ORIGINAL RESEARCH

Design of an epitope-based peptide vaccine against spike protein of human coronavirus: an in silico approach

Arafat Rahman Oany[1]
Abdullah-Al Emran[1,2]
Tahmina Pervin Jyoti[3]

[1]Department of Biotechnology and Genetic Engineering, Life Science Faculty, Mawlana Bhashani Science and Technology University, Tangail, Bangladesh; [2]Translational Research Institute, University of Queensland, Brisbane, Australia; [3]Biotechnology and Genetic Engineering Discipline, Life Science School, Khulna University, Khulna, Bangladesh

Abstract: Human coronavirus (HCoV), a member of *Coronaviridae* family, is the causative agent of upper respiratory tract infections and "atypical pneumonia". Despite severe epidemic outbreaks on several occasions and lack of antiviral drug, not much progress has been made with regard to an epitope-based vaccine designed for HCoV. In this study, a computational approach was adopted to identify a multiepitope vaccine candidate against this virus that could be suitable to trigger a significant immune response. Sequences of the spike proteins were collected from a protein database and analyzed with an in silico tool, to identify the most immunogenic protein. Both T cell immunity and B cell immunity were checked for the peptides to ensure that they had the capacity to induce both humoral and cell-mediated immunity. The peptide sequence from 88–94 amino acids and the sequence KSSTGFVYF were found as the most potential B cell and T cell epitopes, respectively. Furthermore, conservancy analysis was also done using in silico tools and showed a conservancy of 64.29% for all epitopes. The peptide sequence could interact with as many as 16 human leukocyte antigens (HLAs) and showed high cumulative population coverage, ranging from 75.68% to 90.73%. The epitope was further tested for binding against the HLA molecules, using in silico docking techniques, to verify the binding cleft epitope interaction. The allergenicity of the epitopes was also evaluated. This computational study of design of an epitope-based peptide vaccine against HCoVs allows us to determine novel peptide antigen targets in spike proteins on intuitive grounds, albeit the preliminary results thereof require validation by in vitro and in vivo experiments.

Keywords: vaccinomics, HLA, atypical pneumonia, allergenicity, docking

Introduction

Human coronavirus (HCoV) belongs to the *Coronaviridae* family (alphacoronavirus 1) and comprises a large group of enveloped, positive-sense, single-stranded polyadenylated RNA virus.[1,2] It consists of the largest known viral RNA genomes, ranging from 27.6 to 31.6 kb. Usually, coronaviruses are classified into three groups (group I to III), based on their serological cross-reactivity.[3] Their classification is also supported by evolutionary analysis.[1] The group I viruses are animal pathogens, including porcine epidemic diarrhea virus and feline infectious peritonitis virus. The group II viruses are responsible for domestic animal pathogenic infections, and the final group III viruses are responsible for avian species infection.[4] However both the group I and group II viruses are considered HCoV. The protein molecules that usually contribute the structure of all coronaviruses are the spike (S), envelope (E), membrane (M) and nucleocapsid (N). HCoV is usually the causative agent of upper respiratory tract infections and also the causative agent of "atypical pneumonia", which was first identified in the People's Republic of China.[5] As nowadays,

Correspondence: Arafat Rahman Oany
Department of Biotechnology and Genetic Engineering, Life Science Faculty, Mawlana Bhashani Science and Technology University, Tangail-1902, Bangladesh
Tel +880 155 881 9130
Email arafatr@outlook.com

an environmental resistance is shown by these viruses,[6] it is urgent to develop an effective prevention for HCoV. Currently, there is no available treatment or vaccine to cure HCoV infections. Due to the ever rising spread of this viral infection, the development of vaccines or antiviral drugs against HCoVs infections is crucial.

A novel approach integrating immunogenetics and immunogenomics with bioinformatics for the development of vaccines is known as vaccinomics.[7] This approach has been used to address the development of new vaccines. The present conventional approach for vaccine development relies on antigen expression, in sufficient amount, from in vitro culture models; however, many antigens, while expressed sufficiently, may not be good candidates for vaccine. With these conventional approaches, it has not been possible to control different types of outbreaks of viral pathogens, such as recent avian and swine influenza strains, due to their time-consuming development process. Hence, the rapid in silico informatics-based approach has gained much popularity with the recent advancement in the sequencing of many pathogen genomes and protein sequence databases.[8] The "vaccinomics" approach has already proven to be essential for combating diseases such as multiple sclerosis,[9] malaria,[10] and tumors.[11] However, these methods of vaccine development usually work through the identification of human leukocyte antigens (HLA) ligands and T cell epitopes,[12] which specify the selection of the potent vaccine candidates associated with the transporter of antigen presentation (TAP) molecules.[13–16] Allergenicity assessment is one of the vital steps in the development of a peptide vaccine because when we provide the vaccine into the human body, it is detected as a foreign substance. As a result, inflammation occurs, demonstrating an allergic reaction. For the prediction of a B-cell epitope, hydrophilicity is an important criterion which is usually in the beta turns region. These assessments strengthen the possibility of the vaccine candidates. Therefore, our present study was undertaken to design an epitope-based peptide vaccine against HCoVs (229E, NL63, HKU1, EMC, and OC43) using the vaccinomics approach, with the wet lab researcher expected to validate our prediction.

Materials and methods

The flow chart summarizing the protocols for the complete epitope prediction is depicted in Figure 1.

Viral strain selection

ViralZone, a database of the ExPASy Bioinformatics Resource Portal was used for the selection of HCoVs and their associated information, including their genus, family, host, transmission, disease, genome, and proteome.

Protein sequence retrieval

The outer membrane protein (spike protein) sequences of HCoV were retrieved from the UniProtKB database.[17] Then all the sequences were stored as a FASTA format for further analysis.

Evolution analysis

For the analysis of the evolutionary divergence in the membrane proteins of HCoV, a phylogenetic tree was constructed, using the ClustalW2 multiple sequence alignment tool.[18]

Antigenic protein identification

VaxiJen v2.0,[19] a server for the prediction of protective antigens and subunit vaccines, was used for the determination of the most potent antigenic protein. Here, we used the default parameter of this server for the determination of the antigenic protein.

T Cell epitope identification

The NetCTL 1.2 server was used for the identification of the T cell epitope.[20] The prediction method integrated peptide major histocompatibility complex class I (MHC-I) binding; proteasomal C terminal cleavage, and TAP transport efficiency. The epitope prediction was restricted to 12 MHC-I supertypes. MHC-I binding and proteasomal cleavage were performed through artificial neural networks, and the weight matrix was used for TAP transport efficiency. The parameter we used for this analysis was set at threshold 0.5 to maintain sensitivity and specificity of 0.89 and 0.94, respectively. This allowed us to recruit more epitopes for further analysis. A combined algorithm of MHC-I binding, TAP transport efficiency, and proteasomal cleavage efficiency were selected to predict overall scores.

A tool from the Immune Epitope Database[21] was used to predict the MHC-I binding. The stabilized matrix base method (SMM)[22] was used to calculate the half-maximal inhibitory concentration (IC_{50}) values of peptide binding to MHC-I molecules from different prediction methods. For the binding analysis, all the alleles were selected, and the length was set at 9.0 before prediction was done. For the selected epitopes, a web-based tool was used to predict proteasomal cleavage, TAP transport, and MHC-I.[23] This tool combines predictors of proteasomal processing, TAP
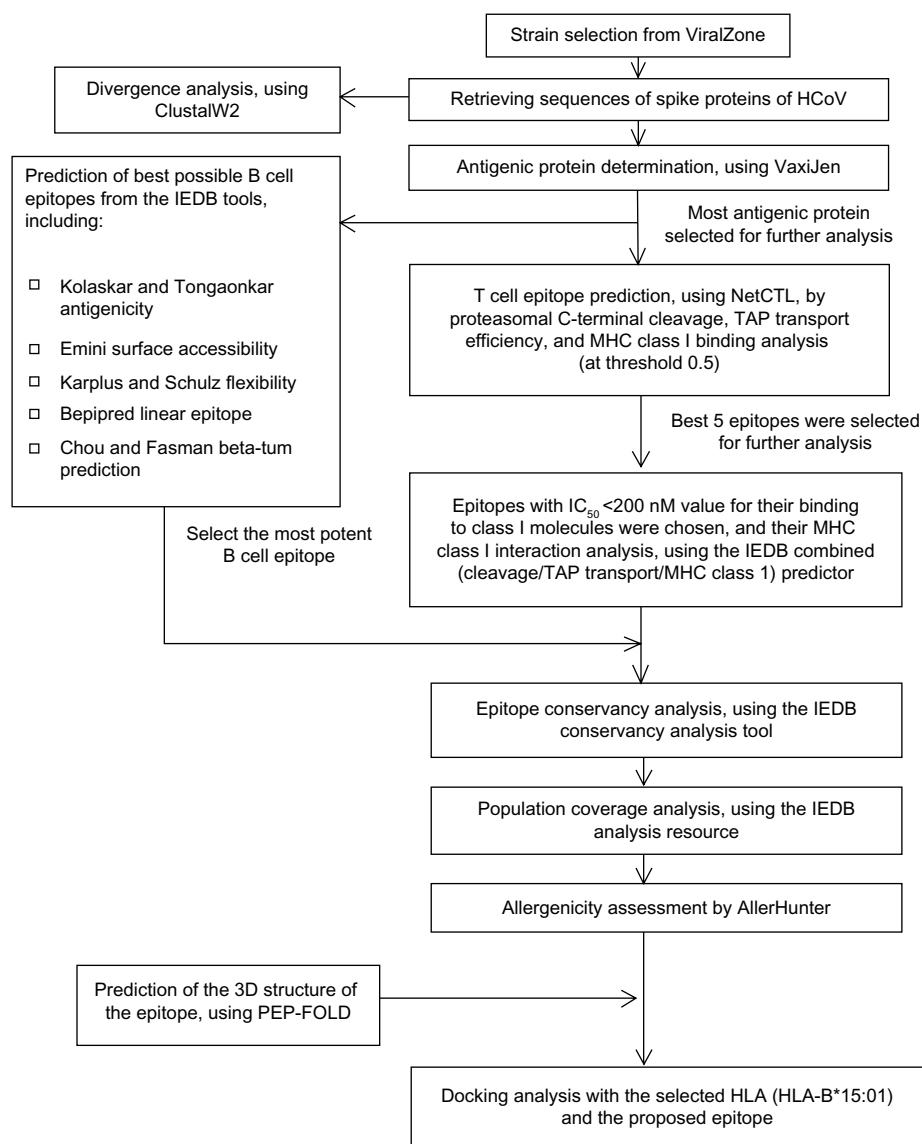
Figure 1 Flow chart summarizing the protocols for the complete epitope prediction.
Abbreviations: 3D, three dimensional; IC$_{50}$, half-maximal inhibitory concentration; HCoV, human coronavirus; HLA, human leukocyte antigen; ; HLA-B, the-major histocompatibility complex, class I, B; IEDB, Immune Epitope Database; MHC, major histocompatibility complex; TAP, transporter of antigen presentation.

transport, and MHC-I binding to produce an overall score for each peptide's intrinsic potential as a T cell epitope. SMM was also implemented for this prediction.

Epitope conservancy analysis

For the analysis of the epitope conservancy, the web-based tool from IEDB[24] analysis resource was used.

Prediction of population coverage

Population coverage for each individual epitope was selected by the IEDB population coverage calculation tool analysis resource. Here we used the allelic frequency of the interacting HLA alleles for the prediction of the population coverage for the corresponding epitope.

Allergenicity assessment

The web-based AllerHunter server[25] was used to predict the allergenicity of our proposed epitope for vaccine development. This server predicts allergenicity through a combinational prediction, by using both integration of the Food and Agriculture Organization (FAO)/World Health Organization (WHO) allergenicity evaluation scheme and support vector machines (SVM)-pairwise sequence similarity. AllerHunter predicts allergens as well as nonallergens with high

specificity. This makes AllerHunter is a very useful program for allergen cross-reactivity prediction.[26,27]

Molecular docking analysis and HLA allele interaction

Design of the three-dimensional (3D) epitope structure

For the docking analysis, the KSSTGFVYF epitope was subjected to PEP-FOLD web-based server[28] for 3D structure conversion, in order to analyze the interactions with different HLAs. This server modeled five 3D structures of the proposed epitope, and the best one was selected for the docking analysis.

Docking analysis

To ensure the binding between HLA molecules and our predicted epitope, a docking study was performed using Molegro Virtual Docker, version 6.0 (CLC bio, Aarhus, Denmark).[29] The HLA-B*15:01 was selected for docking on the basis of the available Protein Data Bank (PDB) structure deposited in the database, which interacted with our proposed epitope. The Protein Data Bank structure 1XR8, of Epstein–Barr virus EBNA-3 complexed with human UbcH6 peptide, was retrieved from the Research Collaboratory for Structural Bioinformatics (RCSB) protein database[30] and simplified to HLA-B*15:01. Finally the docking was established at a grid of X: 24.81, Y: 29.16, and Z: 40.59.

Identification of the B cell epitope

Prediction of potentially immunogenic epitopes in a given protein sequence may significantly reduce wet lab effort needed to discover the epitopes required for the design of vaccines and for immunodiagnostics. The aim of the prediction of the B cell epitope was to find the potential antigen that would interact with B lymphocytes and initiate an immunoresponse.[31] Tools from IEDB were used to identify the B cell antigenicity, including the Kolaskar and Tongaonkar antigenicity scale,[32] Emini surface accessibility prediction,[33] Karplus and Schulz flexibility prediction,[34] and Bepipred linear epitope prediction analysis.[35] The Chou and Fasman beta turn prediction tool[36] was used because the antigenic parts of a protein belong to the beta turn regions.[37]

Results

Divergence analysis of the retrieved sequences

A total of 56 outer membrane protein (spike protein) sequences from the different variants belonging to five types

(229E, NL63, HKU1, EMC, and OC43) of HCoVs were retrieved from the UniProtKB database. Then, the sequences were subjected to multiple sequence alignments in order to construct a phylogenetic tree (Figure S1). The phylogram showed evolutionary divergence among the different strains of HCoV.

Antigenic protein prediction

The VaxiJen server assessed all of the retrieved protein sequences in order to find the most potent antigenic protein. UniprotKB id: B2KKT9 was selected as the most potent antigenic protein, with a highest total prediction score of 0.6016. Then, the protein was used for further analysis.

T cell epitope identification

In a preselected environment, the NetCTL server predicted the potent T cell epitopes from the selected protein sequence. Based on the high combinatorial score, the five best epitopes (Table 1) were selected for further analysis.

MHC-I binding prediction, which was run through SMM, predicted a wide range of MHC-I allele interactions with the five T cell epitopes. The MHC-I alleles for which the epitopes showed higher affinity (IC_{50} <200 nM) were selected for further analysis (Table 2).

MHC-I processing (proteasomal cleavage/TAP transport/MHC-I combined predictor) predicted an overall score for each peptide's intrinsic potential to be a T cell epitope from the protein sequence. Proteasome complex, which cleaved the peptide bonds, thus converted the proteins into peptides. The peptide molecules from proteasome cleavage associated with class-I MHC molecules, and the peptide-MHC molecule then were transported to the cell membrane where they were presented to T helper cells. Here, higher overall score for each peptide denotes higher processing capabilities (Table 2).

Among the five T cell epitopes, a 9 mer epitope, KSSTGFVYF, was found to interact with most MHC-I alleles, including HLA-B*27:20; HLA-B*15:17; HLA-B*15:03; HLA-B*40:13; HLA-A*32:07; HLA-B*58:01;

Table 1 The selected epitopes, on the basis of their overall score predicted by the NetCTL server

Number	Epitopes	Overall score (nM)
1	GSDVNCNGY	2.9177
2	TLQYDVLFY	2.1285
3	YYCFINSTI	1.797
4	KSSTGFVYF	1.7667
5	KTLQYDVLF	1.5846

Table 2 The five potential T cell epitopes, along with their interacting MHC-I alleles and total processing score, and epitope conservancy result

Epitope	Interacting MHC-I allele with an affinity of <200 nM and the total score (proteasome score, TAP score, MHC-I score, processing score)	Epitope conservancy
GSDVNCNGY	HLA-C*12:03 (1.77), HLA-A*32:07 (1.40), HLA-B*27:20 (1.39) HLA-A*68:23 (1.30), HLA-C*05:01 (1.02), HLA-B*40:13 (0.50) HLA-A*01:01 (0.46), HLA-C*07:01 (0.36), HLA-A*32:15 (0.21)	64.29%
KTLQYDVLF	HLA-B*15:17 (1.77), HLA-A*32:07 (1.39), HLA-A*32:01 (1.37) HLA-A*68:23 (1.19), HLA-B*58:01 (0.84), HLA-B*40:13 (0.69) HLA-B*15:03 (0.51), HLA-B*57:01 (0.27), HLA-C*12:03 (0.21) HLA-A*32:15 (0.17), HLA-A*26:02 (0.16)	64.29%
KSSTGFVYF	HLA-B*27:20 (1.96), HLA-B*15:17 (1.65), HLA-B*15:03 (1.47) HLA-B*40:13 (1.34), HLA-A*32:07 (1.29), HLA-B*58:01 (0.95) HLA-C*03:03 (0.71), HLA-A*68:23 (0.61), HLA-C*12:03 (0.56) HLA-A*02:50 (0.54), HLA-A*32:01 (0.52), HLA-A*32:15 (0.50) HLA-C*05:01 (0.44), HLA-C*15:02 (0.41), HLA-B*58:02 (0.39) HLA-B*15:01 (0.36)	64.29%
TLQYDVLFY	HLA-A*80:01 (2.00), HLA-B*27:20 (1.84), HLA-A*32:07 (1.44) HLA-C*12:03 (1.24), HLA-A*29:02 (1.07), HLA-A*68:23 (1.01) HLA-A*32:15 (0.83), HLA-C*03:03 (0.83), HLA-C*14:02 (0.72) HLA-B*40:13 (0.39), HLA-B*15:01 (0.39)	64.29%
YYCFINSTI	HLA-C*14:02 (0.73), HLA-A*24:03 (0.33), HLA-A*32:07 (0.33) HLA-A*02:50 (0.21), HLA-B*27:20 (0.12), HLA-C*12:03(−0.08) HLA-C*03:03(−0.14), HLA-A*23:01(−0.33), HLA-A*68:23(−0.33) HLA-A*32:15 (−0.34), HLA-A*24:02 (−0.70)	64.29%

Abbreviations: HLA, human leukocyte antigen; MHC-I, major histocompatibility complex class I; TAP, transporter of antigen presentation.

HLA-C*03:03; HLA-A*68:23; HLA-C*12:03; HLA-A*02:50; HLA-A*32:01; HLA-A*32:15; HLA-C*05:01; HLA-C*15:02; HLA-B*58:02; HLA-B*15:01 with higher affinity (Table 2).

Epitope conservancy and population coverage analysis

The IEDB conservancy analysis tool analyzed the conservancy of the predicted epitopes, which are shown in Table 2. The population coverage of the predicted epitopes is depicted in Figure 2.

Allergenicity assessment

The sequence-based allergenicity prediction was precisely calculated using the AllerHunter tool, and the predicted queried epitope allergenicity score was 0.02 (sensitivity =93.0%, specificity =79.4%).

Molecular docking analysis

The predicted epitope bound in the groove of the HLA-B*15:01 with an energy of −17.662 kcal/mol. The docking interaction was visualized with the PyMOL molecular graphics system, version 1.5.0.4 (Schrödinger, LLC, Portland, OR, USA), shown in Figure 3.

B cell epitope identification

Here, we predicted amino acid scale-based methods for the identification of potential B-cell epitopes. According to this procedure we used different analysis methods for the prediction of a continuous B cell epitope.

The Kolaskar and Tongaonkar[32] antigenicity prediction method analyzed antigenicity on the basis of the physiochemical properties of amino acids and abundances in experimentally known epitopes. The average antigenic propensity of the protein was 1.058, with maximum of 1.240 and minimum of 0.920. The antigenic determination threshold for the protein was 1.00; all values greater than 1.00 were potential antigenic determinants. We found that seven epitopes satisfied the threshold value set prior to the analysis, and they had the potential to express the B cell response. The results are summarized in Table 3 and Figure 4.

To be a potent B cell epitope, it must have surface accessibility. Hence Emini surface accessibility prediction was obtained. The region 88 to 94 amino acid residues were more accessible. This is described in Figure 5 and Table 4.

The beta turns are often accessible and considerably hydrophilic in nature. These are two properties of antigenic
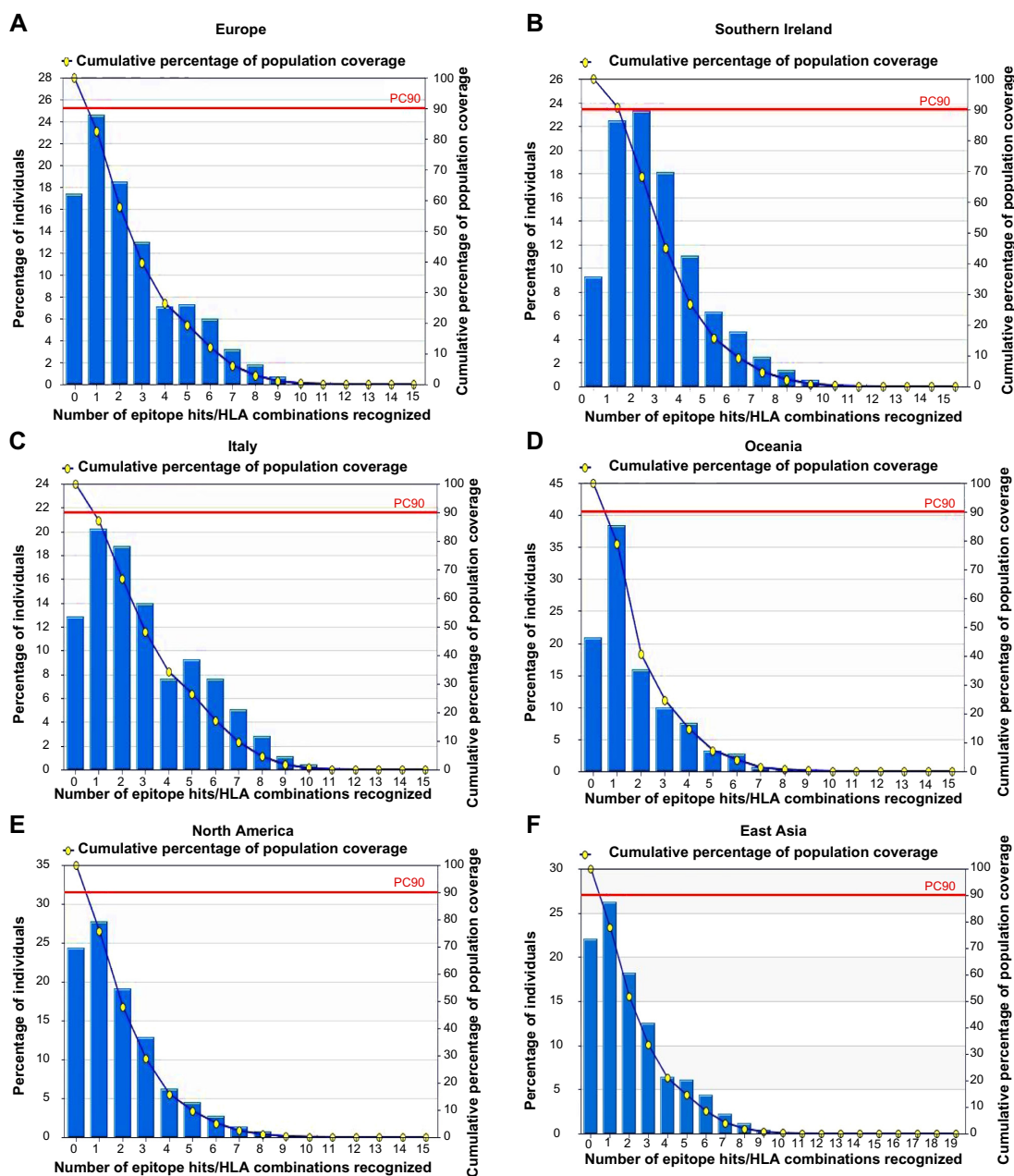
Figure 2 Population coverage, based on MHC-I restriction data. Different HCoV-affected regions were selected for evaluation of the population coverage of the proposed epitopes.
Notes: In the graphs, the line (-o-) represents the cumulative percentage of population coverage of the epitopes; the bars represent the population coverage for each epitope.
Abbreviations: HCoV, human coronavirus; HLA, human leukocyte antigen; MHC-I, major histocompatibility complex class I; PC90, 90% population coverage.

regions of a protein.[38] For this reason, Chou and Fasman beta-turn prediction was done. The region 73–95 (approximately 73–79 and 88–95) was considered as a β-turns region (Figure 6).

From the experimental evidence, it has been found that the flexibility of the peptide is correlated to antigenicity.[39] Hence, the Karplus and Schulz[34] flexibility prediction method was implemented. In this prediction method, the region of

75–95 was found to be the most flexible (Figure 7). Finally, we launched the Bepipred linear epitope prediction tool. This program is based on a Hidden Markov model, the best single method for predicting linear B-cell epitopes. The result of analysis with this method is summarized in Figure 8 and Table 5. By cross-referencing all the data, we predicted that the peptide sequences from 88–94 amino acids are capable of inducing the desired immune response as B cell epitopes.

Table 3 Kolaskar and Tongaonkar antigenicity analysis

Number	Start position	End position	Peptide	Peptide length
1	4	12	CLCPVPGLK	9
2	14	21	STGFVYFN	8
3	26	32	DVNCNGY	7
4	34	40	HNSVADV	7
5	54	84	NLKSGVIVFKTLQYDVLFYCSNSSSGVLDTT	31
6	86	99	PFGPSSQPYYCFIN	14
7	104	126	TTHVSTFVGILPPTVREIVVART	23

Discussion

The development of a new vaccine in a timely fashion is very crucial for defending the ever rising global burden of disease.[40–44] With the advancement of sequence-based

Figure 3 HLA-B*15:01 and epitope KSSTGFVYF interaction analysis. (**A**) The three dimensional structure of the epitope KSSTGFVYF. (**B**) The epitope KSSTGFVYF binds in the groove of the HLA-B*15:01.
Abbreviation: HLA-B, the-major histocompatibility complex, class I, B.

technology, now we have enough information about the genomics and proteomics of different viruses. As a result, with the help of various bioinformatics tools, we can design peptide vaccines based on a neutralizing epitope. For example, the design of an epitope-based vaccine against rhinovirus,[45] dengue virus,[46] chikungunya virus,[47] Saint Louis encephalitis virus,[48] etc has already been suggested. Though epitope-based vaccine design has become a familiar concept, in the case of HCoV there has not yet been much work done. The HCoV is an RNA virus, which tends to mutate more frequently than the DNA viruses.[49] These types of mutation mostly occur at the outer membrane protein, ie, at the spike protein.[50] These types of mutation increase the sustainability of the HCoVs, by ensuring their escape from both the cell-mediated and humoral immune responses.[51] Despite this, spike proteins have the most potential as a target for vaccine design because of their ability to induce a faster and longer-term mucosal immune response than that of the other proteins[52] and for this reason, has gained much popularity with researchers.[53,54] From this aspect, a universal HCoV vaccine needs to be designed, in order to overcome the adverse effects of this viral infection.

At present, vaccines are mostly based on B cell immunity. But recently, vaccine based on T cell epitope has been encouraged as the host can generate a strong immune response by CD8+ T cell against the infected cell.[55] With time, due to antigenic drift, any foreign particle can escape the antibody memory response; however, the T cell immune response often provides long-lasting immunity. Here, we predicted both B cell and T cell epitopes for conferring immunity in different ways, but other recent studies about HCoV represented the T cell epitope only, and we want to express our greater findings here.[56] There are several criteria that need to be fulfilled by a vaccine candidate epitope, and our predicted epitope fulfilled all the criteria. The initial criterion is the conservancy of the epitopes, which was measured by the IEDB conservancy analysis tool. Among
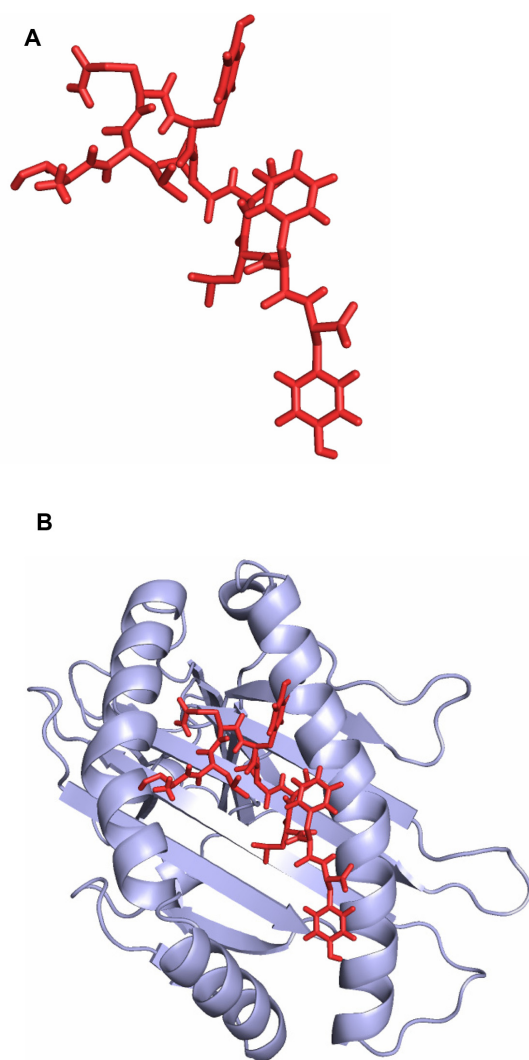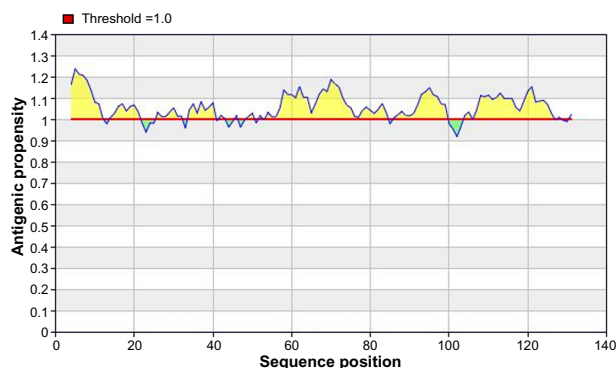
Figure 4 Kolashkar and Tongaonkar antigenicity prediction of the most antigenic protein, B2KKT9.
Notes: The x-axis and y-axis represent the sequence position and antigenic propensity, respectively. The threshold value is 1.0. The regions above the threshold are antigenic, shown in yellow.
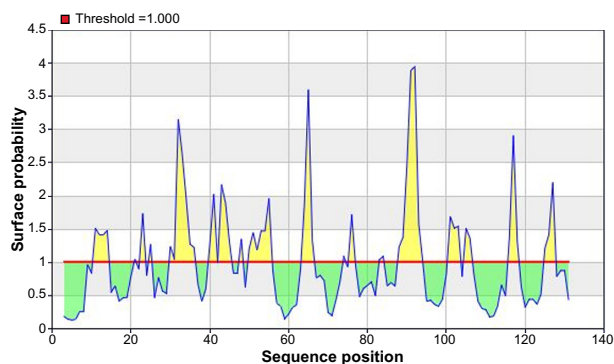

Figure 5 Emini surface accessibility prediction of the most antigenic protein, B2KKT9.
Notes: The x-axis and y-axis represent the sequence position and surface probability, respectively. The threshold value is 1.000. The regions above the threshold are antigenic, shown in yellow.

the five potential T cell epitopes, all possessed the same conservancy, of 64.29%. We also found similar conservancy of the B cell epitope, which was 64.29%. Having the same conservancy for all the epitopes, the KSSTGFVYF epitope possessed the highest amount of interactions with the HLA alleles. A very recent study showed a highly conserved sequence in RNA directed RNA polymerase of HCoVs;[56] nevertheless, our discovery of a spike protein with 64.29% conservancy among the 56 spike proteins has drawn much attention, and we consider this too as a epitope candidate for vaccine development.

Population coverage is another important factor in the development of a vaccine. For the all predicted epitopes, the cumulative percentage of population coverage was measured. We found the highest population coverage in South Ireland, which was 90.73%, followed by Italy and North America, with 87.13% and 75.68% coverage, respectively. The HCoV was first found in Europe,[57,58] hence, we also observed the overall coverage in Europe and found this to be 82.59%. Oceania's region covered 79.08%. We also recorded 80.31% population coverage for the East Asian region, considered as one of the hot spots of this viral infection. It has been reported that in Hong Kong and the People's Republic of China, during 2001–2002 there were about 587 cases (among these, 26 children) of acute respiratory disease caused by different types of HCoV infection. Specifically, in Hong

Kong, each year this virus caused about 224 hospitalizations per 100,000 population aged ≤6 years.[59]

However, allergenicity is one of the prominent obstacles in vaccine development. Today, most vaccines stimulate the immune system into an "allergic" reaction,[60] through induction of type 2 T helper T (Th2) cells and immunoglobulin E (IgE). The AllerHunter score value is the probability that a particular sequence is a cross-reactive allergen. However, the threshold for prediction of allergen cross-reactivity is adjusted such that a sequence is predicted as a cross-reactive allergen if its probability is >0.06. Here, our proposed epitope's allergenicity score was 0.02, and thus it was considered as a nonallergen. According to the FAO/WHO evaluation scheme of allergenicity prediction, a sequence is potentially allergenic if it either has an identity of at least six contiguous amino acids or >35 percent sequence identity over a window of 80 amino acids when compared to known allergens. Hence, our query epitopes did

Table 4 Emini surface accessibility prediction of the peptides

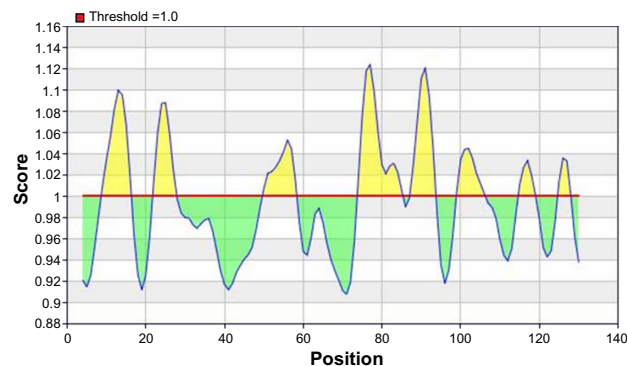| Number	Start position	End position	Peptide	Peptide length
I	30	36	NGYQHNS	7
2	50	55	NSVDNL	6
3	88	94	GPSSQPY	7

Figure 6 Karplus and Schulz flexibility prediction of the most antigenic protein, B2KKT9.
Notes: The x-axis and y-axis represent the position and score, respectively. The threshold is 1.0. The flexible regions of the protein are shown in yellow color, above the threshold value.
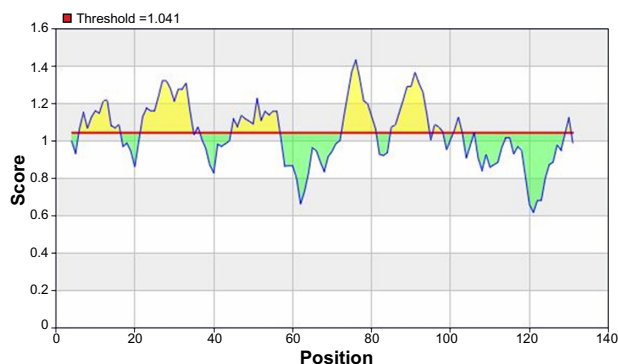
Figure 7 Chou and Fasman beta-turn prediction of the most antigenic protein, B2KKT9.
Notes: The x-axis and y-axis represent the position and score, respectively. The threshold is 1.041. The regions having beta turns in the protein are shown in yellow color, above the threshold value.
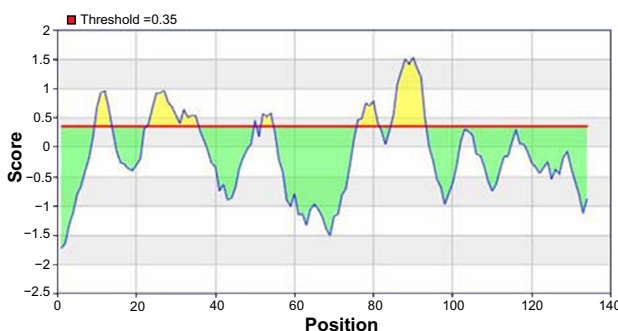

Figure 8 Bepipred linear epitope prediction of the most antigenic protein, B2KKT9.
Notes: The x-axis and y-axis represent the position and score, respectively. The threshold is 0.35. The regions having beta turns are shown in yellow. The highest peak region indicates the most potent B cell epitope.

Table 5 Bepipred linear epitope prediction

Number	Start position	End position	Peptide	Peptide length
1	10	13	GLKS	4
2	23	35	TGSDVNCNGYQHN	13
3	50	50	N	1
4	52	54	VDN	3
5	76	81	SSSGVL	6
6	85	93	IPFGPSSQP	9

Disclosure

The authors report no conflicts of interests in this work.

References

1. González JM, Gomez-Puertas P, Cavanagh D, Gorbalenya AE, Enjuanes L. A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae. *Arch Virol*. 2003;148(11): 2207–2235.
2. Siddell SG. *The Coronaviridae*. New York, NY: Plenum Press; 1995.
3. Drexler JF, Gloza-Rausch F, Glende J, et al. Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. *J Virol*. 2010;84(21): 11336–11349.
4. Cavanagh D, Mawditt K, Welchman Dde B, Britton P, Gough RE. Coronaviruses from pheasants (Phasianus colchicus) are genetically closely related to coronaviruses of domestic fowl (infectious bronchitis virus) and turkeys. *Avian Pathol*. 2002;31(1):81–93.
5. Parry J. WHO investigates China's fall in SARS cases. *BMJ*. 2003; 326(7402):1285.
6. Yoo D, Deregt D. A single amino acid change within antigenic domain II of the spike protein of bovine coronavirus confers resistance to virus neutralization. *Clin Diagn Lab Immunol*. 2001;8(2):297–302.
7. Poland GA, Ovsyannikova IG, Jacobson RM. Application of pharmacogenomics to vaccines. *Pharmacogenomics*. 2009;10(5):837–852.
8. Flower DR. *Bioinformatics for Vaccinology*. Chichester: John Wiley & Sons, Ltd; 2008.
9. Bourdette DN, Edmonds E, Smith C, et al. A highly immunogenic trivalent T cell receptor peptide vaccine for multiple sclerosis. *Mult Scler*. 2005;11(5):552–561.
10. López JA, Weilenman C, Audran R, et al. A synthetic malaria vaccine elicits a potent CD8(+) and CD4(+) T lymphocyte immune response in humans. Implications for vaccination strategies. *Eur J Immunol*. 2001;31(7):1989–1998.
11. Knutson KL, Schiffman K, Disis ML. Immunization with a HER-2/neu helper peptide vaccine generates HER-2/neu CD8 T-cell immunity in cancer patients. *J Clin Invest*. 2001;107(4):477–484.
12. Petrovsky N, Brusic V. Computational immunology: The coming of age. *Immunol Cell Biol*. 2002;80(3):248–254.
13. Brusic V, Bajic VB, Petrovsky N. Computational methods for prediction of T-cell epitopes – a framework for modelling, testing, and applications. *Methods*. 2004;34(4):436–443.
14. Peters B, Bulik S, Tampe R, Van Endert PM, Holzhütter HG. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *J Immunol*. 2003;171(4):1741–1749.
15. Bhasin M, Raghava GP. Analysis and prediction of affinity of TAP binding peptides using cascade SVM. *Protein Sci*. 2004;13(3): 596–607.
16. Nielsen M, Lundegaard C, Lund O, Keşmir C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics*. 2005;57(1–2):33–41.

not fulfill the criteria for the FAO/WHO evaluation scheme of allergenicity prediction and was classified by this scheme as a nonallergen.

However, our predicted in silico results were based on diligent analysis of sequence and various immune databases. This type of study has recently received experimental validation,[61] and for this reason, we have suggested that the proposed epitope would be able to trigger an efficacious immune response as a peptide vaccine in vivo.

Conclusion

Our study has shown that integrated computational approaches could be used for predicting vaccine candidates against pathogens such as HCoV, with previously described, validated procedures.

In this way, in silico studies save both time and costs for researchers and can guide the experimental work, with higher probabilities of finding the desired solutions and with fewer trial and error repeats of assays.

17. Apweiler R, Bairoch A, Wu CH, et al. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2004;32(Suppl 1):D115–D119.

18. Larkin MA, Blackshields G, Brown NP, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23(21):2947–2948.

19. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*. 2007;8:4.

20. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics*. 2007;8:424.

21. Buus S, Lauemøller SL, Worning P, et al. Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach. *Tissue Antigens*. 2003;62(5):378–384.

22. Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics*. 2005;6:132.

23. Tenzer S, Peters B, Bulik S, et al. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell Mol Life Sci*. 2005; 62:1025–1037.

24. 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(1):361.

25. 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.

26. Liao L, Noble WS. Combining pairwise sequence similarity and support vector machines for detecting remote protein evolutionary and structural relationships. *J Comput Biol*. 2003;10(6):857–868.

27. Chen YPP, Wong L, editors. *A Better Gap Penalty for Pairwise-SVM: Proceedings of the 3rd Asia-Pacific Bioinformatics*, Singapore, January 17–21, 2005. Singapore: Scientific Publishing Co; 2005.

28. Thévenet P, Shen Y, Maupetit J, Guyon F, Derreumaux P, Tufféry P. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Res*. 2012;40(Web Server issue):W288–W293.

29. Thomsen R, Christensen MH. MolDock: a new technique for high-accuracy molecular docking. *J Med Chem*. 2006;49(11):3315–3321.

30. Berman HM, Westbrook J, Feng Z, et al. The protein data bank. *Nucleic Acids Res*. 2000;28(1):235–242.

31. Nair DT, Singh K, Siddiqui Z, Nayak BP, Rao KV, Salunke DM. Epitope recognition by diverse antibodies suggests conformational convergence in an antibody response. *J Immunol*. 2002;168(5):2371–2382.

32. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett*. 1990;276(1–2):172–174.

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):836–839.

34. Karplus PA, Schulz GE. Prediction of chain flexibility in proteins. *Naturwissenschaften*. 1985;72(4):212–213.

35. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Res*. 2006;2(1):2.

36. Chou PY, Fasman GD. Empirical predictions of protein conformation. *Annu Rev Biochem*. 1978;47:251–276.

37. Rini JM, Schulze-Gahmen U, Wilson IA. Structural evidence for induced fit as a mechanism for antibody-antigen recognition. *Science*. 1992;255(5047):959–965.

38. Rose GD, Gierasch L, Smith JA. Turns in peptides and proteins. *Adv Protein Chem*. 1985;37:1–109.

39. Novotný J, Handschumacher M, Haber E, et al. Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains). *Proc Natl Acad Sci U S A*. 1986;83(2):226–230.

40. Marshall SJ. Developing countries face double burden of disease. *Bull World Health Organ*. 2004;82(7):556.

41. De Groot AS, Rappuoli R. Genome-derived vaccines. *Expert Rev Vaccines*. 2004;3(1):59–76.

42. Korber B, LaBute M, Yusim K. Immunoinformatics comes of age. *PLoS Comput Biol*. 2006;2(6):e71.

43. Fauci AS. Emerging and re-emerging infectious diseases: influenza as a prototype of the host-pathogen balancing act. *Cell*. 2006;124(4):665–670.

44. 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):404–414.

45. Lapelosa M, Gallicchio E, Arnold GF, Arnold E, Levy RM. In silico vaccine design based on molecular simulations of rhinovirus chimeras presenting HIV-1 gp41 epitopes. *J Mol Biol*. 2009;385(2):675–691.

46. Chakraborty S, Chakravorty R, Ahmed M, et al. A computational approach for identification of epitopes in dengue virus envelope protein: a step towards designing a universal dengue vaccine targeting endemic regions. *In Silico Biol*. 2010;10(5–6):235–246.

47. Islam R, Sakib MS, Zaman A. A computational assay to design an epitope-based peptide vaccine against chikungunya virus. *Future Virol*. 2012;7(10):1029–1042.

48. Hasan MA, Hossain M, Alam MJ. A computational assay to design an epitope-based Peptide vaccine against Saint Louis encephalitis virus. *Bioinform Biol Insights*. 2013;7:347–355.

49. Twiddy SS, Holmes EC, Rambaut A. Inferring the rate and time-scale of dengue virus evolution. *Mol Biol Evol*. 2003;20(1):122–129.

50. Manzin A, Solforosi L, Petrelli E, et al. Evolution of hypervariable region 1 of hepatitis C virus in primary infection. *J Virol*. 1998;72(7):6271–6276.

51. Christie JM, Chapel H, Chapman RW, Rosenberg WM. Immune selection and genetic sequence variation in core and envelope regions of hepatitis C virus. *Hepatology*. 1999;30(4):1037–1044.

52. Ma C, Li Y, Wang L, et al. Intranasal vaccination with recombinant receptor-binding domain of MERS-CoV spike protein induces much stronger local mucosal immune responses than subcutaneous immunization: Implication for designing novel mucosal MERS vaccines. *Vaccine*. 2014;32(18):2100–2108.

53. Yang ZY, Kong WP, Huang Y, et al. A DNA vaccine induces SARS coronavirus neutralization and protective immunity in mice. *Nature*. 2004;428(6982):561–564.

54. Agnihothram S, Gopal R, Yount BL, et al. Platform strategies for rapid response against emerging coronaviruses: MERS-CoV serologic and antigenic relationships in vaccine design. *J Infect Dis* [serial on the Internet]. Nov 2013 [cited May 16, 2014]. Available from: http://jid.oxfordjournals.org/content/early/2013/11/18/infdis.jit609. Accessed July 12, 2014.

55. Shrestha B, Diamond MS. Role of CD8+ T cells in control of West Nile virus infection. *J Virol*. 2004;78(15):8312–8321.

56. Sharmin R, Islam AB. A highly conserved WDYPKCDRA epitope in the RNA directed RNA polymerase of human coronaviruses can be used as epitope-based universal vaccine design. *BMC Bioinformatics*. 2014;15:161.

57. van der Hoek L, Pyrc K, Jebbink MF, et al. Identification of a new human coronavirus. *Nat Med*. 2004;10(4):368–373.

58. Hofmann H, Pyrc K, van der Hoek L, Geier M, Berkhout B, Pöhlmann S. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proc Natl Acad Sci U S A*. 2005;102(22):7988–7993.

59. Chiu SS, Chan KH, Chu KW, et al. Human coronavirus NL63 infection and other coronavirus infections in children hospitalized with acute respiratory disease in Hong Kong, China. *Clin Infect Dis*. 2005;40(12):1721–1729.

60. McKeever TM, Lewis SA, Smith C, Hubbard R. Vaccination and allergic disease: a birth cohort study. *Am J Public Health*. 2004;94(6):985–989.

61. Khan MK, Zaman S, Chakraborty S, et al. In silico predicted mycobacterial epitope elicits in vitro T-cell responses. *Mol Immunol*. 2014;61(1):16–22.
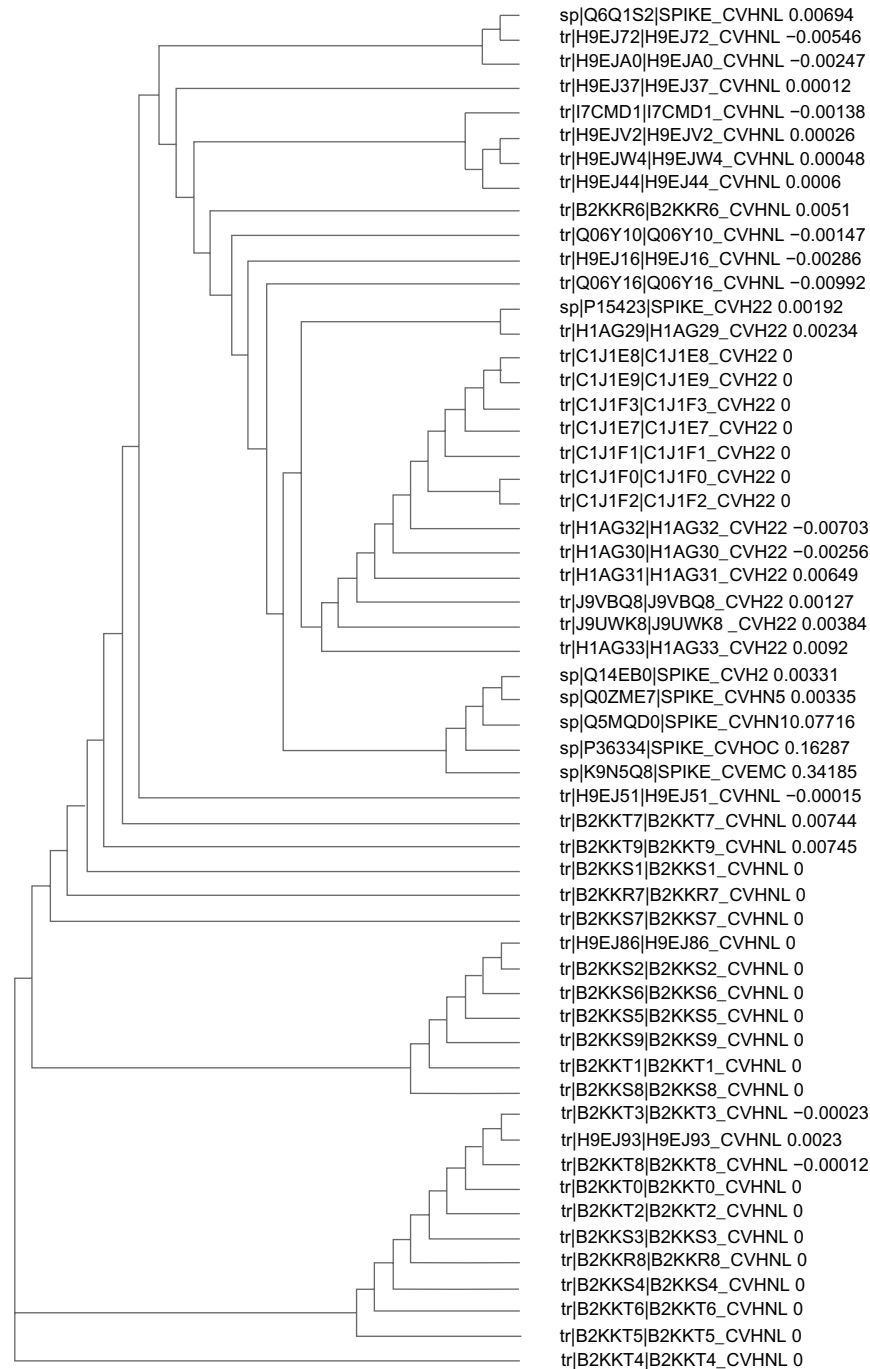
Supplementary materials


```
sp|Q6Q1S2|SPIKE_CVHNL 0.00694
tr|H9EJ72|H9EJ72_CVHNL −0.00546
tr|H9EJA0|H9EJA0_CVHNL −0.00247
tr|H9EJ37|H9EJ37_CVHNL 0.00012
tr|I7CMD1|I7CMD1_CVHNL −0.00138
tr|H9EJV2|H9EJV2_CVHNL 0.00026
tr|H9EJW4|H9EJW4_CVHNL 0.00048
tr|H9EJ44|H9EJ44_CVHNL 0.0006
tr|B2KKR6|B2KKR6_CVHNL 0.0051
tr|Q06Y10|Q06Y10_CVHNL −0.00147
tr|H9EJ16|H9EJ16_CVHNL −0.00286
tr|Q06Y16|Q06Y16_CVHNL −0.00992
sp|P15423|SPIKE_CVH22 0.00192
tr|H1AG29|H1AG29_CVH22 0.00234
tr|C1J1E8|C1J1E8_CVH22 0
tr|C1J1E9|C1J1E9_CVH22 0
tr|C1J1F3|C1J1F3_CVH22 0
tr|C1J1E7|C1J1E7_CVH22 0
tr|C1J1F1|C1J1F1_CVH22 0
tr|C1J1F0|C1J1F0_CVH22 0
tr|C1J1F2|C1J1F2_CVH22 0
tr|H1AG32|H1AG32_CVH22 −0.00703
tr|H1AG30|H1AG30_CVH22 −0.00256
tr|H1AG31|H1AG31_CVH22 0.00649
tr|J9VBQ8|J9VBQ8_CVH22 0.00127
tr|J9UWK8|J9UWK8 _CVH22 0.00384
tr|H1AG33|H1AG33_CVH22 0.0092
sp|Q14EB0|SPIKE_CVH2 0.00331
sp|Q0ZME7|SPIKE_CVHN5 0.00335
sp|Q5MQD0|SPIKE_CVHN10.07716
sp|P36334|SPIKE_CVHOC 0.16287
sp|K9N5Q8|SPIKE_CVEMC 0.34185
tr|H9EJ51|H9EJ51_CVHNL −0.00015
tr|B2KKT7|B2KKT7_CVHNL 0.00744
tr|B2KKT9|B2KKT9_CVHNL 0.00745
tr|B2KKS1|B2KKS1_CVHNL 0
tr|B2KKR7|B2KKR7_CVHNL 0
tr|B2KKS7|B2KKS7_CVHNL 0
tr|H9EJ86|H9EJ86_CVHNL 0
tr|B2KKS2|B2KKS2_CVHNL 0
tr|B2KKS6|B2KKS6_CVHNL 0
tr|B2KKS5|B2KKS5_CVHNL 0
tr|B2KKS9|B2KKS9_CVHNL 0
tr|B2KKT1|B2KKT1_CVHNL 0
tr|B2KKS8|B2KKS8_CVHNL 0
tr|B2KKT3|B2KKT3_CVHNL −0.00023
tr|H9EJ93|H9EJ93_CVHNL 0.0023
tr|B2KKT8|B2KKT8_CVHNL −0.00012
tr|B2KKT0|B2KKT0_CVHNL 0
tr|B2KKT2|B2KKT2_CVHNL 0
tr|B2KKS3|B2KKS3_CVHNL 0
tr|B2KKR8|B2KKR8_CVHNL 0
tr|B2KKS4|B2KKS4_CVHNL 0
tr|B2KKT6|B2KKT6_CVHNL 0
tr|B2KKT5|B2KKT5_CVHNL 0
tr|B2KKT4|B2KKT4_CVHNL 0
```

Figure S1 Phylogenetic tree, showing the evolutionary divergence among the different membrane proteins of human coronaviruses.