Learning from COVID-19 Pandemic: In Silico Vaccine and Cloning Design Against Nipah Virus by Studying and Analyzing the Whole Nipah Virus Proteome.

Peter T. Habib (✉ p.habib911@gmail.com)

Colors Medical Laboratories

Abstract

Nipah virus (NiV) is a zoonotic paramyxovirus of the Henipavirus genus first identified in Malaysia in 1998. Henipavirus have bat reservoir hosts and have been isolated from fruit bats found across Oceania, Asia, and Africa. Bat-to-human transmission is thought to be the primary mode of human NiV infection, although multiple intermediate hosts are described. Human infections with NiV were originally described as a syndrome of fever and rapid neurological decline following contact with swine. More recent outbreaks describe a syndrome with prominent respiratory symptoms and human-to-human transmission. Nearly annual outbreaks have been described since 1998 with case fatality rates reaching greater than 90%. To prevent the spreading of the Nipah virus and turning it into a new pandemic, we must be armed with a ready-made vaccine to save the time consuming that vaccine takes until production. Here we in this paper, we analyzed the whole Nipah virus proteome to find out the most antigenic, non-allergic, and immune inducing epitopes to construct different vaccines that undergone deep investigation to reveal the most appropriate vaccine to immunize humanity from this probably pandemic.

Introduction

Nipah virus (NiV) is a member of the Henipavirus genus of the Paramyxoviridae family and is a zoonotic virus with a high case fatality rate [1]. Our knowledge of the geographic distribution of NiV and the disease it causes, mode of pathogen transmission, and clinical manifestations of infection, have evolved. The first recognized human infection was in the Malaysian village of Kampung Sungai Nipah in 1998, initiating a deadly outbreak that lasted through 1999 [1,2]. Smaller sporadic outbreaks have since recurred nearly annually within South Asia with case fatality rates reaching greater than 90% [3,4]. The original human NiV infections were found to be associated with contact with swine, and it was later confirmed that NiV could be isolated from the nose and oropharynx of pigs [1,2,5,6]. Human infections were characterized by fever for up to 14 days, meningitis, and/or encephalitis, with rapid neurological decline and progression to coma within 24 to 48 hours [5]. Later outbreaks outside the Malay peninsula have been characterized by different transmission dynamics and clinical presentation, including the development of severe respiratory symptoms in addition to neurological complications, with human infection traced to the consumption of horse meat, proximity to other infected humans, and ingestion of raw date palm sap contaminated with the bodily fluids of bats [3,7−14]. Like lyssaviruses, filoviruses, coronaviruses, and the related Hendra virus, NiV is naturally hosted by pteropid bats [15,16]. Fruit bats found across Oceania, South and Southeast Asia, and sub-Saharan Africa are the natural reservoirs of NiV, and almost yearly outbreaks of NiV infections continue to occur throughout South and Southeast Asia [16,17]. The high case fatality rate associated with NiV infection, ubiquitous nature of the reservoir host, increasing deforestation, and expanding modes of transmission coupled with a lack of rapid diagnostics and effective vaccine or therapeutic agents make Nipah virus' pandemic potential increasingly relevant.

Nipah virus belongs to the family Paramyxoviridae—the family that also comprises the human pathogenic viruses HeV, measles virus, mumps virus, respiratory syncytial virus, and human parainfluenza virus. Paramyxoviruses possess a single-stranded, nonsegmented, negative-sense RNA genome fully encapsulated by envelope proteins including a cell receptor binding protein—glycoprotein (G) of henipaviruses, hemagglutinin (H), or hemagglutinin/neuraminidase (HN)—and a separate fusion (F) protein. While cells infected with NiV react strongly with HeV antiserum, cross-neutralization studies have revealed a difference in neutralizing antibodies indicating that NiV and HeV are distinct viruses. Investigations comparing multiple gene regions of NiV to those of other paramyxoviruses have confirmed that HeV and NiV make up a distinct cluster within the Paramyxoviridae family, now classified as the Henipavirus genus [1]. NiV infections are characterized by endothelial syncytium formation, which results in inflammation and hemorrhage. It is postulated that the NiV-G and NiV-F proteins are physically associated and viral fusion results from conformational changes that arise following receptor binding [18]. Guillaume et al. found evidence of a NiV-G protein residue E533 with an important role in receptor binding and similar in structure and function to that of the measles virus attachment hemagglutinin residue R533 [18]. Via 3D modeling of the NiV-G protein, the team identified six other protein residues (W504, E505, N557, Q530, T531, and A532) with what seemed to be important roles in promoting viral fusion as well as in binding to ephrin-B2, a functional receptor for NiV found on epithelial cells and neurons [18−20]. Aguilar and colleagues found that the NiV-F protein is glycosylated at multiple sites, reducing fusion efficiency when compared to mutated F proteins and in contrast to N-glycan function in other paramyxoviruses. However, the authors discovered that the N-glycans of NiV may play a role in protecting the protein from neutralizing antibodies [21]. It was also found that the NiV-G and F proteins are not only necessary for binding and fusion to the host cell, but they are also able to facilitate viral budding. However, their function in this process is minor compared to that of the viral matrix protein M which seems to be integral to the viral organization and budding [22].

Material And Methods

The current experiment was conducted to develop potential vaccines against the Nipah virus, by exploiting the strategies of reverse vaccinology (Figure 01). The methods and python scripts used in this experiment were taken and adapted from the works of Peter, et.al [23,24], and the material was downloaded from NCBI (ID: NC_002728).

Results

Identification, selection, and retrieval of Nipah protein sequences. The Nipah Virus proteins were downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov/). nine protein sequences i.e., nucleocapsid protein (ID: NP_112021.1), P phosphoprotein (ID: NP_112022.1), V protein (ID: NP_112023.1), W protein (ID: YP_007188592.1), C protein (ID: NP_112024.1), matrix protein (ID: NP_112025.1), fusion protein (ID: NP_112026.1), attachment glycoprotein (ID: NP_112027.1), and polymerase (ID: NP_112028.1) were selected for the possible vaccine construction and retrieved from the NCBI database in protein FASTA format. Table (1) lists the proteins sequences with their NCBI accession numbers.

Table 1

ID	Protein
NP_112021.1	Nucleocapsid protein
NP_112022.1	P phosphoprotein
NP_112023.1	V protein
YP_007188592.1	W protein
NP_112024.1	C protein
NP_112025.1	matrix protein
NP_112026.1	fusion protein
NP_112027.1	attachment glycoprotein
NP_112028.1	polymerase

Cytotoxic T lymphocytes epitope prediction. Production of the vaccine subunits, cytotoxic T lymphocyte (CTL) epitope prediction is very important in immune system stimulation. The NetCTL 1.2 server (https://www.cbs.dtu.dk/services/NetCTL/) was used to predict CTL epitopes [25]. MHC-I binding affinity, proteasomal C terminal cleavage, and TAP (Transporter Associated with Antigen Processing) is a major influencing factor in the selection of epitopes. The default thresholds for the CTL prediction were set as is at 0.05, 0.15, and 0.75 for the parameters such as TAP transport efficiency, proteasomal C-terminal cleavage, and epitope identification. The predicted epitopes were classified by the combined value.

Prediction of Helper T lymphocytes epitope. To predict Helper T lymphocytes (HTL) epitopes, we used The IEDB MHC II server (http://tools.iedb.org/mhcii/) [26]. The species/locus was chosen as Human/ HLA-DR, and a 7-allele human leukocyte antigen (HLA) reference set was selected for the HTL epitopes prediction. By default, the 15-meter length of the epitopes was selected, and generated epitopes were classified according to the percentile value. The percentile rank is given after analyzing the epitope score with 5 million peptides with length 15-mer from the SWISSPROT database, compounds with the highest percentile score show a high affinity of MHC-II.

Prediction of interferon-gamma-inducing epitopes. HTLs filtered from the previous step were undergone further analysis to identify whether they can stimulate interferon-gamma (IFN-γ) immune response by using the (15-mer) IFNgamma epitope server (http://crdd.osdd.net/raghava/ifnepitope/scan.php). Server prediction based on Support Vector Machine (SVM) and epitopes were predicted by selecting IFN-γ versus non-IFN-γ stimulating epitopes [27]. Finally, epitopes with positive results for the IFN-γ response go for the in-silico vaccine construction.

Prediction of linear B-cell epitopes. B lymphocytes that produce antibodies can be activated alongside stimulation of humoral immune response. We used ABCpred servers (http://www.imtech.res.in/ raghava/abcpred/) [28] to predict epitopes that initiate the humoral immune response to activate B-cells. 16-mer was the window length, based on a recurrent neural network with a 0.51 threshold value, keeping overlapping filter on. Because of a large number of epitopes predicted, only epitopes having a score of more than 0.9 were only chosen for the construction of the candidate vaccine.

Prediction of antigenicity of protein sequences. It is well-known that significant features of vaccine building blocks must have antigenic properties. ANTIGENpro and VaxiJen v2.0 both were used to predict the antigenicity of epitopes. ANTIGENpro (http://scratch.Proteomics.ics.uci.edu/) uses microarray data to calculate protein antigenicity. Its accuracy with the combined dataset was estimated to be 76% based on cross-validation experiments [29]. VaxiJen 2.0 server (http://www.ddg-pharmfac.net/Vaxijen/ VaxiJen/VaxiJen.html) is also a server to predict the antigenicity with the virus as selected target organism and threshold value of ≥0.4, only probable antigen epitopes were selected for the vaccine construction. VaxiJen 2.0 server is based on auto and cross-covariance (ACC) to evaluate the antigenicity of the vaccine. The VaxiJen algorithm is mainly based on the method of sequence alignment and analyzes protein physicochemical properties to identify them as antigenic [30].

Prediction of allergenicity and toxicity of protein sequences. Allergenicity is a vital factor in the development of the vaccine. AllerTOP v.2.0 and AllergenFP servers evaluate the allergenicity of the proteins. AllerTOP v2.0 an online server (http://www.ddg-pharmfac.net/AllerTOP) develops the k nearest neighbors (kNN), auto- and cross-covariance (ACC) transformation, and amino acid E-descriptors machine learning model for the classification of peptides by analyzing the physiochemical properties of proteins. Its accuracy may reach 85.3% at fivefold cross-validation [31]. AllergenFP (http://ddg-pharmfac.net/AllergenFP/) is an alignment-free based on fingerprint approach for the detection of allergens and non-allergens. The fingerprinting approach is based on a four-step algorithm. firstly, the protein sequences are defined in terms of their properties, including size, hydrophobicity, relative abundance, α helix, and β-strand forming propensities. produced strings are then transformed into vectors of equal size by ACC. The vectors were converted into binary fingerprints and measured according to the Tanimoto coefficient. This method's accuracy reaches 88% and Non-allergen epitopes were chosen. Finally, all the epitopes were checked for toxicity using the ToxinPred server (https://webs.iiitd.edu.in/ raghava/toxinpred/multi_submit.php) [32], and non-toxic epitopes were chosen. All epitope that passes the previous analysis is used for vaccine construction.

Construction of multi-epitope vaccines candidate sequence. Selected CTLs, HTLs, and B-cell epitopes generated by NetCTL 1.2, IEDB MHC II server, and ABCpred server respectively, were used to construct the vaccine. The linear B-cell and HTL epitopes were linked with GPGPG linker and CTL epitopes by AAY linker. Also, Three different adjuvants have been used for vaccines: beta-defensin, ribosomal protein, L7/L12 protein, and HABA protein to increase the immunogenicity of the vaccine, and linked via EAAAK linker [23] to generate different combinations of epitopes to produce all possible vaccines construct. Different combinations generate almost 500 vaccine constructs. The vaccines have undergone filtering according to highly antigenic, immunogenic, non-toxic, and non-allergenic scores.

Physiochemical properties and solubility prediction. Protparam, which is one of expasy server tools (https://web.expasy.org/ protparam/) was used to identify physicochemical properties like isoelectric point (pI), amino acid number and composition, half-life, instability and aliphatic index, molecular weight (MW), and grand average of hydropathicity (GRAVY) of the vaccine constructs [33]. The multi-epitope vaccine solubility was predicted using the Protein−Sol server (http://protein-sol.manchester.ac.uk). The population average for the experimental is 0.45, and thus solubility value greater than 0.45 is predicted to have a higher solubility [34].

Secondary structure prediction. PSIPRED and RaptorX servers were used To predict secondary structures of the vaccine constructs. PRISPRED (http://bioinf.cs.ucl.ac.uk/psipred/), is an online server secondary structure prediction tool [35]. RaptorX server (http://raptorx.uchicago. edu/StructurePropertyPred/predict/) was also used to calculate the secondary structural characteristics of the vaccine. The server uses Deep Convolutional Neural Fields (Deep CNF) machine learning model to calculate secondary structure (SS), disorder regions (DISO), and solvent accessibility (ACC) [36].

Tertiary structure prediction. A Three-dimensional (3D) model of the vaccines was generated using the homology modeling GalaxyWEB server (http://galaxy.seoklab.org/). The GalaxyWEB server is a platform for computerized protein structure and function prediction based on the sequence-to-structure-to-function approach and identifies protein structure with high similarity pattern from the Protein Data Bank (PDB) [37]. GalaxyWEB generates 3D atomic models by performing several sequence alignments and iterative structure assembly. GalaxyWEB automatically refined the 3D model obtained for the vaccines. The refinement approach was experimentally approved in CASP10 based refinement experiments. This provides the relaxation of the structure for repacking and molecular dynamics simulation [38].

Validation of tertiary structure. 3D structure validation is a critical stage of the model construction method because it discovers probably disorders in 3D models predicted [39]. ProSA-web server (https://prosa .services.came.sbg.ac.at/prosa.php) was used for protein 3D structure validation, which calculates protein quality, which is shown in the form of Z plot. If the Z scores are outside the range of the properties for native proteins, it specifies that the structure likely contains errors [40]. To investigate nonbonded atom-atom interactions associated with the ERRAT web-server (http://services.mbi.ucla.edu/ERRAT /) was also used to predict high-resolution crystallography structures. A Ramachandran plot was retrieved via RAMPAGE web-server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) and describe the quality of the modeled structure by displaying the percentage of residues in disallowed and allowed regions [41].

Prediction of discontinuous B-cell epitopes. ElliPro, an online tool (http://tools.iedb.org/ellipro/), was used to predict the validated discontinuous (conformational) B-cell epitopes of the 3D structure. ElliPro uses ellipsoid, measure the residue PI, and adjacent cluster residues to calculate the protein shape. ElliPro provides the user with a score for each epitope termed as an average PI value. PI value of 0.9 means that contains (90%) protein residues while the remaining (10%) residues are outside. ElliPro has the top predictions and provided an area under curve (AUC) value of (0.732).

Molecular docking of the final vaccine with the immune receptor. It is based on the interaction between an antigenic ligand and the immune receptor to produce an immune response. The toll-like receptor or TLR3 (PDB ID: 1ZIW) was retrieved from Protein Databank (PDB) (https://www.rcsb.org). Online servers ClusPro 2.0 (https://cluspro.bu.edu/login.php), HADDOCK server (https://haddock.science.uu.nl/), PatchDock server (https://bioinfo3d.cs.tau.ac.il/PatchDock/php.php), and FireDock server (http://bioinfo3d.cs.tau.ac.il/FireD ock/php.php) were used for molecular docking and docking refinement between the vaccine and TLR3 [42]. Again, the docking was performed for the third time using the HawkDock server (http://cadd.zju.edu.cn/hawkdock/), and subsequently, calculate the Molecular Mechanics/Generalized Born Surface Area (MM-GBSA) score using the same server that predicts the result the affinity score and the lowest score is the better score [43].

Molecular dynamics simulation. After filtering of bad docking scores, The molecular dynamics simulation (MD) study was performed for the vaccines of choice that showed the best molecular docking results. The iMODS web-server (http://imods .Chaconlab.org/) was used for the MD study, a fast, free-accessible, and molecular dynamics simulation server for define, calculating the protein flexibility, and deformability [44].

Immune stimulation. C-ImmSim, an online simulation server (http://150.146.2.1/C-IMMSI M/index.php) predicts the vaccine constructs immune response. C-ImmSim determines the humoral and cellular immune response to the vaccine construct [45]. Three injections of the target vaccine Nipah virus vaccine were administered at different intervals of 4 weeks. All simulation parameters were established at default parameters with periods set at 1, 84, and 168. The volume of simulation and the steps of the simulation were set at 50 and 1000, respectively. And the random seed was 12345 with vaccine injection.

Codon optimization and in silico cloning. A codon optimization approach was adopted to improve the expression of recombinant proteins. Optimization of the codon is essential because genetic code degeneration allows for the encoding of most amino acids by multiple codes. Java Codon Adaptation Tool (JCat) The E.coli (strain K12) codon system was used with the server (http://www. prodoric.de/JCat) to determine protein expression levels by acquiring the CAI values and GC content. The optimum CAI value is 1.0, while>0.8 is somehow a good score, and the GC content range from 30 to 70% is also good because translation and transcriptional efficiency have unfavorable effects across this range [46]. The optimized gene sequence of the vaccine has been cloned in E.coli. The N and C-terminals for the sequence were each added to the E.coli plasmid vectors pET-30a (+) considering NdeI and HindIII restriction sites. Finally, the optimized sequence of the final vaccine construct (with restriction sites) was inserted into the plasmid vector pET-30a (+) using the SnapGene software (https://www. snapgene.com/free-trial/) to confirm the expression of the vaccine.

Table 2: the candidate proteins physiochemical analysis

Protein Name	Antigenicity Score	Antigenicity	Length	Allergenicity	Half-life time	GRAVY Score	The instability index (II)	Molecular weight	Theoretical pI
Nucleocapsid protein	**0.5713**	Antigen	532	NON-ALLERGEN	30 h	-0.236	52.33	58168.07	6.06
P phosphoprotein	**0.5866**	Antigen	709	NON-ALLERGEN	30 h	-0.730	48.52	78302.51	4.61
V protein	**0.6252**	Antigen	456	NON-ALLERGEN	30 h	-0.816	60.47	50325.44	4.66
W protein	**0.6199**	Antigen	449	NON-ALLERGEN	30 h	-0.827	57.56	49464.67	4.84
C protein	**0.3300**	Non-Antigen	166	NON-ALLERGEN	30 h	-0.345	43.67	19735.59	9.44
matrix protein	**0.4033**	Antigen	352	NON-ALLERGEN	30 h	-0.211	29.53	39928.28	9.31
fusion protein	**0.5012**	Antigen	546	NON-ALLERGEN	30 h	0.195	38.00	60281.96	5.85
attachment glycoprotein	**0.5110**	Antigen	602	ALLERGEN	30 h	-0.178	34.56	67039.03	8.58
polymerase	**0.4757**	Antigen	2244	NON-ALLERGEN	30 h	-0.286	41.87	257232.51	7.53

Table 3: List of the CTL epitopes selected with all the antigenicity, non-allergenicity, and non-toxicity criteria.

Peptide	MHC binding affinity	Rescale binding affinity	C-terminal cleavage affinity	Transport efficiency	Prediction score	Antigenicity	Toxicity	Allergenicity
ETDDYNGIY	0.8128	3.451	0.9526	2.571	3.7225	Antigenic	Non-toxic	Non-allergen
VSNTSKHTY	0.5703	2.4214	0.9448	2.981	2.7122	Antigenic	Non-toxic	Non-allergen

Table 4: List of the epitopes selected from MHC-II which are antigenic, non-allergenic, non-toxic, and can induce IFN-γ immune response.

Allele	Start	End	Length	Peptide sequence	Percentile score	Method	IFN-γ Result	Antigenicity	Toxicity	Allergenicity
HLA-DRB3*02:02	474	488	15	PLVVNWRNNTVISRP	1	NetMHCIIpan	Positive	Antigenic	Non-toxic	Non-allergen
HLA-DRB1*15:01	503	517	15	ASLCIGLITFISFII	0.44	Consensus (smm/nn/sturniolo)	Positive	Antigenic	Non-toxic	Non-allergen
HLA-DRB1*15:01	507	521	15	IGLITFISFIIVEKK	0.44	Consensus (smm/nn/sturniolo)	Positive	Antigenic	Non-toxic	Non-allergen
HLA-DRB1*15:01	504	518	15	SLCIGLITFISFIIV	0.44	Consensus (smm/nn/sturniolo)	Positive	Antigenic	Non-toxic	Non-allergen
HLA-DRB3*01:01	1101	1115	15	DLELASFLMDRRVIL	0.51	Consensus (comb.lib./smm/nn)	Positive	Antigenic	Non-toxic	Non-allergen

Table 5: Only the final building vaccine selects linear B-cell epitopes, with a binding score greater than 0.93.

Name	Sequence	Start position	Score	Allergenicity	Antigenicity Score	Antigenicity	Toxicity
polymerase	SYMIYLMNWCDFKKSP	1631	0.95	NON-ALLERGEN	0.4189	ANTIGEN	Non-Toxic
Pphosphoprotein	KGKGERKGKNNPELKP	581	0.95	NON-ALLERGEN	1.4193	ANTIGEN	Non-Toxic
polymerase	AALIPAPIGGFNYLNL	989	0.93	NON-ALLERGEN	0.7346	ANTIGEN	Non-Toxic
attachmentglycoprotein	SFSWDTMIKFGDVLTV	457	0.93	NON-ALLERGEN	0.8741	ANTIGEN	Non-Toxic
Pphosphoprotein	HWSIERSISPDKTEIV	365	0.93	NON-ALLERGEN	0.6304	ANTIGEN	Non-Toxic
polymerase	NHLIYDPDPVSEIDCS	1454	0.92	NON-ALLERGEN	0.5988	ANTIGEN	Non-Toxic
fusionprotein	KSSIESTNEAVVKLQE	148	0.94	NON-ALLERGEN	0.6356	ANTIGEN	Non-Toxic

Discussion

Protein sequences and PDB files. The amino acid sequence of the Nipah Virus was retrieved from the NCBI database (Table 1) in a FASTA format. Hence, then the functional sequences for the nine proteins have undergone the prediction of MHC-II, Cytotoxic T Lymphocytes, and B-cell epitopes to develop a novel subunit vaccine against Nipah Virus. Nine proteins analyzed for physicochemical properties as shown in (Table 2).

Cytotoxic T lymphocytes epitope Prediction. For the nine proteins, 159 CTL (9-mer) epitopes were predicted using the NetCTL 1.2 web-server at the 0.4 default threshold value. Out of all these predicted CTL epitopes, only two epitopes were chosen based on their high scores binding affinity with MHC-I, antigenicity, non-allergenicity, and non-toxicity, as illustrated in (Table 3).

Helper T lymphocytes epitope prediction. The IEDB MHC-II Web Server has determined HTL predicted epitopes as highly binding epitopes for human alleles HLADR. Out of the 357 epitopes forecast for the final vaccine, a total of seven HTL epitopes were selected, as illustrated in (Table 4), based on their binding affinity, antigenicity, and non-toxicity. Human alleles of predicted epitopes are HLA-DRB3*02:02, HLA-DRB1*15:01, HLA-DRB1*15:01, HLA-DRB1*15:01, and HLA-DRB3*01:01.

Interferon-gamma-inducing epitopes prediction. In intracellular pathogen evasion, the IFN gamma plays a significant role and mainly acts as cytokines for cytotoxic T lymphocytes and natural killer cells. The inducing epitopes of the IFN-α predicted using the Support Vector Machine (SVM) technique. The vaccine construction selected four HTL epitopes with IFN-α positive results.

Prediction of linear B-cell epitopes. To predict B cell epitopes, the ABCpred server was used. A total of seven B-cell epitopes, which were finally selected for vaccine construction. high antigenic, non-allergic, and non-toxic, are listed in (Table 5).

Multi-epitope vaccines construction. A novel vaccine was designed with seven B cell epitopes, seven HTL epitopes, and two CTL epitopes that have significant bindings, antigenicity, non-toxicity, and non-allergenicity scores. To improve immunogenicity, the PADRE sequence also improves the CTL response, thus ensuring a potent immune response [47] Three different adjuvants have been used for vaccines: beta-defensin, ribosomal protein, L7/L12 protein, and HABA protein to increase the immunogenicity of the vaccine applied to both the N and C terminals of the vaccine. EAAAK linkers link adjuvant to the epitopes, GPGPG linkers were used to link B-cell and HTL epitopes, and AAY linkers to link CTL epitopes. Antigenicity, non-allergenicity, non-toxicity, solubility, and all criteria have been tested again in the built-up vaccine sequence. The schematic presentation of the current study's final multi-epitope vaccine peptide is defined in the (Figure 1).

Antigenicity and allergenicity prediction of the vaccine's candidate. The 30 vaccine constructs were investigated for the antigenicity by ANTIGENpro and VaxiJen v2.0 web tool, and it was found that 13 of them could be good antigens. In the VaxiJen v2.0 tool, the default threshold of 0.4 was chosen with the virus as an organism. The vaccines NipVac-1, NipVac-2, NipVac-3, NipVac-7, NipVac-8, NipVac-9, NipVac-10, NipVac-16, NipVac-21, NipVac-22, NipVac-26, NipVac-28, NipVac-30 showed an antigenicity score of 0.4298, 0.4644, 0.5142, 0.4584, 0.4197, 0.498, 0.4902, 0.4758, 0.4369, 0.4386, 0.5082, 0.4044, and 0.4554 respectively. The vaccine sequence on both the AllerTOP v.2 and AllergenFP servers is both estimated to be non-allergenic.

Prediction of physicochemical properties and solubility. The final protein's molecular weight was estimated to be ranging from 60 kDa to 80 kDa with a theoretical isoelectric point (pI) score ranging from 5 to 9. It was estimated that the half-life was 1 h in mammalian cells in vitro and more than 30 h in yeast and about 10 h in E. coli in vivo. The instability index (II) value was ranging from 20 to 39, suggesting the protein is extremely stable. (II of>40: instability). The high aliphatic index score ranging from 72 to 90 indicates thermos stability[48]. The Grand average of hydropathicity was found a negative score, which shows a hydrophilic nature of the vaccine constructs.

Prediction of secondary structure and Tertiary structure. α-helix was the most redundant structure in almost all vaccines. Coils come in second, and β-strand was the least. Figures generated for the filtered vaccine are provided in (Figure 2). Tertiary 3D structures designed for the 13 vaccines were predicted by the GalaxyWEB web-server based on ten threading templates (Figure 3). GalaxyWEB auto refined 3D structure after prediction. validation of Tertiary structure. Ramachandran plot analysis was conducted using the RAMPAGE webserver with the refined structure. The VADAR web tool generates almost the same results as Rampag of residues in favored regions and allowed (Figure 4). Both ProSA-web approved the quality and errors in the 3D model (Figures 5)

Discontinuous B-cell epitopes prediction. with values from 0.69 to 0.92 and conformation epitopes ranged in size from 20 to 65 residues, residues predicted to be discontinuous B-cell epitopes. Prediction performed using Ellipro, the score value above 0.75 or more was selected (Figure 1)

Molecular docking. protein-protein docking of the vaccine against the TLR-3 receptor was carried out by various online tools for improving the accurateness of the prediction: i.e., ClusPro 2.0, PatchDock, and HawkDock server. The docked complexes that were created by ClusPro 2.0 and PatchDock tools were further investigated by the PRODIGY tool of HADDOCK web-server and FireDock server, respectively. The PRODIGY tool estimated the binding affinity score (kcal/mol) whereas FireDock predicted the global energy of the docked complexes. However, HawkDock produces ranking scores along with the binding free energy (kcal/mol). The binding free energy was deliberate after the MM-GBSA score in the HawkDock web-server. NipVac-1 was the lowest global energy (-26.47 kcal/mol) shown in (Figure 6), NipVac-16 was the lowest binding free energy (-52.48 kcal/mol), and achieve the highest residue interacted with the receptor (105 members according to ClusPro). Thus, NipVac-1 is then moves to the Molecular dynamic step.

Molecular dynamics simulation. The results of a simulation and analysis of the vaccine structure and docked complex TLR-3 in the normal mode (NMA) are illustrated in (Figure 7). The simulation study was conducted to assess the molecular and atomic movements in the vaccine structure. The deformability graph shows the peaks of the graphs representing the protein regions with deformity (Figure.7C). The eigenvalue of NipVac-1 was 5.29192e-08, respectively as shown in (Figure. 7D). The variance diagram shows the cumulative difference in the color green and the color red. (Figure. 7G). The B Factor Graph shows clearly the relationship between the NMA and the PDB sectors of the docked complex (Figure. 7F). The co-variance map of this complex, in which the correlated motion between a couple of residues is shown by red, white, and blue (Figure. 7B). The complex elastic map shows how the atoms relate to darker grey regions, which shows more steeply (Figure 7E). The NipVac-1 vaccine was used in codon adaptation and cloning design.

Immune stimulation. C-ImmSim studies the successive and effective immune responses of the state of the cell and the memory of immune cells by a mechanism that increases their half-life. The defect of the approach is that few cells substantially increase their half-life and live longer than other cells. ImmSim server immune simulation outcomes confirmed consistency with real immune reactions. The primary response was illustrated by high IgM levels. Also, an increase in the B-cell population was characterized as an increase in immunoglobulin expression (IgG1+IgG2, IgM, and IgG+IgM), resulting in a decrease in antigen concentration (Fig. 8A, H). There is also a clear increase in the population of T (helper) and T C (cytotoxic) cells with memory growth (Fig. 8E, F). IFN-γ production was also identified to have been stimulated after immunization (Fig. 9N). The T cell population results were significantly approachable as the memory developed, and all other immune cell populations were exposed to be consistent.

In-silico codon optimization and cloning. Java Codon Adaption Tool (JCat), for maximum protein expression in E.coli, has been used to optimize the codon use of the vaccine structure. The optimized codon sequence was 1780 nucleotides in length. The CAI value was 0,93, and the GC content average was 56.17 percent, which showed high expression within E. coli plasmid. In the end, the recombinant plasmid sequence was built through the introduction into the plasmid vector pET-30a (+) of the adapted codon sequences using SnapGene (Figure. 9).

Conclusion

New knowledge about Nipah vaccine antigens and new vaccine candidates may result in computational and immunoinformatics approaches presented in the present paper which cannot be acquired simply from pre-clinical, in vitro, and animal studies. The study utilized an immunoinformatics tool to define the highly immunogenic and appropriate carrier vaccine for the Nipah Virus multi-epitope subunit vaccine. Epitopal prediction tools were used to test multiple epitopes of B- and T-cells which were fused to improve immunogenicity with appropriate linkers. The vaccine analysis was shown as very satisfactory to be antigenic, allergic, physicochemical, and tertiary structure properties. TLR-3 and vaccine simulation analysis of molecular dockings and molecular dynamics have been completed to estimate the binding affinity and stability of the complex. Immune cell responses to a rate of antigen clearance are focused on silico immune simulation. The optimization of codon also provided an optimistic CAI value that helps to make in vivo studies soon. We used different immune tools to investigate different vaccine properties in this research. Furthermore, the predicted subunit assessment of the epitopes is highly acceptable to prove as an immunogenic and potential Nipah virus vaccine candidate.

Declarations

Competing interests: The authors have declared that no competing interests exist.

References

1. Chua, K. B., Bellini, W. J., Rota, P. A., Harcourt, B. H., Tamin, A., Lam, S. K., ... & Mahy, B. W. J. (2000). Nipah virus: a recently emergent deadly paramyxovirus. *Science*, *288*(5470), 1432-1435.

2. Enserink, M. (1999). New virus fingered in Malaysian epidemic.

3. Nikolay, B., Salje, H., Hossain, M. J., Khan, A. D., Sazzad, H. M., Rahman, M., ... & Gurley, E. S. (2019). Transmission of Nipah virus—14 years of investigations in Bangladesh. *New England Journal of Medicine*, *380*(19), 1804-1814.

4. Spiropoulou, C. F. (2019). Nipah virus outbreaks: still small but extremely lethal.

5. Centers for Disease Control and Prevention (CDC. (1999). Update: outbreak of Nipah virus–Malaysia and Singapore, 1999. *MMWR. Morbidity and mortality weekly report*, *48*(16), 335-337.

6. Paton, N. I., Leo, Y. S., Zaki, S. R., Auchus, A. P., Lee, K. E., Ling, A. E., ... & Ksiazek, T. G. (1999). Outbreak of Nipah-virus infection among abattoir workers in Singapore. *The Lancet*, *354*(9186), 1253-1256.

7. Chadha, M. S., Comer, J. A., Lowe, L., Rota, P. A., Rollin, P. E., Bellini, W. J., ... & Mishra, A. C. (2006). Nipah virus-associated encephalitis outbreak, Siliguri, India. *Emerging infectious diseases*, *12*(2), 235.

8. Ching, P. K. G., de Los Reyes, V. C., Sucaldito, M. N., Tayag, E., Columna-Vingno, A. B., Malbas Jr, F. F., ... & Foxwell, A. R. (2015). Outbreak of henipavirus infection, Philippines, 2014. *Emerging infectious diseases*, *21*(2), 328.

9. Gurley, E. S., Montgomery, J. M., Hossain, M. J., Bell, M., Azad, A. K., Islam, M. R., ... & Breiman, R. F. (2007). Person-to-person transmission of Nipah virus in a Bangladeshi community. *Emerging infectious diseases*, *13*(7), 1031.

10. Gurley, E. S., Hegde, S. T., Hossain, K., Sazzad, H. M., Hossain, M. J., Rahman, M., ... & Luby, S. P. (2017). Convergence of humans, bats, trees, and culture in Nipah virus transmission, Bangladesh. *Emerging infectious diseases*, *23*(9), 1446.

11. Hsu, V. P., Hossain, M. J., Parashar, U. D., Ali, M. M., Ksiazek, T. G., Kuzmin, I., ... & Breiman, R. F. (2004). Nipah virus encephalitis reemergence, Bangladesh. *Emerging infectious diseases*, *10*(12), 2082.

12. Khan, S. U., Gurley, E. S., Hossain, M. J., Nahar, N., Sharker, M. Y., & Luby, S. P. (2012). A randomized controlled trial of interventions to impede date palm sap contamination by bats to prevent Nipah virus transmission in Bangladesh. *PloS one*, *7*(8), e42689.

13. Luby, S. P., Rahman, M., Hossain, M. J., Blum, L. S., Husain, M. M., Gurley, E., ... & Ksiazek, T. G. (2006). Foodborne transmission of Nipah virus, Bangladesh. *Emerging infectious diseases*, *12*(12), 1888.

14. Luby, S. P., Hossain, M. J., Gurley, E. S., Ahmed, B. N., Banu, S., Khan, S. U., ... & Rahman, M. (2009). Recurrent zoonotic transmission of Nipah virus into humans, Bangladesh, 2001–2007. *Emerging infectious diseases*, *15*(8), 1229.

15. Arinjay, B., Kulcsar, K., Misra, V., Frieman, M., & Mossman, K. (2019). Bats and coronaviruses. *Viruses*, *11*(1), 41.

16. Halpin, K., Hyatt, A. D., Fogarty, R., Middleton, D., Bingham, J., Epstein, J. H., ... & Henipavirus Ecology Research Group. (2011). 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*, *85*(5), 946-951.

17. Chow, V. T., Tambyah, P. A., Yeo, W. M., Phoon, M. C., & Howe, J. (2000). Diagnosis of Nipah virus encephalitis by electron microscopy of cerebrospinal fluid. *Journal of clinical virology*, *19*(3), 143-147.

18. Guillaume, V., Aslan, H., Ainouze, M., Guerbois, M., Wild, T. F., Buckland, R., & Langedijk, J. P. (2006). Evidence of a potential receptor-binding site on the Nipah virus G protein (NiV-G): identification of globular head residues with a role in fusion promotion and their localization on an NiV-G structural model. *Journal of virology*, *80*(15), 7546-7554.

19. Bonaparte, M. I., Dimitrov, A. S., Bossart, K. N., Crameri, G., Mungall, B. A., Bishop, K. A., ... & Broder, C. C. (2005). Ephrin-B2 ligand is a functional receptor for Hendra virus and Nipah virus. *Proceedings of the National Academy of Sciences*, *102*(30), 10652-10657.

20. Negrete, O. A., Levroney, E. L., Aguilar, H. C., Bertolotti-Ciarlet, A., Nazarian, R., Tajyar, S., & Lee, B. (2005). EphrinB2 is the entry receptor for Nipah virus, an emergent deadly paramyxovirus. *Nature*, *436*(7049), 401-405.

21. Aguilar, H. C., Matreyek, K. A., Filone, C. M., Hashimi, S. T., Levroney, E. L., Negrete, O. A., ... & Lee, B. (2006). N-glycans on Nipah virus fusion protein protect against neutralization but reduce membrane fusion and viral entry. *Journal of virology*, *80*(10), 4878-4889..

22. dosSantos Coura, R., & Nardi, N. B. (2007). The state of the art of adeno-associated virus-based vectors in gene therapy. *Virology journal*, *4*(1), 1-7.

23. Habib, P. T. (2021). Vaccine Design, Adaptation, and Cloning Design for Multiple Epitope-Based Vaccine Derived From SARS-CoV-2 Surface Glycoprotein (S), Membrane Protein (M) and Envelope Protein (E): In silico approach.

24. Habib, P. T., Saber-Ayad, M., & Hassanein, S. E. (2021). In Silico Analysis of 716 Natural Bioactive Molecules Form Atlantic Ocean Reveals Candidate Molecule to Inhibit Spike Protein.

25. Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., & Nielsen, M. (2007). Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC bioinformatics*, *8*(1), 1-12.

26. Zhang, Q., Wang, P., Kim, Y., Haste-Andersen, P., Beaver, J., Bourne, P. E., ... & Peters, B. (2008). Immune epitope database analysis resource (IEDB-AR). *Nucleic acids research*, *36*(suppl_2), W513-W518.

27. Dhanda, S. K., Vir, P., & Raghava, G. P. (2013). Designing of interferon-gamma inducing MHC class-II binders. *Biology direct*, *8*(1), 1-15.

28. Saha, S., & Raghava, G. P. S. (2006). Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins: Structure, Function, and Bioinformatics*, *65*(1), 40-48.

29. Magnan, C. N., Zeller, M., Kayala, M. A., Vigil, A., Randall, A., Felgner, P. L., & Baldi, P. (2010). High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics*, *26*(23), 2936-2943.

30. Pandey, R. K., Ojha, R., Aathmanathan, V. S., Krishnan, M., & Prajapati, V. K. (2018). Immunoinformatics approaches to design a novel multi-epitope subunit vaccine against HIV infection. *Vaccine*, *36*(17), 2262-2272.

31. Dimitrov, I., Bangov, I., Flower, D. R., & Doytchinova, I. (2014). AllerTOP v. 2—a server for in silico prediction of allergens. *Journal of molecular modeling*, *20*(6), 1-6.

32. Dimitrov, I., Bangov, I., Flower, D. R., & Doytchinova, I. (2014). AllerTOP v. 2—a server for in silico prediction of allergens. *Journal of molecular modeling*, *20*(6), 1-6.

33. Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., Raghava, G. P., & Open Source Drug Discovery Consortium. (2013). In silico approach for predicting toxicity of peptides and proteins. *PloS one*, *8*(9), e73957.

34. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). The proteomics protocols handbook (pp. 571–608).

35. McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, *16*(4), 404-405.

36. Peng, J., & Xu, J. (2011). RaptorX: exploiting structure information for protein alignment by statistical inference. *Proteins: Structure, Function, and Bioinformatics*, *79*(S10), 161-171.

37. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... & Bourne, P. E. (2000). The protein data bank. *Nucleic acids research*, *28*(1), 235-242.

38. Ko, J., Park, H., Heo, L., & Seok, C. (2012). GalaxyWEB server for protein structure prediction and refinement. *Nucleic acids research*, *40*(W1), W294-W297.

39. Lovell, S. C., Davis, I. W., Arendall III, W. B., De Bakker, P. I., Word, J. M., Prisant, M. G., ... & Richardson, D. C. (2003). Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins: Structure, Function, and Bioinformatics*, *50*(3), 437-450.

40. Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*, *35*(suppl_2), W407-W410.

41. Bhati, S., Kaushik, V. & Singh, J. In silico identifcation of piperazine linked thiohydantoin derivatives as novel androgen antagonist in prostate cancer treatment. Int. J. Pept. Res. Ter. 25, 845–860 (2019).

42. Weng, G., Wang, E., Wang, Z., Liu, H., Zhu, F., Li, D., & Hou, T. (2019). HawkDock: a web server to predict and analyze the protein–protein complex based on computational docking and MM/GBSA. *Nucleic acids research*, *47*(W1), W322-W330.

43. Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert opinion on drug discovery*, *10*(5), 449-461.

44. Lopéz-Blanco, J. R., Garzón, J. I., & Chacón, P. (2011). iMod: multipurpose normal mode analysis in internal coordinates. *Bioinformatics*, *27*(20), 2843-2850.

45. Ali, M., Pandey, R. K., Khatoon, N., Narula, A., Mishra, A., & Prajapati, V. K. (2017). Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Scientific reports*, *7*(1), 1-13.

46. Chen, R. (2012). Bacterial expression systems for recombinant protein production: E. coli and beyond. *Biotechnology advances*, *30*(5), 1102-1107.

47. Wu, C. Y., Monie, A., Pang, X., Hung, C. F., & Wu, T. C. (2010). Improving therapeutic HPV peptide-based vaccine potency by enhancing CD4+ T help and dendritic cell activation. *Journal of biomedical science*, *17*(1), 1-10.

48. Ikai, A. (1980). Thermostability and aliphatic index of globular proteins. *The Journal of Biochemistry*, *88*(6), 1895-1898.
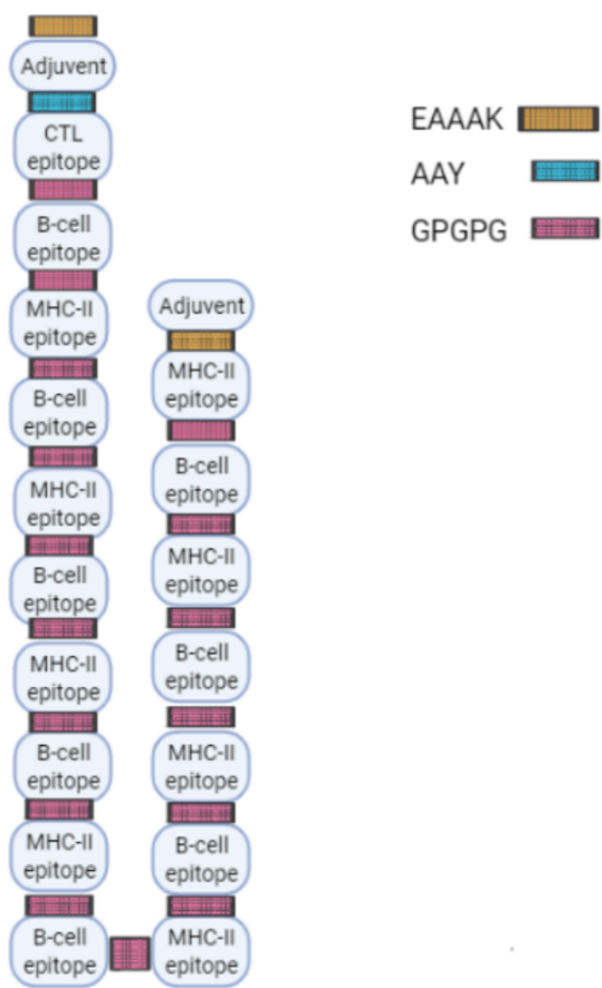
Figures

Figure 1

The final multi-epitope vaccine sequence schematic presentation. A peptide sequence with an adjuvant on both N and C terminals was connected via an EAAAK linker to the multi-epitope sequence (yellow). Use GPGPG linkers (purple) to connect the B cell epitopes and HTL epitopes and CTL epitopes to AAY-connectors (blue).
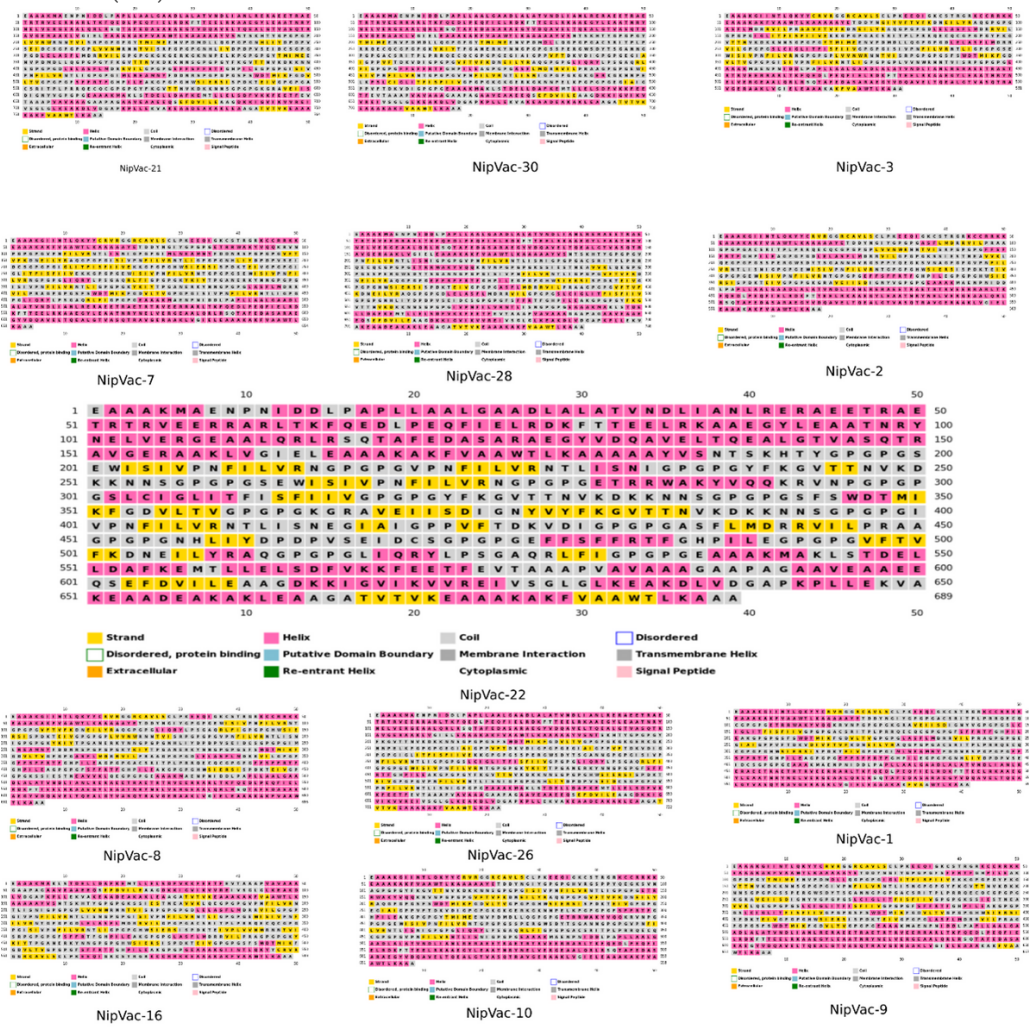

NipVac-21

NipVac-30

NipVac-3

NipVac-7

NipVac-28

NipVac-2

	10		20		30		40		50	
1	E A A A K M A E N P N I D D L P A P L L A A L G A A D L A L A T V N D L I A N L R E R A E E T R A E	50
51	T R T R V E E R R A R L T K F Q E D L P E Q F I E L R D K F T T E E L R K A A E G Y L E A A T N R Y	100
101	N E L V E R G E A A L Q R L R S Q T A F E D A S A R A E G Y V D Q A V E L T Q E A L G T V A S Q T R	150
151	A V G E R A A K L V G I E L E A A A K A K F V A A W T L K A A A A A Y V S N T S K H T Y G P G P G S	200
201	E W I S I V P N F I L V R N G P G P G V P N F I L V R N T L I S N I G P G P G Y F K G V T T N V K D	250
251	K K N N S G P G P G S E W I S I V P N F I L V R N G P G P G E T R R W A K Y V Q Q K R V N P G P G P	300
301	G S L C I G L I T F I S F I I V G P G P G Y F K G V T T N V K D K K N N S G P G P G S F S W D T M I	350
351	K F G D V L T V G P G P G K G R A V E I I S D I G N Y V Y F K G V T T N V K D K K N N S G P G P G I	400
401	V P N F I L V R N T L I S N E G I A I G P P V F T D K V D I G P G P G A S F L M D R R V I L P R A A	450
451	G P G P G N H L I Y D P D P V S E I D C S G P G P G E F F S F F R T F G H P I L E G P G P G V F T V	500
501	F K D N E I L Y R A Q G P G P G L I Q R Y L P S G A Q R L F I G P G P G E A A A K M A K L S T D E L	550
551	L D A F K E M T L L E L S D F V K K F E E T F E V T A A A P V A V A A A G A A P A G A A V E A A E E	600
601	Q S E F D V I L E A A G D K K I G V I K V V R E I V S G L G L K E A K D L V D G A P K P L L E K V A	650
651	K E A A D E A K A K L E A A G A T V T V K E A A A K A K F V A A W T L K A A A	689

| | 10 | 20 | 30 | 40 | 50 | |

▢ Strand	▮ Helix	▢ Coil	▢ Disordered
▢ Disordered, protein binding	▮ Putative Domain Boundary	▮ Membrane Interaction	▮ Transmembrane Helix
▮ Extracellular	▮ Re-entrant Helix	Cytoplasmic	▮ Signal Peptide

NipVac-22

NipVac-8

NipVac-26

NipVac-1

NipVac-16

NipVac-10

NipVac-9

Figure 2

Secondary structure prediction of 13 candidates Vaccines
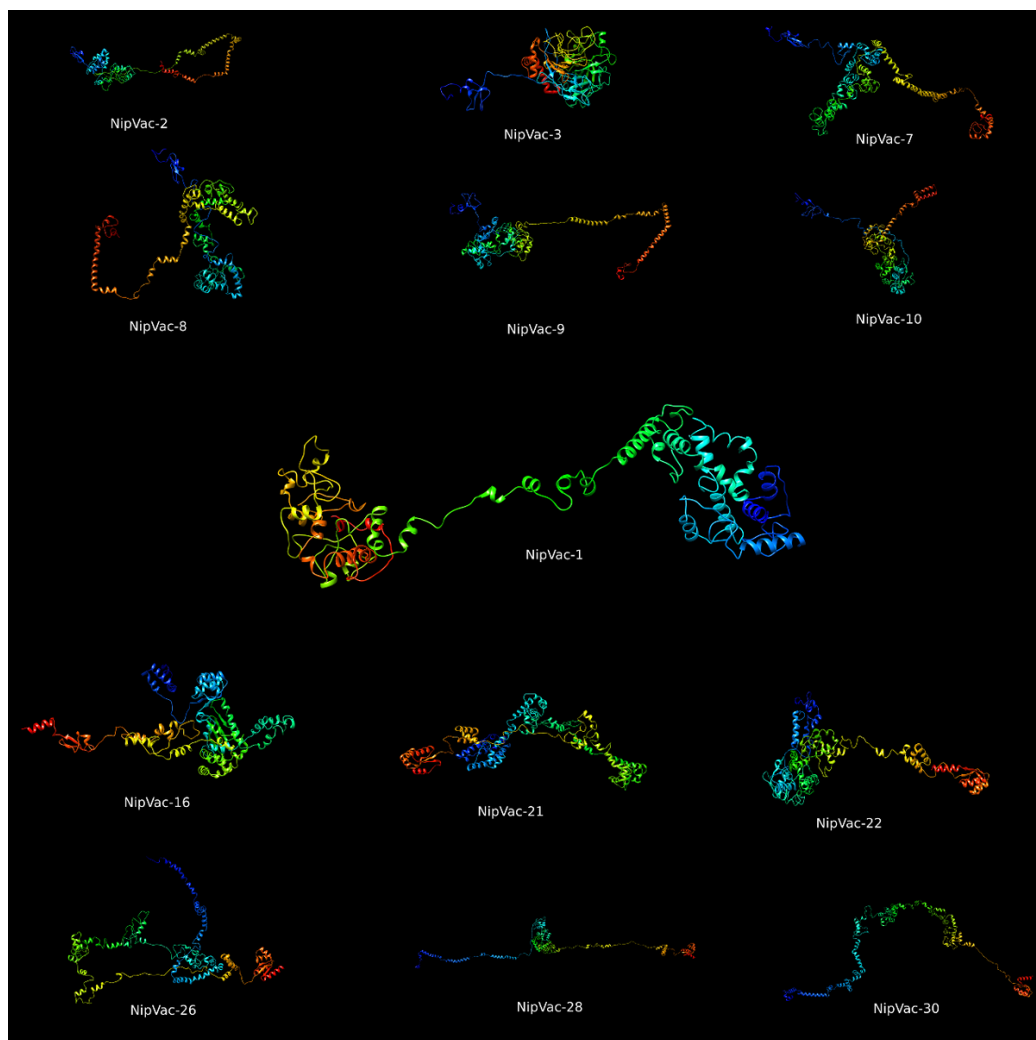
Figure 3

3D Structure of 13 Candidate vaccines
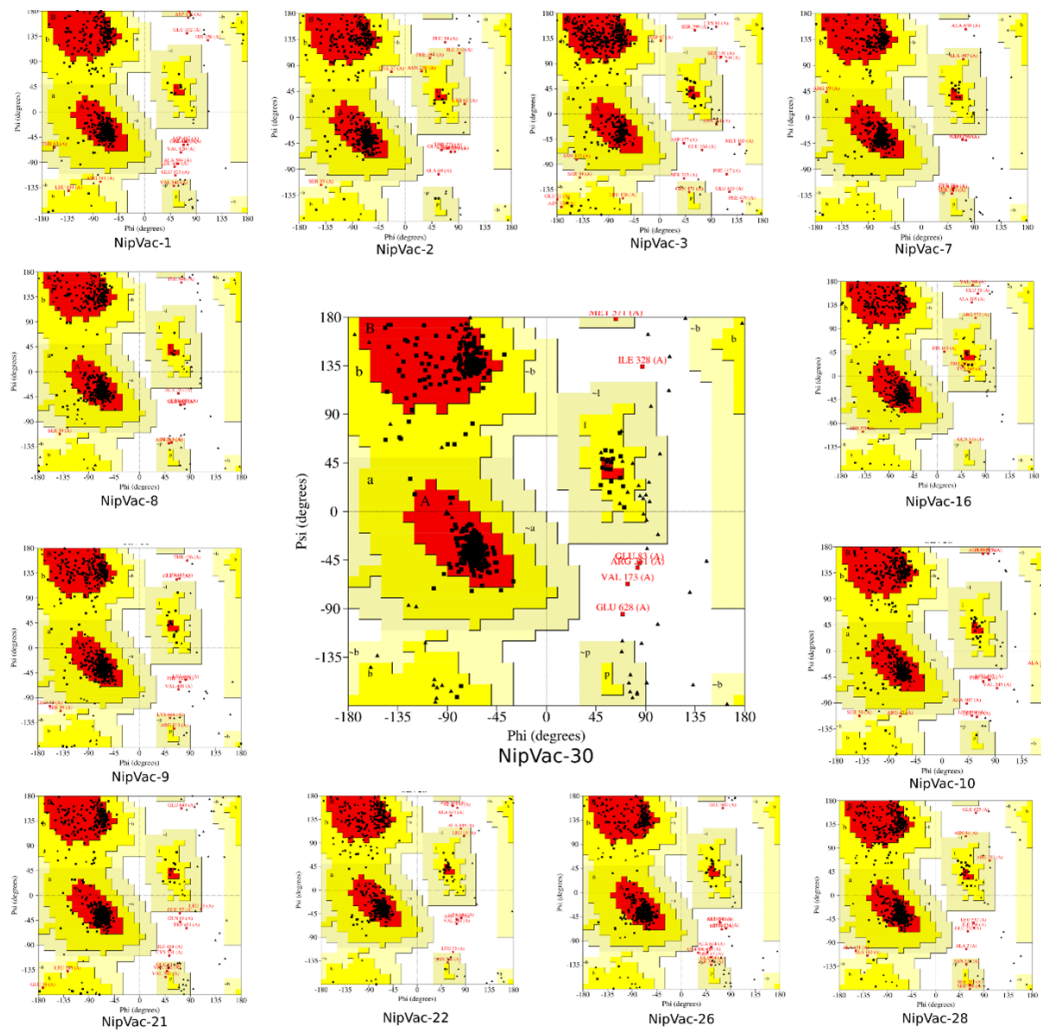
Figure 4

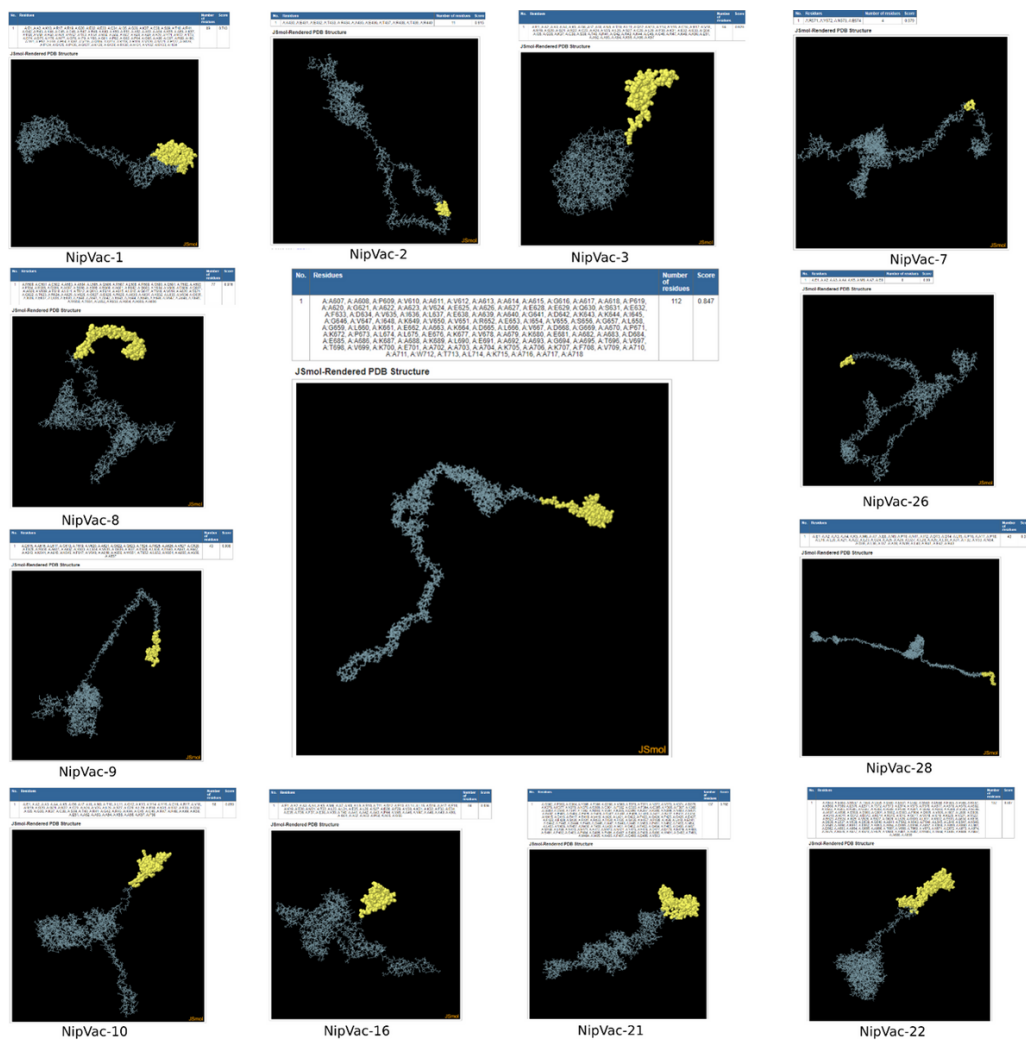Ramachandran plot of the vaccine candidates

NipVac-1

NipVac-2

NipVac-3

NipVac-7

NipVac-8

No.	Residues	Number of residues	Score
1	A:A607, A:A608, A:P609, A:V610, A:A611, A:V612, A:A613, A:A614, A:A615, A:G616, A:A617, A:A618, A:P619, A:A620, A:G621, A:A622, A:A623, A:V624, A:E625, A:A626, A:A627, A:E628, A:E629, A:Q630, A:S631, A:E632, A:F633, A:D634, A:V635, A:I636, A:L637, A:E638, A:A639, A:A640, A:D641, A:D642, A:K643, A:K644, A:I645, A:G646, A:V647, A:I648, A:K849, A:V650, A:V651, A:R652, A:E653, A:I654, A:V655, A:S656, A:G657, A:L658, A:G959, A:L950, A:R651, A:E952, A:K964, A:D665, A:L666, A:V967, A:D968, A:G669, A:A670, A:P671, A:K972, A:P973, A:L974, A:L975, A:E976, A:K977, A:V978, A:A979, A:K980, A:E981, A:A982, A:A983, A:D984, A:E985, A:A986, A:K987, A:A988, A:K989, A:L990, A:E991, A:A992, A:A993, A:G994, A:A995, A:T996, A:V997, A:T996, A:V999, A:K700, A:E701, A:A702, A:A703, A:A794, A:K705, A:A706, A:K707, A:F708, A:V709, A:A710, A:A711, A:W712, A:T713, A:L714, A:K715, A:A716, A:A717, A:A718	112	0.847

NipVac-9

NipVac-26

NipVac-28

NipVac-10

NipVac-16

NipVac-21

NipVac-22

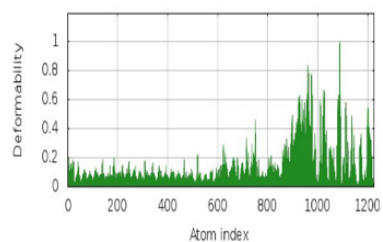**Figure 5**

prediction of interacted amino acid in B-cell epitopes

Figure 6
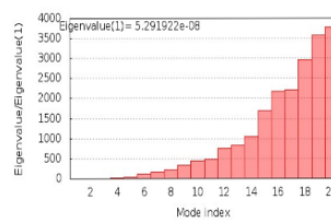
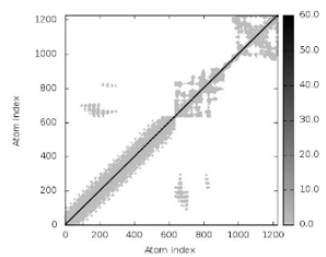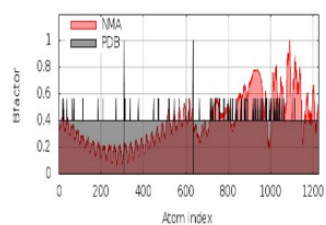molecular docking of NipVac-1 against Toll-Like receptor.

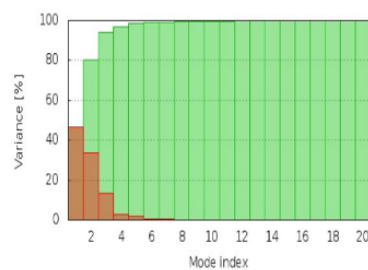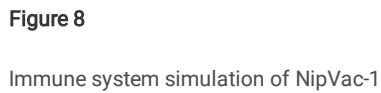**Figure 7**

Molecular Dynamics simulation of NipVac-1

Figure 8
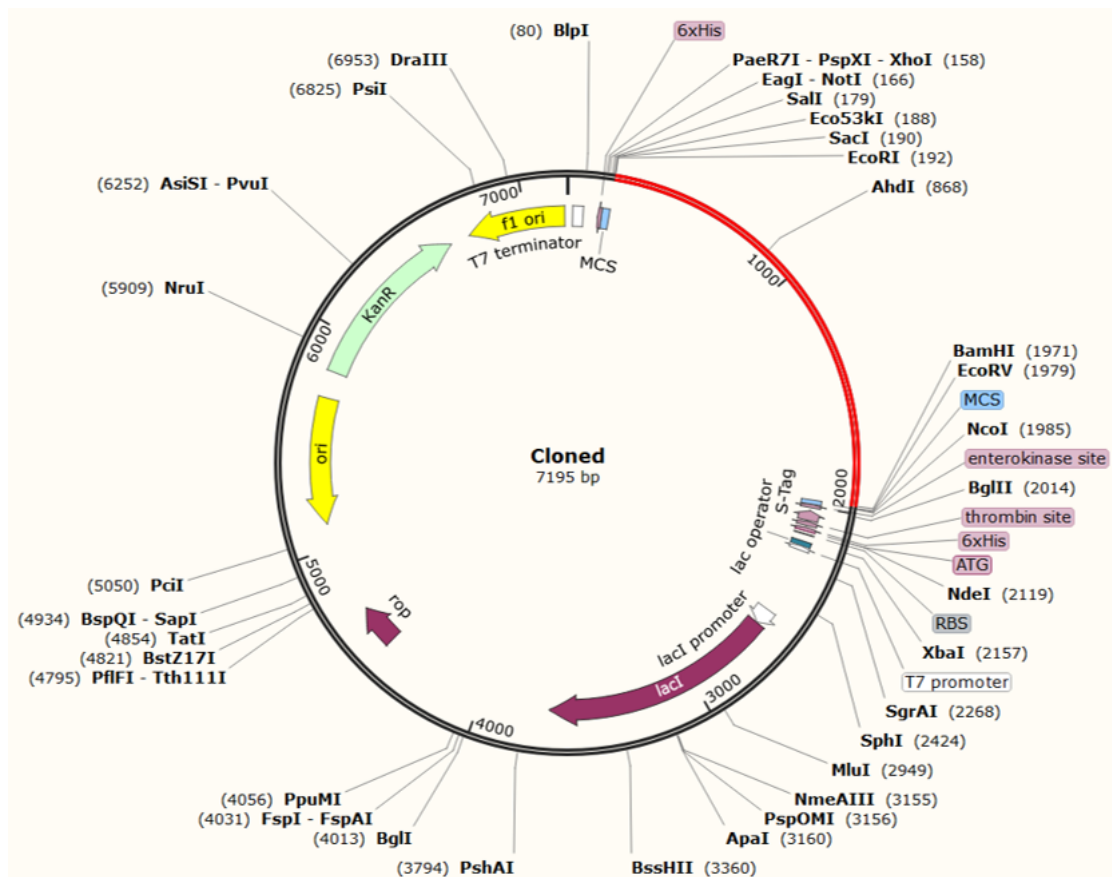
Immune system simulation of NipVac-1

Figure 9

Codon optimization and adaptation of NipVac-1 insert into pET-30a(+) vector