

RESEARCH ARTICLE

# Epitope-Based Vaccine Target Screening against Highly Pathogenic MERS-CoV: An *In Silico* Approach Applied to Emerging Infectious Diseases

Jiandong Shi<sup>1,2</sup>✉, Jing Zhang<sup>1,3</sup>✉, Sijin Li<sup>1,2</sup>, Jing Sun<sup>1,2</sup>, Yumei Teng<sup>1,2</sup>, Meini Wu<sup>1,2</sup>, Jianfan Li<sup>1,2</sup>, Yanhan Li<sup>1,2</sup>, Ningzhu Hu<sup>1,2</sup>, Haixuan Wang<sup>1,2</sup>, Yunzhang Hu<sup>1,2</sup>\*

**1** Institute of Medical Biology, Chinese Academy of Medical Sciences and Peking Union Medical College, Kunming, China, **2** Yunnan Provincial Key Laboratory of Arbo Infectious Disease Control Research (Preparing), Institute of Medical Biology, Chinese Academy of Medical Sciences, Kunming, China, **3** Prenatal Diagnosis Center, Shijiazhuang Obstetrics and Gynecology Hospital, Shijiazhuang, Hebei, China

✉ These authors contributed equally to this work.

\* [huyunzhangym@126.com](mailto:huyunzhangym@126.com)



OPEN ACCESS

**Citation:** Shi J, Zhang J, Li S, Sun J, Teng Y, Wu M, et al. (2015) Epitope-Based Vaccine Target Screening against Highly Pathogenic MERS-CoV: An *In Silico* Approach Applied to Emerging Infectious Diseases. PLoS ONE 10(12): e0144475. doi:10.1371/journal.pone.0144475

**Editor:** Ranjit Ray, Saint Louis University, UNITED STATES

**Received:** October 1, 2015

**Accepted:** November 18, 2015

**Published:** December 7, 2015

**Copyright:** © 2015 Shi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This research was supported by grant from the Applied and Fundamental Research program of Yunnan Province (Grant No. 2013FA025 and 2013BC011), which were received by Yunzhang Hu and Ningzhu Hu. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

## Abstract

Middle East respiratory syndrome coronavirus (MERS-CoV) with pandemic potential is a major worldwide threat to public health. However, vaccine development for this pathogen lags behind as immunity associated with protection is currently largely unknown. In this study, an immunoinformatics-driven genome-wide screening strategy of vaccine targets was performed to thoroughly screen the vital and effective dominant immunogens against MERS-CoV. Conservancy and population coverage analysis of the epitopes were done by the Immune Epitope Database. The results showed that the nucleocapsid (N) protein of MERS-CoV might be a better protective immunogen with high conservancy and potential eliciting both neutralizing antibodies and T-cell responses compared with spike (S) protein. Further, the B-cell, helper T-cell and cytotoxic T lymphocyte (CTL) epitopes were screened and mapped to the N protein. A total of 15 linear and 10 conformational B-cell epitopes that may induce protective neutralizing antibodies were obtained. Additionally, a total of 71 peptides with 9-mer core sequence were identified as helper T-cell epitopes, and 34 peptides were identified as CTL epitopes. Based on the maximum HLA binding alleles, top 10 helper T-cell epitopes and CTL epitopes that may elicit protective cellular immune responses against MERS-CoV were selected as MERS vaccine candidates. Population coverage analysis showed that the putative helper T-cell epitopes and CTL epitopes could cover the vast majority of the population in 15 geographic regions considered where vaccine would be employed. The B- and T-cell stimulation potentials of the screened epitopes is to be further validated for their efficient use as vaccines against MERS-CoV. Collectively, this study provides novel vaccine target candidates and may prompt further development of vaccines against MERS-CoV and other emerging infectious diseases.

## Introduction

Middle East respiratory syndrome (MERS) is a newly emerging acute respiratory system infectious disease, which is characterised by acute pneumonia, respiratory failure, and renal failure and caused by MERS-coronavirus (MERS-CoV) [1,2]. MERS-CoV was first isolated and identified in Saudi Arabia in 2012 as a new member in the lineage C of the genus *Betacoronavirus*, and is a zoonotic virus that is transmitted from bats to camels, and then to humans [3,4,5]. The genome of MERS-CoV is a non-segmented positive-sense, single-stranded RNA (ssRNA) of approximately 30 kb encoding 11 proteins. They include two replicase polyproteins (open reading frames 1ab and 1a); four structural proteins including spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins; and five nonstructural proteins (open reading frames 3, 4a, 4b, 5, and 8b) [6]. MERS-CoV can be transmitted from person-to-person and has as high as 40% mortality [7]. As of June 26 2015, a total of 1356 cases including 484 related deaths have been reported to the World Health Organization [8]. No licensed vaccine or specific treatment is currently available.

As one of the most effective strategies to prevent virus infection, vaccination is considered to be indispensable especially in the absence of effective treatment drugs. The continuous increase of patients and a high mortality rate of MERS-CoV infection highlight the urgent need for safe and effective vaccines development. The majority of recent progresses focused on the viral S glycoprotein and receptor-binding domain (RBD) of S protein [9–19], and showed that S protein-based or RBD-based subunit or vector vaccine that contain multiple neutralizing epitopes with high potentials to induce strong neutralizing antibodies against MERS-CoV infection; thus, it is considered to be a promising target for effective MERS vaccine design [9–19]. Additionally, Almazan *et al.* constructed a full-length infectious cDNA clone of the MERS-CoV genome that lacked of E gene by reverse genetics approach to develop attenuated viruses as vaccine candidates [20]. Theoretically, nearly all viral proteins are potential immunogens and vaccine targets. Thus, in addition to the S protein, other viral proteins of the MERS-CoV might be effective immunogen targets of MERS vaccine. However, few studies have analyzed or tested the possibility of other viral proteins of MERS-CoV as vaccine targets. More importantly, T-cell-based cellular immunity is essential for cleaning MERS-CoV infection, yet the vaccine against the S protein mainly elicits neutralizing antibody response. Further, the high mutation rate of the S protein may result in the escape of neutralizing antibodies against MERS-CoV. Therefore, a highly conserved target that elicit both neutralizing antibody and cellular immunity against MERS-CoV is essential for an effective vaccine development.

Epitopes are also known as the antigenic determinant, which represents the minimal immunogenic region of a protein antigen and precisely elicit a specific immune responses [21]. Numerous studies show that epitope-based vaccines could effectively elicit protective immune responses against diverse pathogens, such as human immunodeficiency virus, influenza virus, hepatitis B virus, and hepatitis C virus [22–25]. Nevertheless, for emerging highly pathogenic pathogens, like MERS-CoV, H5N1, and H7N9, although their genome sequences are available in GenBank database, their immunity associated with protection is currently largely unknown. These gaps in understanding immune protection make developing vaccines against these highly pathogenic viruses more difficult and challenging [26]. Therefore, an immunoinformatics-driven approach to thoroughly screen the dominant immunogens based on available genome sequences data of pathogens is essential and promising for effective vaccine design of emerging infectious diseases [26].

In this study, based on newly available genome sequence data of MERS-CoV, numerous epitope vaccine candidates that could elicit protective humoral and cellular immune responses were obtained by immunoinformatics-driven vaccine target screening strategy. These results

provided new vaccine epitope candidates for MERS vaccine development, and indicated that immunoinformatic-driven immunogen screening is a promising strategy to accelerate vaccine development of the emerging highly pathogenic pathogens.

## Materials and Methods

### Protein sequence retrieval

The entire viral proteome sequences of all MERS-CoV isolates that include structural proteins spike protein (S), envelope protein (E), membrane protein (M), nucleocapsid protein (N), and non-structural proteins ORF3, ORF4a, ORF4b, ORF5 and ORF8b, and two replicase polyproteins ORF1ab and ORF1a, were retrieved from the National Center for Biotechnology Information database (<http://www.ncbi.nlm.nih.gov/>) and used as an input for various bioinformatics tools for antigenicity assessment, secondary structure and 3D model prediction, and epitope prediction. These sequences come from different geographic regions such as Saudi Arabia, England, Qatar, Spain, Germany, Jordan, and Korea with time ranges from 2012 to 2015.

### Antigenic protein screening and structure analysis

Antigenicity refers to the ability of antigen that can be recognized by the immune system. Hence, to find the best potential candidate antigen, the VaxiJen v2.0 server [27] (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) was used for analyzing whole protein antigenicity and determined the most potent antigenic protein. The VaxiJen is a web server, that was developed by Dr. Doytchinova IA and Dr. Flower DR, from Faculty of Pharmacy, Medical University of Sofia, for prediction of protective antigens, tumor antigens and subunit vaccines, with prediction accuracy of 70% to 89% for the discrimination between antigens and non-antigens [27]. In this study, virus was chosen as the target organism with default parameter. A single antigenic protein with the highest antigenicity score was selected as vaccine candidate for further analysis. Antigenicity is closely related to the secondary and tertiary structure of the protein; hence, to determine the antigenicity and function of the target protein, the secondary structure of target protein that included solvent accessibility, transmembrane helices, globular regions, and coiled coil regions were predicted using the ExPASy's secondary structure prediction server ProtParam [28] (<http://web.expasy.org/protparam/>) and a self-optimized prediction method for alignment (SOPMA) [29] ([https://npsa-prabi.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_sopma.html](https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html)) with default parameter. ProtParam is a tool which allows the computation of various parameters including the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) for a given protein. The SOPMA method correctly predicts 69.5% of the secondary structure  $\alpha$ -helix,  $\beta$ -sheet and coil [29]. The RaptorX web server [30] (<http://raptorx.uchicago.edu/>) was used to predict the three-dimensional (3D) structure and binding residues of the chosen protein.

### B-cell epitope prediction

The method of Kolaskar and Tongaonkar [31] at the Immune Epitope Database (IEDB) (<http://tools.immuneepitope.org/bcell/>) was applied to predict linear B-cell epitopes. The Kolaskar and Tongaonkar method has been applied to a large number of proteins to predict B-cell epitopes by the developers. Their results showed that the method can predict epitopes with about 75% accuracy; hence, it is better than most of the known methods [31]. Further, the properties of the B-cell epitopes also include the flexibility, surface accessibility and hydrophobicity; hence, they were predicted by the Karplus and Schulz flexibility prediction [32] and

Emini surface accessibility prediction [33] with a threshold of 1.0 respectively, of the IEDB. Conformational B-cell epitopes were predicted by ElliPro (<http://tools.immuneepitope.org/ellipro/>) [34] with the minimum score value set at 0.7 while the maximum distance selected as 6 Å. The ElliPro allows the prediction and visualization of B-cell epitopes in a given protein sequence or structure. The test for ElliPro on a benchmark dataset of discontinuous epitopes inferred from 3D structures of antibody-antigen complexes has shown the best performance compared with six other structure-based epitope prediction methods [34].

## Helper T-cell epitope prediction

The NetMHCIIpan 3.0 server [35] (<http://www.cbs.dtu.dk/services/NetMHCIIpan/>) was used to predict helper T-cell epitopes. Threshold for strong binding peptides ( $IC_{50}$ ) was set at 50 nM to determine the binding and interaction potentials of helper T-cell epitope peptide and major histocompatibility complex (MHC) class II allele. The NetMHCIIpan-3.0 is the first pan-specific helper T-cell epitope prediction method that covers all human leucocyte antigen (HLA) class II molecules including HLA-DR, HLA-DP, and HLA-DQ. It can predict peptide binding to any HLA class II molecule in a specified protein sequence [34]. The method was evaluated by the developers and shown a significant improvement over molecule-specific methods; hence, it is considered the most accurate MHC class II predictor [36]. Here, the top 10 epitopes with the maximum of binding HLA-DR alleles were selected as putative epitope candidates.

## CTL epitopes prediction

The cytotoxic T lymphocyte (CTL) epitope prediction was made using the IEDB analysis resource NetMHCpan (version 2.8) tool [37,38] and the Consensus tool [39] which combines predictions from ANN aka NetMHC (3.4) [40][41], SMM [42], and Comblib [43]. The NetMHCpan is a high-throughput computational method for genome- and HLA-wide prediction of peptide binding to HLA class I molecules, because it contains all HLA class I alleles. Hence, it can offer a truly global analysis for binding of the antigenic peptide and HLA molecule and promote rational vaccine design [37]. The NetMHC is a prediction method with an average of 75–80% accuracy for peptides binding to HLA class I molecules. It has been employed widely to predict HLA-binding peptides in proteomes of various pathogens including SARS, Influenza and HIV [41]. The percentile rank threshold was set at 0.5. In this study, the top 10 epitopes with the maximum of binding HLA-A alleles were selected as putative epitope vaccine candidates. Further, except for a strong binding affinity, the peptides with strong immunogenicity are more likely to be CTL epitopes than those with weak immunogenicity. Hence, the IEDB immunogenicity prediction tool (<http://tools.immuneepitope.org/immunogenicity/>) was used to predict the immunogenicity of the candidate epitopes [44]. This tool predicts immunogenicity of a peptide-HLA (pHLA) complex based on the amino acid properties and their positions in the defined peptide.

## Conservancy analysis

The epitope conservancy analysis tool ([http://tools.immuneepitope.org/tools/conservancy/iedb\\_input](http://tools.immuneepitope.org/tools/conservancy/iedb_input)) at the IEDB was applied for the epitope conservancy analysis [45]. The conservancy levels were obtained by searching for identities in the given protein sequence. This tool calculates the degree of conservancy of an epitope within a given protein sequence set at different degrees of sequence identity. The degree of conservancy is defined as the fraction of protein sequences containing the epitope at a given identity level.

## Allergenicity assessment

The allergenicity of the epitopes was analyzed by the AllerHunter server [46] (<http://tiger.dbs.nus.edu.sg/AllerHunter>). AllerHunter is a SVM-pairwise system for assessment of allergenicity and allergic cross-reactivity in proteins. It aims to predict allergens and non-allergens with high sensitivity and specificity, without compromising efficiency at classification of proteins with similar sequences to known allergens [46].

## Population coverage prediction

Due to the MHC restriction of T cell response, the peptides with more different HLA binding specificities mean more population coverage in defined geographical regions where the peptide-based vaccine might be employed. The population coverage rate of individual epitope was calculated using the IEDB population coverage tool ([http://tools.immuneepitope.org/tools/population/iedb\\_input](http://tools.immuneepitope.org/tools/population/iedb_input)) [47].

Every epitope and its binding HLA alleles were added, and different geographic areas were also selected.

## Results

### Antigenic protein identification and structure analysis

The selection of immunogen is the first step for vaccine design; hence, to obtain the most probable antigenic protein, the whole viral proteomes of all MERS-CoV isolates were retrieved and screened. A total of 99 MERS-CoV variants from different geographical regions with their entire proteomes were obtained (S1 Table). The antigenicity of each protein is indicated by the overall score produced by the specific protein sequence using the VaxiJen server. On the whole, the scores of the N protein and ORF8b protein of MERS-CoV were higher than the S, E, M, ORF3, ORF4a, ORF4b, ORF5, ORF1ab and ORF1a proteins. Moreover, the ORF8b protein with its Genbank ID: AIL23997.1 and the N protein with its Genbank ID: AGV08499.1 possessed the significantly highest antigenic scores of 0.8218 and 0.7749, respectively, among all the query proteins (S1 Table). However, the ORF8b protein would be rapidly degraded by proteasomes in the absence of the ORF8a protein, an accessory protein of ORF8b [48]. Thus, it was not an ideal antigen candidate although its high antigenic score. In this study, the N protein was chosen as a candidate immunogen and model protein to carry out epitope-based vaccine design. Further, the most probable antigenic protein was analyzed for its secondary structural characteristics, and the properties that included total length of 413 aa, molecular weight of 45048.2 Da, theoretical pI of 10.05, formula of  $C_{1965}H_{3102}N_{594}O_{611}S_7$ , 70 alpha helices, 66 extended strands, 32 beta turns, and 245 random coils are obtained and shown in Table 1 and Fig 1. The grand average of the hydrophobicity rule (GRAVY) of the N protein linear sequence was predicted to be negative (-0.865). This indicated the property of the protein as hydrophilic in nature and most of the residues to be present on the surface. This means that more amino acids tend to be binding residues when interacted with other proteins. The 3D structure showed a maximum of 64% identity of the N protein of MERS-CoV with the best template protein PDB: 2ofzA (RNA Binding Domain of Sars Nucleocapsid Protein), and it is composed of two separate domains or pockets (Fig 2). Protein binding site prediction showed that a total of 12 binding residues including T40, V41, S42, Y44, T45, G46, R97, Y99, Y101, R138, A145 and S173 were mapped on the domain-1. And only 2 binding residues including G267 and L268 were mapped on the domain-2. Obviously, the domain-1 possessed the greater ability to interact with other proteins than the domain-2. This might be associated with distribution of the conformational epitopes on the N protein of MERS-CoV.

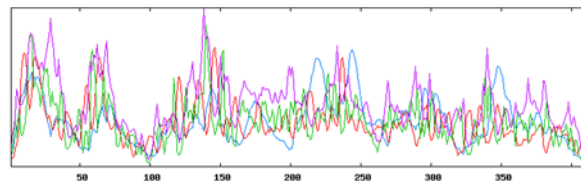
**Table 1. Secondary structural characteristics of the highest antigenic N protein (Genbank ID: AGV08499.1) of MERS-CoV.**

Criteria	Assessment
Number of Amino Acids	413
Molecular Weight	45048.2 Da
Theoretical pI	10.05
Total number of negatively charged residues (Asp + Glu)	33
Total number of positively charged residues (Arg + Lys)	55
Formula	C <sub>1965</sub> H <sub>3102</sub> N <sub>594</sub> O <sub>611</sub> S <sub>7</sub>
Extinction coefficients	47900
Estimated half-life	30 hours (mammalian reticulocytes, in vitro)
Instability index	48.62
Aliphatic index	56.76
Grand average of hydropathicity (GRAVY)	-0.865

doi:10.1371/journal.pone.0144475.t001

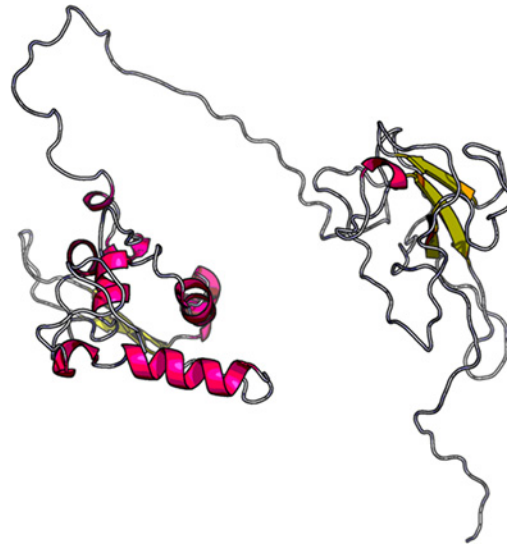
### Identification of B-cell epitopes

One of the key steps in epitope-driven vaccine design is the prediction and identification of the B-cell epitopes in target antigens. Thus, to obtain B-cell epitope candidates in the N protein of MERS-CoV, *in silico* identification of B-cell epitopes based on the IEDB database was performed. Based on the Kolaskar and Tongaonkar's method of the IEDB, a total of 15 linear B-cell epitopes of the N protein of MERS-CoV were identified. The length of the epitopes ranged from 6 to 20 amino acids. These epitopes had 78–91% conservancy level among the specified N protein sequences. Notably, the epitopes <sup>4</sup>PAAPRAVSF<sup>12</sup> and <sup>324</sup>NPVYFLRYSGAIKL<sup>337</sup> were allergic to human; hence, they could not be vaccine candidates. Epitopes' length, sequences, location, conservancy, and allergenicity are shown in Table 2. Further, the results showed that the average antigenic propensity value of the predicted epitopes was 0.993 with a minimum of 0.863 and a maximum of 1.182 (S1 Fig). Since surface accessibility and fragment flexibility are also key features for predicting B-cell epitopes. Thus, the surface accessibility and flexibility were analyzed based on methods of the IEDB. Results of the analysis of the surface accessibility of the predicted peptides showed that the maximum surface probability value was 6.971 at amino acid position from 363 to 368. The sequence of the hexapeptide is <sup>363</sup>KKEKKQ<sup>368</sup>, where 365E is the surface residue. The minimum value of surface probability is 0.074 for peptides <sup>205</sup>GIGAVG<sup>210</sup>, where 207G is the surface residue (S2 Fig). Likewise, results of the analysis of the flexibility of the predicted peptides showed that the maximum flexibility value was 1.160 at amino acid position from 170 to 176, and its sequence is <sup>167</sup>GNSQSS<sup>173</sup>, where 170Q is the flexible residue. The minimum value of flexibility probability is 0.903 for peptides <sup>97</sup>RWYFYTT<sup>103</sup>, where 100F is the flexibility residue (S3 Fig). In addition, a total of 10



**Fig 1. Secondary structure plot of the highest antigenic nucleocapsid (N) protein of MERS-CoV.** Here, helix is indicated by blue, while extended strands, beta turns and random coils are indicated by red, green and yellow, respectively.

doi:10.1371/journal.pone.0144475.g001



**Fig 2. A 3D model of the highest antigenic nucleocapsid (N) protein of MERS-CoV that were modeled by using RaptorX web server.**

doi:10.1371/journal.pone.0144475.g002

conformational B-cell epitopes having a Protrusion Index (PI) score value above 0.7 were obtained by the ElliPro. The highest probability of a conformational epitope was calculated at 97.9% (PI score: 0.979) and shown in Fig 3A. Residues involved in conformational epitopes, their location, number of residues, and scores are shown in Table 3, whereas, their positions on 3D structures are shown in Fig 3A to 3J.

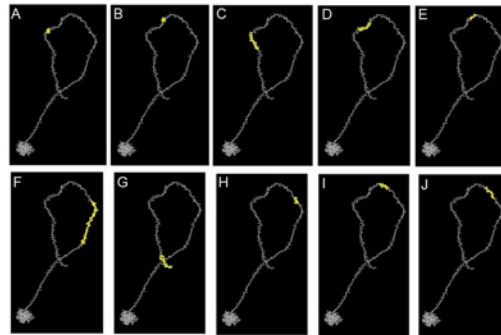
### Identification of helper T-cell epitopes

Since MHC-II restricted CD4<sup>+</sup> T-cells activation is important for inducing and maintaining an efficient antibody response or CTL response, hence, the helper T-cell epitopes in the N protein

**Table 2. Predicted linear B-cell epitopes of the N protein of MERS-CoV with their conservancy and allergenicity.**

Epitope No.	Start	End	Sequence	Length	Conservancy (%)	Allergenicity
1	4	12	PAAPRAVSF	9	87	Yes
2	42	47	SWYTGL	6	91	No
3	50	58	HGKVPLTFP	9	88	No
4	60	66	GQGVPLN	7	91	No
5	92	102	KQLAPRWYFYY	11	90	No
6	108	126	EAALPFRAVKDGIWVHED	19	78	No
7	144	150	SAIVTQF	7	88	No
8	172	180	SSRASSVSR	9	87	No
9	206	225	IGAVGGDLLYLDLLNRLQAL	20	89	No
10	230	240	VKQSQPKVITK	11	89	No
11	261	268	NMVQAFGL	8	90	No
12	296	304	IAELAPTAS	9	89	No
13	324	337	NPVYFLRYSGAIKL	14	90	Yes
14	347	353	WLELLEQ	7	90	No
15	399	405	RPSVQPG	7	89	No

doi:10.1371/journal.pone.0144475.t002



**Fig 3. 3D representation of conformational epitopes (A to J) of the highest antigenic nucleocapsid (N) protein of MERS-CoV.** The epitopes are represented by yellow surface, and the bulk of the N protein is represented in grey sticks.

doi:10.1371/journal.pone.0144475.g003

of MERS-CoV were *in silico* identified. As a result, a total of 71 antigenic peptides with 9-mer core sequences in the N protein of MERS-CoV were identified to be helper T-cell epitopes using the NetMHCIIpan 3.0 server (S2 Table). They can bind a different large number of HLA-DR alleles with an IC<sub>50</sub> value less than 50nM, which indicated a very high binding affinity to HLA-DR molecules. Based on the fact that a good T-cell epitope should interact with as many HLA alleles as possible, the top 10 epitopes with the highest number of binding HLA-DR alleles were selected as putative helper T-cell epitope candidates (Table 4). All of the 10 putative helper T-cell epitopes with numerous binding HLA-DR alleles tend to be good epitope candidates. Among them, the epitope <sup>329</sup>LRYSGAIKL<sup>337</sup> interacting with 357 HLA-DR alleles was the epitope possessing the maximum number of binding HLA-DR alleles. On the contrary, <sup>230</sup>VKQSQPKVI<sup>238</sup> interacting with 94 HLA-DR alleles is the epitope possessing the minimum number of binding HLA-DR alleles. Further, all selected putative epitopes were highly conserved with 89–91% conservancy level, and no epitope was found allergic to humans. Details of predicted putative helper T-cell epitopes along with their respective binding HLA-DR alleles are shown in S2 Table.

**Table 3. Predicted conformational B-cell epitopes of the N protein of MERS-CoV.**

Epitope No.	Residues and their positions	No. of Residues	Score	3D Structure
1	K346, W347, L348, E349	4	0.979	<a href="#">Fig 3A</a>
2	Y327, F328, L329, R330	16	0.893	<a href="#">Fig 3B</a>
3	L350, L351, E352, Q353, N354, I355, D356, A357, Y358, K359, T360, F361, P362, K363, K364, E365	15	0.887	<a href="#">Fig 3C</a>
4	Y331, S332, G333, A334, I335, K336, L337, D338, P339, K340, N341, P342, N343, Y344, N345	7	0.881	<a href="#">Fig 3D</a>
5	D320, D321, H322, G323, N324, P325, V326	41	0.867	<a href="#">Fig 3E</a>
6	H252, K253, R254, T255, S256, T257, K258, S259, F260, N261, M262, V263, Q264, A265, F266, G267, L268, R269, G270, P271, G272, D273, L274, Q275, G276, N277, F278, G279, D280, L281, Q282, L283, N284, K285, L286, G287, T288, E289, D290, P291, R292	15	0.861	<a href="#">Fig 3F</a>
7	R399, P400, S401, V402, Q403, P404, G405, P406, M407, I408, D409, V410, N411, T412, D413	6	0.794	<a href="#">Fig 3G</a>
8	W293, P294, Q295, I296, A297, E298	8	0.778	<a href="#">Fig 3H</a>
9	F312, K313, L314, T315, H316, Q317, N318, N319	13	0.772	<a href="#">Fig 3I</a>
10	L299, A300, P301, T302, A303, S304, A305, F306, M307, G308, M309, S310, Q311	16	0.709	<a href="#">Fig 3J</a>

doi:10.1371/journal.pone.0144475.t003



**Table 4. Helper T-cell epitopes of the N protein of MERS-CoV selected on the basis of maximum HLA binding alleles.**

Epitope core sequence (9 -mer)	Position	No. of binding HLA-DR alleles	Conservancy (%)	Allergenicity
FNMVQAFGL	260–268	173	90	No
LRYSGAIKL	329–337	356	91	No
LQALESQKV	222–230	139	89	No
LNRLQALES	219–227	144	90	No
FMGMSQFKL	306–314	100	91	No
LYDLLNRL	214–222	160	90	No
IKQLAPRWY	91–99	154	89	No
FLRYSQAIK	328–336	142	91	No
LLYDLLNR	213–221	110	90	No
VKQSQPKVI	230–238	94	89	No

doi:10.1371/journal.pone.0144475.t004

### Identification of CTL epitopes

As we all know, MHC-I restricted CD8<sup>+</sup> cytotoxic T lymphocytes (CTLs) plays a crucial role in controlling virus infection. Hence, identification of CTL epitopes is vital for understanding the principles of T cell activation and epitope-driven vaccine design. Herein, a total of 34 immunogenic peptides with 9-mer sequences in the N protein of MERS-CoV were identified to be CTL epitopes using the MHC-I binding predictions of the IEDB with recommended methods ([S3 Table](#)). They can bind a different large number of HLA-A alleles with a very high binding affinity. Top 10 epitopes binding the most HLA-A alleles were selected as putative CTL epitope candidates based on their more broad HLA coverage ([Table 5](#)). Among them, the epitope <sup>92</sup>KQLAPRWYF<sup>100</sup> had the highest number of binding HLA-A alleles (138 alleles), followed by <sup>343</sup>NYNKWLELL<sup>351</sup> (128 alleles). Yet, the epitopes <sup>72</sup>AQNAGYWRR<sup>80</sup> (37 alleles) and <sup>387</sup>RVQGSITQR<sup>395</sup> (31 alleles) had the minimum number of binding HLA-A alleles. Most CTL epitope candidates with a positive score of immunogenicity showed a high potential to elicit strong CTL response. Further, these CTL epitope candidates were highly conserved with 85–91% conservancy level among available N protein sequences of MERS-CoV isolates, and no epitope was found allergenic to humans. Details of predicted putative CTL epitopes along with their respective binding HLA-A alleles are shown in [S3 Table](#).

**Table 5. CTL epitopes of the N protein of MERS-CoV selected on the basis of maximum HLA binding alleles.**

Epitope sequence (9 -mer)	Position	No. of binding HLA-A alleles	Immunogenicity	Conservancy (%)	Allergenicity
AQNAGYWRR	72–80	37	0.28397	91	No
LAPRWYFY	94–102	34	0.35734	90	No
NYNKWLELL	343–351	128	0.06418	90	No
KQLAPRWYF	92–100	138	0.25847	90	No
ELAPTASAF	298–306	53	-0.04059	89	No
NTVSWYTGL	39–47	45	0.11171	90	No
STPAQNAGY	69–77	60	-0.03003	91	No
QLAPRWYFY	93–101	49	0.32573	90	No
HGNPVYFLR	322–330	76	0.11578	89	No
RVQGSITQR	387–395	31	-0.07424	85	No

doi:10.1371/journal.pone.0144475.t005

### Population coverage of the epitopes

HLA distribution varies among different ethnic groups and geographic regions around the world. Thus, population coverage must be taken into account when designing an effective vaccine to cover as much as possible populations. In this study, all indicated alleles in supplementary data were identified as optimum binders with the predicted epitopes and were used to determine the population coverage for these epitopes. The high population coverage was found for all putative helper T-cell epitopes and CTL epitopes in 15 specified geographic regions of the world (Table 6). For helper T-cell epitopes, an excellent percentage of cumulative population coverage of the 10 epitopes was obtained in South Asia (98.68%), closely followed by Central Africa (98.51%), Southeast Asia (98.17%), West Africa (98.10%), East Africa (97.75%), Southwest Asia (97.45%), Oceania (96.78%), West Indies (96.57%), Northeast Asia (96.20%), Europe (95.77%), North Africa (95.05%), North America (94.81%), East Asia (93.89%), and South America (87.18%). And the lowest was in South Africa (62.18%). These results indicated most geographic regions of the world, where more than 90% of their population can be covered by all putative helper T-cell epitopes. For CTL epitopes, a moderate and acceptable percentage of cumulative population coverage of the 10 epitopes was obtained in Central Africa (59.49%), followed by East Asia (50.14%), Oceania (48.85%), Southwest Asia (48.21%), South Asia (47.21%), East Africa (42.51%), Southeast Asia (42.01%), West Africa (41.86%), South Africa (38.95%), North America (35.03%), North Africa (33.02%), South America (32.71%), Northeast Asia (32.47%), and West Indies (30.08%). And the lowest was in Europe (20.62%). These results indicated only partial regions of the world, such as Central Africa and East Asia, where approximately half of the population can be covered by all putative CTL epitopes. Overall, these results suggested that putative helper T-cell epitopes and CTL epitopes can specifically bind with the prevalent HLA molecules in the target population where the vaccine will be employed.

**Table 6. Population coverage rate (%) for all putative helper T-cell epitopes and CTL epitopes of the N protein of MERS-CoV.**

Population/Area	Population coverage (%)	
	All putative helper T-cell epitopes	All putative CTL epitopes
East Asia	93.89	50.14
Northeast Asia	96.20	32.47
South Asia	98.68	47.21
Southeast Asia	98.17	42.01
Southwest Asia	97.45	48.21
Europe	95.77	20.62
East Africa	97.75	42.51
West Africa	98.10	41.86
Central Africa	98.51	59.49
North Africa	95.05	33.02
South Africa	62.18	38.95
West Indies	96.57	30.08
North America	94.81	35.03
South America	87.18	32.71
Oceania	96.78	48.85

doi:10.1371/journal.pone.0144475.t006

## Discussion

Emerging infectious pathogens such as MERS-CoV, H5N1, and H7N9 are highly pathogenic for humans. For these pathogens, immunity associated with protection remain largely unknown. Gaps in understanding the protective immunity against these pathogens make developing vaccines for newly emerging infectious diseases more difficult and challenging [26]. Thus, the genome-wide screening of vaccine targets based on newly available genome sequences data of pathogens is essential and urgent for developing efficiently vaccines against these newly emerging highly pathogenic pathogens.

This study aim to screen and investigate the most antigenic protein of the MERS-CoV and to find out the B- and T-cell epitopes that were mapped on the most antigenic protein by using a strategy namely genome-wide screening of vaccine epitopes. Based on advances in bioinformatics, the immunoinformatics approach could be an effective and essential strategy for vaccine development of emerging highly pathogenic pathogens. In this study, an immunoinformatics-driven approach was used to screen vital dominant immunogens against MERS-CoV. The results showed that the N protein was a better antigenic protein with the highest antigenic scores compared with other viral proteins in entire viral proteome. However, nearly all current studies of MERS vaccine focused on the S protein of MERS-CoV [9–19], because the S protein mediated recognition of host cell DPP4 receptor for MERS-CoV and induced significant immune response [49]. In fact, not just the S protein, other MERS-CoV viral proteins might be effective MERS vaccine candidates. However, few studies have emphasized this possibility. Further, T-cell-based cellular immunity is essential for cleaning MERS-CoV infection, yet the vaccine against the S protein mainly elicit neutralizing antibody response. More importantly, high mutation rate of the S protein of MERS-CoV may cause escape of neutralizing antibodies against S protein. Therefore, an ideal target should be highly conserved that elicit both neutralizing antibody and cellular immunity against MERS-CoV, which is more essential for an effective MERS vaccine development. The N protein of human CoV is abundantly produced during infection and exhibits strong immunogenicity and conservancy, which can act as an ideal immunogen to elicit both cellular and humoral immune responses [50]. It is noteworthy that a large number of previous studies have shown the feasibility of the N protein as an immune target antigen or vaccine product [51–60]. Ma *et al.* demonstrated that a SARS-CoV vaccine based on the N gene that was expressed by DNA plasmid and adenovirus vector could induce detectable antibody and IFN- $\gamma$  [51]. Similar studies of the DNA vaccine based on the SARS-CoV N gene have shown potential inducing specific humoral and cellular immunity in BALB/c mice [52,53]. Moreover, multiple immunodominant B-cell epitopes, helper T-cell epitopes and CTL epitopes were mapped on the N protein of SARS-CoV [54–58]. Additionally, the antigenicity of the N protein from other viruses has been demonstrated [59,60]. And the antibodies against the N proteins of diverse human CoVs have been revealed by Gao *et al.* [61]. Based on these findings, it concludes that the N protein of MERS-CoV might be a putative and valuable immunogen for vaccine development. Further, the screened epitopes from the N protein in this study also may be some valuable epitope-based vaccine candidates for further *in vitro* and *in vivo* tests for their antigenic and immunogenetic potentials.

The purpose of vaccination is to induce immunity against specific pathogens by selectively stimulating antigen-specific B-cells or CTLs, and helper T-cells. Theoretically, a vaccine should contain two classes of antigenic epitopes: a helper T-cell epitope and a B-cell epitope or a CTL epitope. Based on a combination of these epitopes, the vaccine is able to either induce specific humoral or cellular immune against specific pathogens [21]. Therefore, the B-cell, CTL, and helper T-cell epitopes were screened systemically in the N protein of MERS-CoV to obtain

putative epitope vaccine candidates. In this study, one B-cell epitope, <sup>108</sup>EAALPFRAVKD-GIVVWHED<sup>126</sup>, showed a lower conservancy of 78% among different MERS-CoV isolates compared with other 14 epitopes, suggesting that it might be not an idea epitope candidate based on the fact that an epitope should be as conservative as possible to provide broader protection among different virus strains. In addition, various continuous and discontinuous B-cell epitopes were mapped on the N protein (Table 2 and Fig 3). All discontinuous B-cell epitopes were located on the surface of the domain-2 of the N protein, showing the accessibility for the entered virus. Moreover, the diverse T-cell epitopes including helper T-cell epitopes and CTL epitopes were delineated by dissecting the N proteins (Tables 4 and 5). Notably, neutralizing antibodies are not far enough to completely clean infectious pathogens. CTLs are needed because they play a central role in the generation of a protective immune response against diverse pathogen infection [62]. In this study, although the CTL epitope <sup>93</sup>QLAPRWYFY<sup>101</sup> showed the highest immunogenic score of 0.32573, it possessed the minimal HLA binding alleles among all predicted CTL epitopes. On the contrary, the epitope <sup>92</sup>KQLAPRWYF<sup>100</sup> had the maximum HLA binding alleles, but less immunogenic score of 0.25847. This inconsistency of immunological characteristics among different epitopes suggested that various parameters needed to be considered when screened epitopes. Helper T-cell epitopes are critical to the generation of vigorous humoral and CTL responses. However, the response to T-cell epitopes is restricted by HLA proteins. Moreover, HLA is highly polymorphic in diverse ethnic populations. Therefore, to induce broad immune responses in diverse ethnic genetically diverse human populations, the HLA specificity of T-cell epitopes must be considered first as a major criteria for screening of the epitopes [63,64]. Based on the above analysis, to get more population coverage, the epitope candidates should bind more HLA alleles. In this study, the 10 ten helper T-cell epitopes and CTL epitopes that bind the maximum number of HLA alleles were selected as putative vaccine candidates. Further analysis revealed that all putative helper T-cell and CTL epitopes have an ideal population coverage and may provide broad immune protection for different geographic regions around the world. Additionally, as an ideal epitope, it should be highly conserved among different MERS-CoV isolates. Thus, conservancy level should be determined for all putative epitope candidates. The epitopes from this study were highly conserved in designated MERS-CoV isolates, suggesting that they could be ideal epitope vaccine candidates to elicit protective neutralizing antibodies and cellular immune responses against MERS-CoV.

## Conclusions

In conclusion, this study indicated that immunoinformatics-driven genome-wide screening of vaccine targets of emerging highly pathogenic pathogens is a promising strategy to accelerate their vaccine development. Based on this strategy, the B-cell epitopes, helper T-cell epitopes and CTL epitopes in the N protein of MERS-CoV were mapped and selected as putative MERS vaccine candidates. However, the B- and T-cell stimulation potentials of the screened epitopes are needed to be tested by *in vitro* and *in vivo* experiments along with this *in silico* study for their efficient use as vaccines against MERS-CoV. The present study provides new and valuable epitope candidates and prompts the future vaccine development of MERS and other emerging infectious diseases.

## Supporting Information

**S1 Fig. Kolaskar and Tongaonkar's antigenicity prediction of the N protein of MERS-CoV.** X-axis represents the amino acids, whereas Y-axis represents the antigenic propensity. Average antigenic score was 0.993. Area displayed by the threshold line is considered as potential B-cell

antigenic regions.  
(PDF)

**S2 Fig. Emini surface accessibility prediction of the N protein of MERS-CoV.** X-axis represents the amino acids, whereas Y-axis represents accessible propensity. Average antigenic score was 1.000. Area displayed by the threshold line is considered as potential B-cell accessible regions.  
(PDF)

**S3 Fig. Karplus and schulz flexibility prediction of the N protein of MERS-CoV.** X-axis represents the amino acids, whereas Y-axis represents flexible propensity. Average antigenic score was 1.028. Area displayed by the threshold line is considered as potential B-cell flexible regions.  
(PDF)

**S1 Table. Antigenic scores of the entire proteomes of all MERS-CoV isolates.**  
(XLS)

**S2 Table. 9-mer peptide core sequences in the N protein of MERS-CoV are predicted to be helper T-cell epitopes using the NetMHCIIpan 3.0 web server.**  
(DOC)

**S3 Table. 9-mer peptide sequences in the N protein of MERS-CoV are predicted to be CTL epitopes using the IEDB MHC-I binding prediction web server.**  
(DOC)

## Author Contributions

Conceived and designed the experiments: JDS JZ YZH. Performed the experiments: JDS JZ SJL YMT JS MNW JFL YHL. Analyzed the data: JDS JZ. Contributed reagents/materials/analysis tools: NZH HXW. Wrote the paper: JDS.

## References

1. Bermingham A, Chand MA, Brown CS, Aarons E, Tong C, Langrish C, et al. Severe respiratory illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle East, September 2012. *Euro Surveill.* 2012; 17(40):20290. PMID: [23078800](#)
2. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med.* 2012; 367(19):1814–20. doi: [10.1056/NEJMoa1211721](#) PMID: [23075143](#)
3. Coleman CM, Frieman MB. Coronaviruses: important emerging human pathogens. *J Virol.* 2014; 88(10):5209–12. doi: [10.1128/JVI.03488-13](#) PMID: [24600003](#)
4. Ithete NL, Stoffberg S, Corman VM, Cottontail VM, Richards LR, Schoeman MC, et al. Close relative of human Middle East respiratory syndrome coronavirus in bat, South Africa. *Emerg Infect Dis.* 2013; 19(10):1697–9. doi: [10.3201/eid1910.130946](#) PMID: [24050621](#)
5. Briese T, Mishra N, Jain K, Zalmout IS, Jabado OJ, Karesh WB, et al. Middle East respiratory syndrome coronavirus quasispecies that include homologues of human isolates revealed through whole-genome analysis and virus cultured from dromedary camels in Saudi Arabia. *MBio.* 2014; 5(3):e01146–14. doi: [10.1128/mBio.01146-14](#) PMID: [24781747](#)
6. Mackay IM, Arden KE. Middle East respiratory syndrome: An emerging coronavirus infection tracked by the crowd. *Virus Res.* 2015; 202:60–88. doi: [10.1016/j.virusres.2015.01.021](#) PMID: [25656066](#)
7. Zumla A, Hui DS, Perlman S. Middle East respiratory syndrome. *Lancet.* 2015; 386(9997): 995–1007. doi: [10.1016/S0140-6736\(15\)60454-8](#) PMID: [26049252](#)
8. World Health Organization. Middle East respiratory syndrome coronavirus (MERS-CoV)—Republic of Korea. 26 June 2015. Available: <http://www.who.int/csr/don/26-june-2015-mers-korea/en/>. Accessed 2015 Jul 16.

9. Song F, Fux R, Provacia LB, Volz A, Eickmann M, Becker S, et al. Middle East respiratory syndrome coronavirus spike protein delivered by modified vaccinia virus Ankara efficiently induces virus-neutralizing antibodies. *J Virol*. 2013; 87(21):11950–4. doi: [10.1128/JVI.01672-13](https://doi.org/10.1128/JVI.01672-13) PMID: [23986586](https://pubmed.ncbi.nlm.nih.gov/23986586/)
10. Wang L, Shi W, Joyce MG, Modjarrad K, Zhang Y, Leung K, et al. Evaluation of candidate vaccine approaches for MERS-CoV. *Nat Commun*. 2015; 6:7712. doi: [10.1038/ncomms8712](https://doi.org/10.1038/ncomms8712) PMID: [26218507](https://pubmed.ncbi.nlm.nih.gov/26218507/)
11. Guo X, Deng Y, Chen H, Lan J, Wang W, Zou X, et al. Systemic and mucosal immunity in mice elicited by a single immunization with human adenovirus type 5 or 41 vector-based vaccines carrying the spike protein of Middle East respiratory syndrome coronavirus. *Immunology*. 2015; 145(4):476–84. doi: [10.1111/imm.12462](https://doi.org/10.1111/imm.12462) PMID: [25762305](https://pubmed.ncbi.nlm.nih.gov/25762305/)
12. Du L, Zhao G, Kou Z, Ma C, Sun S, Poon VK, et al. Identification of a receptor-binding domain in the S protein of the novel human coronavirus Middle East respiratory syndrome coronavirus as an essential target for vaccine development. *J Virol*. 2013; 87(17):9939–42. doi: [10.1128/JVI.01048-13](https://doi.org/10.1128/JVI.01048-13) PMID: [23824801](https://pubmed.ncbi.nlm.nih.gov/23824801/)
13. Kim E, Okada K, Kenniston T, Raj VS, AlHajri MM, Farag EA, et al. Immunogenicity of an adenoviral-based Middle East Respiratory Syndrome coronavirus vaccine in BALB/c mice. *Vaccine*. 2014; 32(45):5975–82. doi: [10.1016/j.vaccine.2014.08.058](https://doi.org/10.1016/j.vaccine.2014.08.058) PMID: [25192975](https://pubmed.ncbi.nlm.nih.gov/25192975/)
14. Tang J, Zhang N, Tao X, Zhao G, Guo Y, Tseng CT, et al. Optimization of antigen dose for a receptor-binding domain-based subunit vaccine against MERS coronavirus. *Hum Vaccin Immunother*. 2015; 11(5):1244–50. doi: [10.1080/21645515.2015.1021527](https://doi.org/10.1080/21645515.2015.1021527) PMID: [25874632](https://pubmed.ncbi.nlm.nih.gov/25874632/)
15. Volz A, Kupke A, Song F, Jany S, Fux R, Shams-Eldin H, et al. Protective Efficacy of Recombinant Modified Vaccinia Virus Ankara Delivering Middle East Respiratory Syndrome Coronavirus Spike Glycoprotein. *J Virol*. 2015; 89(16):8651–6. doi: [10.1128/JVI.00614-15](https://doi.org/10.1128/JVI.00614-15) PMID: [26018172](https://pubmed.ncbi.nlm.nih.gov/26018172/)
16. Mou H, Raj VS, van Kuppeveld FJ, Rottier PJ, Haagmans BL, Bosch BJ. The receptor binding domain of the new Middle East respiratory syndrome coronavirus maps to a 231-residue region in the spike protein that efficiently elicits neutralizing antibodies. *J Virol*. 2013; 87(16):9379–83. doi: [10.1128/JVI.01277-13](https://doi.org/10.1128/JVI.01277-13) PMID: [23785207](https://pubmed.ncbi.nlm.nih.gov/23785207/)
17. Du L, Zhao G, Kou Z, Ma C, Sun S, Poon VK, et al. Identification of a receptor-binding domain in the S protein of the novel human coronavirus Middle East respiratory syndrome coronavirus as an essential target for vaccine development. *J Virol*. 2013; 87(17):9939–42. doi: [10.1128/JVI.01048-13](https://doi.org/10.1128/JVI.01048-13) PMID: [23824801](https://pubmed.ncbi.nlm.nih.gov/23824801/)
18. Du L, Kou Z, Ma C, Tao X, Wang L, Zhao G, et al. A truncated receptor-binding domain of MERS-CoV spike protein potently inhibits MERS-CoV infection and induces strong neutralizing antibody responses: implication for developing therapeutics and vaccines. *PLoS One*. 2013; 8(12):e81587. doi: [10.1371/journal.pone.0081587](https://doi.org/10.1371/journal.pone.0081587) PMID: [24324708](https://pubmed.ncbi.nlm.nih.gov/24324708/)
19. Zhang N, Jiang S, Du L. Current advancements and potential strategies in the development of MERS-CoV vaccines. *Expert Rev Vaccines*. 2014; 13(6):761–74. doi: [10.1586/14760584.2014.912134](https://doi.org/10.1586/14760584.2014.912134) PMID: [24766432](https://pubmed.ncbi.nlm.nih.gov/24766432/)
20. Almazán F, DeDiego ML, Sola I, Zuñiga S, Nieto-Torres JL, Marquez-Jurado S, et al. Engineering a replication-competent, propagation-defective Middle East respiratory syndrome coronavirus as a vaccine candidate. *MBio*. 2013; 4(5):e00650–13. doi: [10.1128/mBio.00650-13](https://doi.org/10.1128/mBio.00650-13) PMID: [24023385](https://pubmed.ncbi.nlm.nih.gov/24023385/)
21. Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. *Nat Rev Drug Discov*. 2007; 6(5):404–14. PMID: [17473845](https://pubmed.ncbi.nlm.nih.gov/17473845/)
22. Firbas C, Jilma B, Tauber E, Buerger V, Jelovcan S, Lingnau K, et al. Immunogenicity and safety of a novel therapeutic hepatitis C virus (HCV) peptide vaccine: a randomized, placebo controlled trial for dose optimization in 128 healthy subjects. *Vaccine*. 2006; 24(20):4343–53. PMID: [16581161](https://pubmed.ncbi.nlm.nih.gov/16581161/)
23. Staneková Z, Varečková E. Conserved epitopes of influenza A virus inducing protective immunity and their prospects for universal vaccine development. *Virology*. 2010; 7:351. doi: [10.1186/1743-422X-7-351](https://doi.org/10.1186/1743-422X-7-351) PMID: [21118546](https://pubmed.ncbi.nlm.nih.gov/21118546/)
24. He L, Cheng Y, Kong L, Azadnia P, Giang E, Kim J, et al. Approaching rational epitope vaccine design for hepatitis C virus with meta-server and multivalent scaffolding. *Sci Rep*. 2015; 5:12501. doi: [10.1038/srep12501](https://doi.org/10.1038/srep12501) PMID: [26238798](https://pubmed.ncbi.nlm.nih.gov/26238798/)
25. Sominskaya I, Skrastina D, Dislers A, Vasiljev D, Mihailova M, Ose V, et al. Construction and immunological evaluation of multivalent hepatitis B virus (HBV) core virus-like particles carrying HBV and HCV epitopes. *Clin Vaccine Immunol*. 2010; 17(6):1027–33. doi: [10.1128/CVI.00468-09](https://doi.org/10.1128/CVI.00468-09) PMID: [20410327](https://pubmed.ncbi.nlm.nih.gov/20410327/)
26. Terry FE, Moise L, Martin RF, Torres M, Pilotte N, Williams SA, et al. Time for T? Immunoinformatics addresses vaccine design for neglected tropical and emerging infectious diseases. *Expert Rev Vaccines*. 2015; 14(1):21–35. doi: [10.1586/14760584.2015.955478](https://doi.org/10.1586/14760584.2015.955478) PMID: [25193104](https://pubmed.ncbi.nlm.nih.gov/25193104/)
27. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*. 2007; 8:4. PMID: [17207271](https://pubmed.ncbi.nlm.nih.gov/17207271/)

28. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol.* 1999; 112:531–52. PMID: [10027275](#)
29. Combet C, Blanchet C, Geourjon C, Deléage G. NPS@: network protein sequence analysis. *Trends Biochem Sci.* 2000; 25(3):147–50. PMID: [10694887](#)
30. Källberg M, Margaryan G, Wang S, Ma J, Xu J. RaptorX server: a resource for template-based protein structure modeling. *Methods Mol Biol.* 2014; 1137:17–27. doi: [10.1007/978-1-4939-0366-5\\_2](#) PMID: [24573471](#)
31. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett.* 1990; 276(1–2):172–4. PMID: [1702393](#)
32. Vihinen M, Torkkila E, Riikonen P. Accuracy of protein flexibility predictions. *Proteins.* 1994; 19(2):141–9. PMID: [8090708](#)
33. Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol.* 1985; 55(3):836–9. PMID: [2991600](#)
34. Wilson JA, Hart MK. Protection from Ebola virus mediated by cytotoxic T lymphocytes specific for the viral nucleoprotein. *J Virol.* 2001; 75(6):2660–4. PMID: [11222689](#)
35. Karosiene E, Rasmussen M, Blicher T, Lund O, Buus S, Nielsen M. NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. *Immunogenetics.* 2013; 65(10):711–24. doi: [10.1007/s00251-013-0720-y](#) PMID: [23900783](#)
36. Zhang L, Udaka K, Mamitsuka H, Zhu S. Toward more accurate pan-specific MHC-peptide binding prediction: a review of current methods and tools. *Brief Bioinform.* 2012; 13(3):350–64. doi: [10.1093/bib/bbr060](#) PMID: [21949215](#)
37. Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, et al. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS One.* 2007; 2(8):e796. PMID: [17726526](#)
38. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics.* 2009; 61(1):1–13. doi: [10.1007/s00251-008-0341-z](#) PMID: [19002680](#)
39. Kim Y, Ponomarenko J, Zhu Z, Tamang D, Wang P, Greenbaum J, et al. Immune epitope database analysis resource. *Nucleic Acids Res.* 2012; 40(Web Server issue):W525–30. doi: [10.1093/nar/gks438](#) PMID: [22610854](#)
40. Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lamberth K, Buus S, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci.* 2003; 12(5):1007–17. PMID: [12717023](#)
41. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. *Nucleic Acids Res.* 2008; 36(Web Server issue):W509–12. doi: [10.1093/nar/gkn202](#) PMID: [18463140](#)
42. Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics.* 2005; 6:132. PMID: [15927070](#)
43. Sidney J, Assarsson E, Moore C, Ngo S, Pinilla C, Sette A, et al. Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. *Immunome Res.* 2008; 4:2. doi: [10.1186/1745-7580-4-2](#) PMID: [18221540](#)
44. Calis JJ, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, et al. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput Biol.* 2013; 9(10):e1003266. doi: [10.1371/journal.pcbi.1003266](#) PMID: [24204222](#)<http://www.ncbi.nlm.nih.gov/pubmed/24204222>
45. Bui HH, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics.* 2007; 8:361. PMID: [17897458](#)
46. Muh HC, Tong JC, Tammi MT. AllerHunter: a SVM-pairwise system for assessment of allergenicity and allergic cross-reactivity in proteins. *PLoS One.* 2009; 4(6):e5861. doi: [10.1371/journal.pone.0005861](#) PMID: [19516900](#)
47. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics.* 2006; 7:153. PMID: [16545123](#)
48. Le TM, Wong HH, Tay FP, Fang S, Keng CT, Tan YJ, et al. Expression, post-translational modification and biochemical characterization of proteins encoded by subgenomic mRNA8 of the severe acute respiratory syndrome coronavirus. *FEBS J.* 2007; 274(16):4211–22. PMID: [17645546](#)
49. Wang N, Shi X, Jiang L, Zhang S, Wang D, Tong P, et al. Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res.* 2013; 23(8):986–93. doi: [10.1038/cr.2013.92](#) PMID: [23835475](#)

50. Zheng N, Xia R, Yang C, Yin B, Li Y, Duan C, et al. Boosted expression of the SARS-CoV nucleocapsid protein in tobacco and its immunogenicity in mice. *Vaccine*. 2009; 27(36):5001–7. doi: [10.1016/j.vaccine.2009.05.073](https://doi.org/10.1016/j.vaccine.2009.05.073) PMID: [19523911](https://pubmed.ncbi.nlm.nih.gov/19523911/)
51. Ma C, Yao K, Zhou F, Zhu M. Comparative immunization in BALB/c mice with recombinant replication-defective adenovirus vector and DNA plasmid expressing a SARS-CoV nucleocapsid protein gene. *Cell Mol Immunol*. 2006; 3(6):459–65. PMID: [17257500](https://pubmed.ncbi.nlm.nih.gov/17257500/)
52. Hu H, Tao L, Wang Y, Chen L, Yang J, Wang H. Enhancing immune responses against SARS-CoV nucleocapsid DNA vaccine by co-inoculating interleukin-2 expressing vector in mice. *Biotechnol Lett*. 2009; 31(11):1685–93. doi: [10.1007/s10529-009-0061-y](https://doi.org/10.1007/s10529-009-0061-y) PMID: [19579009](https://pubmed.ncbi.nlm.nih.gov/19579009/)
53. Dutta NK, Mazumdar K, Lee BH, Baek MW, Kim DJ, Na YR, et al. Search for potential target site of nucleocapsid gene for the design of an epitope-based SARS DNA vaccine. *Immunol Lett*. 2008; 118(1):65–71. doi: [10.1016/j.imlet.2008.03.003](https://doi.org/10.1016/j.imlet.2008.03.003) PMID: [18440652](https://pubmed.ncbi.nlm.nih.gov/18440652/)
54. Liu SJ, Leng CH, Lien SP, Chi HY, Huang CY, Lin CL, et al. Immunological characterizations of the nucleocapsid protein based SARS vaccine candidates. *Vaccine*. 2006; 24(16):3100–8. PMID: [16494977](https://pubmed.ncbi.nlm.nih.gov/16494977/)
55. Zhao J, Huang Q, Wang W, Zhang Y, Lv P, Gao XM. Identification and characterization of dominant helper T-cell epitopes in the nucleocapsid protein of severe acute respiratory syndrome coronavirus. *J Virol*. 2007; 81(11):6079–88. PMID: [17392374](https://pubmed.ncbi.nlm.nih.gov/17392374/)
56. Zhou M, Xu D, Li X, Li H, Shan M, Tang J, et al. Screening and identification of severe acute respiratory syndrome-associated coronavirus-specific CTL epitopes. *J Immunol*. 2006; 177(4):2138–45. PMID: [16887973](https://pubmed.ncbi.nlm.nih.gov/16887973/)
57. Han Z, Zhao F, Shao Y, Liu X, Kong X, Song Y, et al. Fine level epitope mapping and conservation analysis of two novel linear B-cell epitopes of the avian infectious bronchitis coronavirus nucleocapsid protein. *Virus Res*. 2013; 171(1):54–64. doi: [10.1016/j.virusres.2012.10.028](https://doi.org/10.1016/j.virusres.2012.10.028) PMID: [23123213](https://pubmed.ncbi.nlm.nih.gov/23123213/)
58. Gupta V, Tabiin TM, Sun K, Chandrasekaran A, Anwar A, Yang K, et al. SARS coronavirus nucleocapsid immunodominant T-cell epitope cluster is common to both exogenous recombinant and endogenous DNA-encoded immunogens. *Virology*. 2006; 347(1):127–39. PMID: [16387339](https://pubmed.ncbi.nlm.nih.gov/16387339/)
59. Liu DQ, Ge JW, Qiao XY, Jiang YP, Liu SM, Li YJ. High-level mucosal and systemic immune responses induced by oral administration with Lactobacillus-expressed porcine epidemic diarrhea virus (PEDV) S1 region combined with Lactobacillus-expressed N protein. *Appl Microbiol Biotechnol*. 2012; 93(6):2437–46. doi: [10.1007/s00253-011-3734-0](https://doi.org/10.1007/s00253-011-3734-0) PMID: [22134641](https://pubmed.ncbi.nlm.nih.gov/22134641/)
60. Wilson JA, Hart MK. Protection from Ebola virus mediated by cytotoxic T lymphocytes specific for the viral nucleoprotein. *J Virol*. 2001; 75(6):2660–4. PMID: [11222689](https://pubmed.ncbi.nlm.nih.gov/11222689/)
61. Gao X, Zhou H, Wu C, Xiao Y, Ren L, Paranhos-Baccalà G, et al. Antibody against nucleocapsid protein predicts susceptibility to human coronavirus infection. *J Infect*. 2015; 71(5):599–602. doi: [10.1016/j.jinf.2015.07.002](https://doi.org/10.1016/j.jinf.2015.07.002) PMID: [26165609](https://pubmed.ncbi.nlm.nih.gov/26165609/)
62. Esser MT, Marchese RD, Kierstead LS, Tussey LG, Wang F, Chirmule N, et al. Memory T cells and vaccines. *Vaccine*. 2003; 21(5–6):419–30. PMID: [12531640](https://pubmed.ncbi.nlm.nih.gov/12531640/)
63. Maenaka K, Jones EY. MHC superfamily structure and the immune system. *Curr Opin Struct Biol*. 1999; 9(6):745–53. PMID: [10607669](https://pubmed.ncbi.nlm.nih.gov/10607669/)
64. Stern LJ, Wiley DC. Antigenic peptide binding by class I and class II histocompatibility proteins. *Structure*. 1994; 2(4):245–51. PMID: [8087551](https://pubmed.ncbi.nlm.nih.gov/8087551/)