

**Two highly conserved LAEDDTNAQKT and LTKIGTEI epitopes in G
glycoprotein can be used for effective epitope based vaccine design
against Henipavirus**

**Md. Masud Parvege, Monzilur Rahman, Yead Morshed Nibir, Mohammad Shahnoor
Hossain***

Department of Genetic Engineering & Biotechnology
University of Dhaka, Dhaka-1000, Bangladesh
Phone: 880-2-9661900-7824; Fax: 880-2-8615583

*Corresponding Author
Mohammad Shahnoor Hossain, Ph.D.

Assistant Professor

mshahnoor@du.ac.bd

Phone: 880-2-9661900-7824; Fax:880-2-9667222

Author Contribution: Conceived and designed the experiments: MSH, MR and MMP. Analyzed the data: MMP, MR and YMN. Wrote the manuscript: MMP and MR. Agree with manuscript results and conclusions: MMP, MR, YMN and MSH. Jointly developed the structure and arguments for the paper: MMP and MR. Made critical revisions and approved final version: MSH. All authors reviewed and approved of the final manuscript.

Abstract

Nipah virus and Hendra virus, two members of the genus Henipavirus, are newly emerging zoonotic pathogens which cause acute respiratory illness and severe encephalitis in human. Lack of the effective antiviral therapy endorses the urgency for the development of vaccine against these deadly viruses. In this study, we employed various computational approaches to identify epitopes which has the potential for vaccine development. By analyzing the immune parameters of the conserved sequences of G glycoprotein using various databases and bioinformatics tools, we identified two potential epitopes which could be used as peptide vaccines. Using different B cell epitope prediction servers, four conserved B cell epitopes were identified. Immunoinformatics analyses revealed that LAEDDTNAQKT is a highly flexible and accessible B-cell epitope to antibody. Highly conserved putative CTL epitopes were analyzed for their binding with the HLA-C 12*03 molecule. Docking simulation assay revealed that LTDKIGTEI has significantly lower binding energy, which bolstered its potential as epitope-based vaccine design. Finally, cytotoxicity analysis has also justified their potential as promising epitope-based vaccine candidate. In sum, our computational analysis indicates that either LAEDDTNAQKT or LTDKIGTEI epitope could be promising as a component for the development of universal vaccine against all kinds of Henipavirus. Further *in vivo* and *in vitro* studies are necessary to validate the obtained findings.

Keywords: Henipavirus, G glycoprotein, conserved epitopes, universal vaccine

Running header: Epitope based vaccine design of Henipavirus

Introduction

Henipavirus, a genus of the paramyxovirus family, is responsible for several epidemics in some regions of Asia and Australia in the last few years. Viruses belonging to this group are deadly pathogens and are characterized by their zoonotic nature, broad host range, and ease of transmission from person to person^{1,2}. Among three established species of the genus Henipavirus, Nipah Virus (NiV) and Hendra Virus (HeV) are responsible for causing lung disease and encephalitis in pigs, horses and humans, while the other species, Cedar Virus (CedPV) is harmless²⁻¹¹. Pteroid fruit bats and microbats are the common reservoir of Henipavirus¹²⁻¹⁵. Upon contact with the bats, domestic animals develop infection and further transmit the virus to human. Environmental changes, destruction of bat habitats through deforestation, and genetic diversification have been implicated as the reasons behind the emergence and re-emergence of the viruses in recent time.

Henipaviruses are pleomorphic in nature and their size ranges from 40 to 600 nm in diameter¹⁶. The genome of the Henipavirus is ~18.2 kb long, non-segmented, single stranded negative sense RNA, which encodes six genes corresponding to six structural proteins¹⁷. The core of the virus is composed of RNA, which is tightly bound to nucleocapsid (N) protein and associated with large (L) and phosphoprotein (P). F (fusion) and G (attachment) proteins are embedded within the lipid membrane. The role of the G glycoprotein is to attach the virus with the surface of a host cell through a highly conserved mammalian protein, EFN B2^{16,18}. The F protein helps to fuse the viral membrane with the host cell membrane, which allows the release of the virion into the cell.

Since the first emergence in 1994, Henipavirus outbreaks have been observed at least once a year¹⁹. During all outbreak events, the horses were infected first and subsequently transmitted to human. Approximately, one hundred horses were affected by the virus with a fatality rate of 70%. In case of human, four out of seven infections became fatal along with influenza-like symptoms in all patients. While some of them recovered from the viral infection, others either developed pneumonitis or encephalitis or even multi-organ failure²⁰. NiV was identified relatively later in 1999 during an outbreak in pig farms. The event was responsible for 257 human infections, including 105 confirmed deaths²¹. Eight more outbreaks of NiV have occurred since then in Bangladesh and India²². In an outbreak in February, 2011 in the northern region of Bangladesh, 17 out of 24 NiV infected persons died²³⁻²⁵. The number of outbreaks and re-emergence of Henipaviruses imply its high potency as pathogen which can evolve further by the interplay between the genome of the virus species during growth and multiplication in the common host.

The deadly nature of Henipaviruses become more frightening due to the unavailability of any approved vaccine or drug either to induce protection against the virus or to control viral infection. None of the several attempts that have been made could develop a successful treatment option. For example, Ribavirin was tested to treat animal models infected with NiV and HeV, but showed no significant increase in the survival rate of the animals^{26,27}. Monoclonal antibody has also been an attractive treatment option. A neutralizing antibody, mAb m102.4, targeting the glycoprotein domain of the virus that binds to ephrin-B₂ and ephrin-B₃ receptor of host, has already been tested²⁸⁻³⁰. In vivo experiment has shown enhanced survival of the animals challenged with HeV and NiV³¹.

In addition to these post-exposure treatment options, some recent studies have been able to successfully induce immunity against Hendravirus by injecting animals with a fragment of G protein working as a subunit vaccine³²⁻³⁵. Assessment of the vaccine in non-human primates revealed its potential for human use, but it has not been studied in human so far³⁶. Nevertheless, no universal vaccine has yet been developed that works against all Henipaviruses irrespective of their genetic differences.

With the advent of various bioinformatics tools and the availability of enormous sequence data, epitope based vaccine design against highly conserved antigenic protein has become easier. Computational prediction of conserved epitope for the development of novel vaccine will not only save time, but also reduces the cost associated with the whole process. In the present study, we employed various bioinformatics tools to find out the highly conserved peptide and mapped the evolutionary conserved epitopes. We predicted two highly conserved epitopes which can be used to develop universal vaccines to prevent all types of Henipavirus infections.

Materials and Methods

Protein sequence retrieval

Sequences of the G glycoprotein of Henipaviruses were retrieved from the NCBI database³⁷ and a total 40 sequences were found in the database. All the sequences were downloaded from the database in FASTA format. Among 40 sequences, 21 belongs to Hendra Virus (HeV) and 19 belong to Nipah Virus (NiV) respectively.

Variability analysis of G glycoproteins

To determine the degree of conservation, all the retrieved sequences were aligned by using EBI-Clustal Omega program³⁸ and a multiple sequence alignment (MSA) was generated. The MSA was visualized using Jalview³⁹. The absolute site variability in the MSA was calculated by Shannon entropy analysis⁴⁰ using Protein Variability Server (PVS)⁴¹. PVS utilizes several variability metrics to compute absolute variation in a MSA.

Prediction of antigenicity

In order to develop a subunit vaccine, it is imperative to identify those proteins which can induce protection from subsequent challenges. A reference G glycoprotein having accession number, NP_047112, was tested for its antigenicity. VaxiJen v2.0 server⁴² and Kolaskar & Tongaonkar method⁴³ were used to analyze the antigenic property of the given sequence. VaxiJen utilizes an alignment independent approach to predict the antigenicity of a given protein, which is solely based on physiochemical properties of amino acids.

Linear B-cell prediction

B cell epitope prediction can be done with several software packages and each one has its own strength and weakness^{44,45}. To minimize the false positive predictions, four popular and widely used tools were utilized. For the prediction of B-cell epitopes, sequence was submitted to BepiPred⁴⁶, IEDB linear B cell prediction⁴⁷, BCPRED⁴⁸, and ABCpred⁴⁹ web servers. BepiPred 1.0⁴⁶ server predicts the epitopes based on a hidden Markov model along with a propensity scale and the threshold for the prediction was set 0.35. IEDB linear B cell epitope prediction tool⁴⁷ was used as default parameters. Non-overlapping and fixed length epitopes of 14 amino acids were predicted using 75% specific criteria using BCPREDS server⁴⁸. Seventy five percent threshold along with a window length of 16 amino acids was set as parameter for the prediction of epitopes in ABCpred server⁴⁹. The predicted epitopes from these four servers were compared manually and epitopes that were recognized by all the servers were selected for further analysis.

Surface accessible regions prediction

When an antibody or a cell surface receptor binds to an epitope, it will induce an immune response in turn. An ideal epitope should be accessible to an antibody or a cell surface receptor⁵⁰. In order to determine the surface accessible regions, Emni surface accessible prediction tool of IEDB was used^{47,51}. Later, the accessible regions were compared with the selected B cell epitopes. Epitopes that were found to have surface accessibility were selected further for conservancy analysis.

Conservancy analysis of B cell epitope

An ideal epitope should be conserved so that it provides a broader protection against multiple strains, in some cases, even species. Amino acid residues that are crucial for critical functions are believed to have lower variability even under immune pressure. Conservancy analysis of the surface accessible epitopes with all the glycoprotein sequences was analyzed by the IEDB epitope conservancy tool⁵².

Prediction of flexibility and hydrophilicity of B cell epitopes

Earlier studies reported that hydrophilicity and flexibility of a peptide correlate with its antigenicity⁵³. For this, conserved epitopes were subjected to Karplus and Schulz flexibility prediction⁵⁴, and Parker hydrophilicity prediction tools⁵⁵ for flexibility and hydrophilicity analyses respectively⁵⁶.

T cell epitope prediction and conservancy analysis

T cell epitopes were identified by NetCTL prediction method^{57,58} of IEDB. The threshold was set at 0.50 allowing the sensitivity and specificity of 0.89 and 0.94 respectively. NetCTL predicts T cell epitopes by integrating predictions of proteasomal cleavage, TAP transport efficiency, and MHC class I affinity. This combined algorithm provides overall prediction scores and based on this scores, top 10 epitopes were taken for further analysis. MHC-I alleles interacting each of the selected epitopes were determined by MHC-I prediction server⁵⁹ of IEDB. Stabilized matrix method (SMM)⁵⁹ was utilized for the determination of half maximal inhibitory concentration (IC₅₀) of peptide binding to MHC-I alleles. For the binding analysis, all the available MHC class I alleles were selected and the

peptide lengths were set at 9 amino acids. Selected T cell epitopes were subjected to IEDB conservancy analysis tool⁵².

Prediction of the 3D structure of conserved T cell epitopes and selected HLA-C 12*03

For the docking simulation assay, 3D structures both of the peptides and corresponding alleles were required. The 3D structures of the selected peptides were generated using the PEP-FOLD Peptide Structure Prediction server^{60,61}. The best models provided by the server were chosen for the docking study. HLA-C 12*03 was found to interact with most of the predicted T cell epitopes. For this reason, the 3-D model of HLA-C 12*03 was generated using SWISS MODEL server⁶². The HLA-C 12*03 -3D model was evaluated by PROCHECK software⁶³ and ProSA web tool⁶⁴. Ramachandran plot⁶³, constructed using the PROCHECK software assesses the stereo-chemical quality of 3-D structure analyzing residue-by-residue geometry and overall structure geometry. ProSA tool⁶⁴ provides a Z-score which is a measurement for the quality of the model.

Docking simulation assay of conserved T cell epitopes with the HLA-C 12*03 allele

To investigate the interaction of the T cell epitopes with the corresponding MHC class I molecules, docking analysis was performed. Computer-simulated ligand docking is a powerful technique for evaluating relative binding affinity of the ligand towards its receptor⁶⁵. The AutoDOCK tool from the MGL software package was utilized for the docking purpose^{66,67}. At first, both the allele (HLA-C 12*03) and ligand (epitope) files were converted into PDBQT format to use them for the docking study. The grid/space box center was set at -15.059, -3.063, and -26.955° Å in the *x*-, *y*-, and *z*-axes, respectively,

which allows the predicted epitopes to bind to the binding groove of the selected MHC class I. The size was set at 20, 40, and 40° Å in the x , y , and z dimensions, respectively. The 3D structure of MHC class I H-2Kb molecule complexes with the octapeptide KVITFIDL (PKB1) was employed as a control for the binding assay. KVITFIDL structure was first separated from H-2Kb and then it was docked with H-2Kb and selected the MHC class I molecule by setting similar parameters. Binding energy for control peptide was calculated and compared with the energy found for the predicted T cell epitopes. All the analyses were done at 1.0 Å spacing and complexes were visualized using Discovery studio 4.1⁶⁸.

Toxicity prediction of the peptides

The ease of manufacture, high specificity, and high penetration of peptides presents the peptide based therapy promising. But toxicity of peptides bottlenecks the success of peptide based therapy. An ideal epitope should have no or less toxicity, but high antigenicity⁶⁹. Hence, to determine the toxicity of the selected epitopes, ToxinPred web server was used⁷⁰.

Results

Glycoprotein G is conserved in all pathogenic Henipavirus strains

The degree of variability or similarity of specific proteins provides important insights about its evolution, structure function, and immunology. To determine the degree of

conservation, MSA by Clustal Omega³⁸ and protein variability analysis⁴¹ were performed. From the MSA, G glycoprotein was found to be well conserved in all sequences (Additional Figure 1). The absolute variability computed by PVS⁴¹ revealed 11 fully conserved regions, which comprise more than 94% of the length of G protein (Figure 1).

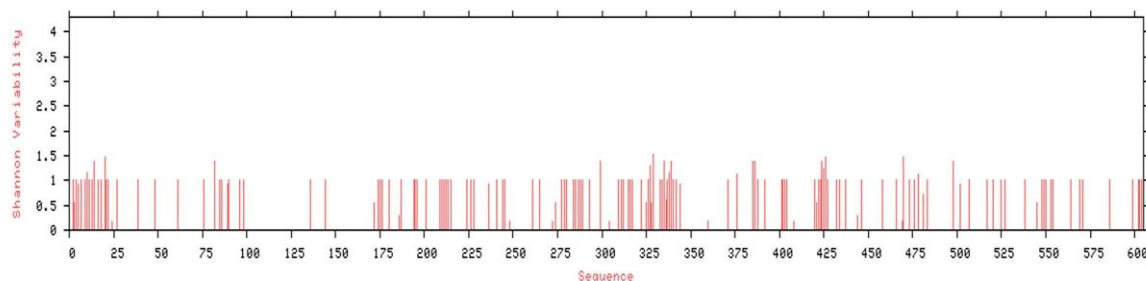


Figure 1: Protein variability index of G protein determined by using PVS server. The conservancy threshold was 1.0 in this analysis. X axis indicates the amino acid position in sequences and Y axis indicates the Shanon entropy

G protein is highly antigenic

To be a vaccine candidate, a protein must be antigenic enough to provoke sufficient immune response. Evaluation of G protein sequence having accession number NP_047112 by VaxiJen server⁴² identified it as a probable antigen with a value of 0.53. Kolaskar & Tongaonkar antigenicity prediction tool⁴³ revealed that most of the residues in the protein were above the threshold value (Figure 2).

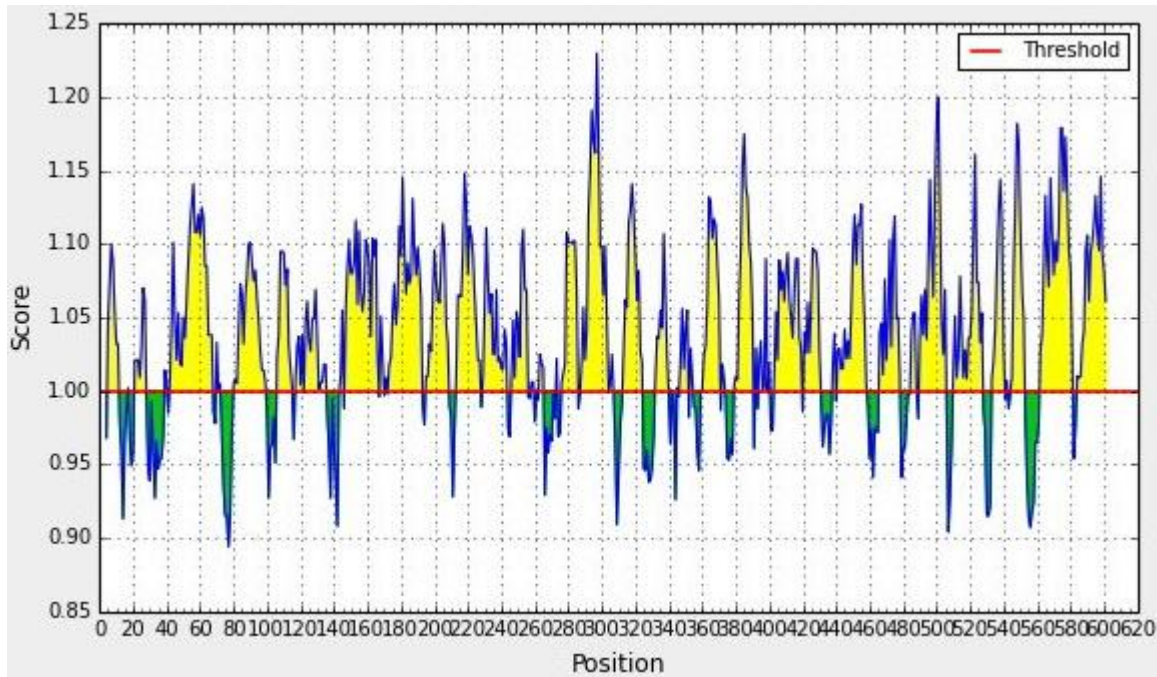


Figure 2: The G glycoprotein was found to be highly antigenic. The threshold is 1.00 and residues in yellow regions are antigenic in nature

Nine B cell epitopes were recognized by all the prediction tools

Several B cell epitope prediction software packages are currently available. Each software has its own dataset and utilizes a particular method for prediction. Since, epitopes predicted for a given protein differ from one software to another^{44,45}, perfect identification of immunogenic regions in a given antigen is difficult, and prediction of false positive epitopes is a typical problem⁷¹. Hence, we utilized four different software packages for the B cell epitope prediction. Epitopes recognized by the four prediction tools were taken into account for further analysis. Nine antigenic epitopes of protein were found common to the four prediction tools (Table 1). The locations of these epitopes are 1: 74-80, 2 : 163-176, 3:271-279, 4: 322-332 5: 342-359, 6: 375-380, 7: 529-534 , 8: 551-561, 9: 582-587.

Table 1: Nine B cell epitopes found common in four prediction servers along with their lengths and sequence positions

No	Sequence	Length	Start	End
1	TRTTDNQ	7	74	80
2	PNPLPFREYRPISQ	12	163	176
3	WTPPNPSTI	9	271	279
4	RPKSDSGDYNQ	11	322	332
5	RGKYDKVMPYGPSGIKQG	18	342	359
6	FQYNDS	6	375	380
7	NQTAEN	6	529	534
8	LAEDDTNAQKT	11	551	561
9	DTGDSV	6	582	587

Thirteen surface accessible peptides were predicted

An ideal epitope should have access to an antibody so that the antibody can bind to the epitope to induce immune response⁵⁰. At threshold cutoff 1.0, the surface accessibility of the G protein was determined by Emni surface accessibility prediction tool⁵¹, and 13 peptides were found to have scores above the threshold. These peptides were compared with the selected nine epitopes (Table 2, Figure 3). Seven epitopes were found to have surface accessibility properties. These seven epitopes were analyzed for conservancy.

Table 2: Predicted surface accessible antigenic sites by using Emni surface accessibility prediction analysis

No	Peptide	Length	Sequence position
1	IKNYYG	6	25-30
2	NYTRTTDN	8	72-79
3	KISQST	6	130-135
4	PLPFREYRPI	10	165-174
5	VWTPPNPS	8	270-277
6	TYHEDFY	7	285-291
7	RPKSDSGDYNQK	12	322-333
8	KVERGKYDKV	10	339-348
9	PRTEFQYNDS	10	371-380

10	KYSKAE	6	388-393
11	QASYSW	6	455-460
12	NSNQTAE	7	527-533
13	AEDDTNAQKT	10	552-561

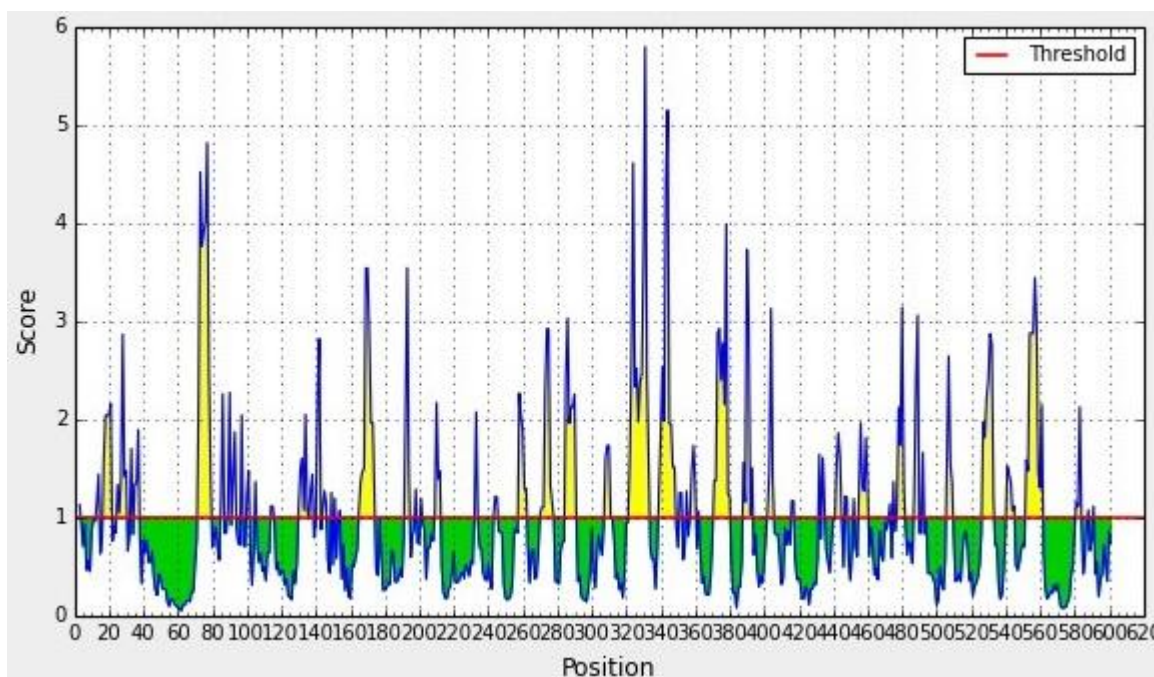


Figure 3: Surface accessibility of G protein. The horizontal red line indicates surface accessibility cutoff and yellow color regions above this cutoff are surface accessible epitopes

Four B cell epitopes are highly conserved among all sequences

The use of conserved epitopes provides broader protection across multiple strains, or even species, than epitopes derived from highly variable genomic regions. So, in an epitope based vaccine setting, an ideal epitope should be highly conserved. The conservancy of the seven B cell epitopes was evaluated by the IEDB conservancy analysis tool⁵². Among seven, four epitopes were found to be fully conserved in all G protein sequences (Table 3).

Table 3: Consensus sequences between the predicted B cell epitopes and Emni surface peptides along with their conservancy using IEDB conservancy analysis

No	Sequence	Length	Identity %
1	NYTRTTDNQ	9	100
2	PLPFREYRPI	10	100
3	VWTPPNPSTI	10	85
4	RPKSDSGDYNQ	11	52.50
5	KVERGKYDKV	10	52.50
6	NSNQTAEN	8	100
7	LAEDDTNAQKT	11	100

LAEDDTNAQKT is highly flexible and accessible

Flexibility and accessibility are two key features of an epitope to induce an immune response⁵³. Among four conserved epitopes, LAEDDTNAQKT and NYTRTTDNQ were found to be highly flexible in Karplus and Schulz flexibility prediction⁵⁴ analysis (Figure 4). Then, these two epitopes were assessed for their hydrophilicity by IEDB Parker hydrophilicity analysis⁵⁵. LAEDDTNAQKT epitope was found to be hydrophilic in nature,

whereas NYTRTTDNQ was hydrophobic (Figure 5).

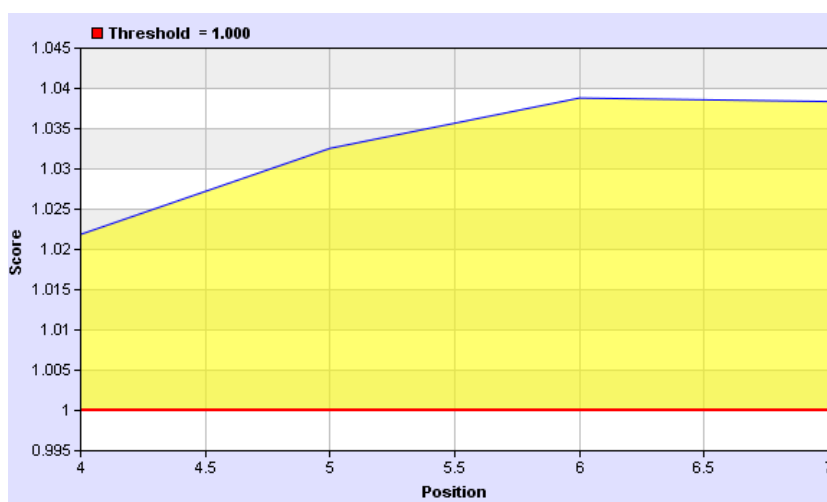


Figure 4: Flexibility of the LAEDDTNAQKT epitope. Amino acids of this epitope were found to be above the threshold level

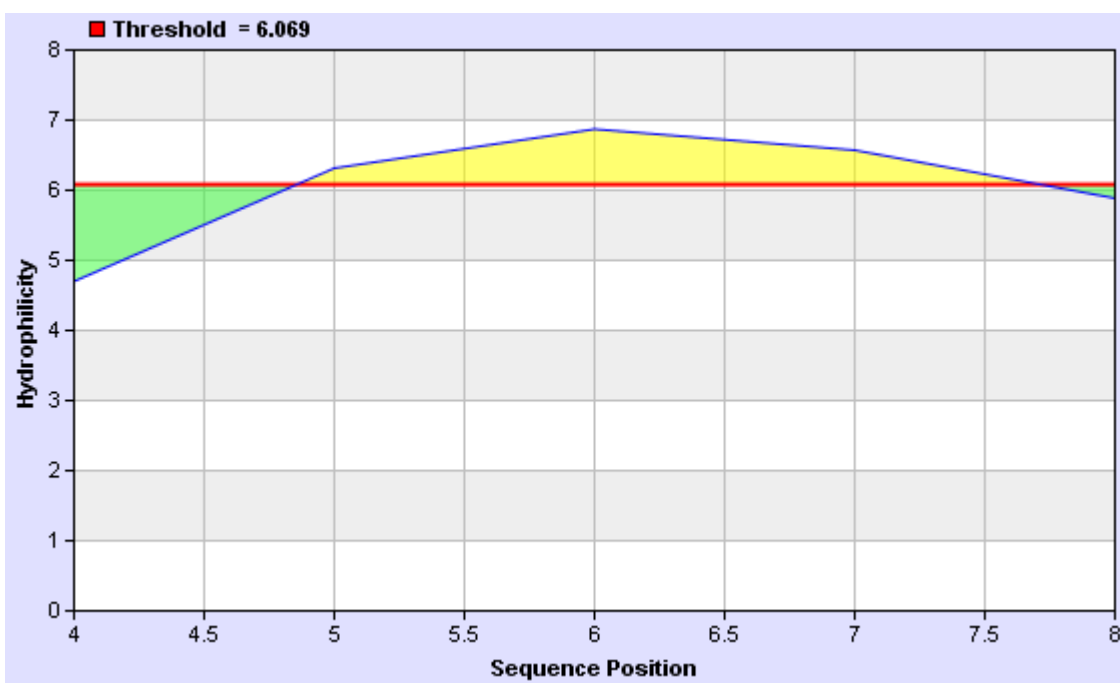


Figure 5: Hydrophilicity of the LAEDDTNAQKT epitope. Most of the residues of this selected epitope were found to be hydrophilic in nature. Residues above the cutoff 6.069 (horizontal red line) are in the yellow region

LTDKIGTEI and NSLGQPVFY are highly conserved T cell epitopes

According to selected parameter settings, NetCTL server⁵⁷ identified 27 potential T cell epitopes, but only 10 epitopes were chosen on the basis of high combinatorial scores. MHC class I alleles that interact with the selected epitopes were determined by MHC-I binding prediction server based on IC_{50} values^{59,72}. The result is summarized in the Table 4. Conservancy analysis of the 10 T cell epitopes revealed that LTDKIGTEI and NSLGQPVFY epitopes were fully conserved (Table 4).

Table 4: Predicted epitopes along with their interacting MHC-I alleles .

Epitopes	Prediction Score	Interacting MHC-I allele with an affinity of $IC_{50} < 200$ nM	Conservancy %
LTDKIGTEI	1.8091	C*12:03, A*68:23, C*05:01, A*32:07,	100

		A*32:07, C*15:02, C*14:02, B*27:20, A*02:50	
NSLGQPVFY	1.2151	C*03:03, A*32:07, B*27:20, A*32:15, A*68:23, C*12:03, B*40:13	100
AVDNGFFAY	3.2584	C*05:01, C*12:03, A*68:23, A*80:01, A*32:07, A*32:15, B*35:01, A*29:02, B*15:02, A*30:02, B*27:20, A*11:01	52.50
SSTYHEDFY	2.5667	A*32:07, B*27:20, A*68:23, C*12:03, B*40:13, C*03:03, A*32:15, B*15:02, B*15:17	52.50
STYHEDFY	2.3544	B*15:17, A*68:23, A*32:07, A*29:02, C*12:03, C*03:03, B*27:20, A*11:01, B*40:13, A*02:17, A*80:01, A*32:15, A*30:02, C*14:02, A*26:02, A*68:01, B*15:03	52.50
DSGDYNQKY	2.2951	C*12:03, A*32:07, A*68:23, B*40:13, B*27:20, C*07:01, A*26:02, A*32:15	52.50
ITKVERGKY	2.1445	A*68:23, C*12:03, A*30:02, B*15:17, A*32:07, B*27:20, A*32:15, C*03:03	52.50
CSSTYHEDF	1.7539	B*27:20, C*03:03, A*32:07, A*68:23, B*15:03, B*40:13, C*12:03, A*32:15, B*15:17, B*58:01, C*15:02, C*07:01	52.50
ITDCFLLEN	1.7215	C*12:03, A*68:23, B*27:20, C*05:01, A*32:07, C*15:02, A*02:11, B*40:13, A*02:06, A*02:50	52.50
TTDNQALIK	1.2856	C*12:03, A*68:23, C*05:01, A*32:07, A*11:01, A*32:15, B*27:20, C*14:02, B*40:13	52.50

3D structures of the predicted epitope peptides and HLA-C 12*03 allele were predicted and validated

For the molecular docking simulation assay, the 3-D structure of both the epitopes and MHC molecule are required. PEP- FOLD Peptide Structure Prediction server was used for generating 3-D structure of two T cell epitopes LTDKIGTEI and NSLGQPVFY (Figure 6 A and B). As both epitopes were predicted to interact with HLA-C 12*03 allele, the 3-D structure of it was generated by homology modeling⁶². The model was further validated

using Ramachandran plot and Z-score. Ramachandran plot generated by Procheck software showed that 80% residues were in the favorable region (Figure 6 D). To check whether the input structure is within the range of scores typically found for native protein of similar size, ProSAz-score was calculated⁶⁴. z-score, indicative of overall model quality, was -8.81 that ascertained the quality of the generated model (Figure 6 E).

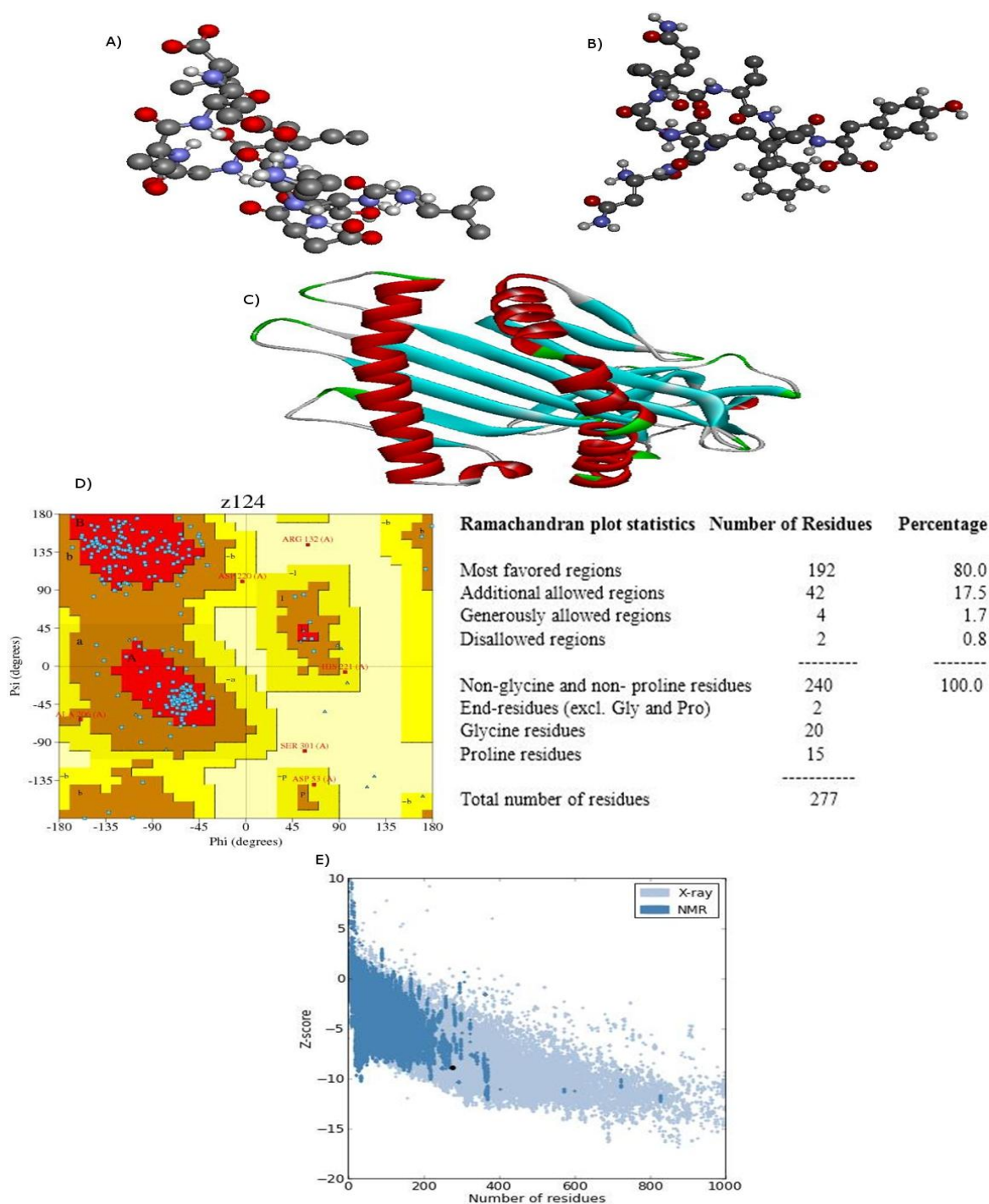


Figure 6: Predicted 3-D structures using homology modelling of the T cell epitopes LTDKIGTEI (A) and NSLGQPVFY (B), and MHC class I molecule HLA-C 12*03 (C). Ramachandran plot along with statistics showing residues in the most favorable and disallowed regions (D). Z-score for quality of the 3D structure HLA-C 12*03 (E)

T cell epitope LTDKIGTEI bound with the HLA-C 12*03 allele with a binding energy of -6.8 kcal/mol

Computer-simulated ligand docking is a rapid and powerful technique to evaluate the relative binding affinity of the ligand towards its receptors. At first binding models for both of the conserved T cell epitopes with the HLA-C 12*03 were generated using the AutoDOCK tool⁶⁶. The energy values calculated for binding both of the epitopes, LTDKIGTEI and NSLGQPVFY, to the binding groove of the HLA-C 12*03 were -6.8 and -2.1 kcal/mol, respectively (Figure 7). In case of binding groove of H-2KB (control MHC allele), the binding energy were -7.5 and -2.2 kcal/mol, respectively. Control peptide “KVITFIDL” bound to the HLA-C 12*03 and H-2kb alleles with the binding energy of -6.9 and -7.7 kcal/mol, respectively (Figure 7). As lower binding energy favors the formation of stable interaction, it is appreciable to expect that LTDKIGTEI will interact with MHC-I molecules in vivo readily.

LTDKIGTEI and NSLGQPVFY are non-cytotoxic

Because of the advantages of high specificity, high penetration, and ease of manufacturing over small molecules, peptides have emerged as promising therapeutics against many fatal diseases⁷³. But, one of the bottlenecks that impede the efficacy of peptide based therapies is their toxicity. Hence, toxicity for the selected epitopes was assessed by ToxinPred tool⁷⁰. Both of the selected epitopes were found to be non-toxic to cell rendering their potential as candidate vaccines.

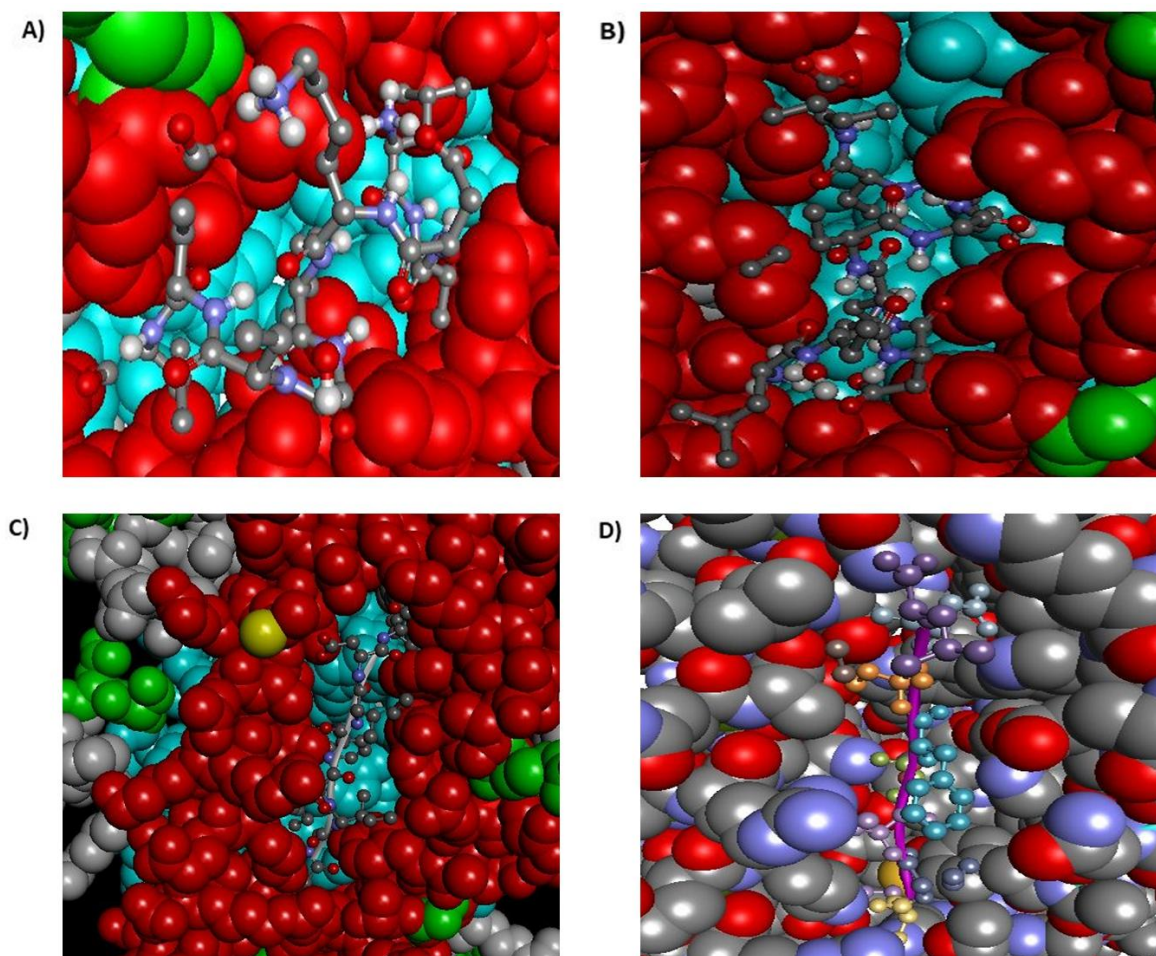


Figure 7: Docking simulation assay of the binding of predicted and control epitopes to MHC class I molecule HLA-C 12*03 and H-2Kb. Binding of “LTDKIGTEI” to the binding grooves (A) of the predicted structure of HLA-C 12*03 (binding energy: -6.8 kcal/mol) and (B) of the 3D structure of H-2Kb (binding energy: -7.5 kcal/mol); (C) binding of control peptide “KVITFIDL” to the predicted 3D structure of HLA-C 12*03 (-6.9 kcal/mol) and (D) H-2Kb (-7.7 kcal/mol)

Discussion

Henipavirus is an emerging zoonotic pathogen which causes severe encephalitis and respiratory illness in humans¹. The frequent outbreaks of this pathogen have exceeded the mortality rate of 70%. The swift human to human transmission, unavailability of

therapeutic options, and widespread distribution of reservoirs have rendered Henipavirus a profound global public health concern⁷⁴. Therefore, it is inevitable to take preventive measures. The development of a universal vaccine for all of the pathogenic members of Henipavirus genus is much more promising and economical solution rather than individual vaccine to each member. The concept of designing universal vaccine to prevent virus infection has already been reported. In case of influenza virus, a universal vaccine against matrix 2 protein has been designed which is able to fight against all Influenza serotypes⁷⁵. An endeavor to design a universal vaccine targeting spike protein against the members of coronaviruses was also reported⁷⁶.

At present, vaccination is widely conceived as the most effective method for the prevention of infectious diseases. Conventional development of vaccines relied on the inactivated or live attenuated pathogens. But, now the increasing understanding of antigen recognition at the molecular level has opened a new window for vaccine design and development, which surmounts the drawbacks of the traditional vaccine development process. The rationale behind the epitope based vaccine is the chemical synthesis of identified B cell and T cell epitopes, that are immunodominant and can generate specific immune responses⁶⁵. Epitopes have become suitable vaccine candidates owing to its comparative ease of construction and production, absence of infectious potential, and chemical stability^{77,78}.

In the current study, we aimed to identify potential epitopes which can be used to design universal vaccines for all three members of Henipavirus. We chose G protein over other proteins for epitope prediction as it facilitates the attachment of Henipavirus with the host cell membrane²⁸. Most of the vaccines are based on B cell immunity, but at present T cell

epitope-based vaccine has been drawn much more interest as CD8+T cells generate strong immune responses in the host against viral infection⁷⁹. For this reason, we identified B cell epitopes as well as T cell epitopes.

Surface glycoproteins always undergo frequent mutations to evade host defense. First step in an epitope based vaccine design is to identify the sequences of low variability. Hence, at first, we determined the degree of conservation of G protein. MSA and absolute variability site analysis demonstrated that G protein is highly conserved in Henipavirus. Several B cell epitope prediction methods have been developed in recent years, but performances of prediction of these methods are still far from ideal⁸⁰. To avoid the false positive, we utilized four different B cell epitope prediction methods and only the nine epitopes which were found common to all prediction were selected for further analysis. B cell epitopes must be accessible enough to bind to antibodies for eliciting immune responses⁵⁰. Here, we identified the surface peptides and compared these peptides with the selected B cell epitopes. From the conservation analysis of seven B cell epitopes, it was found that four B cell epitopes are fully conserved. The flexibility and accessibility are two key requirements for a B cell epitope to be a vaccine. LAEDDTNAQKT and NYTRTTDNQ, two conserved B cell epitopes were found to be highly flexible in nature, but only LAEDDTNAQKT epitope was found to be hydrophilic.

The T cell epitope prediction server, NetCTL, identified 27 CD8+ epitopes and based on the prediction scores, top 10 T cell epitopes were evaluated further by immune-informatics tools. To identify interacting MHC-I alleles for each of the selected epitopes, an elegant machine learning method called support vector machine (SVM) was utilized along with the

binding affinity. Two conserved T cell epitopes, LTDKIGTEI and NSLGQPVFY, were found to interact with the HLA-C 12*03 allele with high affinity. We further utilized computer-simulated molecular docking to investigate the intermolecular interaction. From the results of docking simulation assay, it was found that the binding energy of the LTDKIGTEI epitope is - 6.8 kcal/mol, which is almost similar to the binding energy of the control peptide. As lower binding energy favors stable intermolecular interaction, LTDKIGTEI could be a potential epitope. We strongly believe that nontoxic property of these two epitopes can bypass the toxicity related problem of peptide based vaccine. Stability and antigenicity can be further enhanced conjugating these epitopes with the adjuvant.⁸¹. To address the questions regarding actual immunogenicity, stability, efficacy, and delivery strategy inside the recipient bodies of these epitopes, in vitro and in vivo experiments are essential.

Conclusion

This study shows potential epitopes to design an epitope-based universal vaccine for all of the members of Henipavirus. Our results are based on sequence analyses of a surface glycoprotein and the predicted epitopes would be candidate targets for the universal vaccine design. The actual effectiveness of these two epitopes for eliciting immune response is needed to be tested both in in vitro and in vivo experiments.

Acknowledgments

This work was supported by the research grant provided to Dr. Mohammad Shahnoor Hossain from Biotechnology Research Centre, University of Dhaka.

Disclosure

There is no conflict of interest in this work

References

1. Williamson MM, Torres-Velez FJ. Henipavirus: a review of laboratory animal pathology. *Veterinary pathology*. Sep 2010;47(5):871-880.
2. Murray K, Selleck P, Hooper P, et al. A morbillivirus that caused fatal disease in horses and humans. *Science (New York, N.Y.)*. Apr 7 1995;268(5207):94-97.
3. Goh KJ, Tan CT, Chew NK, et al. Clinical Features of Nipah Virus Encephalitis among Pig Farmers in Malaysia. *New England Journal of Medicine*. 2000;342(17):1229-1235.
4. O'Sullivan JD, Allworth AM, Paterson DL, et al. Fatal encephalitis due to novel paramyxovirus transmitted from horses. *Lancet*. Jan 11 1997;349(9045):93-95.
5. Tan CT, Chua KB. Nipah virus encephalitis. *Current infectious disease reports*. Jul 2008;10(4):315-320.
6. Tan CT, Goh KJ, Wong KT, et al. Relapsed and late-onset Nipah encephalitis. *Annals of Neurology*. 2002;51(6):703-708.
7. Tyler KL. Emerging viral infections of the central nervous system: Part 2. *Archives of Neurology*. 2009;66(9):1065-1074.
8. Hooper P, Zaki S, Daniels P, Middleton D. Comparative pathology of the diseases caused by Hendra and Nipah viruses. *Microbes and infection / Institut Pasteur*. Apr 2001;3(4):315-322.
9. Hossain MJ, Gurley ES, Montgomery JM, et al. Clinical presentation of nipah virus infection in Bangladesh. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. Apr 1 2008;46(7):977-984.
10. Middleton DJ, Westbury HA, Morrissy CJ, et al. Experimental Nipah virus infection in pigs and cats. *Journal of comparative pathology*. Feb-Apr 2002;126(2-3):124-136.
11. Marsh GA, de Jong C, Barr JA, et al. Cedar virus: a novel Henipavirus isolated from Australian bats. *PLoS pathogens*. 2012;8(8):e1002836.
12. Yob JM, Field H, Rashdi AM, et al. Nipah virus infection in bats (order Chiroptera) in peninsular Malaysia. *Emerging infectious diseases*. May-Jun 2001;7(3):439-441.
13. Li Y, Wang J, Hickey AC, et al. Antibodies to Nipah or Nipah-like viruses in bats, China. *Emerging infectious diseases*. Dec 2008;14(12):1974-1976.
14. Hayman DT, Suu-Ire R, Breed AC, et al. Evidence of henipavirus infection in West African fruit bats. *PloS one*. 2008;3(7):e2739.

15. Halpin K, Hyatt AD, Fogarty R, et al. Pteropid bats are confirmed as the reservoir hosts of henipaviruses: a comprehensive experimental study of virus transmission. *The American journal of tropical medicine and hygiene*. Nov 2011;85(5):946-951.
16. Bonaparte MI, Dimitrov AS, Bossart KN, et al. Ephrin-B2 ligand is a functional receptor for Hendra virus and Nipah virus. *Proceedings of the National Academy of Sciences of the United States of America*. Jul 26 2005;102(30):10652-10657.
17. Wang L-F, Harcourt BH, Yu M, et al. Molecular biology of Hendra and Nipah viruses. *Microbes and Infection*. 4// 2001;3(4):279-287.
18. Negrete OA, Levroney EL, Aguilar HC, et al. EphrinB2 is the entry receptor for Nipah virus, an emergent deadly paramyxovirus. *Nature*. Jul 21 2005;436(7049):401-405.
19. Wang LF, Yu M, Hansson E, et al. The exceptionally large genome of Hendra virus: support for creation of a new genus within the family Paramyxoviridae. *Journal of virology*. Nov 2000;74(21):9972-9979.
20. Blum LS, Khan R, Nahar N, Breiman RF. In-depth assessment of an outbreak of Nipah encephalitis with person-to-person transmission in Bangladesh: implications for prevention and control strategies. *The American journal of tropical medicine and hygiene*. Jan 2009;80(1):96-102.
21. Field H, Young P, Yob JM, Mills J, Hall L, Mackenzie J. The natural history of Hendra and Nipah viruses. *Microbes and infection / Institut Pasteur*. Apr 2001;3(4):307-314.
22. Chadha MS, Comer JA, Lowe L, et al. Nipah virus-associated encephalitis outbreak, Siliguri, India. *Emerging infectious diseases*. Feb 2006;12(2):235-240.
23. Chua KB, Bellini WJ, Rota PA, et al. Nipah virus: a recently emergent deadly paramyxovirus. *Science (New York, N.Y.)*. May 26 2000;288(5470):1432-1435.
24. Hughes JM, Wilson ME, Luby SP, Gurley ES, Hossain MJ. Transmission of Human Infection with Nipah Virus. *Clinical Infectious Diseases*. December 1, 2009 2009;49(11):1743-1748.
25. Chua KB, Chua BH, Wang CW. Anthropogenic deforestation, El Nino and the emergence of Nipah virus in Malaysia. *The Malaysian journal of pathology*. Jun 2002;24(1):15-21.
26. Gurley ES, Montgomery JM, Hossain MJ, et al. Person-to-person transmission of Nipah virus in a Bangladeshi community. *Emerging infectious diseases*. Jul 2007;13(7):1031-1037.
27. Homaira N, Rahman M, Hossain MJ, et al. Nipah virus outbreak with person-to-person transmission in a district of Bangladesh, 2007. *Epidemiology and infection*. Nov 2010;138(11):1630-1636.
28. Bishop KA, Stantchev TS, Hickey AC, et al. Identification of Hendra virus G glycoprotein residues that are critical for receptor binding. *Journal of virology*. Jun 2007;81(11):5893-5901.
29. Negrete OA, Wolf MC, Aguilar HC, et al. Two key residues in ephrinB3 are critical for its use as an alternative receptor for Nipah virus. *PLoS pathogens*. Feb 2006;2(2):e7.
30. Pasquale EB. Eph-ephrin bidirectional signaling in physiology and disease. *Cell*. Apr 4 2008;133(1):38-52.
31. Plotkin SA. Vaccination against the major infectious diseases. *Comptes rendus de l'Academie des sciences. Serie III, Sciences de la vie*. Nov 1999;322(11):943-951.
32. Guillaume V, Contamin H, Loth P, et al. Nipah virus: vaccination and passive protection studies in a hamster model. *Journal of virology*. Jan 2004;78(2):834-840.
33. Guillaume V, Contamin H, Loth P, et al. Antibody prophylaxis and therapy against Nipah virus infection in hamsters. *Journal of virology*. Feb 2006;80(4):1972-1978.

34. Guillaume V, Wong KT, Looi RY, et al. Acute Hendra virus infection: Analysis of the pathogenesis and passive antibody protection in the hamster model. *Virology*. May 10 2009;387(2):459-465.
35. Weingartl HM, Berhane Y, Caswell JL, et al. Recombinant nipah virus vaccines protect pigs against challenge. *Journal of virology*. Aug 2006;80(16):7929-7938.
36. Mire CE, Geisbert JB, Agans KN, et al. A recombinant Hendra virus G glycoprotein subunit vaccine protects nonhuman primates against Hendra virus challenge. *Journal of virology*. May 2014;88(9):4624-4631.
37. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic acids research*. Jan 2009;37(Database issue):D26-31.
38. Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*. 2011;7:539.
39. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. May 1 2009;25(9):1189-1191.
40. Shannon CE. A Mathematical Theory of Communication. *Bell System Technical Journal*. 1948;27(3):379-423.
41. Garcia-Boronat M, Diez-Rivero CM, Reinherz EL, Reche PA. PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic acids research*. Jul 1 2008;36(Web Server issue):W35-41.
42. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC bioinformatics*. 2007;8:4.
43. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS letters*. Dec 10 1990;276(1-2):172-174.
44. Blythe MJ, Flower DR. Benchmarking B cell epitope prediction: underperformance of existing methods. *Protein science : a publication of the Protein Society*. Jan 2005;14(1):246-248.
45. Yang X, Yu X. An introduction to epitope prediction methods and software. *Reviews in medical virology*. Mar 2009;19(2):77-96.
46. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome research*. 2006;2:2.
47. Vita R, Zarebski L, Greenbaum JA, et al. The immune epitope database 2.0. *Nucleic acids research*. Jan 2010;38(Database issue):D854-862.
48. El-Manzalawy Y, Dobbs D, Honavar V. Predicting linear B-cell epitopes using string kernels. *Journal of molecular recognition : JMR*. Jul-Aug 2008;21(4):243-255.
49. Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins*. Oct 1 2006;65(1):40-48.
50. Caoili SEC. B-cell epitope prediction for peptide-based vaccine design: towards a paradigm of biological outcomes. Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology; 2010; Niagara Falls, New York.
51. Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *Journal of virology*. Sep 1985;55(3):836-839.

52. Bui HH, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC bioinformatics*. 2007;8:361.
53. Novotny J, Handschumacher M, Haber E, et al. Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains). *Proceedings of the National Academy of Sciences of the United States of America*. Jan 1986;83(2):226-230.
54. Karplus PA, Schulz GE. Prediction of chain flexibility in proteins. *Naturwissenschaften*. 1985/04/01 1985;72(4):212-213.
55. Parker JM, Guo D, Hodges RS. 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*. Sep 23 1986;25(19):5425-5432.
56. Tenzer S, Peters B, Bulik S, et al. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cellular and molecular life sciences : CMLS*. May 2005;62(9):1025-1037.
57. Larsen MV, Lundegaard C, Lamberth K, et al. An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *European journal of immunology*. Aug 2005;35(8):2295-2303.
58. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC bioinformatics*. 2007;8:424.
59. Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC bioinformatics*. 2005;6:132.
60. Thevenet P, Shen Y, Maupetit J, Guyon F, Derreumaux P, Tuffery P. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic acids research*. Jul 2012;40(Web Server issue):W288-293.
61. Maupetit J, Derreumaux P, Tuffery P. PEP-FOLD: an online resource for de novo peptide structure prediction. *Nucleic acids research*. Jul 2009;37(Web Server issue):W498-503.
62. Biasini M, Bienert S, Waterhouse A, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research*. July 1, 2014 2014;42(W1):W252-W258.
63. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*. 1993;26(2):283-291.
64. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*. Jul 2007;35(Web Server issue):W407-410.
65. Patronov A, Doytchinova I. *T-cell epitope vaccine design by immunoinformatics*. Vol 32013.
66. Morris GM, Huey R, Lindstrom W, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry*. Dec 2009;30(16):2785-2791.

67. Morris GM, Goodsell DS, Halliday RS, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of computational chemistry*. 1998;19(14):1639-1662.
68. Discovery Studio Modeling Environment, Release 4.0, San Diego: Accelrys Software Inc., 2013.
69. Vlieghe P, Lisowski V, Martinez J, Khrestchatisky M. Synthetic therapeutic peptides: science and market. *Drug Discovery Today*. 1// 2010;15(1-2):40-56.
70. Gupta S, Kapoor P, Chaudhary K, et al. In silico approach for predicting toxicity of peptides and proteins. *PloS one*. 2013;8(9):e73957.
71. Gao J, Faraggi E, Zhou Y, Ruan J, Kurgan L. BEST: Improved Prediction of B-Cell Epitopes from Antigen Sequences. *PloS one*. 2012;7(6):e40104.
72. Peters B, Bulik S, Tampe R, Van Endert PM, Holzhutter HG. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *Journal of immunology (Baltimore, Md. : 1950)*. Aug 15 2003;171(4):1741-1749.
73. Thundimadathil J. Cancer treatment using peptides: current therapies and future prospects. *Journal of amino acids*. 2012;2012:967347.
74. Eaton BT, Broder CC, Middleton D, Wang LF. Hendra and Nipah viruses: different and dangerous. *Nature reviews. Microbiology*. Jan 2006;4(1):23-35.
75. Pica N, Palese P. Toward a universal influenza virus vaccine: prospects and challenges. *Annual review of medicine*. 2013;64:189-202.
76. Jones EV, Klepfer S, Miller TJ, Reed AP. Universal coronavirus vaccine. Google Patents; 1993.
77. Naz RK, Dabir P. Peptide vaccines against cancer, infectious diseases, and conception. *Frontiers in bioscience : a journal and virtual library*. 2007;12:1833-1844.
78. Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. *Nat Rev Drug Discov*. 05//print 2007;6(5):404-414.
79. Shrestha B, Diamond MS. Role of CD8+ T cells in control of West Nile virus infection. *Journal of virology*. Aug 2004;78(15):8312-8321.
80. Greenbaum JA, Andersen PH, Blythe M, et al. Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. *Journal of molecular recognition : JMR*. Mar-Apr 2007;20(2):75-82.
81. Olesen OF, Lonnroth A, Mulligan B. Human vaccine research in the European Union. *Vaccine*. Jan 29 2009;27(5):640-645.