Immunome Research
ISSN: 1745-7580

Idris et al., Immunome Res 2017, 13:3
DOI: 10.4172/1745-7580.1000141

Research Article Open Access

Immunoinformatics Predication and Modelling of a Cocktail of B- and T-cells Epitopes from Envelope Glycoprotein and Nucleocapsid Proteins of Sin Nombre Virus

Idris AB[1*], Mahmoud SM[2], Mohamed Elamin S[3], Mustafa YY[1], Osman AA[4], Adam ME[1], Ali LB[3], Abbas MH[5], Abu-haraz AH[6], Abdelrahman KA[7#] and Salih MA[6#]

[1]*Department of Medical Microbiology, Faculty of Medical Laboratory Sciences, University of Khartoum-Khartoum, Sudan*
[2]*National Public Health Laboratory, Khartoum, Sudan*
[3]*Faculty of Pharmacy, University of Khartoum, Khartoum, Sudan*
[4]*Faculty of Medicine and Health Sciences, Omdurman Islamic University, Khartoum, Sudan*
[5]*Ahfad University for Women School of Pharmacy, Khartoum, Sudan*
[6]*Department of Biotechnology, Africa City of Technology, Khartoum, Sudan*
[7]*University of Medical Science and Technology, Khartoum, Sudan*
[#]*Contributed equally*

Abstract

Sin Nombre virus is a category A pathogen with a reported mortality rate ranging from 30% to 50%. It was responsible for the 2012 Yosemite National Park outbreak. Until now, Specific therapy is not available for the treatment of HCPS caused by SNV. Despite many efforts to develop safe and effective vaccines against SNV, included conventional approaches as well as molecular vaccine approaches, to date there are no vaccines proven to be highly efficacious against SNV. In our study, we analyzed envelope glycoprotein and nucleocapsid of SNV by using immunoinformatics tools housed in IEDB resources; in order to determine the most conserved and immunogenic epitopes for B- and T-cells. Then the predicted epitopes were assessed for the population coverage against the whole world population with the MHC-I and MHC-II restricted alleles. Among predicted epitopes for B-cell, the best candidates for glycoprotein and nucleocapsid were the epitope 743CKKYAYPWQT752 and the epitope 271QVDESKVS278, respectively. For glycoprotein $CD8^+$ T cell predicted epitopes, the epitopes 208MTLPVTCFL216 and 458YTFTSLFSL466 were selected. Interestingly, the best candidates epitopes for nucleocapsid were the epitopes 25YILSFALPI133 and 239FLAARCPFL247 which had high affinity to interact with both MHC classes, I and II, and they had an excellent population coverage for Class I and II alleles throughout the world. To the best of our knowledge, our study for the first time has predicted a cocktail of B- and T-cell epitopes for designing an effective vaccine against HCPS caused by SNV

Keywords: Sin Nombre virus; Epitope-based vaccines; Immune epitope database; In silico tools

Introduction

Hantaviruses (Bunyaviridas, Hantavirus) are rodent-borne emerging viruses [1] which cause over 100,000-200,000 cases per annually worldwide [2-5]. They are broadly classified into the New World Hantaviruses, which includes those causing hantavirus cardiopulmonary syndrome HCPS, and the Old World Hantaviruses, which are associated with another disease, hemorrhagic fever with renal syndrome (HFRS) [2-5]. This classification is determined by the geographic distribution of the rodent reservoirs [4].

HCPS is a severe respiratory disease, rare but often fatal, presents as a wide clinical spectrum ranging from brief febrile prodrome with headache, myalgia and thrombocytopenia, to rapidly progressive pulmonary edema characterized by increased vascular permeability due to dysregulation of the endothelial barrier function. The vascular leakage leads to non-cardiogenic pulmonary edema and in serious cases followed by respiratory failure, multi-organ failure, cardiogenic shock and death [5-10]. A different of pathogenic mechanisms have been suggested, including immune cell-mediated injury, cytokine-mediated injury and enhanced VEGF responses from intercellular junctions resulting from highly specific virus–integrin interactions [8]. Since most deaths are caused by myocardial dysfunction and hypoperfusion rather than hypoxia, some investigators prefer using the term hantavirus cardiopulmonary syndrome rather than the previous term hantavirus pulmonary syndrome [3,10]. All of the etiologic agents associated with cardiopulmonary syndrome, at least 20 hantavirus genotypes, are found in the Western Hemisphere with the primary important etiologic agents include Sin Nombre Hantavirus (SNV) in the United States and Canada [11,12].

SNV is a Category A pathogen [13] that causes approximately 692 cases of HCPS according to the Centers for Disease Control and Prevention, from 1993-2016, [14] with a reported mortality rate ranging from 30% to 50%. [2,3, 6]. SNV is an enveloped virus with a trisegmented, negative-sense RNA genome. The large (L) segment encodes an RNA-dependent RNA polymerase; the medium (M) segment encodes a glycoprotein precursor, GPC, which is co-translationally cleaved into Gn and Gc transmembrane glycoproteins; the small (S) segment encodes a nucleocapsid protein (Np). Both glycoprotein and nucleocapsid have been shown to contribute to protective immunity [15-19]. SNV was first discovered within the 1993s outbreak, [2,3,15-17]. SNV-associated

***Corresponding author:** Abeer Babiker Idris, Department of Medical Microbiology, Faculty of Medical Laboratory Sciences, University of Khartoum-Khartoum, Sudan, Tel: +00249963064355; E-mail: abeer.babiker89@gmail.com

Received August 21, 2016; **Accepted** September 30, 2017; **Published** October 03, 2017

Citation: Idris AB, Mahmoud SM, Mohamedelamin S, Mustafa YY, Osman AA, et al. (2017) Immunoinformatics Predication and Modelling of a Cocktail of B- and T-cells Epitopes from Envelope Glycoprotein and Nucleocapsid Proteins of Sin Nombre Virus. Immunome Res 13: 141. doi: 10.4172/17457580.1000141

HPS is a severe and unpredictable public health threat, as affirmed by the 2012 outbreak at Yosemite National Park, California that resulted in 3 fatalities and more than 260,000 park visitors potentially exposed to a lethal virus [20]. SNV infections have been identified among persons with occupational exposure to North American deer mice (Peromyscus maniculatus) [21], which is the natural reservoir of SNV [10,16].

Similar to other viruses that cause HCPS, SNV has a serious potential applicability as a biological weapon because of its high lethality and known aerosol route of transmission [5]. Moreover, it causes a considerable economic harm and stigma to affected communities. Until now, Specific therapy is not available for the treatment of HCPS caused by SNV [2,3,5]. The main treatment of severe HCPS cases is completely supportive, often in intensive care unit surroundings [2]. Ribavirin, a broad-spectrum nucleoside analogue antiviral drug, is of some benefit as therapy in HCPS but it does not seem to have any clinical application in HCPS patients due to the lack of conclusive clinical data [3,4]. However, many efforts have been investigated in developing safe and effective vaccine against HCPS, included conventional approaches as well as molecular vaccine approaches using recombinant subunit vaccines utilizing vaccine delivery systems and plasmid DNA delivered by gene gun [2,3,12,13]. These approaches do not seem to have made much progress in this area, to date there is no vaccine proven to be highly efficacious against HCPS especially SNV [1-3,20]. Thus, epitope-based vaccines (EVs) have a priority for selection than other traditional vaccines. They offer many of advantages such as: biosafety, selecting of conserved or immunodominant epitopes for activating cellular or humoral responses and bio-processing of these epitopes are introduced economically and rapidly. In addition, the geographic distributions of the viruses are often well defined and the ethnic populations in need of vaccination can be determined [22] Indeed, many studies showed the immunological success of peptide-based vaccines against infectious diseases in experimental animals as well as in clinical trials, which demonstrated the responses to peptide vaccines against infectious diseases, for example malaria [23-27] hepatitis B, and HIV infections, but unfortunately, epitope mapping using experimental methods are costly and laborious.

More recently, immunoinformatics-based approaches in predicting most conserved and immunogenic epitopes from the viral genome sequence databases could give a significant amount of information for EVs developments. Several of recent publications explain in an excellently detail the values and benefits gained by the use of immunoinformatics and predictions in applied immunology and vaccinology [28]. Moreover, There are many studies used immunoinformatics-based approaches for designing EVs for different infectious and serious diseases [29,30]. In general, any active vaccine candidate must have to contain, at least, two antigenic epitopes; one to induce specific B cell or CTL responses while other induce specific Th cell response [19]. In this study, we aim to analyze envelope glycoprotein and nucleocapsid of SNV by using immunoinformatics tools housed in IEDB resources; in order to determine the most conserved and immunogenic epitopes for B cell and T cell which could trigger humoral as well as cell mediated immune response.

Materials and Methods

Retrieval of protein sequences

Sin Nombre virus's glycoprotein and nucleocapsid sequences were obtained from National Centre for Biotechnology and Information (http://www.ncbi.nlm.nih.gov/protein/) databases in Nov 2016. Accession numbers of glycoprotein and nucleocapsid of retrieved strains and their date and area of collections are listed in Tables 1 and 2, respectively.

Determinations of conserve regions

The obtained sequences were aligned using multiple sequence aliment (MSA) with the aid of Clustal was implemented in the BioEdit program, version 7.0.9.0; in order to find the conserved regions among glycoprotein and neuclocapsid variants of the strains [31].

B-cell epitopes prediction

B cell epitope is part of an immunogen which interacts with B-lymphocytes. This interaction leads to differentiation of B-lymphocyte into an antibody-secreting plasma cell and memory cell [32].

The characteristic features of B cell epitopes are hydrophilicity and accessibility [19]. Therefore, the classical propensity scale methods and hidden Markov model programmed softwares were used from the IEDB (http://www.iedb.org/) analysis resource for the following aspects [33].

- **linear B-cell Epitopes Prediction**

Using Bepipred tool [34] from the conserved regions with a default threshold value of 0.35.

- **Prediction of surface accessibility**

Using Emini surface accessibility prediction tool [35] from the conserved region holding the default threshold value 1.000 or higher.

- **Prediction of epitopes antigenicity sites**

Using kolaskar and tongaonker antigenicity method [36] for detection of the antigenic sites with a default threshold value of 1.042.

T Cell epitope prediction

GenBank Protein Accession No.	Country	Date of collection
NP_941974*	USA	20-OCT-2015
AFV71283	USA	05-JAN-2013
AFV71282	USA	05-JAN-2013
ALI59819	UK	13-OCT-2015
AAA75530	USA	03-JUN-2011
AAC42202	USA	21-NOV-1995
2124409B	USA	16-JUL-1996

* Reference strain.

Table 1: Accession numbers of glycoprotein retrieved strains and their date and area of collection.

GenBank Protein Accession No.	Country	Date of collection
NP_941975*	USA	20-OCT-2015
AIA08880	USA	26-MAY-2014
AIA08877	USA	26-MAY-2014
AFV71285	USA	05-JAN-2013
AFV71284	USA	05-JAN-2013
AFV71288	USA	02-NOV-2012
AFV71287	USA	02-NOV-2012
AFV71286	USA	02-NOV-2012
ALI59820	UK	13-OCT-2015
AAA75529	USA	03-JUN-2011
AAC42203	USA	21-NOV-1995
2124409C	USA	16-JUL-1996

* Reference strain

Table 2: Accession numbers of nucleocapsid retrieved strains and their date and area of collection

CD8+ T-cell epitopes prediction: For the prediction of CD8+ T-cell epitopes that bind to MHC-1, we used an artificial neural network (ANN) prediction method [37] from Immune Epitope Database (IEDB) and the predicted IC50 value was less than 100 nm. Prior to the prediction, peptide length was set to 9 amino acids and only frequently occurring alleles were selected. Conserved epitopes having IC_{50} value less than 100 nm were used for further analysis.

CD4+ T-cell epitopes prediction: The NN-align method under MHC-II binding prediction tool in IEDB was used for MHC-II binding [38]. Human allele references set were selected. The conserved predicted CD4+ T-cell epitopes having higher binding affinity to interact with alleles at IC_{50} less than 1000 nm was chosen for further analysis.

Population coverage analysis

All CD8+ T-cell and CD4+ T-cell predicted epitopes were assessed for population coverage against the whole world population by using the IEDB population coverage calculation tool [39].

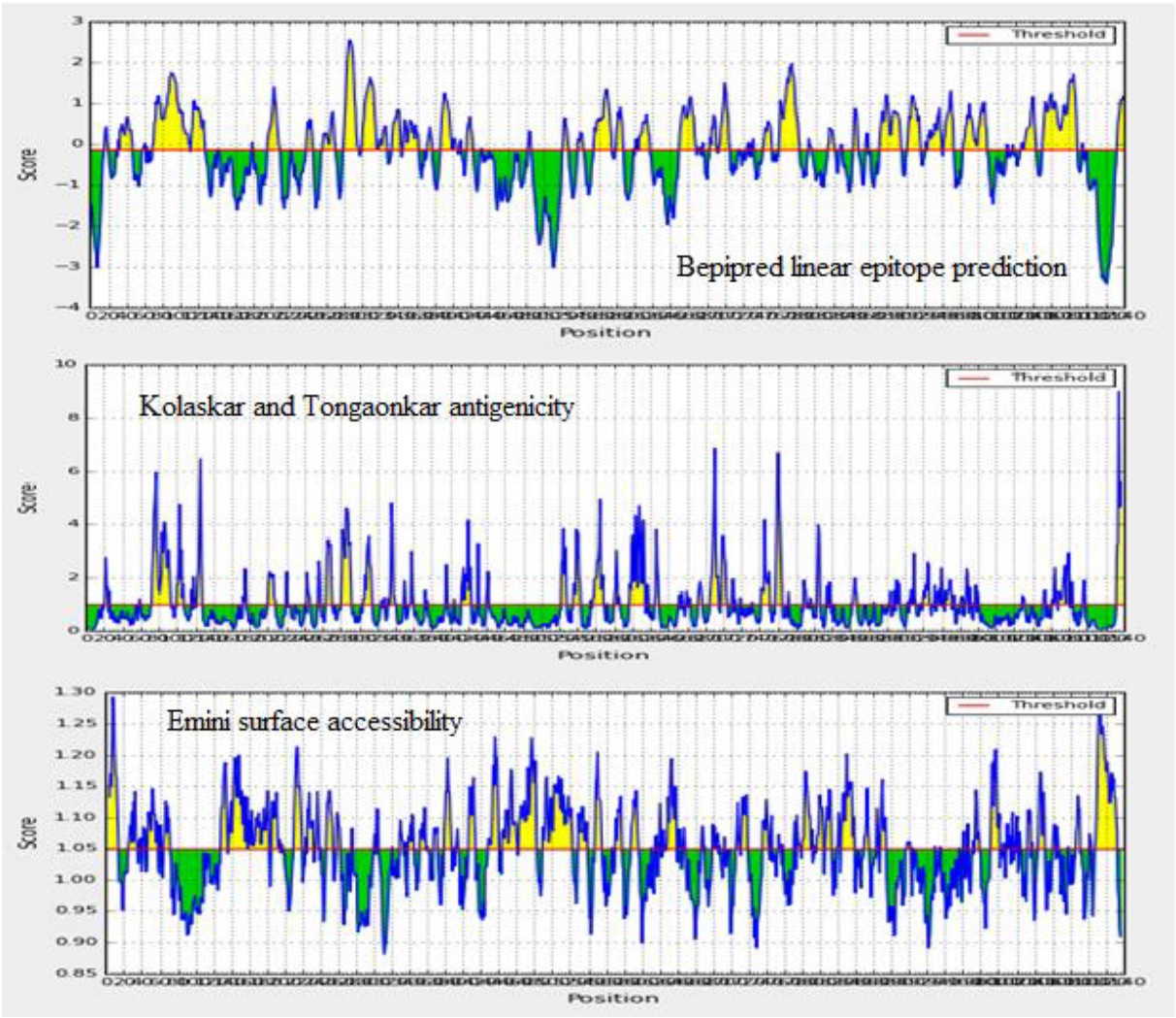
3D structure modeling

Sin Number virus's glycoprotein and nucleocapsid 3D structures were obtained by RaptorX Property (http://raptorx2.uchicago.edu/StructurePropertyPred/predict/), which is a web server predicting structure property of a protein sequence without using any template information [40]. For visualization of the 3D structure, we used UCSF Chimera (version 1.8). Chimera currently available within the Chimera package and available from the chimera web site (http://www.cgl.ucsf.edu/cimera) [41]. The 3D structure verifies the service accessibility and hydrophilicity of predicted B lymphocyte epitopes, and identifies all predicted T cell epitopes at the structural level [29].

Results

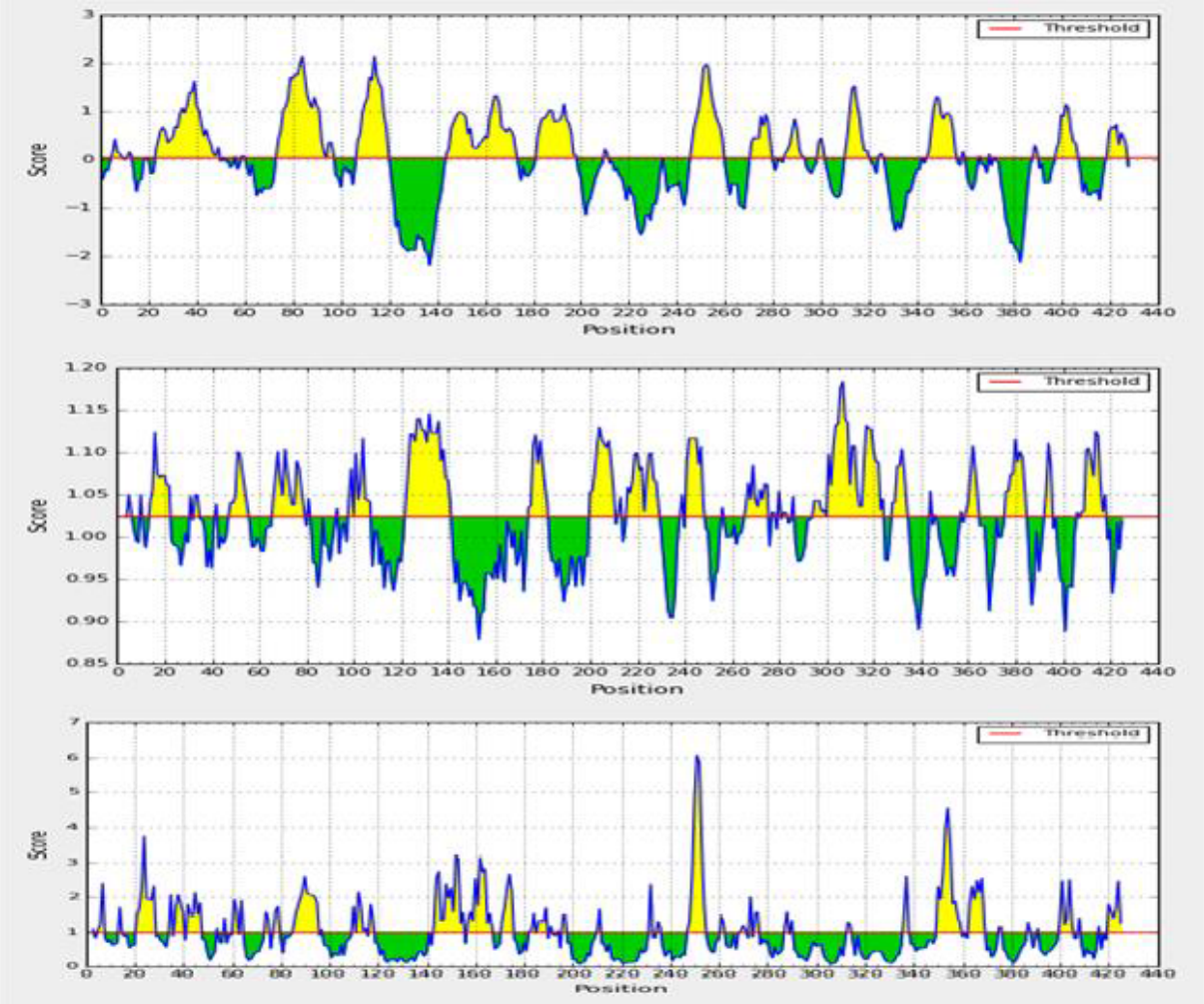
Prediction of B-cell epitope

The glycoprotein and Nucleocapsid protein was exhibited to Bepipred linear epitope prediction, Kolaskar and Tongaonkar antigenicity and Emini surface accessibility prediction methods in IEDB, to predict the probability of specific regions in the protein to bind B cell receptor, being on the surface, being immunogenic in each. The results are illustrated in Figures 1-3 and also in Tables 3 and 4.



Yellow areas above threshold (red line) are proposed to be a part of B cell epitope whereas green areas are not.

Figure 1: Prediction of B-cell epitopes for glyc oprotein by different scales.



Yellow areas above threshold (red line) are proposed to be a part of B cell epitope, whereas green areas are not.

Figure 2: Prediction of B-cell epitopes for nuclocapsid by different scales.

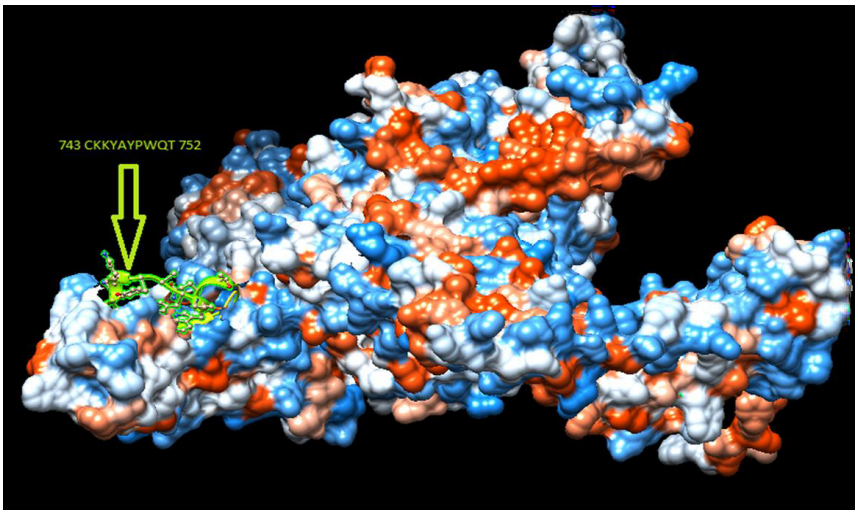


Figure 3: B-cell epitopes proposed for glycoprotein (visualization by UCSF chimera).

Prediction of cytotoxic T-lymphocyte epitopes and interaction with MHC Class I

The sequences of glycoprotein and nucleocapsid analyzed by IEDB MHC-1 binding prediction tool to predict T cell epitopes submit interacting with different types of MHC Class I alleles. Based on Artificial neural network (ANN) with half-maximal inhibitory concentration (IC_{50}) ≤ 100. The results of top epitopes are shown in Figures 3-5 and also in Tables 5 and 6.

Prediction of T helper cell epitopes and interaction with MHC Class II

The sequences of glycoprotein and nucleocapsid analyzed by IEDB MHC-II binding prediction tool based on NN-align with (IC_{50}) ≤ 1000. The results of two top epitopes are shown in Figure 6 and in Tables 7 and 8.

Analysis of the population coverage

Epitopes with high affinity to interact with MHC-I and II international alleles were applied to population coverage analysis in IEDB. There were 22 and 70 epitopes in glycoprotein, and 10 and 44 in nucleocapsid in MHC-I and II have given a high percentage of population coverage. The results were shown in Tables 9 and 10 and proposed epitopes in both MHC class I and II were shown in Figures 7 and 8 and also in Tables 11 and 12.

Discussion

Vaccination, which is a form of preventative medicine, has an extremely powerful impact on the defense against infectious agents that cause disease and death, and sometimes vaccines have the ability to completely eradicate disease from the globe (e.g. smallpox) [42,43]. None the less, there are still many human diseases have no effective vaccines [42]. One of these diseases is HCPS especially caused by SNV. Despite SNV is not transmitted from person-to-person; SNV is highly pathogenic (case fatality rate 30 to 50%) [44] regardless of age, [45] health status, [46] or access to advanced medical care, and was responsible for the 2012 Yosemite National Park outbreak [21,47]. The absence of effective vaccines, post exposure prophylactics, or therapeutics to prevent HCPS caused by SNV participated in the worry experienced by about 270,000 Yosemite visitors who got notice of possible exposure [46]. This brings us to the urgent need for the

Peptide	Start	End	Length	Antigencity	Emini
TAGL	16	19	4	1.024	0.505
TVGLGQGYVTGSVE	33	46	14	1.054	0.194
TCNIPPTTFEAAYKSR	112	127	16	1.015	1.554
IALTQPGHTYDTMT	196	209	14	1.001	1.079
SCTENSF	239	245	7	1.009	0.733
SEPLFVPTMEDYRS	257	270	14	1.015	2.279
LNPRGEDHDPDQNGQGLM	279	296	18	0.949	4.552
GPVTAKVPSTETTETMQGIAFAGAPMYS	300	327	28	1.004	0.39
VRKADPEYVFS	333	343	11	1.033	0.625
GIIAESNHSVCDKKTVPLT	345	363	19	1.054	0.241
GEIEKI	372	377	6	0.968	0.797
AGPGASCEAYSETGIFNIS	387	405	19	1.014	0.127
KFRGSEQ	416	422	7	0.949	2.953
LHGW	483	486	4	1.03	0.466
HYSTESK	522	528	7	0.997	4.189
VEYQKTM	538	544	7	2.181	1.011
CETAKELETHKKSCPEGQCP	556	575	20	1.035	1.558
ITESTESAL	580	588	9	1.001	0.928
QEN	602	604	3	0.881	2.547
ASADTPLMESGWSDTAHGV	650	668	19	0.995	0.451
ASSSSYSY	684	691	8	1.879	1.056
KLVNPANQEETLP	694	706	13	1.014	2.742
*CKKYAYPWQT	743	752	10	1.054	2.348
DYQYETSWGCNPPDCPGVGTGCTA	760	783	24	1.029	0.248
SVGK	793	796	4	1.05	0.836
GTEQTCKHI	814	822	9	1.017	0.896
LVT	828	830	3	1.181	0.466
SKLQPG	842	847	6	1.024	1.6
LEQ	856	858	3	1.039	1.306
CVFGDPGDIMSTTSGMRCPEHTGSF	871	895	25	1.006	0.081
ATTPTCEYQGNTVSG	902	916	15	1.008	0.962
RDSFQSFNVTEPHITSN	924	940	17	0.994	2.346
LEWIDPDSSIKD	942	953	12	0.993	1.382
VSFQDLSDNPCKVDLHTQSIDGAWGSG	963	989	27	1.031	0.203
LRGSNTVKVVGKGGHSG	1027	1043	17	1.019	0.32
DTDCTEEGLAASPPHLDRVTGYNQIDSDKVYDDGAPPCT	1051	1089	39	1.019	1.916

*Selected epitopes.

Table 3: List of predicted B cell epitopes for glycoprotein

Peptide	Start	End	Length	Antigencity	Emini
KEVQ	5	8	4	1.045	1.52
QKLKDAERAVELDPDDVNKSTLQSRRA	23	49	27	1.003	5.71
LASKPVDPTGIEPDDHLKEK	74	93	20	1.014	1,732
SLR	95	97	3	1.045	0.965
IDLEEPSGQTADWK	107	120	14	0.971	1.263
RGRQTIKENKGTRIRFKDDSSYEEVNGIRK	144	173	30	0.95	11.914
PTAQSTMKADEITPGRF	182	198	17	0.969	1.197
FLPEQKDPRDAALA	246	259	14	1.022	1.141
*QVDESKVSD	271	279	9	1.035	1.48
ADARAESAT	285	293	9	0.974	1.02
ATP	299	301	3	1.012	1.005
APDRCPPTALY	310	320	11	1.072	0.665
MPE	324	326	3	0.914	1.181
KSVGTSEEKLKK	345	356	12	0.988	3.126
GDDMDPELRE	398	407	10	0.919	2.363
EISNQEPLK	419	427	9	0.989	1,789

* Peptide from (271-279) gives higher score if it is shorten (271 to 278) in all tools.

Table 4: List of predicted B cell epitopes for nucleocapsid protein.

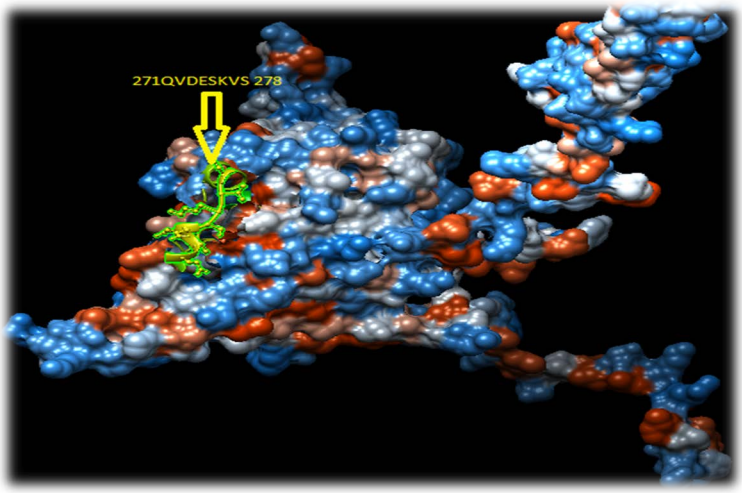


Figure 4: B cell epitope proposed for nucleocapsid (visualization by UCSF chimera).

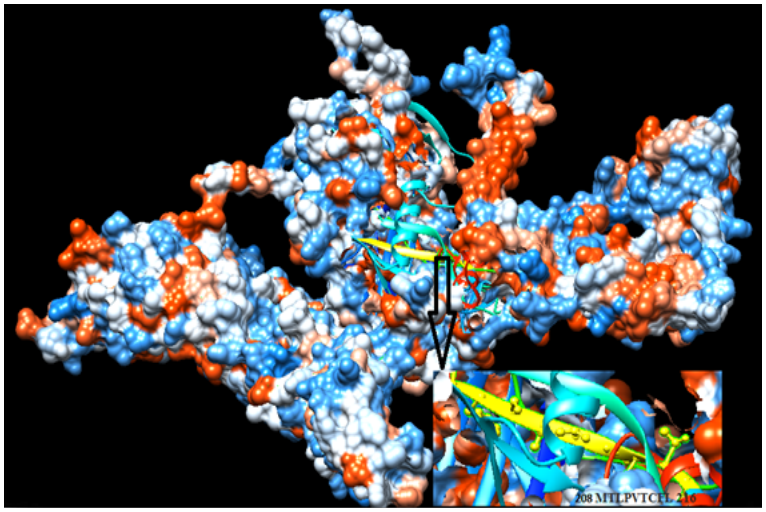


Figure 5: Glycoprotein epitope 208**MTLPVTCFL**216 suggested for cytotoxic T cell interaction (MHC 1) (visualization by UCSF chimera).

Peptide	Start	End	Allele	ic50	Percentile
MTLPVTCFL	208	216	HLA-A*02:01	41.88	0.5
			HLA-A*02:06	7.21	0.1
			HLA-A*30:01	52.78	0.5
			HLA-A*68:02	5.58	0.2
			HLA-B*57:01	76.91	0.3
			HLA-B*58:01	15.35	0.3
			HLA-C*03:03	74.45	0.4
			HLA-C*14:02	34.97	0.2
YTFTSLFSL	458	466	HLA-A*02:01	6.44	0.1
			HLA-A*02:06	4.57	0.1
			HLA-A*68:02	9.22	0.2
			HLA-B*39:01	80.18	0.3
			HLA-C*03:03	32.15	0.4
			HLA-C*12:03	22.83	0.2
			HLA-C*14:02	39.92	0.2

Table 5: The most potential 2 CD8+ T-cell epitopes of glycoprotein with interacting MHC-1 alleles.

development of an effective vaccine against SNV. In our study, we analyzed envelope glycoprotein and nucleocapsid of SNV, and predicted most conserved and immunogenic B- and T-cell epitopes which could trigger humoral as well as cellular mediated immune responses in order to develop an effective epitope-based vaccine. This concept of epitope-based vaccine, in which a cocktail of B- and T-cell epitopes are used to enhance protection against viral infection, was applied by Steward et al. in a study on the Respiratory syncytial virus (RSV), where RSV-specific humoral and cellular immunities were induced after immunization with a cocktail of peptides. Following challenge infection, a 190-fold reduction in RSV titer was observed in the lungs of immunized mice [48]. Thus the combination of humoral and cellular immunity is more promising at clearing viral infection than humoral or cellular immunity alone. A number of studies have demonstrated the strong antigenicity of both the membrane glycoprotein and nucleocapsid protein. However, neutralizing antibodies have been shown to be sufficient to confer protection in passive immunization of experiments using glycoproteins-specific monoclonal and polyclonal antibodies [2,15,49]. The nucleocapsid is a multifunctional protein that plays several roles in virus replication and assembly and is the major antigen to which the host mounts the humoral immune response. It is highly conserved among various Hantavirus genotypes and is expressed early with huge amounts in the cell during infection. The antibody to the nucleocapsid protein is demonstrated early in the immune response [50,51].

The conserved predicted epitopes for B- and T-cell were retrieved from all declared strains of SNV glycoprotein and nucleocapsid in NCBI databases until Nov 2016. As obviously clear from the result, the conserved regions cross retrieved strains accompanied with a good level of confidence. To design an effective peptide antigen for B cell, epitopes were predicted from the conserved sequences by Bepipred prediction tools should be passed both Emini surface accessibility and Kolaskar and Tongaonkar antigenicity thresholds, and peptide sequences should best be within 8-22 amino acids in length [19]. However, 20 predicted epitopes of glycoprotein were satisfied both Emini surface accessibility and Kolaskar and Tongaonkar antigenicity tests in IEDB. The linear epitope 743CKKYAYPWQT752 showed high succeeded score in both tests as illustrated in Table 3. For nucleocapsid, 9 predicted epitopes for B cell were succeeded both Emini surface accessibility and Kolaskar and Tongaonkar antigenicity tests. The linear epitope 271QVDESKVSD279 was 9 amino acids residue long, when we decreased the epitope length to 8 amino acid 271QVDESKVS278 we got the highest score in accessibility with a succeeded score in antigenicity tests, as illustrated

Peptide	Start	End	Allele	ic50	Percentile
YILSFALPI	125	133	HLA-A*02:01	6.98	0.1
			HLA-A*02:06	10.49	0.2
			HLA-A*32:01	45.63	0.3
FLAARCPFL	239	247	HLA-A*02:01	3.63	0.1
			HLA-A*02:06	8.16	0.1
			HLA-B*08:01	82.87	0.2
			HLA-C*03:03	44.74	0.4

Table 6: The most potential 2 CD8+ T-cell epitopes of nucleocapsid with interacting MHC-1 alleles.

in Table 4. Sagadevan et al. predicted IMASKSVGS/TAEEKLKKKSAF from nucleocapsid protein of Hantaviruses causing HCPS, based on the mean percent prediction probability score, to be the best candidate B-cell epitope for developing immunoassays in the detection of antibodies to hantaviruses causing HCPS [51]. But this peptide has a low value of antigenicity and does not pass the threshold of Kolaskar and Tongaonkar antigenicity test in IEDB. Thus it may be useful for the development of immunodiagnostic tools towards pan-detection of Hantavirus antibodies causing HCPS but not effective vaccines.

T-cell mediated immunity plays a crucial role in the defense against viral infections [52]. The inclusion of T-cell epitopes in the vaccine development induces a strong and long lasting immune response and antigenic drift where antigen can easily escape the antibody memory response [19,22]. While CD8+ cytotoxic T-cells generally recognize intracellular peptides displayed by HLA class I molecules, CD4+ T-helper cells generally recognize peptides from the extracellular space, displayed by HLA class II molecules (CD4+ T-cell epitopes) [22]. A potential problem in the development of CTL epitope-based vaccines is the large degree of MHC polymorphism and the need for the understanding of HLA restrictions in the target population for vaccines [53]. In our study, we chose the most frequently occurring alleles in IEDB for MHC binding prediction. For glycoprotein, among 168 CD8+ T cell conserved predicted epitopes, we found the epitope 208MTLPVTCFL216 had higher affinity to interact with 8 alleles (HLA-A*02:06, HLA-A*02:01, HLA-A*30:01, HLA-A*68:02, HLA-B*58:01, HLA-B*57:01, HLA-C*03:03, HLA-C*14:02) and the epitope 458YTFTSLFSL466 had the affinity to interact with 7 alleles (HLA-A*02:06, HLA-A*02:01, HLA-A*68:02, HLA-B*39:01, HLA-C*03:03, HLA-C*12:03, HLA-C*14:02). These two epitopes had very good population coverage for class I alleles throughout the world, as shown in Table 5 and Table 12. Also we found 674 CD4+ T cell predicted conserved epitopes interacting with MHC-II alleles, as shown in Table 7. Epitopes 50ILLTQVADL58 and 165LLSSRIQVI173 had high affinity to interact with 22 and 21 alleles, respectively. They interacted with HLA-DPA1*01/2/3, -DPA1*02, -DQB1*02/3/4/5/6, -DRB1*03/4/7, -DRB3*01, -DRB4*01, -DRB5*01. Interestingly, these two epitopes covered 99.65% of the world population. For nucleocapsid, we found 66 CD8+ T cell and 228 CD4+ T cell conserved predicted epitopes, as shown in Tables 6 and 7. Most importantly, the epitopes 125YILSFALPI133 and 239FLAARCPFL247 had high affinity to interact with both MHC classes, I and II, and also they had excellent population coverage for Class I and II alleles throughout the world (Table 11). Epitope FLAARCPFL was the same epitope predicted by Sathish et al. in 2017 as a suitable epitope for vaccine specific to the SNV genotype [54].

The HLA-A*02 supertype is expressed in all main ethnicities in the 39-46% range. Since many peptides that bind A*02:01 also display degenerate binding (binding to multiple alleles), an A2 supertype multiepitope vaccine could be developed to allow broad, non-

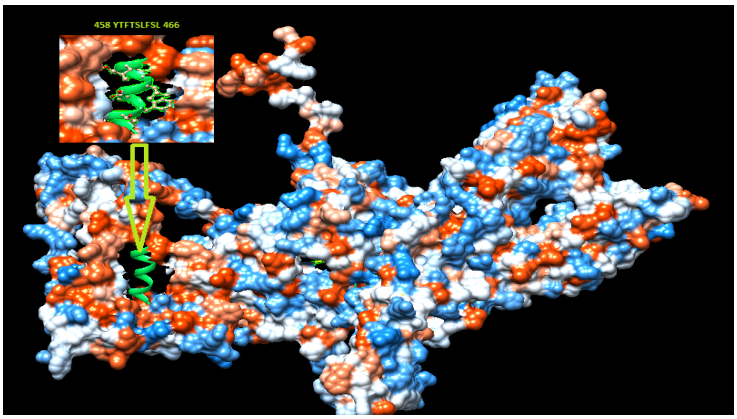


Figure 6: Glycoprotein epitope 458**YTFTSLFSL**466 suggested for cytotoxic T cell interaction (MHC 1) (visualization by UCSF chimera).

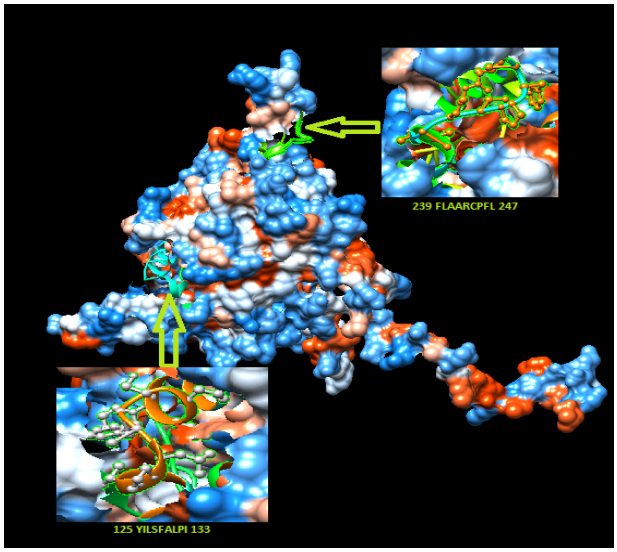


Figure 7: Nucleocapsid protein epitopes suggested for both MHC-I and MHC-II interaction (visualization by UCSF chimera).

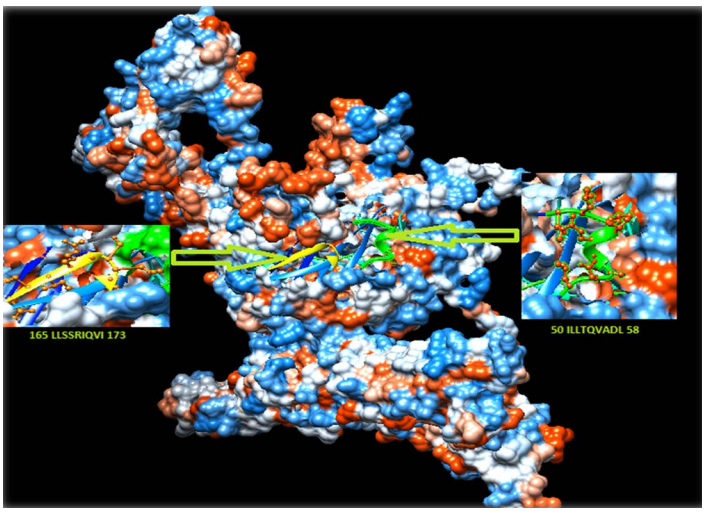


Figure 8: Glycoprotein epitopes suggested for T helper cell interaction (MHC-II) (visualized by USCF chimera).

Core sequence	Start	End	Peptide sequence	Allele	IC_{50}	Rank
ILLTQVADL	50	58	EITPILLTQVADLKI	HLA-DPA1*01:03/DPB1*02:01	442.9	20.15
				HLA-DPA1*02:01/DPB1*01:01	46	4.8
				HLA-DPA1*03:01/DPB1*04:02	37.3	4.42
				HLA-DQA1*01:02/DQB1*06:02	142.5	10.3
				HLA-DQA1*04:01/DQB1*04:02	495.9	7.75
				HLA-DQA1*05:01/DQB1*02:01	344.5	7.78
				HLA-DRB1*01:01	58.9	23.25
				HLA-DRB1*03:01	357.2	11.51
				HLA-DRB1*04:01	242.5	17.41
				HLA-DRB1*04:04	36.3	3.74
				HLA-DRB1*04:05	436.3	27.16
				HLA-DRB1*13:02	38.6	2.76
				HLA-DRB1*15:01	639.9	32.87
			ILLTQVADLKIESSC	HLA-DPA1*02:01/DPB1*01:01	82.3	8.76
				HLA-DPA1*03:01/DPB1*04:02	71.5	7.65
				HLA-DQA1*05:01/DQB1*02:01	696.4	15.12
				HLA-DRB1*04:04	141.3	15.79
				HLA-DRB1*04:05	897	39.36
				HLA-DRB1*13:02	244.6	11.19
				HLA-DRB4*01:01	84.8	6.61
			ITPILLTQVADLKIE	HLA-DPA1*01:03/DPB1*02:01	469.3	20.77
				HLA-DPA1*03:01/DPB1*04:02	37.5	4.44
				HLA-DQA1*04:01/DQB1*04:02	556.8	8.77
				HLA-DQA1*05:01/DQB1*02:01	415.4	9.37
				HLA-DRB1*01:01	55.8	22.59
				HLA-DRB1*03:01	214.6	8.37
				HLA-DRB1*04:04	34.3	3.45
				HLA-DRB1*04:05	367.8	24.6
				HLA-DRB1*09:01	298.6	17.64
				HLA-DRB1*13:02	40.1	2.86
				HLA-DRB1*15:01	656	33.25
			PILLTQVADLKIESS	HLA-DPA1*01:03/DPB1*02:01	776.7	26.88
				HLA-DPA1*02:01/DPB1*01:01	71.4	7.65
				HLA-DPA1*03:01/DPB1*04:02	48	5.56
				HLA-DQA1*05:01/DQB1*02:01	575.5	12.74
				HLA-DRB1*04:04	51.2	5.85
				HLA-DRB1*04:05	602.4	32.3
				HLA-DRB1*13:02	95.2	5.86
			SVEITPILLTQVADL	HLA-DPA1*01:03/DPB1*02:01	377.8	18.5
				HLA-DQA1*04:01/DQB1*04:02	298.3	4.26
				HLA-DQA1*05:01/DQB1*02:01	409.6	9.24
				HLA-DRB1*01:01	89.3	28.63
				HLA-DRB1*04:01	688.5	34.92
				HLA-DRB1*04:04	49.5	5.62
				HLA-DRB1*04:05	628.6	33.02
				HLA-DRB1*07:01	55.6	9.43
				HLA-DRB1*13:02	73.2	4.8
			TPILLTQVADLKIES	HLA-DPA1*01:03/DPB1*02:01	603.3	23.68
				HLA-DPA1*03:01/DPB1*04:02	40.6	4.79
				HLA-DQA1*04:01/DQB1*04:02	673.6	10.73
				HLA-DQA1*05:01/DQB1*02:01	519.3	11.57
				HLA-DRB1*01:01	73.3	26.02
				HLA-DRB1*03:01	206.9	8.2
				HLA-DRB1*04:04	35.5	3.62
				HLA-DRB1*04:05	430.1	26.94
				HLA-DRB1*09:01	644.1	30

				HLA-DRB1*13:02	59.1	4.03
				HLA-DRB1*15:01	825.1	36.92
			VEITPILLTQVADLK	HLA-DPA1*01:03/DPB1*02:01	439.7	20.07
				HLA-DQA1*04:01/DQB1*04:02	392.3	5.94
				HLA-DQA1*05:01/DQB1*02:01	335.1	7.57
				HLA-DRB1*01:01	71.3	25.66
				HLA-DRB1*04:01	304.5	20.64
				HLA-DRB1*04:04	42.3	4.6
				HLA-DRB1*04:05	585.1	31.81
				HLA-DRB1*07:01	83.4	12.53
				HLA-DRB1*13:02	64.2	4.3
LLSSRIQVI	165	173	CMIGLLSSRIQVIYE	HLA-DPA1*01/DPB1*04:01	244	9.67
				HLA-DPA1*01:03/DPB1*02:01	98.7	8.29
				HLA-DPA1*02:01/DPB1*01:01	98.7	10.28
				HLA-DQA1*01:02/DQB1*06:02	82	5.66
				HLA-DQA1*05:01/DQB1*03:01	177.6	21.11
				HLA-DRB1*03:01	62.4	3.44
				HLA-DRB1*07:01	23.3	4.5
				HLA-DRB1*13:02	96	5.89
				HLA-DRB4*01:01	63.2	4.85
			GLLSSRIQVIYEKTY	HLA-DPA1*01/DPB1*04:01	403.7	13.14
				HLA-DPA1*01:03/DPB1*02:01	202.8	12.99
				HLA-DPA1*02:01/DPB1*01:01	232	19.59
				HLA-DPA1*03:01/DPB1*04:02	35	4.16
				HLA-DQA1*01:02/DQB1*06:02	120.6	8.72
				HLA-DQA1*05:01/DQB1*03:01	480.5	35.13
				HLA-DRB1*01:01	163.3	37.53
				HLA-DRB1*03:01	218.1	8.46
				HLA-DRB1*04:05	446.7	27.52
				HLA-DRB1*07:01	111.5	15.21
				HLA-DRB1*08:02	576.2	13.97
				HLA-DRB1*13:02	247.9	11.28
				HLA-DRB4*01:01	46.7	3.42
			IGLLSSRIQVIYEKT	HLA-DPA1*01/DPB1*04:01	306.5	11.14
				HLA-DPA1*01:03/DPB1*02:01	116.7	9.22
				HLA-DPA1*02:01/DPB1*01:01	100.2	10.42
				HLA-DPA1*03:01/DPB1*04:02	19.1	2.17
				HLA-DQA1*01:02/DQB1*06:02	83.1	5.75
				HLA-DQA1*05:01/DQB1*03:01	269.8	26.51
				HLA-DRB1*01:01	89	28.59
				HLA-DRB1*03:01	105.2	5.21
				HLA-DRB1*04:05	384.8	25.27
				HLA-DRB1*07:01	82.6	12.46
				HLA-DRB1*08:02	523.1	12.72
				HLA-DRB1*13:02	165.5	8.68
				HLA-DRB4*01:01	42	2.99
			LLSSRIQVIYEKTYC	HLA-DPA1*01/DPB1*04:01	573.6	16.1
				HLA-DPA1*01:03/DPB1*02:01	291.3	15.99
				HLA-DPA1*02:01/DPB1*01:01	253.5	20.78
				HLA-DPA1*03:01/DPB1*04:02	178.4	13.86
				HLA-DQA1*01:02/DQB1*06:02	216.1	14.94
				HLA-DQA1*05:01/DQB1*03:01	577.7	38.19
				HLA-DRB1*01:01	227.2	43.13
				HLA-DRB1*03:01	475.1	13.68
				HLA-DRB1*04:05	529.2	30.2
				HLA-DRB1*07:01	163.8	19.2
				HLA-DRB1*13:02	386.8	14.91

				HLA-DRB4*01:01	48.4	3.57
			MIGLLSSRIQVIYEK	HLA-DPA1*01/DPB1*04:01	235.5	9.46
				HLA-DPA1*01:03/DPB1*02:01	104.8	8.61
				HLA-DPA1*02:01/DPB1*01:01	95.3	9.98
				HLA-DQA1*01:02/DQB1*06:02	62.7	4.03
				HLA-DQA1*05:01/DQB1*03:01	186.5	21.71
				HLA-DRB1*03:01	50	2.86
				HLA-DRB1*07:01	42.5	7.59
				HLA-DRB1*08:02	351.6	8.44
				HLA-DRB1*13:02	99.2	6.03
				HLA-DRB4*01:01	50.1	3.71
			RSCMIGLLSSRIQVI	HLA-DPA1*01/DPB1*04:01	151.4	7.08
				HLA-DPA1*01:03/DPB1*02:01	99.6	8.33
				HLA-DPA1*02:01/DPB1*01:01	86.4	9.15
				HLA-DPA1*03:01/DPB1*04:02	30.6	3.65
				HLA-DQA1*01:02/DQB1*06:02	75	5.08
				HLA-DQA1*04:01/DQB1*04:02	993.9	15.71
				HLA-DQA1*05:01/DQB1*03:01	234	24.59
				HLA-DRB1*03:01	202.2	8.09
				HLA-DRB1*07:01	9.7	1.57
				HLA-DRB1*13:02	110.7	6.53
				HLA-DRB4*01:01	81.4	6.35
			SCMIGLLSSRIQVIY	HLA-DPA1*01/DPB1*04:01	172.7	7.75
				HLA-DPA1*01:03/DPB1*02:01	89.2	7.75
				HLA-DPA1*02:01/DPB1*01:01	91	9.59
				HLA-DQA1*04:01/DQB1*04:02	977	15.47
				HLA-DQA1*05:01/DQB1*03:01	181.2	21.36
				HLA-DRB1*03:01	93.6	4.78
				HLA-DRB1*07:01	13.6	2.49
				HLA-DRB1*13:02	98	5.98
				HLA-DRB4*01:01	71.7	5.57

Table 7: The most potential 2 CD4+ T-cell epitopes of glycoprotein with interacting MHC-II alleles.

Core sequence	Start	End	Peptide sequence	Allele	IC_{50}	Rank
YILSFALPI	125	133	GLYILSFALPIILKA	HLA-DPA1*01/DPB1*04:01	47.3	2.93
				HLA-DQA1*05:01/DQB1*02:01	543	12.07
				HLA-DRB1*01:01	6.3	1.99
				HLA-DRB1*03:01	98.4	4.96
				HLA-DRB1*04:01	106.4	8.61
				HLA-DRB1*04:04	109.9	12.9
				HLA-DRB1*04:05	162.3	14.22
				HLA-DRB1*07:01	10.9	1.85
				HLA-DRB1*09:01	38	2.24
				HLA-DRB1*11:01	69.3	10.81
				HLA-DRB1*13:02	18.8	1.34
				HLA-DRB1*15:01	45.4	4.56
				HLA-DRB3*01:01	267.2	8.14
				HLA-DRB5*01:01	66.4	12
			IGLYILSFALPIILK	HLA-DPA1*01/DPB1*04:01	40	2.55
				HLA-DPA1*01:03/DPB1*02:01	33.8	3.78
				HLA-DPA1*02:01/DPB1*01:01	35	3.48
				HLA-DPA1*03:01/DPB1*04:02	34.9	4.15
				HLA-DQA1*05:01/DQB1*02:01	414	9.34
				HLA-DRB1*01:01	5.2	0.97
				HLA-DRB1*03:01	41.5	2.45
				HLA-DRB1*04:01	103.4	8.38
				HLA-DRB1*04:04	107.7	12.67

				HLA-DRB1*04:05	130	11.99
				HLA-DRB1*07:01	8.9	1.38
				HLA-DRB1*09:01	43.2	2.68
				HLA-DRB1*11:01	63.7	10.18
				HLA-DRB1*13:02	15.1	1.04
				HLA-DRB1*15:01	22.4	1.82
				HLA-DRB3*01:01	118.8	5
				HLA-DRB5*01:01	56.5	10.81
			KSIGLYILSFALPII	HLA-DPA1*01/DPB1*04:01	62.7	3.7
				HLA-DPA1*01:03/DPB1*02:01	34.3	3.83
				HLA-DPA1*03:01/DPB1*04:02	96.6	9.44
				HLA-DQA1*05:01/DQB1*03:01	345.9	30.11
				HLA-DRB1*01:01	6.2	1.9
				HLA-DRB1*03:01	117	5.6
				HLA-DRB1*04:01	162.1	12.61
				HLA-DRB1*04:04	75.9	9.15
				HLA-DRB1*04:05	106.2	10.14
				HLA-DRB1*07:01	6.5	0.83
				HLA-DRB1*09:01	58.2	3.87
				HLA-DRB1*11:01	253.2	22.84
				HLA-DRB1*13:02	26.7	1.94
				HLA-DRB3*01:01	76.1	3.76
				HLA-DRB5*01:01	141.8	18.5
			LYILSFALPIILKAL	HLA-DPA1*01/DPB1*04:01	50.2	3.08
				HLA-DQA1*05:01/DQB1*02:01	543.2	12.08
				HLA-DRB1*01:01	8.3	3.8
				HLA-DRB1*03:01	264.9	9.56
				HLA-DRB1*04:01	137.5	10.92
				HLA-DRB1*04:04	112.4	13.15
				HLA-DRB1*04:05	225.2	17.98
				HLA-DRB1*07:01	12.5	2.22
				HLA-DRB1*09:01	38.7	2.29
				HLA-DRB1*11:01	95.4	13.34
				HLA-DRB1*13:02	24	1.73
				HLA-DRB1*15:01	89.8	9.03
				HLA-DRB3*01:01	609.2	13.2
				HLA-DRB5*01:01	84.2	13.85
			SIGLYILSFALPIIL	HLA-DPA1*01/DPB1*04:01	49.9	3.06
				HLA-DPA1*03:01/DPB1*04:02	72.6	7.73
				HLA-DQA1*05:01/DQB1*02:01	310.4	6.98
				HLA-DRB1*01:01	5.6	1.33
				HLA-DRB1*03:01	56.1	3.15
				HLA-DRB1*04:01	128	10.24
				HLA-DRB1*04:05	113.6	10.73
				HLA-DRB1*07:01	6.8	0.9
				HLA-DRB1*09:01	44	2.74
				HLA-DRB1*11:01	124.3	15.64
				HLA-DRB1*13:02	18.5	1.32
				HLA-DRB3*01:01	86.4	4.08
				HLA-DRB5*01:01	90.1	14.42
			WKSIGLYILSFALPI	HLA-DPA1*01/DPB1*04:01	69.1	3.99
				HLA-DPA1*01:03/DPB1*02:01	46.7	4.87
				HLA-DPA1*03:01/DPB1*04:02	113.5	10.51
				HLA-DQA1*05:01/DQB1*02:01	476.2	10.68
				HLA-DQA1*05:01/DQB1*03:01	369.7	31.08
				HLA-DRB1*01:01	6.7	2.37
				HLA-DRB1*03:01	277.3	9.85
				HLA-DRB1*04:01	223.8	16.38

				HLA-DRB1*04:05	124.8	11.6
				HLA-DRB1*07:01	7	0.94
				HLA-DRB1*09:01	67.7	4.58
				HLA-DRB1*11:01	474.9	30.37
				HLA-DRB1*13:02	49.1	3.42
				HLA-DRB3*01:01	70.7	3.57
				HLA-DRB5*01:01	199.8	21.97
			YILSFALPIILKALY	HLA-DPA1*01/DPB1*04:01	117.7	5.94
				HLA-DQA1*05:01/DQB1*02:01	608.2	13.4
				HLA-DRB1*01:01	12.3	6.87
				HLA-DRB1*03:01	732.1	17.41
				HLA-DRB1*04:01	181.9	13.87
				HLA-DRB1*04:04	243.9	23.25
				HLA-DRB1*04:05	369.7	24.68
				HLA-DRB1*07:01	16	3.01
				HLA-DRB1*11:01	117.3	15.11
				HLA-DRB1*13:02	33.2	2.39
				HLA-DRB1*15:01	81.1	8.25
				HLA-DRB5*01:01	103.4	15.59
FLAARCPFL	239	247	DDFLAARCPFLPEQK	HLA-DPA1*01/DPB1*04:01	508.6	15.04
				HLA-DPA1*01:03/DPB1*02:01	342.9	17.53
				HLA-DPA1*02:01/DPB1*01:01	840.2	39.47
				HLA-DPA1*03:01/DPB1*04:02	905.3	30.6
				HLA-DQA1*05:01/DQB1*03:01	376.1	31.34
				HLA-DRB1*01:01	234.4	43.68
				HLA-DRB1*04:01	952.1	41.86
				HLA-DRB1*04:04	516.7	35.27
				HLA-DRB1*04:05	729.2	35.62
				HLA-DRB1*07:01	449.4	31.75
				HLA-DRB1*09:01	301.7	17.77
				HLA-DRB1*11:01	783.9	37.19
				HLA-DRB1*15:01	645.2	33
				HLA-DRB3*01:01	597.1	13.04
				HLA-DRB5*01:01	927.5	42.07
			DFLAARCPFLPEQKD	HLA-DPA1*01:03/DPB1*02:01	440	20.08
				HLA-DQA1*05:01/DQB1*03:01	667	40.67
				HLA-DRB1*01:01	241.7	44.23
				HLA-DRB1*04:04	487.3	34.26
				HLA-DRB1*04:05	752.1	36.16
				HLA-DRB1*07:01	626.6	36.66
				HLA-DRB1*09:01	450.8	23.8
				HLA-DRB1*15:01	657.7	33.29
			ERIDDFLAARCPFLP	HLA-DPA1*01/DPB1*04:01	569.6	16.03
				HLA-DPA1*01:03/DPB1*02:01	276	15.53
				HLA-DPA1*03:01/DPB1*04:02	845	29.72
				HLA-DQA1*05:01/DQB1*02:01	470.5	10.56
				HLA-DQA1*05:01/DQB1*03:01	269.8	26.51
				HLA-DRB1*01:01	204.6	41.28
				HLA-DRB1*04:01	564.1	31.06
				HLA-DRB1*04:04	422.3	31.9
				HLA-DRB1*04:05	570.4	31.39
				HLA-DRB1*07:01	104.3	14.56
				HLA-DRB1*09:01	297	17.57
				HLA-DRB3*01:01	260.7	8.03
				HLA-DRB5*01:01	617.7	35.98
			FLAARCPFLPEQKDP	HLA-DPA1*01:03/DPB1*02:01	753.7	26.5
				HLA-DRB1*01:01	248.5	44.75

				HLA-DRB1*04:04	747.1	41.84
				HLA-DRB1*07:01	958.8	43.51
				HLA-DRB1*09:01	860.8	35.73
				HLA-DRB1*15:01	756.8	35.52
			IDDFLAARCPFLPEQ	HLA-DPA1*01/DPB1*04:01	385.7	12.79
				HLA-DPA1*01:03/DPB1*02:01	267.1	15.24
				HLA-DPA1*02:01/DPB1*01:01	767.5	37.9
				HLA-DPA1*03:01/DPB1*04:02	760.4	28.43
				HLA-DQA1*05:01/DQB1*02:01	735.8	15.86
				HLA-DQA1*05:01/DQB1*03:01	261.8	26.08
				HLA-DRB1*01:01	155.2	36.71
				HLA-DRB1*03:01	508.1	14.2
				HLA-DRB1*04:01	693.5	35.08
				HLA-DRB1*04:04	519.6	35.36
				HLA-DRB1*04:05	675.5	34.26
				HLA-DRB1*07:01	268.4	24.85
				HLA-DRB1*09:01	236.6	14.79
				HLA-DRB1*11:01	483.5	30.61
				HLA-DRB1*15:01	417.2	26.43
				HLA-DRB3*01:01	297.8	8.67
				HLA-DRB5*01:01	786.8	39.54
			MERIDDFLAARCPFL	HLA-DPA1*01/DPB1*04:01	574.6	16.11
				HLA-DPA1*01:03/DPB1*02:01	235.2	14.16
				HLA-DPA1*02:01/DPB1*01:01	667.3	35.49
				HLA-DQA1*05:01/DQB1*02:01	581.7	12.86
				HLA-DQA1*05:01/DQB1*03:01	316.7	28.81
				HLA-DRB1*01:01	173.5	38.51
				HLA-DRB1*04:01	476.1	27.94
				HLA-DRB1*04:04	476.6	33.88
				HLA-DRB1*04:05	388	25.39
				HLA-DRB1*07:01	70.4	11.2
				HLA-DRB1*09:01	321	18.67
				HLA-DRB3*01:01	255.4	7.92
				HLA-DRB5*01:01	463	32.02
			RIDDFLAARCPFLPE	HLA-DPA1*01/DPB1*04:01	435	13.74
				HLA-DPA1*01:03/DPB1*02:01	245.2	14.51
				HLA-DPA1*02:01/DPB1*01:01	960	41.81
				HLA-DPA1*03:01/DPB1*04:02	820.3	29.36
				HLA-DQA1*05:01/DQB1*02:01	515.7	11.5
				HLA-DQA1*05:01/DQB1*03:01	257.4	25.85
				HLA-DRB1*01:01	191.8	40.18
				HLA-DRB1*03:01	340.5	11.16
				HLA-DRB1*04:01	574.8	31.42
				HLA-DRB1*04:04	423.4	31.94
				HLA-DRB1*04:05	670.6	34.13
				HLA-DRB1*07:01	165.2	19.29
				HLA-DRB1*09:01	258.6	15.83
				HLA-DRB1*15:01	333.1	23.32
				HLA-DRB3*01:01	263.7	8.08
				HLA-DRB5*01:01	705.2	37.92

Table 8: The most potential 2 CD4+ T-cell epitopes of nucleocapsid with interacting MHC-II alleles.

Epitope	Coverage	Total HLA hits	Epitope	Coverage	Total HLA hits
	Class I			**Class II**	
VIYEKTYCV	49.21%	4	ILLTQVADL	99.65%	22
QLIEGLCFI	40.60%	2	VHLIAPVQT	93.85%	18
FIPTHTIAL	42.41%	3	LLSSRIQVI	99.65%	21
MTLPVTCFL	55.00%	8	VIYEKTYCV	96.71%	13
TLPVTCFLV	40.60%	2	IYEKTYCVT	96.21%	8
FQGYYICFI	40.60%	2	YCVTGQLIE	97.55%	10
GLMRIAGPV	40.60%	2	VTGQLIEGL	98.99%	13
SSFSTLVRK	40.03%	4	QLIEGLCFI	97.48%	9
IIAESNHSV	42.53%	3	IEGLCFIPT	97.64%	10
LTWTGFLAV	40.60%	2	LCFIPTHTI	92.89%	17
YTFTSLFSL	55.70%	7	CFIPTHTIA	90.56%	6
SLFSLIPGV	40.60%	2	FIPTHTIAL	93.13%	16
ITFCFGWLL	44.77%	4	TLPVTCFLV	93.82%	7
FRYKSRCYV	43.99%	4	LPVTCFLVA	91.01%	7
YAYPWQTAK	47.31%	5	CFLVAKKLG	91.56%	7
KAYKIVSLK	43.03%	5	FLVAKKLGT	94.54%	15
KQWCTTSCV	40.60%	2	KKLGTQLKL	95.97%	9
GLTECANFI	40.60%	2	KLAVELEKL	98.96%	13
NLLRGSNTV	40.60%	2	LAVELEKLI	97.70%	12
GILNGNWVV	40.60%	2	FQGYYICFI	99.28%	18
ILNGNWVVV	42.66%	2	FIGKHSEPL	98.81%	17
LLFSFFCPV	49.90%	4	LFVPTMEDY	96.79%	8
			IAFAGAPMY	91.31%	13
			FAGAPMYSS	97.78%	16
			PMYSSFSTL	96.39%	9
			MYSSFSTLV	95.28%	15
			YVFSPGIIA	93.67%	14
			VPLTWTGFL	92.38%	5
			LTWTGFLAV	99.26%	14
			FLAVSGEIE	99.06%	17
			KFRGSEQRI	92.17%	7
			VIGQCIYTF	95.33%	12
			CIYTFTSLF	99.08%	18
			IYTFTSLFS	97.86%	13
			YTFTSLFSL	99.24%	13
			TFTSLFSLI	96.21%	8
			FTSLFSLIP	89.54%	14
			VAHSLAVEL	98.54%	19
			GWATTALLI	99.31%	14
			LITFCFGWL	96.39%	9
			FCFGWLLIP	97.43%	10
			LLTFSCSHY	96.15%	11
			YSTESKFKV	98.16%	13
			KFKVILERV	98.46%	14
			VILERVKVE	96.23%	10
			ILERVKVEY	90.67%	4
			YQKTMGSMV	93.92%	15
			TLGVFRYKS	90.19%	5
			FRYKSRCYV	96.40%	13
			CYVGLVWGI	97.29%	10
			YVGLVWGIL	97.14%	13
			LVWGILLTT	95.81%	8
			ILLTTELII	99.45%	15
			LLTTELIIW	97.48%	9
			IIWAASADT	90.00%	13

WAASADTPL	91.43%	12
FALASSSSY	98.78%	17
FHFQLDKQV	99.50%	19
FQLDKQVVH	93.65%	11
VVHAEIQNL	93.73%	15
YPWQTAKCF	93.57%	10
CFFEKDYQY	92.71%	8
FFEKDYQYE	93.01%	6
VYLDKLRSV	96.48%	13
YKIVSLKYT	95.55%	15
LKYTRKVCI	99.27%	19
QPGDTLLFL	95.83%	9
LFLGPLEQG	98.37%	10
YGATVTNLL	94.51%	16
FTKSGEWLL	99.24%	17
ILILSILLF	93.35%	7

Table 9: Glycoprotein epitopes with highest population coverage percentage.

Epitope	Coverage	Total HLA hits	Epitope	Coverage	Total HLA hits
	Class I			**Class II**	
GLYILSFAL	39.08%	1	ITLHEQQLV	83.18%	9
YILSFALPI	44.14%	3	DAERAVELD	79.44%	6
ILSFALPII	39.08%	1	SRRAAVSAL	92.35%	9
IILKALYML	40.60%	2	RRAAVSALE	86.21%	9
IILKALYML	40.60%	2	LKRELADLI	94.55%	18
GVIGFSFFV	42.53%	3	ADLIAAQKL	84.17%	8
FLAARCPFL	51.18%	4	LIAAQKLAS	96.44%	17
HLYVSMPTA	39.08%	1	LRYGNVLDV	99.07%	16
LRYGNVLDV	33.31%	2	WKSIGLYIL	94.46%	15
IRKPRHLYV	33.31%	2	LYILSFALP	93.34%	7
			YILSFALPI	99.77%	23
			ILKALYMLS	99.61%	18
			LKALYMLST	98.72%	16
			LRRTQSMGI	81.74%	12
			IIILYMSHW	91.28%	11
			LYMSHWGRE	95.37%	10
			LRELAQTLV	84.55%	11
			LRRTQSMGI	81.74%	12
			AFYQSYLRR	96.21%	9
			FYQSYLRRT	98.79%	16
			KSAFYQSYL	93.67%	9
			IMASKSVGT	91.14%	13
			ELGAFFAIL	96.65%	10
			LGAFFAILQ	96.05%	15
			GAFFAILQD	94.82%	10
			AFFAILQDM	89.50%	9
			FFAILQDMR	99.50%	20
			IFADIATPH	89.33%	14
			AESATIFAD	96.21%	12
			ESATIFADI	97.74%	10
			ARAESATIF	92.52%	8
			LATNRAYFI	97.70%	14
			FVKDWMERI	99.52%	19
			FLAARCPFL	99.74%	22
			VMGVIGFSF	97.33%	10
			VIGFSFFVK	97.77%	11

IGFSFFVKD	93.55%	8
FRTIACGLF	91.66%	10
ILKALYMLS	99.61%	18
LKALYMLST	98.72%	16
LYILSFALP	93.34%	7
YILSFALPI	99.77%	23
WKSIGLYIL	94.46%	15
LKRELADLI	94.55%	18

Table 10: Nucleocapsid epitopes with high population coverage percentage.

Epitope	Coverage	Total HLA hits	Epitope	Coverage	Total HLA hits
	Class I			**Class II**	
YILSFALPI	44.14%	3	YILSFALPI	99.77%	23
FLAARCPFL	51.18%	4	FLAARCPFL	99.74%	22
Epitope set	54.09%		Epitope set	99.77%	

Table 11: Proposed epitopes in both MHC class I and II of nucleocapsid.

Epitope	Coverage	Total HLA hits	Epitope	Coverage	Total HLA hits
	Class I			**Class II**	
MTLPVTCFL	55.00%	8	ILLTQVADL	99.65%	22
YTFTSLFSL	55.70%	7	LLSSRIQVI	99.65%	21
Epitope set	61.06%		Epitope set	99.89%	

Table 12: Proposed epitopes in both MHC class I and II of glycoprotein.

ethnically biased population coverage [53] Of greater significance is the finding that all CD8+ T cell predicted epitopes showed interactions with HLA-A*02 and most of HLA-A*30, -A*68, -B*08, -B*39, -B*57 and C*04:01 which are more predominant HLA alleles in the US and North America where infection and outbreak of SNV have occurred [54]. On the other hand, it was found that all the CD4+ T cell epitopes interacted with HLA-DRB1*03:02, -DRB1*03:01 and -DRB1*07:01 which are among most common haplotype observed within American populations, that reported by Martin et al. in 2007.

Conclusion

Bioinformatics approaches are a fundamental component of a high throughput pipeline for *in silico* mapping of thousands of potential epitopes, minimizing the time and cost involved in the experimental testing of such peptides [53]. Target populations for the vaccine are inhabitants of SNV-endemic geographic areas and people with occupational risks of infection [1]. Analysis for the proteasomes and transporter associated with antigen processing (TAP) efficiency is required; in order to predict the candidate epitopes based on the processing of the peptides *in vivo*. Along with *in silico* study, further *in vivo* and *in vitro* experiments are needed to prove the effectiveness of triggering and mounting an immune response. Our study for the first time has predicted a cocktail of B- and T-cell epitopes for designing an effective vaccine against HCPS caused by SNV.

References

1. Krüger D, Schönrich G, Klempa B (2011) Human pathogenic hantaviruses and prevention of infection. Hum Vaccin 7: 685-693.
2. Maes P, Clement J, Ranst M (2009) Recent approaches in hantavirus vaccine development. Expert Rev Vaccines 8: 67-76.
3. Hussein I, Haseeb A, Haque A, Mir M (2011) Recent Advances in Hantavirus Molecular Biology and Disease. Adv Appl Microbiol 74: 35-75.
4. Levinson W (2012) Review of medical microbiology and immunology.12th ed. US: Mc Graw Hill; 2012
5. Larson R, Brown D, Ye C, Hjelle B (2005) Peptide Antagonists That Inhibit Sin Nombre Virus and Hantaan Virus Entry through the 3-Integrin Receptor. J Virol 79: 7319-7326.
6. Bondu V, Schrader R, Gawinowicz M, McGuire P, Lawrence D, et al. (2015) Elevated Cytokines, Thrombin and PAI-1 in Severe HCPS Patients Due to Sin Nombre Virus. Viruses 7: 559-589.
7. MacNeil A, Nichol S, Spiropoulou C (2011) Hantavirus pulmonary syndrome. Virus Res 162: 138-147.
8. Koster F, Mackow E (2012) Pathogenesis of the hantavirus pulmonary syndrome. Future Virology 7: 41-51.
9. Bharadwaj M, Nofchissey R, Goade D, Koster F, Hjelle B (2000) Humoral Immune Responses in the Hantavirus Cardiopulmonary Syndrome. J Infect Dis 182: 43-48.
10. Ye C, Prescott J, Nofchissey R, Goade D, Hjelle B (2004) Neutralizing Antibodies and Sin Nombre Virus RNA after Recovery from Hantavirus Cardiopulmonary Syndrome. Emerg Infect Dis 10: 478-482.
11. Mattar S, Garzon D, Tadeu L, Faccini-Martínez AA, Mills JN (2014) Serological diagnosis of hantavirus pulmonary syndrome in a febrile patient in Colombia. Int J Infect Dis 25: 201-203.
12. Bharadwaj M, Mirowsky K, Ye C, Botten J, Masten B, et al. (2002) Genetic vaccines protect against Sin Nombre hantavirus challenge in the deer mouse (Peromyscus maniculatus). J Gen Virol 83: 1745-1751.
13. Schmaljohn C (2009) Vaccines for Hantaviruses. Vaccine 27: D61–D64.
14. US Centers for Disease Control and Prevention (2016). Atlanta, GA: CDC.
15. Xiao R, Yang S, Koster F, Ye C, Stidley C, et al. (2006) Sin Nombre Viral RNA Load in Patients with Hantavirus Cardiopulmonary Syndrome. J Infect Dis 194: 1403-1409.
16. Carver S, Mills J, Parmenter C, Parmenter R, Richardson K, et al. (2015) Toward a Mechanistic Understanding of Environmentally Forced Zoonotic Disease Emergence: Sin Nombre Hantavirus. Bioscience 65: 651-666.
17. Schmaljohn C, Hjelle BB (1997) Hantaviruses: A Global Disease Problem. Emerg Infect Dis 3: 95-104.
18. Oyarzuna P, Ellisa J, Gonzalez-Galarzac F, Jonesc A, Middletond D, et al. (2015) A bioinformatics tool for epitope-based vaccine design that accounts for human ethnic diversity: Application to emerging infectious diseases. Vaccine 33:1267-1273.

19. Shawan M, Al Mahmud H, Hasan M, Parvin AA, Rahman M, et al. (2014) In silico modeling and immunoinformatics probing disclose the epitope based peptide vaccine against Zika virus envelope glycoprotein. Indian J Pharm Biol Res 2: 44-57.

20. Hoopera J, Josleyna M, Ballantyneb J, Brocatoa R (2013) A novel Sin Nombre virus DNA vaccine and its inclusion in a candidate pan-hantavirus vaccine against hantavirus pulmonary syndrome (HPS) and hemorrhagic fever with renal syndrome (HFRS). Vaccine 31: 4314-4321.

21. Wilken JA, Jackson R, Materna BL, Windham GC, Enge B, et al. (2015) Assessing prevention measures and Sin Nombre hantavirus seroprevalence among workers at Yosemite National Park. Am J Ind Med 58: 658-667.

22. Oyarzuna P, Ellisa J, Gonzalez-Galarzac F, Jonesc A, Middletond D, et al. (2015) A bioinformatics tool for epitope-based vaccine design that accounts for human ethnic diversity: Application to emerging infectious diseases. Vaccine 33: 1267-1273.

23. Kashala O, Amador R, Valero M, Morino A, Barbosa A, et al. (2002) Safety, tolerability and immunogenicity of new formulations of the Plasmodium falciparum malaria peptide vaccine SPf66 combined with the immunological adjuvant QS-21. Vaccine 20: 2263-2277.

24. Nardin EH, Oliveira GA, Calvo-Calle JM, Castro ZR, Nussenzweig RS, et al. (2000) Synthetic malaria peptide vaccine elicits high levels of antibodies in vaccinees of defined HLA genotypes. J Infect Dis 182: 1486-1496.

25. Engler O, Dai W, Sette A, Hunziker IP, Reichen J, et al. (2001) Peptide vaccines against hepatitis B virus: from animal model to human studies. Mol Immunol 38: 457-465.

26. Gahery H, Daniel N, Charmeteau B, Ourth L, Jackson A, et al. (2006) New CD4+ and CD8+ T cell responses induced in chronically HIV type-1-infected patients after immunizations with an HIV type 1 lipopeptide vaccine. AIDS Res Hum Retroviruses 22: 684-694.

27. Pinto L, Berzofsky J, Fowke K, Little RF, Merced-Galindez F, et al. (1999) HIV-specific immunity following immunization with HIV synthetic envelope peptides in asymptomatic HIV-infected patients. AIDS 13: 2003-2012.

28. Lundegaard C, Lund O, Kesmir C, Brunak S, Nielsen M (2007) Modeling the adaptive immune system: predictions and simulations. Bioinformatics 23: 3265-3275.

29. Badawi M, Osman M, Alla AF, Ahmedani A, Abdalla M, et al. (2016) Highly Conserved Epitopes of ZIKA Envelope Glycoprotein May Act as a Novel Peptide Vaccine with High Coverage: Immunoinformatics Approach. Ameran J Infectious Diseases and Microbiology 4: 46-60.

30. Badawi M, Alla AF, Alam S, Mohamed W, Osman D, et al. (2016) Immunoinformatics Predication and in silico Modeling of Epitope-Based Peptide Vaccine Against virulent Newcastle Disease Viruses. Ameran J Infectious Diseases and Microbiology 4: 61-71

31. Hall TA (2016) "BioEdit: a user-friendly biologicalsequence alignment editor and analysis program for Windows 95/98/NT". Nucl Acids Symp 41: 95-98.

32. Hasan A, Hossain M, Alam M (2013) A Computational Assay to Design an Epitope-Based Peptide Vaccine Against Saint Louis Encephalitis Virus. Bioinform Biol Insights 7: 347-355.

33. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, et al. (2014) The immune epitope database (IEDB) 3.0. Nucleic Acids Res 43: D405-12.

34. Larsen J, Lund O, Nielsen M (2006) Improved method for predicting linear B-cell epitopes. Immunome Res 2: 2.

35. Emini EA, Hughes JV, Perlow DS, Boger J (1985) Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J Virol 55: 836-839.

36. Kolaskar AS, Tongaonkar PC (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett 276: 172-174.

37. Buus S, Lauemøller SL, Worning P, Kesmir C, Frimurer T, et al. (2003) Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach. Tissue Antigens 62: 378-384.

38. Nielsen M, Lund O (2009) NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. BMC Bioinformatics10: 296.

39. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, et al. (2006) Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics 7: 153.

40. Wang S, Li W, Liu S, Xu J (2016) RaptorX-Property: a web server for protein structure property prediction. Nucleic Acids Res 44: W430-W435.

41. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera-a visualization system for exploratory research and analysis. J Comput Chem 25: 1605-1612.

42. Hans D, Young P, Fairlie D (2006) Current status of short synthetic peptides as vaccines. Med Chem 2: 627-646.

43. Patronov A, Doytchinova I (2013) T-cell epitope vaccine design by immunoinformatics. Open Biol 3: 120139.

44. Safronetza D, Prescotta J, Feldmannb F, Haddocka E, Rosenkeb R, et al. (2014) Pathophysiology of hantavirus pulmonary syndrome in rhesus macaques. PNAS 111: 7114-7119.

45. Ramos M, Overturf G, Crowley M, Rosenberg R, Hjelle B, et al. (2001) Infection with Sin Nombre Hantavirus: clinical presentation and outcome in children and adolescents. Pediatrics 108: e27.

46. Hooper JW, Brocato RL, Kwilas SA, Hammerbeck CD, Josleyn MD, et al. (2014) DNA vaccine–derived human IgG produced in transchromosomal bovines protect in lethal models of hantavirus pulmonary syndrome. Sci Transl Med 6: 264ra162.

47. Kwilas S, Kishimori J, Josleyn M, Jerke K, Ballantyne J, et al. (2014) Hantavirus Pulmonary Syndrome (HPS) DNA vaccine delivered using a spring-powered Jet Injector Elicits a potent neutralizing antibody response in rabbits and nonhuman primates. Curr Gene Ther 14: 200-210.

48. Hsu S, Chargelegue D, Obeid O, Steward M (1999) Synergistic effect of immunization with a peptide cocktail inducing antibody, helper and cytotoxic T-cell responses on protection against respiratory syncytial virus. J Gen Virol 80: 1401-1405.

49. Kalaiselvan SS, Sankar S, Ramamurthy M, Ghosh A, Nandagopal B, et al. (2017) Prediction of pan-specific B-cell epitopes from nucleocapsid protein of hantaviruses causing hantavirus cardiopulmonary syndrome. J Cell Biochem 118: 2320-2324.

50. Klavinskis L, Whitton J, Oldstone M (1989) Molecularly engineered vaccine which expresses an immunodominant T-cell epitope induces cytotoxic T lymphocytes that confer protection from lethal virus infection. J Virol 63: 4311-4316.

51. Keogh E, Fikes J, Southwood S, Celis E, Chesnut R, et al. (2001) Identification of new epitopes from four different tumor-associated antigens: recognition of naturally processed epitopes correlates with HLA-A*0201-binding affinity. J Immunol 167: 787-796.

52. Sankar S, Ramamurthy M, Nandagopal B, Sridharan G (2017) T-cell epitopes predicted from the Nucleocapsid protein of Sin Nombre virus restricted to 30 HLA alleles common to the North American population. Bioinformation 13: 94-100.

53. Ben-Yedidia T, Arnon R (2007) Epitope-based vaccine against influenza. Expert Rev Vaccines 6: 939-948.

54. Mori M, Beatty PG, Graves M, Boucher KM, Milford EL (1997) HLA gene and haplotype frequencies in the North American population: the National Marrow Donor Program Donor Registry. Transplantation 64: 1017-1027.

55. Maiers M, Gragert L, Klitz W (2007) High-resolution HLA alleles and haplotypes in the United States population. Hum Immunol 68: 779-788.