection appears to be a consistent feature of henipaviruses. In addition, it is also reported that NiV transmission occurs from human to human [7, 13]. According to CDC (centers for disease control and prevention), treatment is limited to supportive care. Approved vaccine or drugs have still remained unavailable for human use. As of now, a subunit vaccine targeting Hendra G protein has been recently used in Australia to protect horses from the infection of Hendra virus (HeV) (<http://www.cdc.gov/vhf/nipah/prevention/index.html>).

NiV has two major envelope glycoproteins which are attachment protein G and fusion protein F, responsible for the viral attachment and entry to host cell, respectively. Initial attachment is mediated by the N-glycans of G protein, whereas F protein is mainly responsible for the viral

fusion with host cell [14, 15]. Prior to this, G protein triggers a conformational change in F protein to mediate the viral fusion by an unknown mechanism [16]. In this study, we aimed to use both G and F proteins for designing epitope-based peptide vaccine against NiV on the basis of the available knowledge on immunity to other paramyxoviruses [17–19]. Both G and F proteins are the major targets of the neutralizing antibodies and vaccine-induced protection of this virus [20, 21]. An experimental protective efficacy study in pigs using canary pox virus vector has suggested that together they are the most potent candidate of immunogens [18].

In spite of serious health complications, the scientific knowledge of the epidemiology and ecology of this virus is limited to design an universal vaccine. Bioinformatics, especially immunoinformatics, is an emerging field in vaccine design. The combination of experimental and *in silico* methods is crucial to solve complex problems such as revealing immune responses and vaccine design [22]. Available bioinformatics tools provide the searching option to scan for probable epitope candidate from huge sets of protein antigens which are encoded by complete viral genomes. This computational vaccine design approach has proven very effective in combating diseases such as multiple sclerosis [23], malaria [24], and tumors [25]. The most critical step in any computational vaccine designing approach is the identification of HLA ligands and T-cell epitopes [26]. Through T-cell epitope prediction tools for the identification of allele-specific binding peptides, it is also possible to reduce the number of potential peptides considered as vaccine candidates.

2 Methods

2.1 Retrieval of Protein Sequences

NCBI (National Center for Biotechnology Information) (<http://www.ncbi.nlm.nih.gov/>) and PATRIC (pathosystems resource integration center) (<http://www.viprbrc.org/>) database were explored to retrieve the required sequences of G and F proteins of NiV. Different Nipah endemic countries like Bangladesh, India, Malaysia, and Singapore were considered to collect sequences.

2.2 Multiple Sequence Alignment

Clustal Omega software can align virtually any number of sequences quickly and accurately. In this study, EBI-Clustal Omega program (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) was used to perform multiple sequence alignment for the retrieved sequences [27]. The sequence alignment

was the basis to find conserved regions from G and F proteins of NiV.

2.3 Immunogenicity of Conserved Peptide

Immune Epitope Database (IEDB) has different methods of T-cell epitope prediction for the purpose of assessing immunogenicity of a peptide. To identify T-cell epitopes from the conserved peptide, NetCTL prediction method was utilized (<http://tools.immuneepitope.org/stools/netchop/netchop.do>) [28, 29]. To set the sensitivity and specificity levels at 0.7 and 0.985, respectively, the threshold was set at 1.0. The IEDB MHC class-I binding prediction tool (<http://tools.immuneepitope.org/mhci/>) was used to identify MHC class-I alleles for the final set of CTL peptides. Through there were several methods to predict the MHC class-I alleles, stabilized matrix method (SMM) was used in this study [30]. All the available alleles were selected prior to prediction. The specified length of epitope was 9.0.

TAP score, proteasomal score, processing score, and MHC-I binding score for each epitope were evaluated on the basis of SMM method available from the IEDB “Proteasomal cleavage/TAP transport/MHC class-I combined prediction” tool (<http://tools.immuneepitope.org/processing/>) [30–32]. Conservancy of the finally selected epitopes within the sequences of G and F proteins was determined through the epitope conservancy analysis tool [33].

2.4 Assessment of HLA–Epitope Interaction by Molecular Docking Study

Protein Data Bank (PDB) database was explored to retrieve the 3D structure of MHC class-I molecule (HLA-C*12:03; PDB ID: 2FSE). PEPstr peptide tertiary structure prediction server (<http://www.imtech.res.in/raghava/pepstr/>) has designing facility for 3D structures of epitopes, and this facility was utilized to design the 3D structure of the best epitope ITFISFIIV [34].

The docking study was executed through the AutoDock Tools (ADT) and AutoDock Vina (Vina) from MGL software packages (version 1.5.6) [35, 36]. PDBQT files of both the protein (HLA-C*12:03) and ligand (ITFISFIIV) were generated prior to docking study. The parameter of center grid box was set at 22.474, 5.416, 34.437 Å in *x*, *y*, and *z* axis respectively, whereas the size was set at 40, 38 and 52 Å in *x*, *y*, and *z* directions respectively. All of the above parameters were set at 1.0 Å spacing. Vina was employed for conducting the final docking experiment. To run a control experiment along with immune dominant determinant of human type-II collagen (PDB ID: 2FSE) was used as an epitope with all the above parameter unchanged.

2.5 Population Coverage

Population coverage is used to see whether the final set of epitopes and their HLA alleles are feasible for the significant percentage of people or not for the endemic regions. The IEDB population coverage calculation tool (http://tools.immuneepitope.org/tools/population/iedb_input) was used to calculate population coverage for the final set of epitopes and their alleles [37].

2.6 Prediction of B-Cell Epitope

B-cell epitope prediction tool (<http://tools.immuneepitope.org/bcell/>) was utilized to predict the B-cell antigenicity of conserved peptide on the basis of the Kolaskar and Tongaonkar [38] method which has the ability to predict antigenic determinants with approximately 75 % accuracy.

3 Results and Discussion

3.1 Multiple Sequence Alignment

A total of 18 G protein sequences and 15 F protein sequences from different NiV strains from different geographic regions were retrieved from NCBI and PATRIC databases to perform multiple sequence alignment. A peptide region of 503–544 amino acids was found to be remained conserve in all given G protein sequences (Supplementary Fig. 1). For F protein, 461–546 amino acids region remained conserved in all given sequences

(Supplementary Fig. 2). Both the conserved peptides were present on the C-terminal region of their respective protein.

3.2 Immunogenicity of Conserved Peptides

For the conserved region of G protein, NetCTL prediction tool did not predict any potential T-cell epitopes. So, further analysis was not performed for this peptide. Results produced from the NetCTL prediction tool for the conserved region of F protein are depicted in Fig. 1. NetCTL tool predicted 78 overlapping CTL peptides. But only six of them were finally selected on the basis of a total score ≥ 1 (Table 1). MHC-I binding prediction tool provided 481 possible MHC-I allele interactions with the six finally selected CTL peptides. Criteria like IC_{50} value ≤ 100 was set to determine the potential CTL peptide-MHC-I allele interaction. It resulted in the final set that includes only 19 MHC class-I alleles which is given in Table 2. To become a successful candidate for designing peptide vaccine, potent binding interaction with MHC-I allele is must for a

Table 1 CTL epitopes predicted by NetCTL prediction tool

Sequence of epitope	Overall score
PTSSGDLYY	4.5960
IVEKKRNTY	2.1352
QSLQSKDY	1.1371
RPTSSGDLY	0.6185
SMLSMIILY	0.5963
ITFISFIIV	0.5122

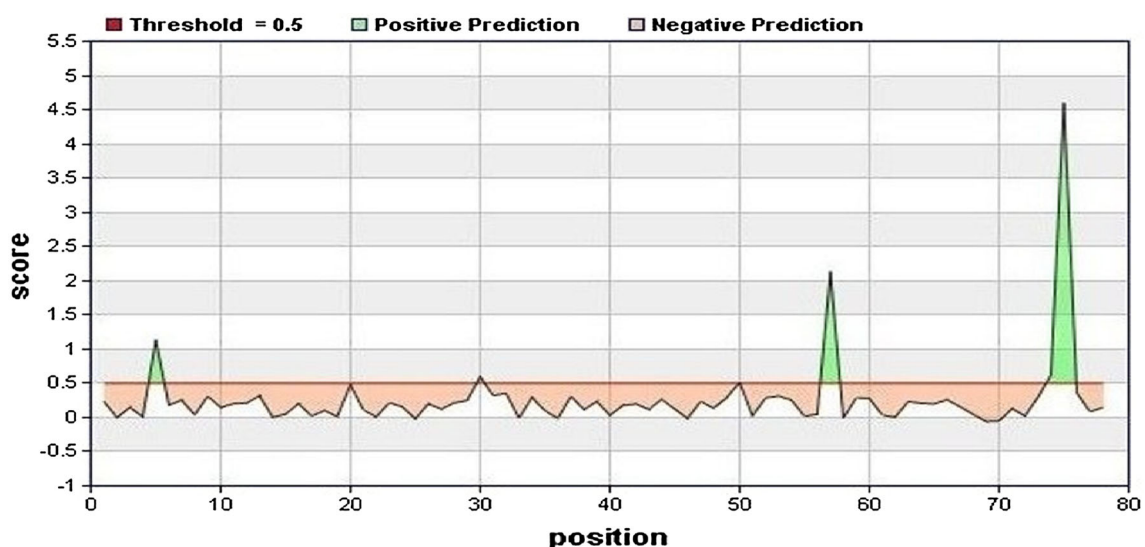
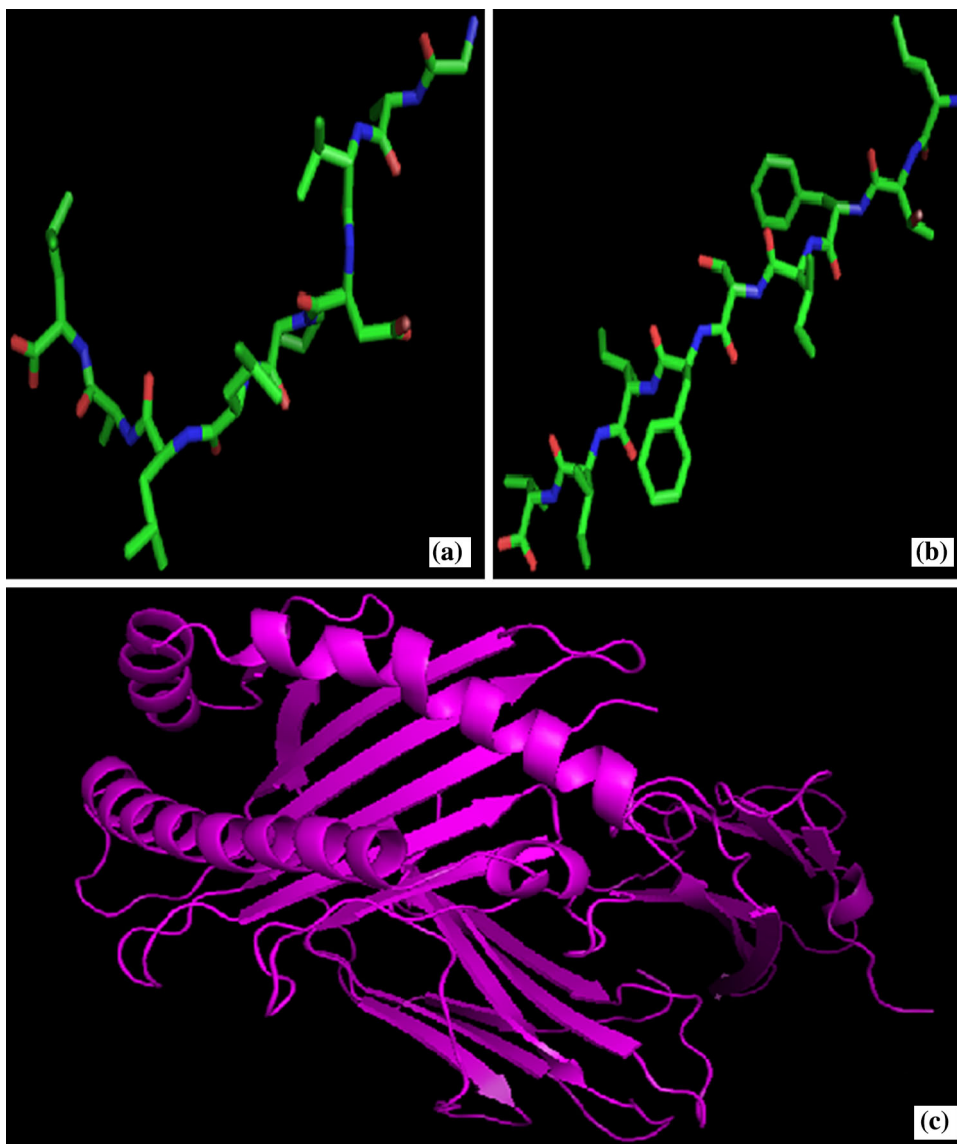


Fig. 1 T-cell epitopes predicted by NetCTL prediction tool. While most of the potential epitopes were failed to cross the threshold level (0.50), sharp peaks represent the epitopes that cross threshold level. X-axis represents the amino acids, whereas Y-axis represents the score

Table 2 Epitopes and their HLA class-I alleles

Epitope	HLA alleles
PTSSGDLYY	HLA-A*68:23, HLA-A*32:07, HLA-A*80:01, HLA-A*29:02, HLA-C*12:03, HLA-C*03:03, HLA-A*01:01, HLA-B*27:20
IVEKRRNTY	HLA-A*68:23, HLA-C*12:03, HLA-A*32:07, HLA-C*03:03, HLA-B*27:20, HLA-C*05:01
QSLQQSKDY	HLA-A*32:07, HLA-C*03:03, HLA-B*27:20, HLA-B*15:17, HLA-A*68:23, HLA-C*12:03, HLA-B*40:13, HLA-A*32:15
RPTSSGDLY	HLA-A*32:07, HLA-A*68:23, HLA-B*27:20, HLA-C*12:03, HLA-B*40:13, HLA-C*03:03, HLA-B*35:01
SMLSMIILY	HLA-A*32:07, HLA-A*29:02, HLA-A*68:23, HLA-B*27:20, HLA-B*15:03, HLA-B*40:13, HLA-A*80:01, HLA-C*12:03
ITFISFIIV	HLA-A*68:23, HLA-C*12:03, HLA-A*68:02, HLA-A*32:07, HLA-B*40:13, HLA-B*27:20, HLA-A*32:15, HLA-A*02:11, HLA-A*02:06, HLA-A*02:50, HLA-A*02:16

Fig. 2 Three dimensional (3D) structures of both experimental and predicted epitopes and MHC Class-I molecule. **(a)** 3D Structure of experimental epitope (immunodominant determinant of human type-II collagen (PDB ID: 2FSE). **(b)** Predicted 3D structure of designed epitope ITFISFIIV. **(c)** Visualization of the crystal structure of MHC class-I molecule HLA-C 12:03 (PDB ID: 2FSE) using PyMOL



T-cell epitope. The obtained results have demonstrated that all the finally chosen T-cell epitope has significant interactions with MHC-I alleles (Fig. 2).

Additionally, proteasomal score, TAP score, MHC-I binding score, and processing score were also considered to determine the immunogenicity of final epitopes

(Supplementary Table 1). In our study, the above parameters were found in favor of the six final epitopes. The six T-cell epitopes were also found to have 100 % sequence identity with the given F proteins which is also a necessary prerequisite to become potential epitope candidate in designing peptide vaccine. The result of epitope conservancy analysis is summarized in Table 3.

3.3 Evaluation of HLA–Epitope Interactions

A great deal of interactions with MHC-I alleles increases the chance of a peptide to become the best candidate for vaccine designing. Among the six final T-cell epitopes, a 9-mer ITFISFIIV epitope was found to have the highest interactions with MHC-I alleles. Therefore, we used it in molecular docking study to further validate its potentiality to interact with MHC-I allele. The 3D structures of experimental epitope, ITFISFIIV epitope, and MHC class-I molecule (HLA-C*12:03) are presented in Fig. 2 which were used in docking study. In Vina docking protocol, binding energy expressed in kcal/mol plays the major role for predicting best conformer of epitope ITFISFIIV.

Table 3 Conservancy of epitopes in source sequences

Epitope no.	Epitope sequence	Sequence identity ≥ 100 %
1	PTSSGDLYY	100.00
2	IVEKKRNTY	100.00
3	QSLQSKDY	100.00
4	RPTSSGDLY	100.00
5	SMLSMILY	100.00
6	ITFISFIIV	100.00

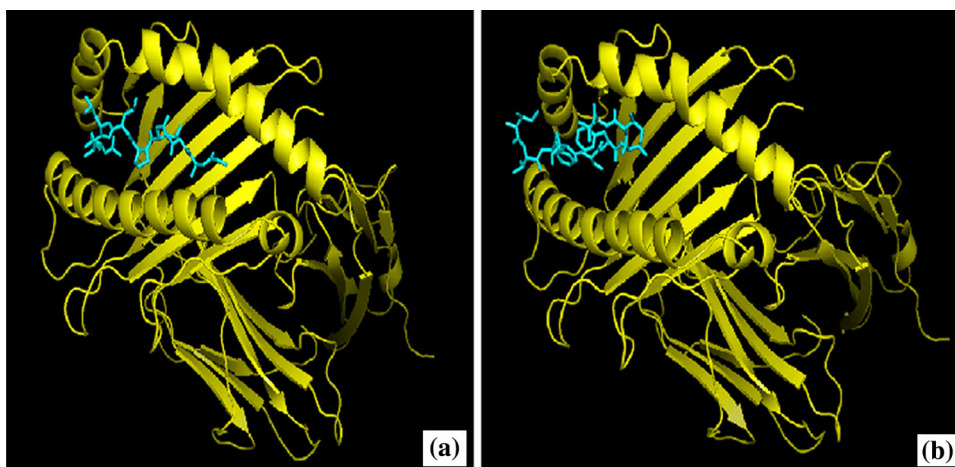
The binding energy of epitope ITFISFIIV with the binding groove of HLA-C*12:03 was found to be -7.1 kcal/mol which was compared with the binding energy of experimental epitope (-7.0 kcal/mol) to HLA-C*12:03 and found to be similar to each other. This almost similar binding energy of both the simulations indicates the satisfactory accuracy of the predicted epitope to form interaction with MHC-I allele. Figure 3 represents a comparative analysis at binding of both designed and experimental epitopes.

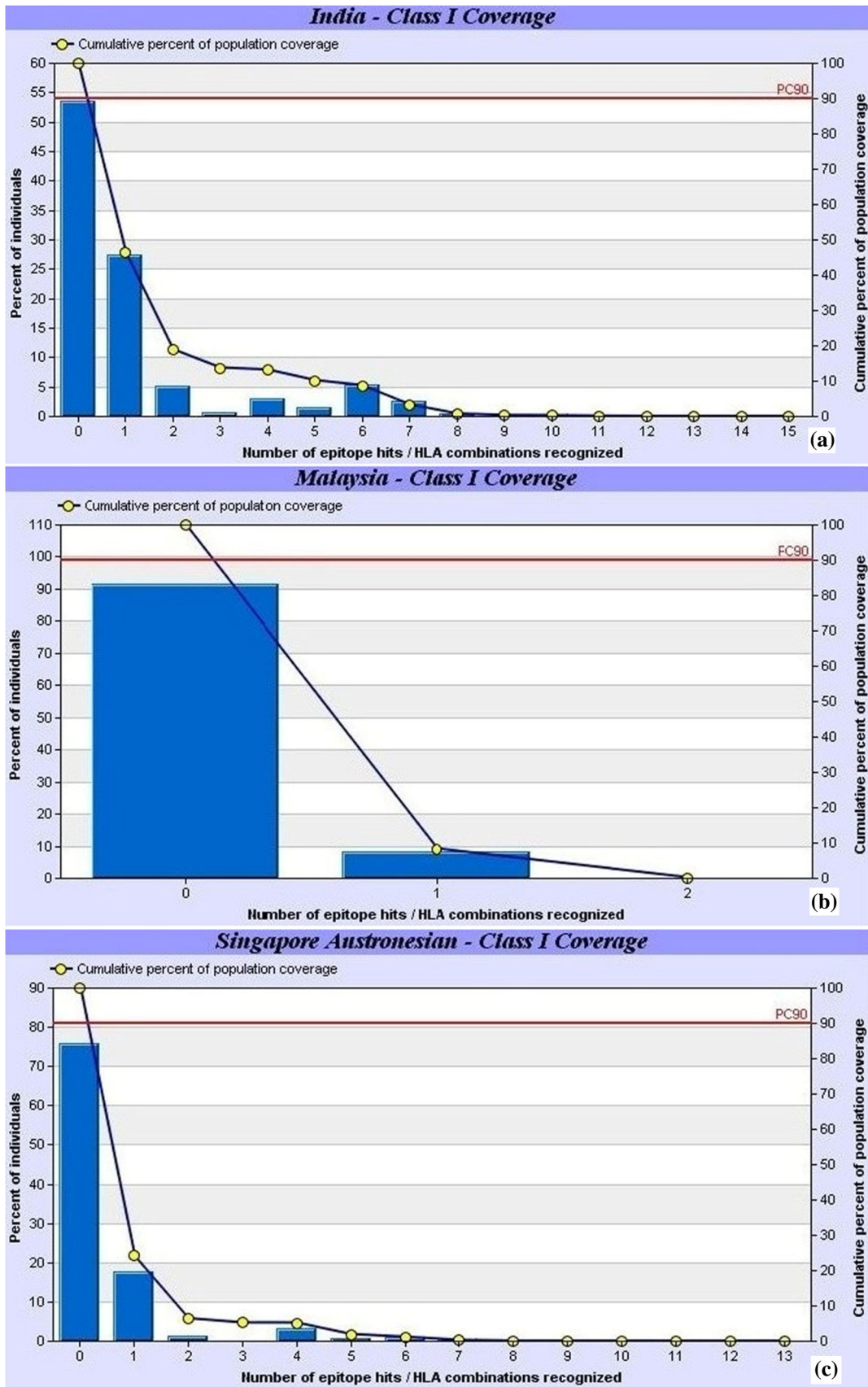
3.4 Population Coverage

Population from different genetic backgrounds has different MHC allele frequencies. Therefore, for the efficient vaccine development one should consider the candidate epitope which would be able to interact with HLA alleles prevailing in the affected region where the designed vaccine would be implemented. In this study, we performed population coverage study for our predicted T-cell epitopes and their respective MHC-I alleles in a preliminary manner. We selected only three countries while predicting population coverage, namely: India, Malaysia, and Singapore. The following parameters were computed by population coverage tool: (1) projected population coverage (2) number of epitope hits/HLA combination recognized by the populations, and (3) minimum number of epitope hits/HLA combinations recognized by 90 % of the population (PC90). The highest population coverage was found in India 46.5 % (Fig. 4a), and the lowest one was in Malaysia 8.37 % (Fig. 4b). In Singapore population coverage was 24.13 % (Fig. 4c). Although in our study we did not get considerable population coverage for Malaysia and Singapore, 46.5 % population coverage from India for the predicted T-cell epitopes and their respective MHC-I

Fig. 3 Binding of both the designed and experimental epitope at the binding groove of HLA-C*12:30.

(a) Experimental epitope bound at the binding groove of the HLA-C 12:03. (b) Designed epitope bound at the binding groove of HLA-C 12:03. *Stick* represents the epitopes, whereas *cartoon* represents the HLA molecule





◀**Fig. 4** Population coverage based on MHC restriction data. In the graph, the *line* (-o-) represents the cumulative percentage of the population coverage of the epitopes. The *individual bar* represents the population coverage for each epitope. (a) The highest population coverage 46.45 % was seen for India. (b) Population coverage in Malaysia was 8.37 %, and (c) for Singapore population coverage was 24.13 %

alleles is remarkable, further validating the efficacy of our proposed epitope vaccine.

3.5 B-Cell Antigenicity

The results of IEDB B-cell epitope prediction toll is illustrated in Figs. 5 and 6. A total of two B-cell epitopes were identified from the conserved peptide of F protein, whereas three potential B-cell epitopes were found in the conserved peptide of G protein (Table 4). All residues over mean antigenicity are considered as potentially antigenic. From the result of F protein antigenicity prediction, it was observed that the MHC-I restricted epitope ITFISFIIV superimposed with one of the predicted B-cell epitopes also. Therefore, we considered it as the best epitope candidate from NiV F protein for the development of epitope-

based peptide vaccine against this virus which can trigger both the cell mediated and humoral immunity in host body. Although it was mentioned earlier that no T-cell epitope was identified from the conserved region of G protein, but three of B-cell epitopes were found from this peptide. This concludes that the conserved peptide of G protein could be used to design vaccine targeting B-cell-mediated immune response.

4 Conclusions

Designing epitope-based peptide vaccine has the growing interest for the viral vaccination due to the recent advances in protein data and sequencing technologies. It is advantageous to work with *in silico* findings before attempting laboratory trial to some extent. In this study, the results suggest that it is possible to design a universal peptide vaccine to prevent all the strains of NiV. Additionally, they may assist to develop unique detection method of NiV. We suggest further *in vitro* and *in vivo* studies to determine the actual potency of identified epitopes to stimulate immune response.

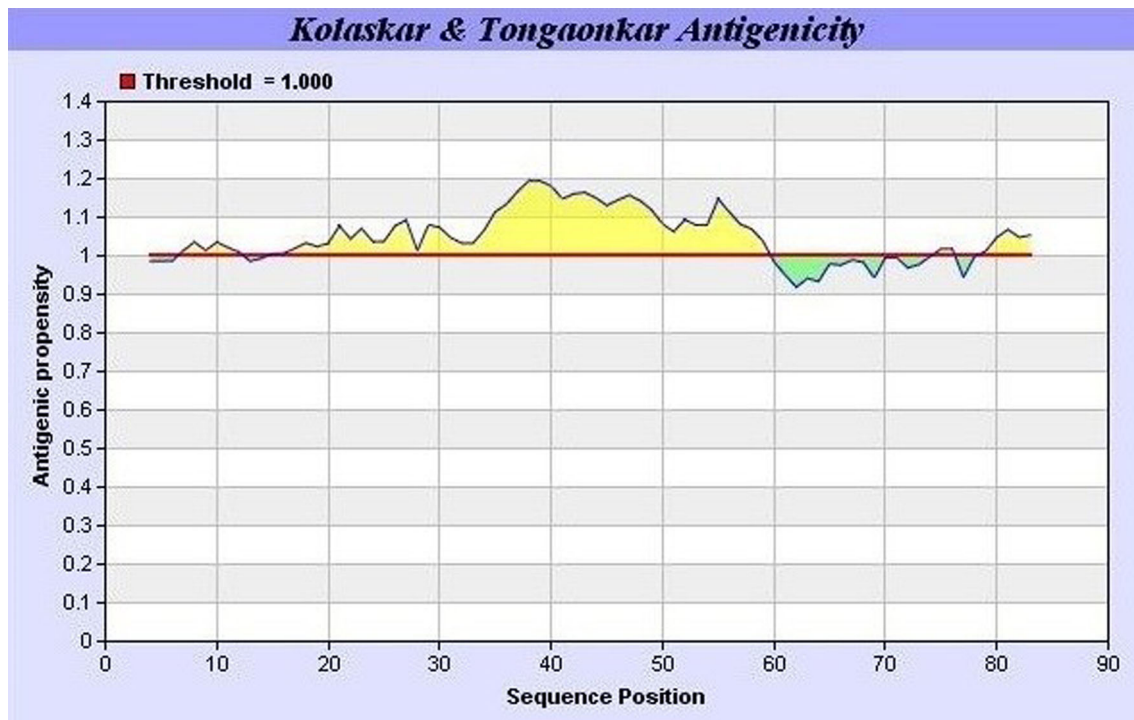


Fig. 5 B-cell antigenicity for the conserved region of F protein. X-axis represents the amino acids, whereas Y-axis represents antigenic propensity. Average antigenic score was 1.048. Area displayed by the *threshold line* is considered as potential B-cell antigenic regions

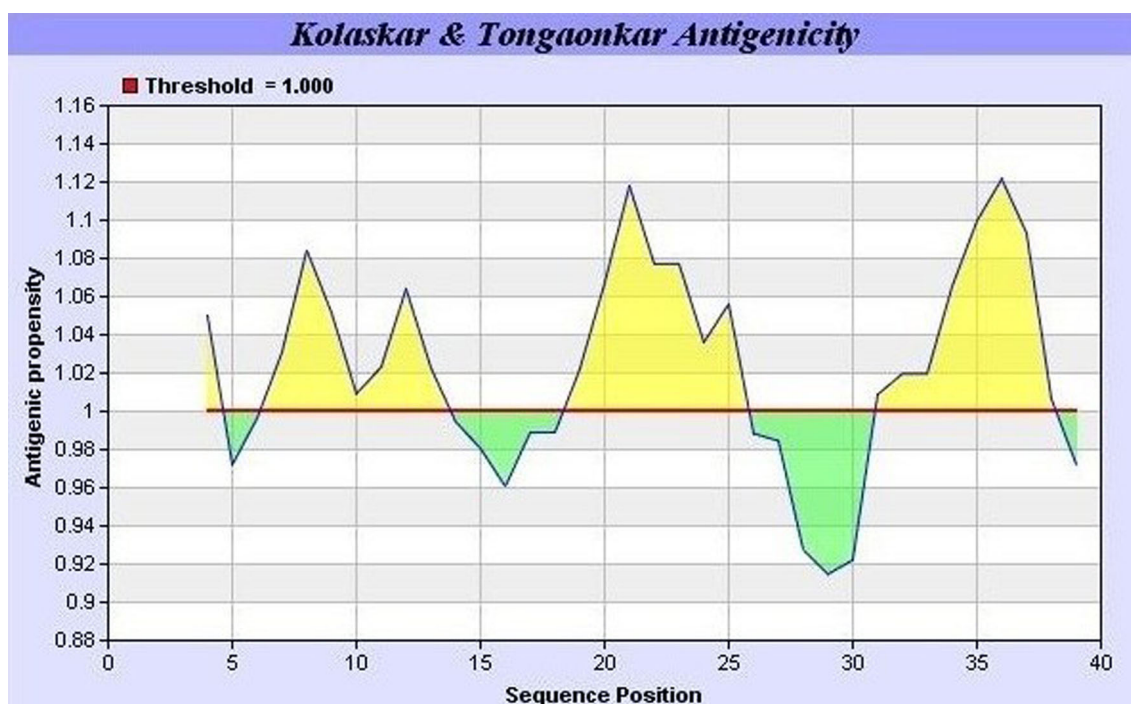


Fig. 6 B-cell antigenic propensity for the conserved region of G protein. X-axis represents the amino acids, whereas Y-axis represents the antigenic propensity. Average antigenic propensity was 1.023.

Area displayed by the *threshold line* is considered as potential B-cell antigenic regions

Table 4 B-cell epitopes from the conserved regions of F and G proteins

No. of epitopes	Sequence of epitopes	Length
F protein		
1	LQQSKD	6
2	KEAQRLLDTVNP SLISLSMILY VLSIASLCIGLITFISFIIVE	45
G protein		
1	NDAFLID	7
2	SAGVFLD	7
3	ENPVFTVF	8

Acknowledgments No financial support was taken for this research. No writing assistance was utilized in the production of this manuscript.

Conflict of interest No reported conflicts. The results presented in this paper has not been published or presented in any meeting or conferences or scientific publications.

References

- Paton NI, Leo YS, Zaki SR, Auchus AP, Lee KE, Ling AE, Chew SK, Ang B, Rollin PE, Umapathi T (1999) Outbreak of Nipah-virus infection among abattoir workers in Singapore. *Lancet* 354:1253–1256
- Chua KB (2003) Nipah virus outbreak in Malaysia. *J Clin Virol* 26:265–275
- Chadha MS, Comer JA, Lowe L, Rota PA, Rollin PE, Bellini WJ, Ksiazek TG, Mishra AC (2006) Nipah virus-associated encephalitis outbreak, Siliguri, India. *Emerg Infect Dis* 12:235–240
- Luby SP, Hossain MJ, Gurley ES, Ahmed B-N, Banu S, Khan SU, Homaira N, Rota PA, Rollin PE, Comer JA (2009) Recurrent zoonotic transmission of Nipah virus into humans, Bangladesh, 2001–2007. *Emerg Infect Dis* 15:1229
- Arankalle VA, Bandyopadhyay BT, Ramdasi AY, Jadi R, Patil DR, Rahman M, Majumdar M, Banerjee PS, Hati AK, Goswami RP (2011) Genomic characterization of nipah virus, west bengal, India. *Emerg Infect Dis* 17:907–909
- Chua KB, Lam SK, Tan CT, Hooi PS, Goh KJ, Chew NK, Tan KS, Kamarulzaman A, Wong KT (2000) High mortality in Nipah encephalitis is associated with presence of virus in cerebrospinal fluid. *Ann Neurol* 48:802–805
- Gurley ES, Montgomery JM, Hossain MJ, Bell M, Azad AK, Islam MR, Molla MAR, Carroll DS, Ksiazek TG, Rota PA (2007) Person-to-person transmission of Nipah virus in a Bangladeshi community. *Emerg Infect Dis* 13:1031–1037
- Hossain MJ, Gurley ES, Montgomery JM, Bell M, Carroll DS, Hsu VP, Formenty P, Croisier A, Bertherat E, Faiz M (2008) Clinical presentation of Nipah virus infection in Bangladesh. *Clin Infect Dis* 46:977–984
- Yob JM, Field H, Rashdi AM, Morrissy C, van der Heide B, Rota P, bin Adzhar A, White J, Daniels P, Jamaluddin A (2001) Nipah virus infection in bats (order Chiroptera) in peninsular Malaysia. *Emerg Infect Dis* 7:439–441
- Wang L-F, Harcourt BH, Yu M, Tamin A, Rota PA, Bellini WJ, Eaton BT (2001) Molecular biology of Hendra and Nipah viruses. *Microbes Infect* 3:279–287
- Williamson M, Hooper P, Selleck P, Gleeson L, Daniels P, Westbury H, Murray P (1998) Transmission studies of Hendra virus (equine morbilli-virus) in fruit bats, horses and cats. *Aust Vet J* 76:813–818

12. Middleton D, Westbury H, Morrissy C, Van der Heide B, Russell G, Braun M, Hyatt A (2002) Experimental Nipah virus infection in pigs and cats. *J Comp Pathol* 126:124–136
13. Wacharapluesadee S, Boongird K, Wanghongsa S, Ratanasetyuth N, Supavonwong P, Saengsen D, Gongal G, Hemachudha T (2010) A longitudinal study of the prevalence of Nipah virus in *Pteropus lylei* bats in Thailand: evidence for seasonal preference in disease transmission. *Vector-Borne Zoonotic Dis* 10:183–190
14. Tamin A, Harcourt BH, Ksiazek TG, Rollin PE, Bellini WJ, Rota PA (2002) Functional properties of the fusion and attachment glycoproteins of Nipah virus. *Virology* 296:190–200
15. Bowden TA, Crispin M, Harvey DJ, Aricescu AR, Grimes JM, Jones EY, Stuart DI (2008) Crystal structure and carbohydrate analysis of Nipah virus attachment glycoprotein: a template for antiviral and vaccine design. *J Virol* 82:11628–11636
16. Aguilar HC, Ataman ZA, Aspericueta V, Fang AQ, Stroud M, Negrete OA, Kammerer RA, Lee B (2009) A novel receptor-induced activation site in the Nipah virus attachment glycoprotein (G) involved in triggering the fusion glycoprotein (F). *J Biol Chem* 284:1628–1635
17. Guillaume V, Contamin H, Loth P, Georges-Courbot M-C, Lefevre A, Marianneau P, Chua K, Lam S, Buckland R, Deubel V (2004) Nipah virus: vaccination and passive protection studies in a hamster model. *J Virol* 78:834–840
18. Weingartl HM, Berhane Y, Caswell JL, Loosmore S, Audonnet J-C, Roth JA, Czub M (2006) Recombinant Nipah virus vaccines protect pigs against challenge. *J Virol* 80:7929–7938
19. Mire CE, Versteeg KM, Cross RW, Agans KN, Fenton KA, Whitt MA, Geisbert TW (2013) Single injection recombinant vesicular stomatitis virus vaccines protect ferrets against lethal Nipah virus disease. *Virol J* 10:353–366
20. Reading S, Dimmock N (2007) Neutralization of animal virus infectivity by antibody. *Arch Virol* 152:1047–1059
21. Plotkin SA (2010) Correlates of protection induced by vaccination. *Clin Vaccine Immunol* 17:1055–1065
22. Korber B, LaBute M, Yusim K (2006) Immunoinformatics comes of age. *PLoS Comput Biol* 2:e71
23. Bourdette D, Edmonds E, Smith C, Bowen J, Guttmann CR, Nagy Z, Simon J, Whitham R, Lovera J, Yadav V (2005) A highly immunogenic trivalent T cell receptor peptide vaccine for multiple sclerosis. *Multi Scler* 11:552–561
24. López JA, Weilenman C, Audran R, Roggero MA, Bonelo A, Tiercy JM, Spertini F, Corradin G (2001) A synthetic malaria vaccine elicits a potent CD8+ and CD4+ T lymphocyte immune response in humans. Implications for vaccination strategies. *Eur J Immunol* 31:1989–1998
25. Knutson KL, Schiffman K, Disis ML (2001) Immunization with a HER-2/neu helper peptide vaccine generates HER-2/neu CD8 T-cell immunity in cancer patients. *J Clin Invest* 107:477–484
26. Petrovsky N, Brusic V (2002) Computational immunology: the coming of age. *Immunol Cell Biol* 80:248–254
27. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539
28. Larsen MV, Lundegaard C, Lamberth K, Buus S, Brunak S, Lund O, Nielsen M (2005) An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur J Immunol* 35:2295–2303
29. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M (2007) Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinform* 8:424
30. Peters B, Sette A (2005) Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinform* 6:132
31. Peters B, Bulik S, Tampe R, Van Endert PM, Holzhütter H-G (2003) Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *J Immunol* 171:1741–1749
32. Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, Schatz M, Kloetzl P-M, Rammensee H-G, Schild H, Holzhütter H-G (2005) Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell Mol Life Sci CMLS* 62:1025–1037
33. Bui H-H, Sidney J, Li W, Fuseder N, Sette A (2007) Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinform* 8:361
34. Kaur H, Garg A, Raghava G (2007) PEPstr: a de novo method for tertiary structure prediction of small bioactive peptides. *Protein Pept Lett* 14:626–631
35. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 30:2785–2791
36. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31:455–461
37. Bui H-H, Sidney J, Dinh K, Southwood S, Newman M, Sette A (2006) Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinform* 7:153
38. Kolaskar A, Tongaonkar PC (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* 276:172–174