Chapter 4

A Computational Vaccine Designing Approach for MERS-CoV Infections

Hiba Siddig Ibrahim and Shamsoun Khamis Kafi

Abstract

The aim of this study was to use IEDB software to predict the suitable MERS-CoV epitope vaccine against the most known world population alleles through four selecting proteins such as S glycoprotein and envelope protein and their modification sequences after the pandemic spread of MERS-CoV in 2012. IEDB services is one of the computational methods; the output of this study showed that S glycoprotein, envelope (E) protein, and S and E protein modified sequences of MERS-CoV might be considered as a protective immunogenic with high conservancy because they can elect both neutralizing antibodies and T-cell responses when reacting with B-cell, T-helper cell, and cytotoxic T lymphocyte. NetCTL, NetChop, and MHC-NP were used to confirm our results. Population coverage analysis showed that the putative helper T-cell epitopes and CTL epitopes could cover most of the world population in more than 60 geographical regions. According to AllerHunter results, all those selected different protein showed non-allergen; this finding makes this computational vaccine study more desirable for vaccine synthesis.

Key words Middle East respiratory syndrome coronavirus, Severe acute respiratory syndrome coronavirus, Federal Drug Administration, Immuno epitope database, FAO, AllerHunter

1 Introduction

Vaccine development was considered as the most important subjects to protect from a highly infectious disease especially when treatment is not available; nowadays, a new way for vaccine design was done by a new aspects called immune-informatics that depends on software program to determine the most immunogenic parts of the organisms (epitopes) like these software that were used in this study to try to develop more powerful immunogenic MERS-CoV vaccine because the previous MERS-CoV vaccine can be either inactivated coronavirus, live attenuated coronavirus, S protein-based, DNA vaccines, and combination vaccines against coronaviruses; as we know coronaviruses were first described in the 1960s from the nasal cavities of patients with common cold. These strains of coronaviruses were called HC-229E and HC-OC43; in 2003,

Namrata Tomar (ed.), *Immunoinformatics*, Methods in Molecular Biology, vol. 2131,
https://doi.org/10.1007/978-1-0716-0389-5_4,

following the outbreak of severe acute respiratory syndrome (SARS) that resulted in over 8000 infections, about 10% of which resulted in death, but in 24 September 2012, a first report of isolated new novel coronavirus like SARS-CoV by Egyptian virologist Dr. Ali Mohamed Zaki in Jeddah, Saudi Arabia, from the lungs of a 60-year-old male patient with acute pneumonia and acute renal failure becomes a new discovery that was recently called MERS-CoV; this finding was posted on ProMED-mail [1–3]. MERS-CoV belong to group C β-coronaviruses that characterize 30 KB genome, ssRNA virus, positive sense with 10 predicting open reading frames (ORFs) like E, M, S, enveloped. MERS-CoV can grow in a culture media; the genome size, organization, and sequence analysis revealed that the NCoV is most closely related to bat coronaviruses BtCoV-HKU4 and BtCoV-HKU5; a partial spike gene sequencing of South African Neoromicia bats was considered as close relative to MERS-Cov as illustrated by nucleotide percentage distance substitution model and the complete deletion option in MEGA; this makes the possibility of a common coronavirus vaccine more desirable [3–5].

This study depended on using S and E with modified S and E protein sequences through in silico approach to develop MERS-CoV vaccine in addition to study the side effects of mutation in those selected sequences on vaccine development. Spike glycoprotein is characterized by a trimeric, envelope-anchored, type I fusion glycoprotein that interfaces with human dipeptidyl peptidase 4 (DPP4) receptor; to mediate viral entry, it is composed of 2 subunits; they are S1, which contains the receptor-binding domain and determines cell tropism, and S2, the location of the cell fusion machinery, while E protein was considered as part of virus cell membrane [4, 6].

This study showed that S, E and their modified sequences can be considered safe and most promising MERS-CoV vaccine without any kinds of allergic reactions.

2 Materials and Methods

2.1 Protein Sequence Retrieval

A total number of 130 spike (S) glycoproteins and 41 envelope (E) proteins of MERS-CoV were retrieved from NCBI (http://www.ncbi.nlm.nih.gov/protein/) database in September 2016, which was actually collected from different parts of the world, such as Saudi Arabia, China, Thailand, United Kingdom, Qatar, Tunisia, and South Africa. The accession numbers of retrieved strains were listed in Supplementary Tables 1 and 2. All methods below were applied for S, E, modified S & E proteins; modified S and E proteins were made by randomly changing some amino acids in their reference sequences; *see* Table 1 envelope protein (E) with Table 2 spike glycoprotein (S) gene bank accession numbers.

Table 1
Gene Bank Accession No of Envelope protein

Accession No of E protein	Date and place of collection	Type of specimen
YP_009047209.1	13-Jun-2012	
AKJ80142.1	27-May-2015/China	Nasopharyngeal swab
AIZ74456.1	07-May-2013/France	Sputum on Vero E6
AIZ74443.1	07-May-2013/France	Induced sputum
AIZ74434.1	07-May-2013/France	Induced sputum
AIZ74422.1	26-Apr-2013/France	Broncho-alveolar lavage
AIZ74406.1	26-Apr-2013/France	Broncho-alveolar lavage
AID50423.1	10-Feb-2013/United Kingdom	Throat swab
AID50423.1	10-Feb-2013/United Kingdom	Throat swab
ALD51909.1	17-Jun-2015/Thailand	Sputum
AMQ49075.1	24-Aug-2015/Saudi Arabia	Respiratory secretions
AMQ49064.1	27-Aug-2015/Saudi Arabia	Respiratory secretions
AMQ49053.1	24-Aug-2015/Saudi Arabia	Respiratory secretions
AMQ49020.1	12-Jul-2015/Saudi Arabia	Respiratory secretions
AMQ49042.1	24-Aug-2015/Saudi Arabia	Respiratory secretions
AMQ49031.1	24-Aug-2015/Saudi Arabia	Respiratory secretions
ALW82736.1	02-Feb-2015/Saudi Arabia	
ALW82714.1	05-Feb-2015/Saudi Arabia	Respiratory secretions
ALW82758.1	10-Feb-2015/Saudi Arabia	Respiratory secretions
ALW82747.1	13-Feb-2015/Saudi Arabia	Respiratory secretions
ALW82696.1	15-Feb-2015/Saudi Arabia	Respiratory secretions
ALW82685.1	07-Feb-2015/Saudi Arabia	Respiratory secretions
ALW82674.1	27-Mar-2015/Saudi Arabia	Respiratory secretions
AFY13312.1	11-Sep-2012/United Kingdom	
AIG13101.1	2011/South Africa	
AHY21474.1	Mammalian cell line Vero CCL81	
AHY22569.1	Nov-2013/Saudi Arabia	nasal swab (camel)
AHB33331.1	07-May-2013/France	Vero E6 isolate/sputum
AHC74092.1	13-Oct-2013/Qatar	
AHC74103.1	17-Oct-2013/Qatar	
AHI48522.1	02-May-2013/Saudi Arabia	

(continued)

Table 1
(continued)

Accession No of E protein	Date and place of collection	Type of specimen
AHI48566.1	05-Aug-2013/Saudi Arabia	
AHI48544.1	28-Aug-2013/Saudi Arabia	
AHI48533.1	17-Jul-2013/Saudi Arabia	
AHI48555.1	12-Jun-2013/Saudi Arabia	
AHI48588.1	02-Jul-2013/Saudi Arabia	
AHI48577.1	15-Aug-2013/Saudi Arabia	
AHI48599.1	12-Jun-2013/Saudi Arabia	
AHI48610.1	01-Mar-2013/Saudi Arabia	

2.2 In Silico PCR

(http://insilico.ehu.es/PCR_virus/) In silico PCR amplification is a program that made amplification against sequenced viruses, by mimicking PCR amplification and primers confirmatory tools too; here it was used for the above viruses by using store gene bank sequence; it contains 1783 sequences from 1421 completely sequenced viruses (last update: 31 May 2010).

2.3 Determination of Conserved Regions

The retrieved sequences, which were collected from NCBI, were used as a platform to obtain the conserved regions by using multiple sequence alignment (MSA). Sequences were aligned with the aid of ClustalW as implemented in the BioEdit program, version 7.0.9.0.

2.4 B-Cell Epitope Prediction

B-cell epitope is characterized by being hydrophilic, accessible, flexible, antigenic propensity and in a beta turn region. Thus, the classical propensity scale methods and hidden Markov model programmed software from IEDB analysis resource (http://www.iedb.org/) were used for the following aspects:

2.4.1 Prediction of Linear B-Cell Epitopes

BepiPred from immune epitope database and analysis resource (http://toolsiedb.ofg/bcell/) was used for linear B-cell epitope prediction from the conserved region with a default threshold value of 0.350. BepiPred combines the predictions of a hidden Markov model and the propensity scale of Parker et al. as it is described in Larsen et al. (Immunome Research, 2006).

2.4.2 Prediction of Surface Accessibility

By Emini surface accessibility prediction tool of the immune epitope database (IEDB), the surface-accessible epitopes were predicted from the conserved regions holding the default threshold value 1.000 or higher.

Table 2
Gene Bank Accession No of S glycoprotein

Accession No of S glycoprotein	Date and place of collection	Type of specimen
YP_009047204.1	13-Jun-2012	
AHX00721.1	30-Dec-2013/Saudi Arabia	Camel
AHX00711.1	30-Dec-2013/Saudi Arabia	Dromedary
AHX00731.1	30-Nov-2013/Saudi Arabia	Dromedary
AHZ90568.1	08-May-2013/Tunisia	Serum
AHX71946.1	16-Feb-2014/Qatar	Camelus dromedaries
ALJ54521.1	12-May-2015/Saudi Arabia	Respiratory secretions
ALJ54520.1	13-Jun-2015/Saudi Arabia	Respiratory secretions
ALJ54519.1	07-Jun-2015/Saudi Arabia	Respiratory secretions
ALJ54518.1	04-Jun-2015/Saudi Arabia	Respiratory secretions
ALJ54517.1	03-Jun-2015/Saudi Arabia	Respiratory secretions
ALJ54516.1	02-Jun-2015/Saudi Arabia	Respiratory secretions
ALJ54515.1	01-Jun-2015/Saudi Arabia	Respiratory secretions
ALJ54514.1	29-May-2015/Saudi Arabia	Respiratory secretions
ALJ54513.1	25-Apr-2015/Saudi Arabia	Respiratory secretions
ALJ54512.1	27-May-2015/Saudi Arabia	Respiratory secretions
ALJ54511.1	27-May-2015/Saudi Arabia	Respiratory secretions
ALJ54510.1	28-May-2015/Saudi Arabia	Respiratory secretions
ALJ54509.1	28-May-2015/Saudi Arabia	Respiratory secretions
ALJ54508.1	29-May-2015/Saudi Arabia	Respiratory secretions
ALJ54507.1	29-May-2015/Saudi Arabia	Respiratory secretions
ALJ54506.1	23-May-2015/Saudi Arabia	Respiratory secretions
ALJ54505.1	22-May-2015/Saudi Arabia	Respiratory secretions
ALJ54504.1	20-May-2015/Saudi Arabia	Rrespiratory secretions
ALJ54503.1	17-May-2015/Saudi Arabia	Respiratory secretions
ALJ54502.1	12-May-2015/Saudi Arabia	Respiratory secretions
ALJ54501.1	21-Mar-2015/Saudi Arabia	Respiratory secretions
ALJ54500.1	10-May-2015/Saudi Arabia	Respiratory secretions
ALJ54499.1	09-May-2015/Saudi Arabia	Respiratory secretions
ALJ54498.1	09-May-2015/Saudi Arabia	Respiratory secretions
ALJ54497.1	09-May-2015/Saudi Arabia	Respiratory secretions

(continued)

Table 2
(continued)

Accession No of S glycoprotein	Date and place of collection	Type of specimen
ALJ54496.1	16-Apr-2015/Saudi Arabia	Respiratory secretions
ALJ54495.1	13-Apr-2015/Saudi Arabia	Respiratory secretions
ALJ54494.1	04-Apr-2015/Saudi Arabia	Respiratory secretions
ALJ54493.1	04-Apr-2015/Saudi Arabia	Respiratory secretions
ALJ54492.1	30-Mar-2015/Saudi Arabia	Respiratory secretions
ALJ54491.1	25-Mar-2015/Saudi Arabia	Respiratory secretions
ALJ54490.1	24-Mar-2015/Saudi Arabia	Respiratory secretions
ALJ54489.1	08-Mar-2015/Saudi Arabia	Respiratory secretions
ALJ54488.1	04-Mar-2015/Saudi Arabia	Respiratory secretions
ALJ54487.1	04-Mar-2015/Saudi Arabia	Respiratory secretions
ALJ54486.1	28-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54485.1	25-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54484.1	14-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54483.1	13-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54482.1	13-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54481.1	13-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54480.1	10-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54479.1	01-Apr-2015/Saudi Arabia	Respiratory secretions
ALJ54478.1	29-Mar-2015/Saudi Arabia	Respiratory secretions
ALJ54477.1	29-Mar-2015/Saudi Arabia	Respiratory secretions
ALJ54476.1	21-Mar-2015/Saudi Arabia	Respiratory secretions
ALJ54475.1	20-Mar-2015/Saudi Arabia	Respiratory secretions
ALJ54474.1	09-Mar-2015/Saudi Arabia	Respiratory secretions
ALJ54473.1	05-Mar-2015/Saudi Arabia	Respiratory secretions
ALJ54472.1	01-May-2015/Saudi Arabia	Respiratory secretions
ALJ54471.1	08-May-2015/Saudi Arabia	Respiratory secretions
ALJ54470.1	10-May-2015/Saudi Arabia	Respiratory secretions
AID55078.1	2014/Saudi Arabia	
AID55077.1	2014/Saudi Arabia	
AID55076.1	2014/Saudi Arabia	
AID55075.1	2014/Saudi Arabia	

(continued)

Table 2
(continued)

Accession No of S glycoprotein	Date and place of collection	Type of specimen
AID55074.1	2014/Saudi Arabia	
AID55073.1	22-Apr-2014/Saudi Arabia	
AID55072.1	15-Apr-2014/Saudi Arabia	
AID55071.1	21-Apr-2014/Saudi Arabia	
AID55070.1	14-Apr-2014/Saudi Arabia	
AID55069.1	12-Apr-2014/Saudi Arabia	
AID55068.1	07-Apr-2014/Saudi Arabia	
AID55067.1	2014/Saudi Arabia	
AID55066.1	2014/Saudi Arabia	
ALJ54469.1	13-May-2015/Saudi Arabia	Respiratory secretions
ALJ54468.1	10-May-2015/Saudi Arabia	Respiratory secretions
ALJ54467.1	12-May-2015/Saudi Arabia	Respiratory secretions
ALJ54466.1	12-Mar-2015/Saudi Arabia	Respiratory secretions
ALJ54465.1	07-Mar-2015/Saudi Arabia	Respiratory secretions
ALJ54464.1	08-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54463.1	01-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54462.1	Saudi Arabia	Respiratory secretions
ALJ54461.1	10-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54460.1	21-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54459.1	21-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54458.1	23-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54457.1	23-Feb-2015/Saudi Arabia	Respiratory secretions
AID55098.1	2014/Saudi Arabia	
AID55097.1	2014/Saudi Arabia	
AID55096.1	2014/Saudi Arabia	
AID55095.1	2014/Saudi Arabia	
AID55094.1	2014/Saudi Arabia	
AID55093.1	2014/Saudi Arabia	
AID55092.1	2014/Saudi Arabia	
AID55091.1	2014/Saudi Arabia	
AID55090.1	2014/Saudi Arabia	

(continued)

Table 2
(continued)

Accession No of S glycoprotein	Date and place of collection	Type of specimen
AID55089.1	2014/Saudi Arabia	
AID55088.1	2014/Saudi Arabia	
AID55087.1	2014/Saudi Arabia	
AID55086.1	2014/Saudi Arabia	
AID55085.1	2014/Saudi Arabia	
AID55084.1	2014/Saudi Arabia	
AID55083.1	2014/Saudi Arabia	
AID55082.1	2014/Saudi Arabia	
AID55081.1	2014/Saudi Arabia	
AID55080.1	2014/Saudi Arabia	
AID55079.1	2014/Saudi Arabia	
ALJ54478.1	29-Mar-2015Saudi Arabia	Respiratory secretions
ALJ54477.1	29-Mar-2015/Saudi Arabia	Respiratory secretions
ALJ54473.1	05-Mar-2015/Saudi Arabia	Respiratory secretions
ALJ54472.1	01-May-2015/Saudi Arabia	Respiratory secretions
ALJ54471.1	08-May-2015/Saudi Arabia	Respiratory secretions
ALJ54470.1	10-May-2015/Saudi Arabia	Respiratory secretions
ALJ54469.1	13-May-2015/Saudi Arabia	Respiratory secretions
ALJ54468.1	10-May-2015/Saudi Arabia	Respiratory secretions
ALJ54467.1	12-May-2015/Saudi Arabia	Respiratory secretions
ALJ54466.1	12-Mar-2015/Saudi Arabia	Respiratory secretions
ALJ54465.1	07-Mar-2015/Saudi Arabia	Respiratory secretions
ALJ54464.1	08-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54463.1	01-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54462.1	30-Jan-2015/Saudi Arabia	Respiratory secretions
ALJ54461.1	10-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54460.1	21-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54459.1	21-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54458.1	23-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54457.1	23-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54456.1	26-Feb-2015/Saudi Arabia	Respiratory secretions

(continued)

Table 2
(continued)

Accession No of S glycoprotein	Date and place of collection	Type of specimen
ALJ54454.1	28-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54455.1	28-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54453.1	06-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54452.1	14-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54451.1	14-Feb-2015/Saudi Arabia	Respiratory secretions
ALJ54450.1	12-Feb-2015/Saudi Arabia	Respiratory secretions

2.4.3 Prediction of Epitope Antigenicity Sites

The Kolaskar and Tongaonkar antigenicity method was used to determine the antigenic sites with a default threshold value of 1.045.

2.4.4 Prediction of Epitope Hydrophilicity

Parker hydrophilicity prediction tool was used to determine the hydrophilicity of the conserved regions; the threshold default value was 1.286.

2.4.5 Prediction of Beta Turn Sites

Chou and Fasman beta turn prediction method was used with the default threshold 1.009 to determine the sites that contain beta turns.

2.4.6 Prediction of Flexibility

Karplus and Schulz flexibility prediction tools were used for the prediction of chain flexibility in proteins (selection of peptide antigen) with default threshold value 0.992.

Thresholds of all tools were provided by IEDB and it is mainly calculated by the software as the average score of the tested protein for each corresponding tools.

2.5 T-Cell Epitope Prediction

Scanning an antigen sequence for amino acid patterns indicative of:

2.5.1 MHC Class I Binding Predictions

Analysis of peptide binding to MHC class I molecules was assessed by the IEDB MHC I prediction tool http://tools.iedb.org/mhci/n; for MHC-I binding prediction, several alleles were used including HLA-A, HLA-B, HLA-C, and HLA-E that have been reported as frequent around the world. MHC-I peptide complex presentation to T lymphocytes undergo several steps. The attachment of cleaved peptides to MHC molecules step was predicted. Consensus method which combines ANN, SMM, and scoring matrices derived from combinatorial peptide libraries (Comblib_Sidney2008) was used. 9-mer epitope lengths were selected. All internationally conserved epitopes that bind to alleles at score equal or less than 1.0 percentile rank (low percentile rank = good binders) were selected for further

analysis as in selecting thresholds (cutoffs) for MHC class I and II binding predictions, http://help.iedb.org/entries/23854373-Selecting-thresholds-cut-offs-for-MHC-class-I-and-II-binding-predictions.

Note: For S glycoprotein, the sequence was divided into ten parts due to software limitations, no more than 200 FASTA sequences interring [7–11].

2.5.2 MHC Class II Binding Predictions

Analysis of peptide binding to MHC class II molecules was assessed by the IEDB MHC II prediction tool http://tools.immuneepitope.org/mhcii/. For MHC-II binding prediction, the reference set of alleles was used, which include HLA-DQ, HLA-DP, and HLA-DR that are most frequent around the world. MHC class II groove has the ability to bind to peptides with different lengths. There are seven prediction methods in the IEDB MHC II prediction tool; NetMHCIIpan was used in this study; the conserved epitopes that bind to alleles at scores equal or less than 10 percentile rank were selected for further analysis as in selecting thresholds (cutoffs) for MHC class I and II binding predictions, http://help.iedb.org/entries/23854373-Selecting-thresholds-cut-offs-for-MHC-class-I-and-II-binding-predictions [7, 11–14].

2.5.3 Proteasomal Cleavage/TAP Transport/ MHC Class I Combined Predictor

This tool combines predictors of proteasomal processing, TAP transport, and MHC binding to produce an overall score for each peptide's intrinsic potential of being a T-cell epitope selected; in this study NetMHCpan was used with immunoproteasomal cleavage prediction; there are two types of proteasomes, the constitutively expressed "housekeeping" type and immunoproteasomes that are induced by IFN-γ secretion. Results can be displayed in proteasome score, TAP score, MHC score, processing score, total score, and IC50 score. Explanations of prediction output:

Proteasome cleavage

The scores can be interpreted as logarithms of the total amount of cleavage site usage liberating the peptide C-terminus; it depends on a lot of other factors, e.g., the amount of source protein degraded.

TAP transport

The TAP score estimates an effective −log (IC50) values for the binding to TAP of a peptide or its N-terminal prolonged precursors.

MHC binding

The MHC binding prediction is identical to Class I with output −log (IC50) values.

Processing

This score combines the proteasomal cleavage and TAP transport predictions. It predicts a quantity proportional to the amount of peptide present in the ER, where a peptide can bind to multiple MHC molecules. This allows predicting T-cell epitope candidates independent of MHC restriction.

Total

This score combines the proteasomal cleavage, TAP transport, and MHC binding predictions. It predicts a quantity proportional to the amount of peptide presented by MHC molecules on the cell surface. High scores mean high efficiency.

2.5.4 Neural Network-Based Prediction of Proteasomal Cleavage Sites (NetChop) and T-Cell Epitopes (NetCTL and NetCTLpan)

NetChop that was used here is a predictor of proteasomal processing based upon a neural network. NetCTL and NetCTLpan are predictors of T-cell epitopes along a protein sequence. The positive predictions threshold, 0.5, 0.75, and 1, sequentially for all methods above are displayed in green, while the red color for prediction below the threshold.

2.5.5 MHC-NP: Prediction of Peptides Naturally Processed by the MHC

MHC-NP employs data obtained from MHC elution experiments in order to assess the probability that a given peptide is naturally processed and binds to a given MHC molecule. This tool used in this study was the winner of the second Machine Learning Competition in Immunology; it is composed of three groups of peptides, binders, nonbinders, and eluted peptides that considered as naturally processed peptides, so greater probe score considered naturally processing peptide.

2.6 Epitope Analysis Tools

2.6.1 Population Coverage Calculation

All potential MHC I and MHC II binders from spike glycoprotein, E protein, and S and E modified sequences were assessed for a population coverage against the whole world population especially Saudi Arabia with other reported MERS-CoV countries. Calculations are achieved using the selected MHC-I and MHC-II interacted alleles by the IEDB population coverage calculation tool http://tools.iedb.org/tools/population/iedb_input; it computes projected population coverage, average number of epitope hits/HLA combinations recognized by the population, and minimum number of epitope hits/HLA combinations recognized by 90% of the population (PC90).

2.7 Homology Modeling

The complete 3D structure of spike glycoprotein and envelope protein was obtained by phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2) which uses advanced remote homology detection methods to build 3D models. UCSF Chimera (version 1.8) was used to visualize the 3D structure, which is currently available within the chimera package and available from the chimera website (http://www.cgl.ucsf.edu/cimera). Homology modeling was achieved for further verification of the service accessibility and hydrophilicity of B-lymphocyte epitopes predicted, as well as visualization of all predicted T-cell epitopes in the structural level.

In addition to the above methods, three other software were used to determine the effect that was induced in S and E reference sequences among the amino acid (SNP, single nucleotide polymorphism).

2.8 Confirmation of Amino Acid Change in Spike Glycoprotein (S) and Envelope Protein (E) Sequence

2.8.1 PolyPhen-2

(Polymorphism Phenotyping v2) (http://genetics.bwh.harvard.edu/pph2/index.shtml) is an online bioinformatics program to automatically predict the consequence of an amino acid change on the structure and function of a protein was assessed here. Basically, this program searches for 3D protein structures, multiple alignments of homologous sequences, and amino acid contact information in several protein structure databases and then calculates position-specific independent count scores (PSIC) for each of two variants and then computes the PSIC score difference between two variants; PolyPhen scores were assigned as probably damaging (2.00 or more), possibly damaging (1.40–1.90), potentially damaging (1.0–1.50), and benign (0.00–0.90). Basically PolyPhen accepts input in form of SNPs or protein sequences [18].

2.8.2 I-Mutant Suite

I used I-Mutant version 3.0 (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) to predict the protein stability changes upon single-site mutations. I-Mutant3.0 basically can evaluate the stability change of a single-site mutation starting from the protein structure or from the protein sequences. This program was trained on some data set derived from ProTherm which is considered to be the most comprehensive database of experimental data on protein mutations [18].

2.8.3 Project Hope Mutation

(http://www.cmbi.ru.nl/hope/) Hope Version 1.1.0, HOPE is an easy-to-use web service that analyzes the structural effects of a point mutation in a protein sequence.

2.8.4 SNPs and GO

(http://snps.biofold.org/snps-and-go//snps-and-go.html) were used to predict disease-associated variations through using GO terms by collected information in a unique framework that derived from protein sequence, 3D structure, protein sequence profile, and protein function, beside gene ontology annotation to predict if a given variation can be classified disease-related or neutral. It calculates the result according to the three methods used depending on SVM type and data such as:

PANTHER — output of the PANTHER algorithm.

PhD-SNP — SVM input is the sequence and profile at the mutated position.

SNPs and GO — SVM input is all the input in PhD-SNP, PANTHER, and GO term features, by giving disease probability (if >0.5 mutation is predicted disease).

2.9 Peptide Search Tool

The peptide search tool was used to find all UniProtKB sequences that exactly match a query peptide sequence (http://www.uniprot.org/peptidesearch/). This means we can easily synthesis the

desired peptides in the laboratory by cloning methods and so on to study peptide impact on immune system via injected laboratory animals with peptide sequence of any organisms.

2.10 AllerHunter

(http://tiger.dbs.nus.edu.sg/AllerHunter/index.html) is a cross-reactive allergen prediction program built on a combination of support vector machine (SVM) and pairwise sequence similarity. Results of prediction of query sequence(s) can be achieved by using AllerHunter and FAO/WHO evaluation scheme; in AllerHunter sequence can be considered as a cross-reactive allergen if it has a probability of $\geqq 0.06$, while in the guideline of the FAO/WHO, they stated that a sequence is potentially allergenic if it either has an identity of at least 6 contiguous amino acids OR >35 percent sequence identity over a window of 80 amino acids when compared to known allergens.

2.11 AlgPred: Prediction of Allergenic Proteins and Mapping of IgE Epitopes

(http://www.imtech.res.in/raghava/algpred/index.html)
AlgPred used to predict allergenic protein and mapping of IgE epitopes by:

1. It allows prediction of allergens based on similarity of known epitope with any region of protein.
2. The mapping of IgE epitope(s) feature of server allows user to locate the position of epitope in their protein.
3. Server search MEME/MAST allergen motifs using MAST and assign a protein allergen if it has any motif.
4. It allows predicting allergens based on SVM modules using amino acid or dipeptide composition.
5. It facilitates BLAST search against 2890 allergen-representative peptides (ARPs) obtained from Bjorklund et al. (2005) and assigns a protein allergen if it has a BLAST hit.
6. Hybrid option of server allows predicting allergen using combined approach (SVMc + IgE epitope + ARPs BLAST + MAST).

2.12 VaxiJen v2.0

(http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen_help.html) VaxiJen is the first server for alignment-independent prediction of protective antigens. It was developed to allow antigen classification solely based on the physicochemical properties of proteins without recourse to sequence alignment.

3 Results

3.1 Prediction of B-Cell Epitopes

Spike glycoprotein, E protein, and modified S and E protein were subjected to BepiPred linear epitope prediction, Emini surface accessibility, Kolaskar and Tongaonkar antigenicity, Parker hydrophobicity, Chou and Fasman beta turn prediction methods, and

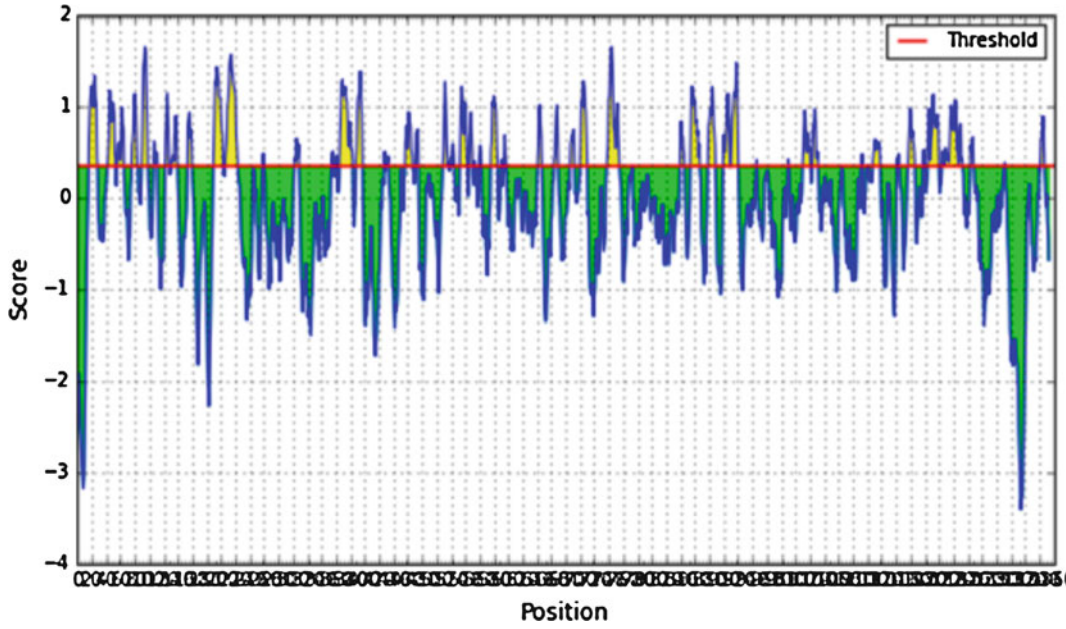

Fig. 1 BepiPred linear epitope prediction of S glycoprotein, the desired epitope residue showed in yellow color. The red horizontal line indicates surface accessibility threshold (0.35)

Karplus and Schulz flexibility in IEDB, as the results in Figs. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, and 24.

3.1.1 BepiPred Linear Epitope Prediction Method

The average binder score of spike glycoprotein to B cell was 0.35; all values equal or greater than the default threshold 0.35 were predicted to be potential B-cell binders.

3.1.2 Emini Surface Accessibility Prediction

The average surface accessibility areas of the protein were scored as 1.000; all values equal or greater than the default threshold 1.0 were regarded potentially in the surface. A total number of positive S glycoprotein peptide represent 481 peptide out of 1349, while in E protein represents 23 out of 77 and in S and E modified sequence represents 485 out 485 and 17out of 77 peptides sequentially.

3.1.3 Kolaskar and Tongaonkar Antigenicity

The default threshold of antigenicity of the protein was 1.045; all values greater than 1.045 were considered as potential antigenic determinants. The positive result number of selected S glycoprotein peptide represents 655 out of 1348, while in E protein represents 55 out of 76 and in S and E modified sequence represents 668 out of 668 and 47 out of 76 peptides sequentially.

3.1.4 Parker Hydrophilicity Prediction

The average hydrophilicity score of the protein was 1.286; all values equal or greater than the default threshold 1.286 were potentially hydrophilic. The positive result number of S glycoprotein peptide

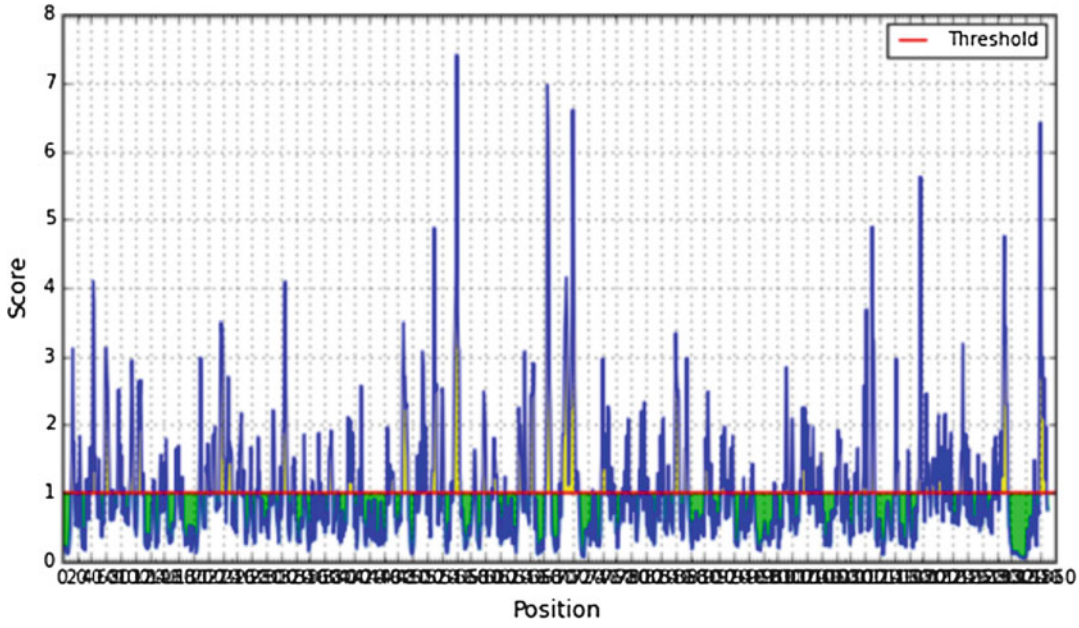

Fig. 2 Emini surface accessibility prediction of S glycoprotein. The desired epitope residue for surface accessibility showed in yellow color, while green color was below threshold (1.000)

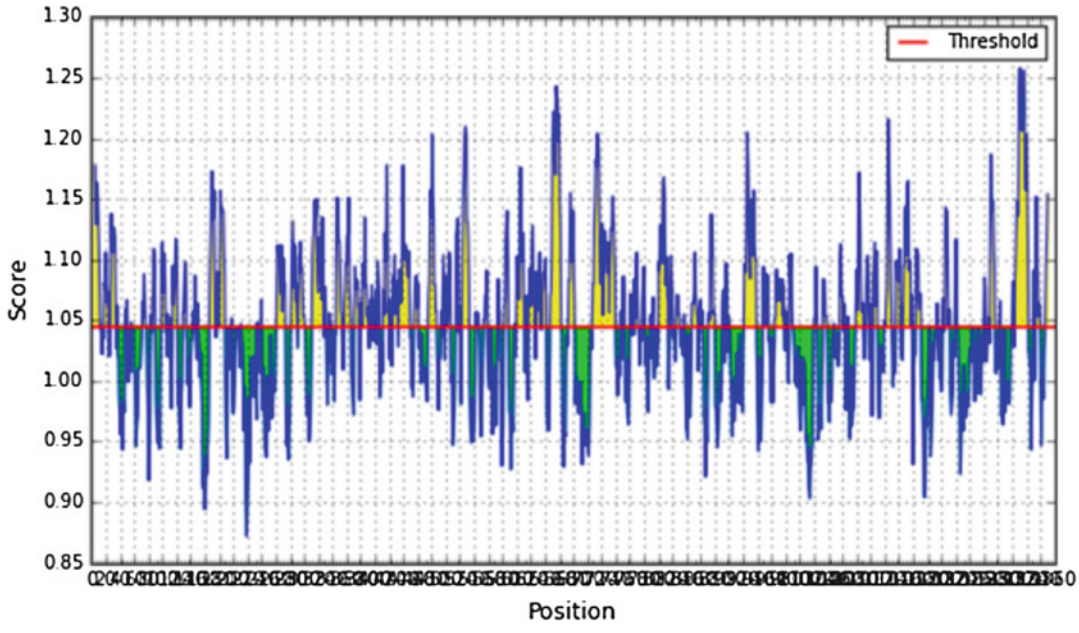

Fig. 3 Kolaskar and Tongaonkar antigenicity prediction of S glycoprotein. The desired epitope residue for antigenicity showed in yellow color, while the green color below the red horizontal line indicates less antigenicity below (1.045)

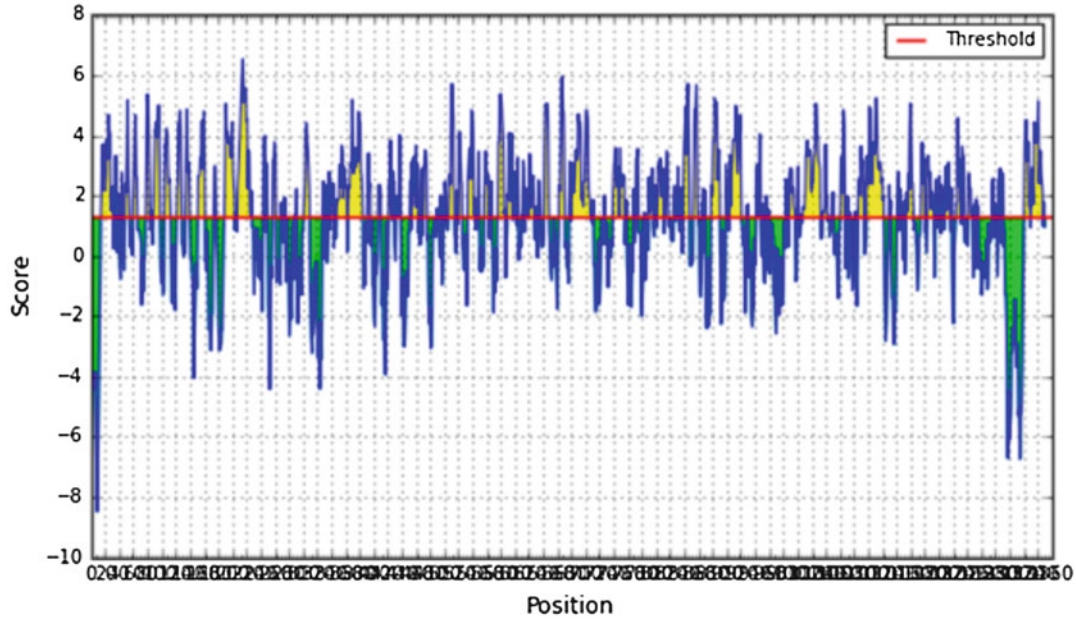

Fig. 4 Parker hydrophilicity prediction of S glycoprotein. The desired epitope residue showed in yellow color. The red horizontal line indicates parker hydrophilicity threshold (1.286)

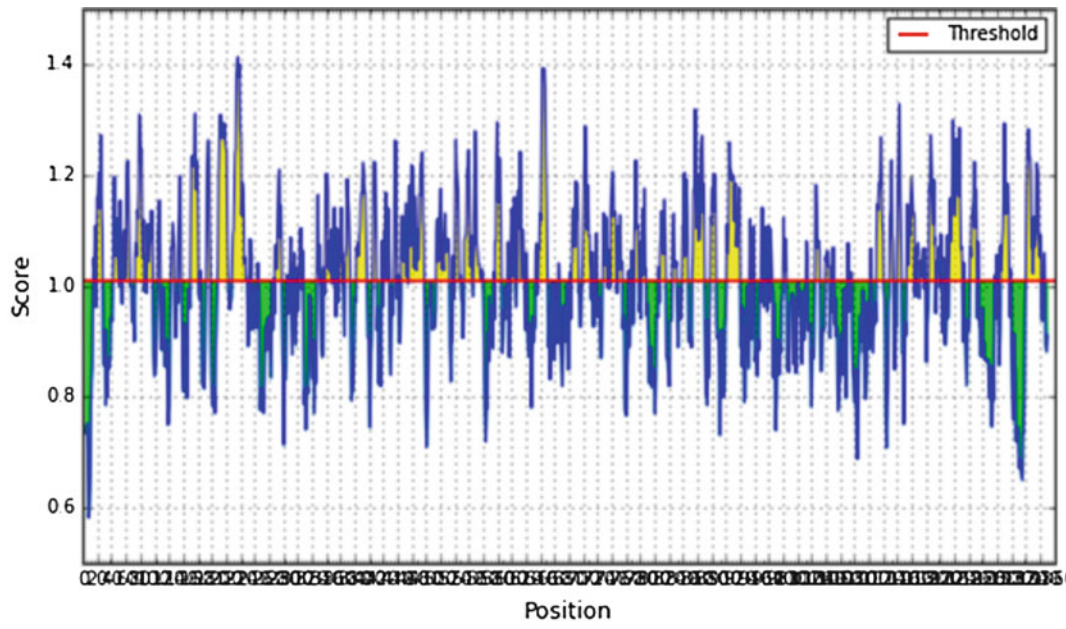

Fig. 5 Chou and Fasman beta turn prediction of S glycoprotein. The desired epitope residue showed in yellow color. The red horizontal line indicates beta turn prediction threshold (1.009)

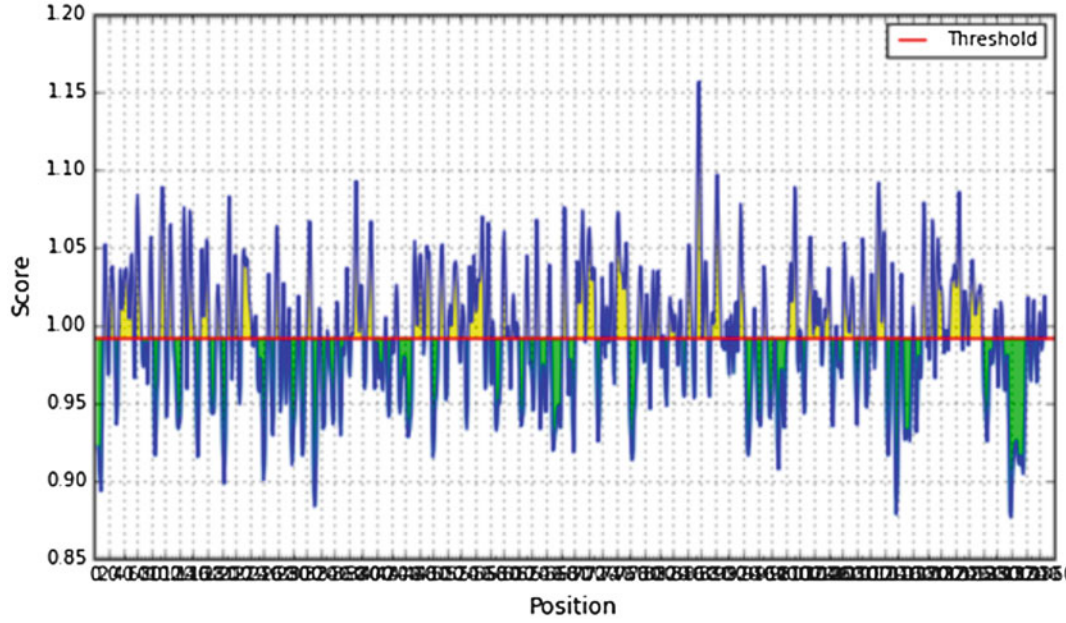

Fig. 6 Karplus and Schulz flexibility prediction of S glycoprotein. The desired epitope residue showed in yellow color. The red horizontal line indicates surface accessibility threshold (0.35)

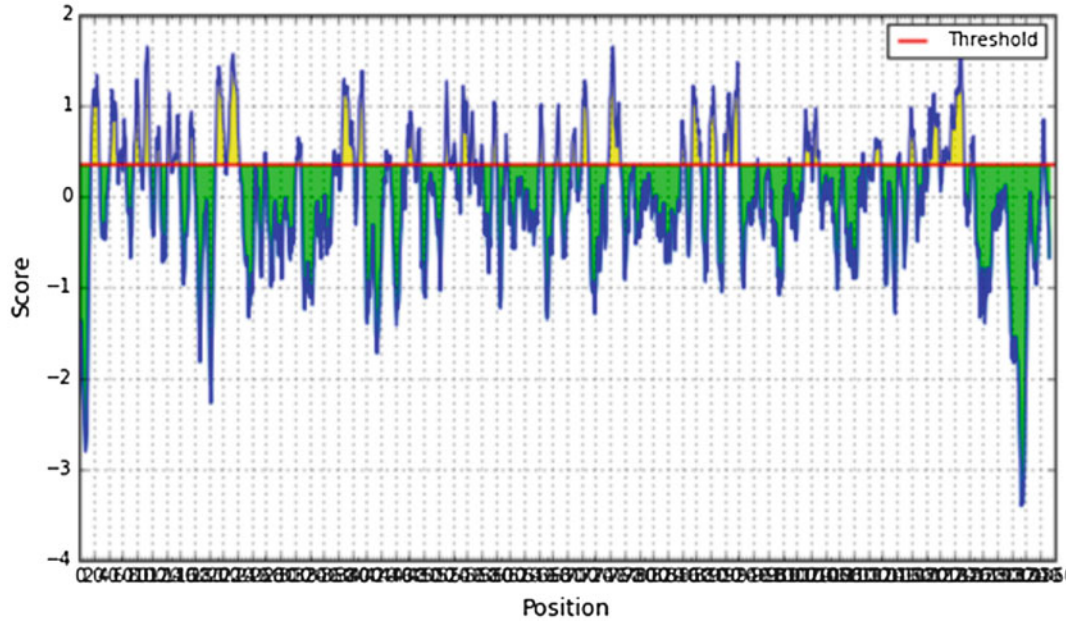

Fig. 7 BepiPred linear epitope prediction of S glycoprotein modified sequence. The desired epitope residue showed in yellow color. The red horizontal line indicates BepiPred Linear Epitope threshold (0.35)

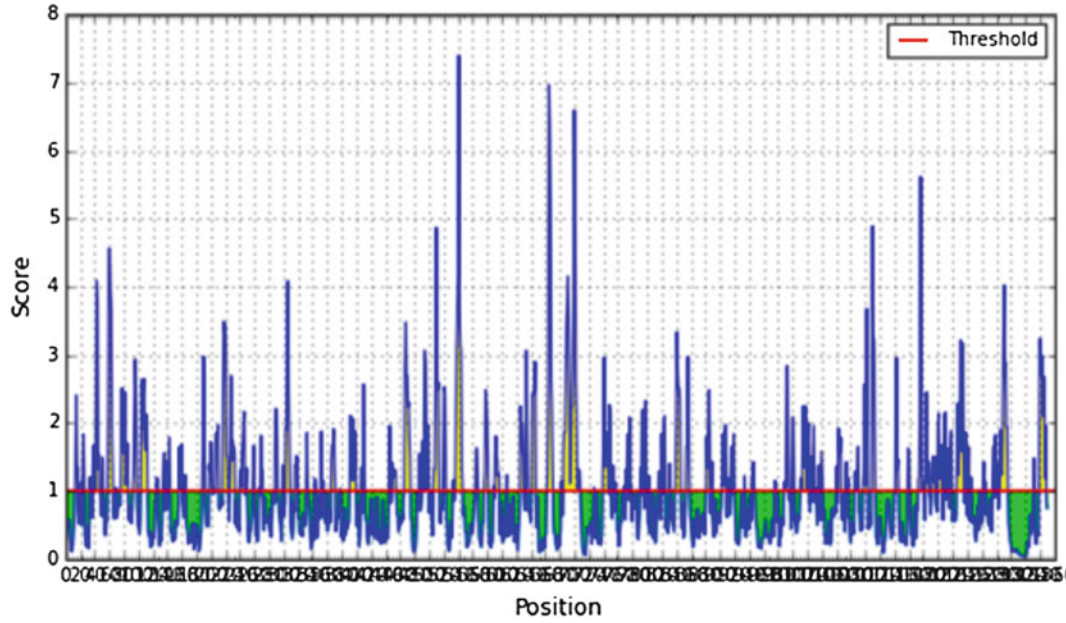

Fig. 8 Emini surface accessibility prediction of S glycoprotein modified sequence. The desired epitope residue showed in yellow color, while green color below the red horizontal line indicates surface accessibility threshold $\leq$ (1.000)

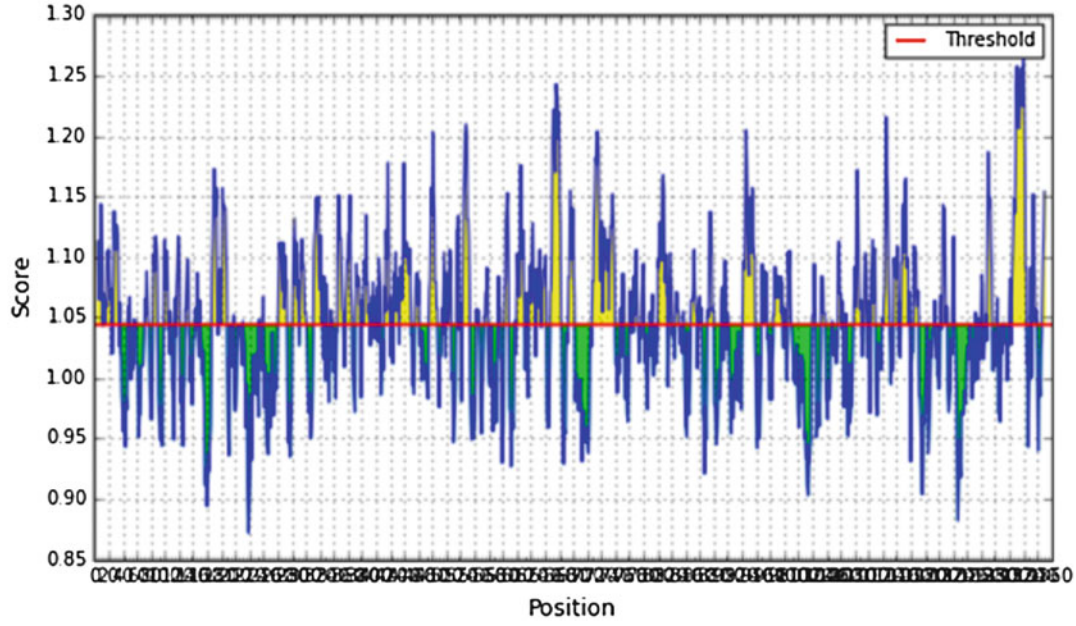

Fig. 9 Kolaskar and Tongaonkar antigenicity prediction of S glycoprotein modified sequence. The desired epitope residue showed in yellow color. The red horizontal line indicates antigenicity threshold $\leq$ (1.045)

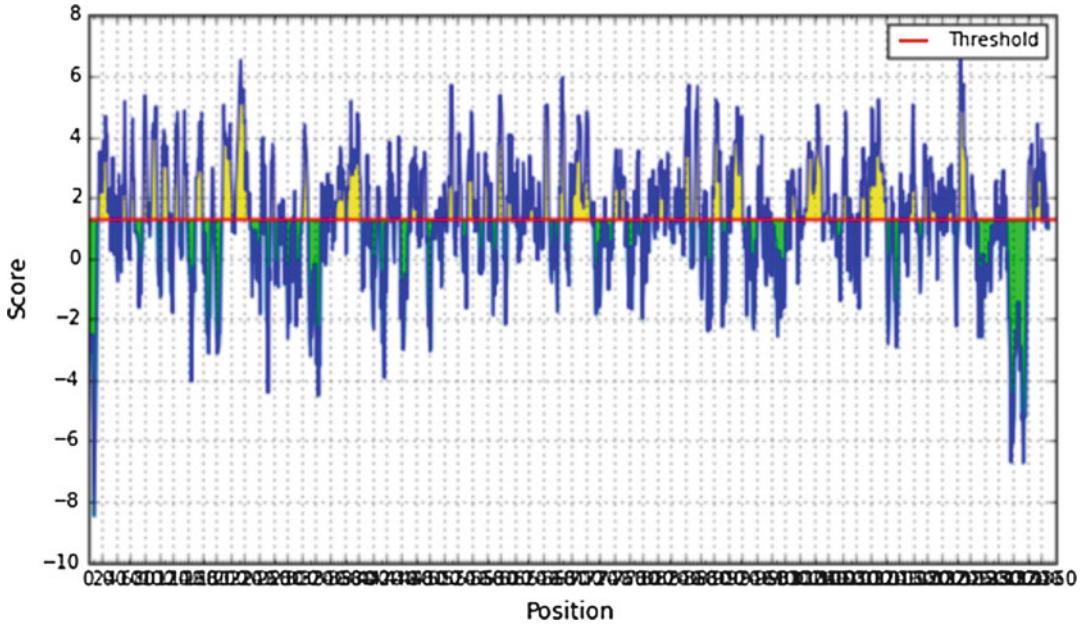

Fig. 10 Parker hydrophilicity prediction of S glycoprotein modified sequence. The desired epitope residue showed in yellow color, while green color below the red horizontal line indicates hydrophilicity threshold $\leq$ (1.286)

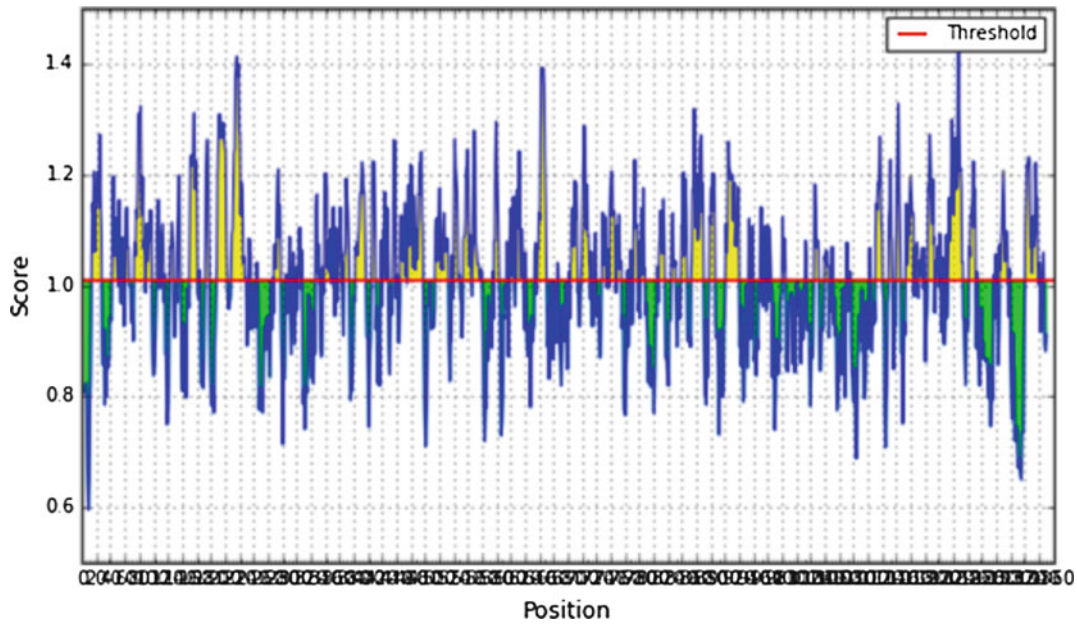

Fig. 11 Chou and Fasman beta turn prediction of S glycoprotein modified sequence. The desired epitope residue showed in yellow color. The red horizontal line indicates beta turn threshold (1.009)

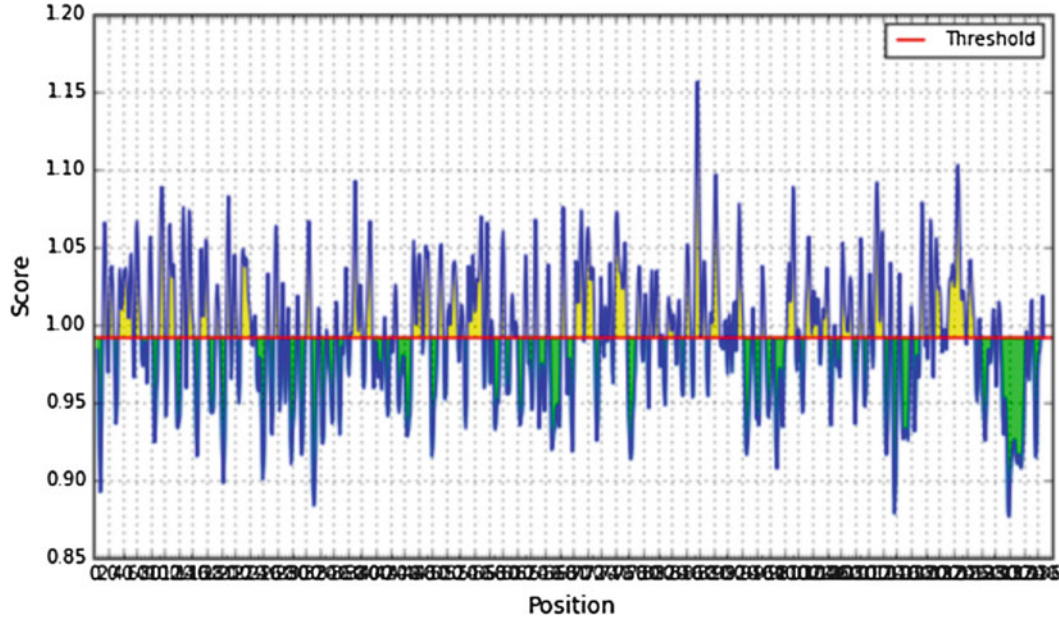

Fig. 12 Karplus and Schulz flexibility prediction of S glycoprotein modified sequence. The desired epitope residue showed in yellow color, while green color below the red horizontal line indicates flexibility threshold $\leq$ (0.992)

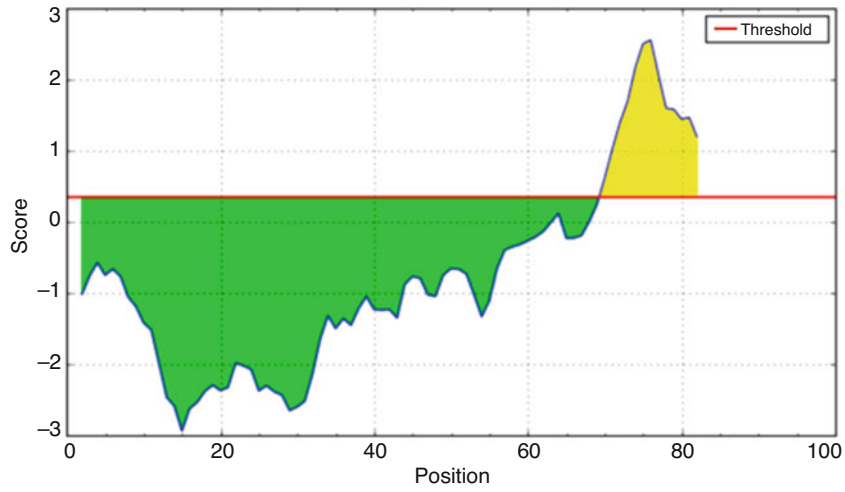

Fig. 13 BePipred linear epitope prediction of E protein. The desired epitope residue showed in yellow color. The red horizontal line indicates Bepipred Linear Epitope threshold $\leq$ (0.35)

represents 693 out of 1348, while in E protein represents 18 out of 76 and in S and E modified sequence represents 690 out of 695 and 20 out of 76 peptides sequentially.

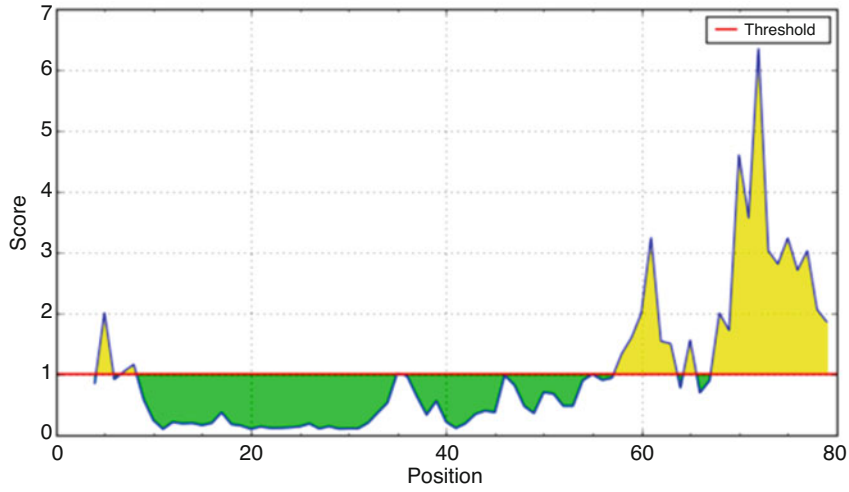

Fig. 14 Emini surface accessibility prediction of E protein. The desired epitope residue showed in yellow color, while green color below the red horizontal line indicates surface accessibility threshold (1.000)

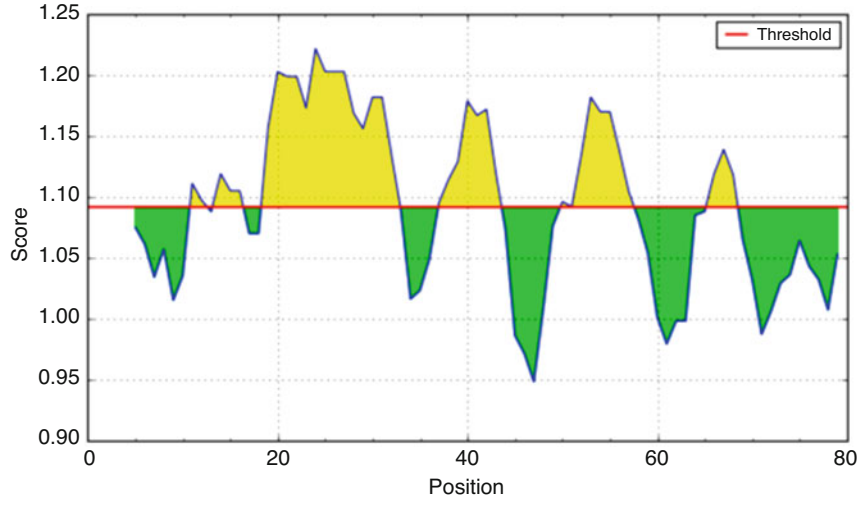

Fig. 15 Kolaskar and Tongaonkar antigenicity prediction of E protein. The desired epitope residue showed in yellow color, while green color below the red horizontal line indicates antigenicity threshold (1.045)

3.1.5 Chou and Fasman Beta Turn Prediction

To determine the site that contains beta turns, the default threshold was 1.009; all values equal or greater than the default threshold were considered beta turn sites. The positive result number of selected peptide represents 668 out of 1348 in S glycoprotein, while it represents 19 out of 76 in E protein and 673 out of 673 with 21 out of 76 in both S and E modified sequence sequentially.

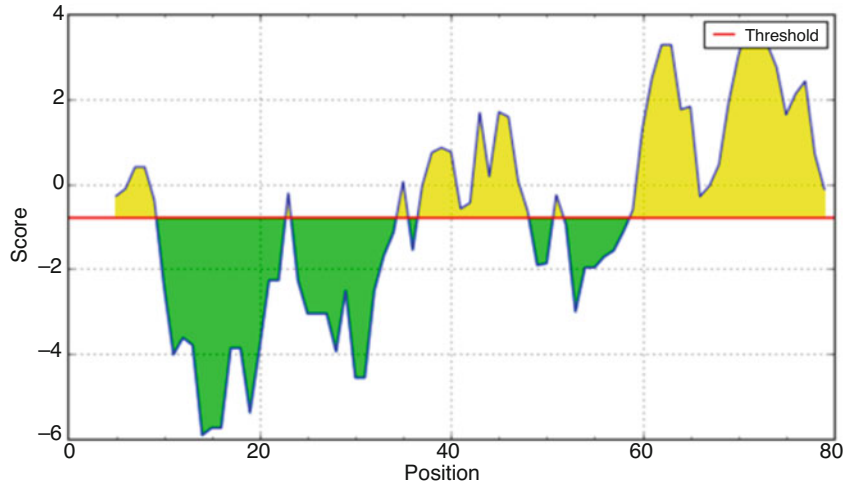

Fig. 16 Parker hydrophilicity prediction of E protein the desired epitope residue showed in yellow color. The red horizontal line indicates hydrophilicity threshold $\leq$ (1.286)

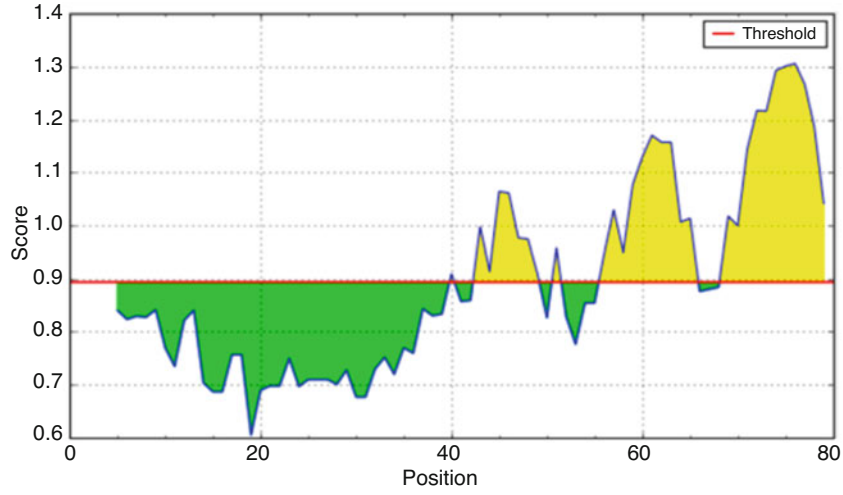

Fig. 17 Chou and Fasman beta turn prediction of E protein. The desired epitope residue showed in yellow color. The red horizontal line indicates beta turn threshold $\leq$ (1.009)

3.1.6 Karplus and Schulz Flexibility Prediction

The default threshold value 0.992 determined chain flexibility in proteins, so all values equal or greater than the default threshold were considered as chain flexibility of protein. The positive results of selected peptide represent 679 out of 1347 in S glycoprotein, and it represents 24 out of 24 in E protein beside represented 680 out of 681 and 24 out of 75 in S and E modified sequences sequentially.

The most common B-cell epitope for E protein is YVKFQDS in a position 69, while for E protein modified sequence, they are

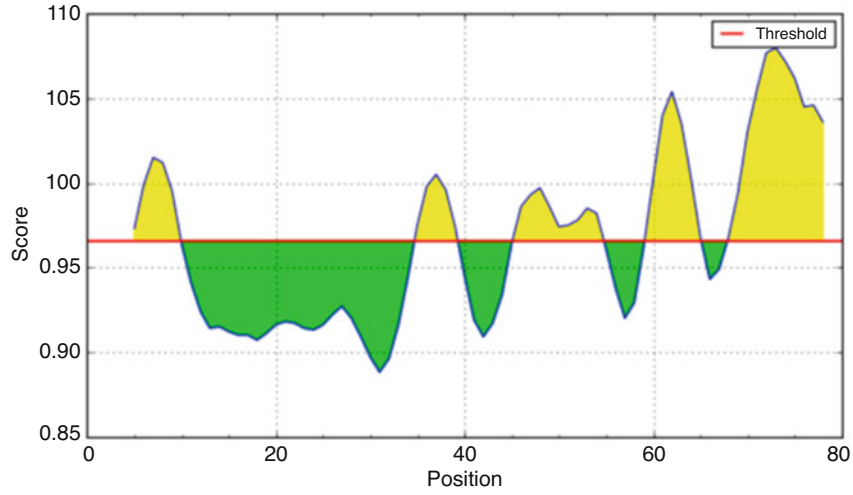

Fig. 18 Karplus and Schulz flexibility prediction of E protein. The desired epitope residue showed in yellow color, while green color below the red horizontal line indicated flexibility below threshold (0.992)

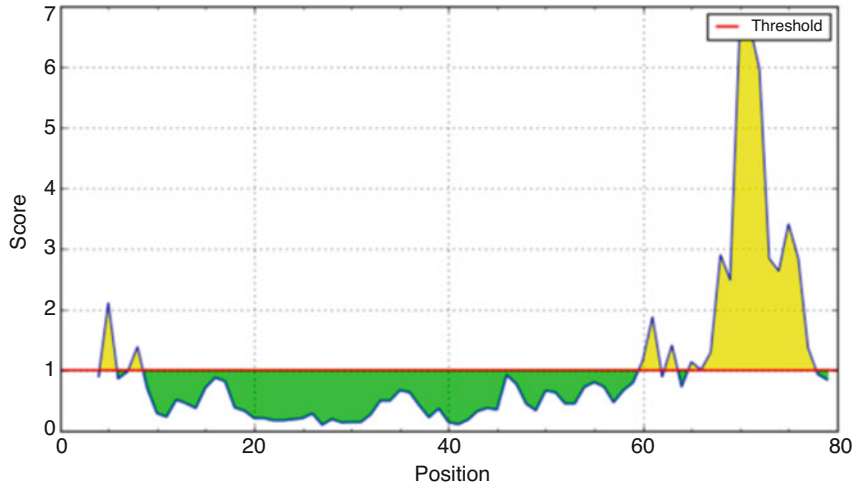

Fig. 19 BepiPred linear epitope prediction of E protein modified sequence. The desired epitope residue showed in yellow color. The red horizontal line indicates BepiPred Linear Epitope threshold (0.35)

VYVPQQD, YVPQQDS, and PPLPED/PPLPEDV in positions 68, 69, and 77 respectively.

The most common B-cell epitopes for both S and modified S are DVGPDSV, PDSVKSA, DSVKSAC, PRPIDVS, HTPATDC, AKPSGSV, KPSGSVV, SGTPPQV, GTPPQVY, TPPQVYN, QLSPLEG, YGPLQTP, PRSVRSV, RSVRSVP, SVKSSQS, VKSSQSS, SQSSPII, and SLNTKYV in the following positions 23, 26, 27, 48, 211, 371, 372, 393, 394, 395, 547, 707, 750, 751, 855, 856, 859 (or 857 in modified S), and 1202 sequentially; but QVDQLNS and VDQLNSS in positions 772 and

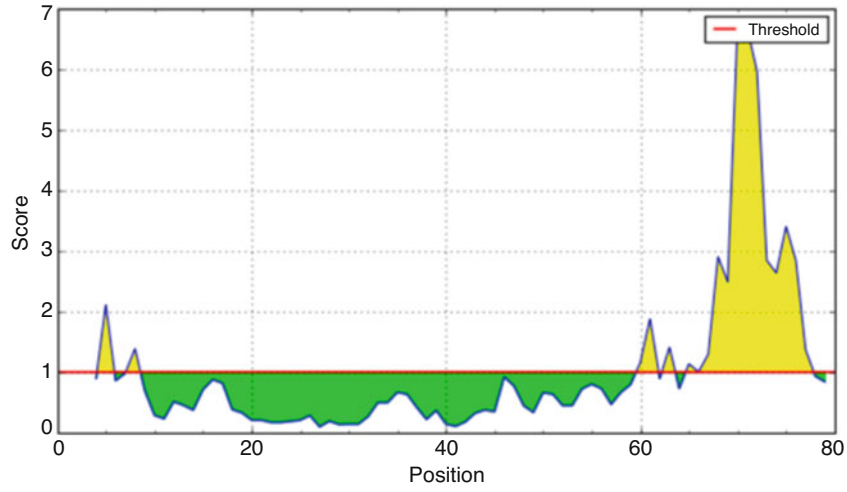

Fig. 20 Emini surface accessibility prediction of E protein modified sequence. The desired epitope residue showed in yellow color, above the red horizontal line threshold (1.000)

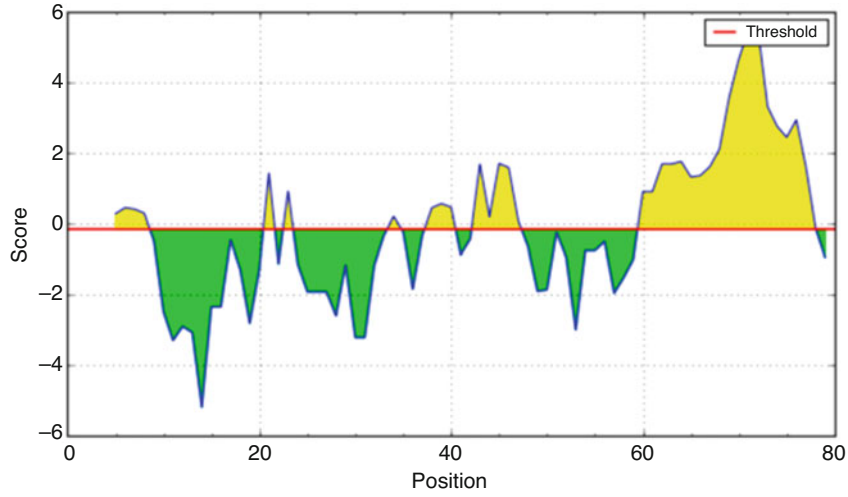

Fig. 21 Kolaskar and Tongaonkar Antigenicity prediction of E protein modified sequence. The desired epitope residue showed in yellow color, while green color indicates antigenicity below threshold (1.045)

773 are ordinary only found in S glycoprotein, while LTPTSSY, TPTSSYV, PTSSYVD, TSSYVDV, DHGDYYV, YSQDVKQ, ANQYSPC, NQYSPCV, and YYRKQLS in a positions 15, 16, 17, 18, 83, 108, 523, 524, and 543 sequentially are only found in S glycoprotein modified sequence.

3.2 T-Cell Epitope Prediction

Spike glycoprotein, E protein, and S and E modified sequence were subjected to consensus method for MHC-I binding, NetMHCII-pan for MHC-II binding, NetMHCpan for proteasomal cleavage/TAP transport/MHC class I combined predictor, NetChop and

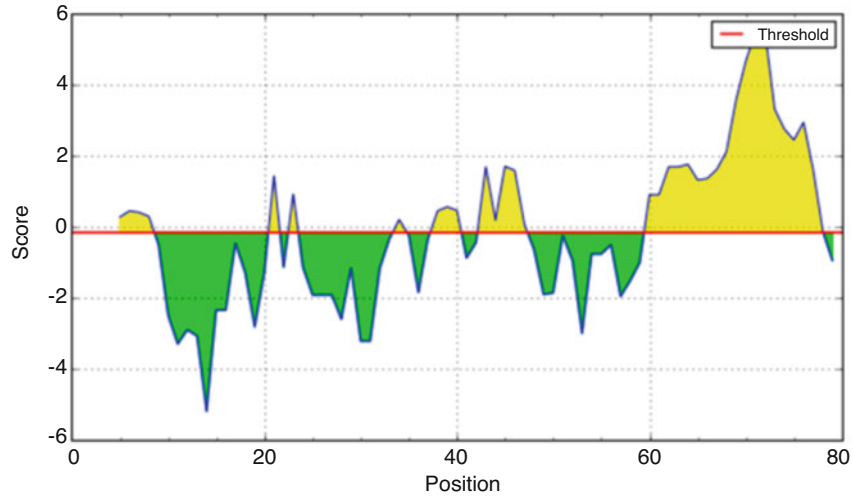

Fig. 22 Parker hydrophilicity prediction of E protein modified sequence. The desired epitope residue showed in yellow color. The red horizontal line indicates hydrophilicity threshold $\leq$ (1.286)

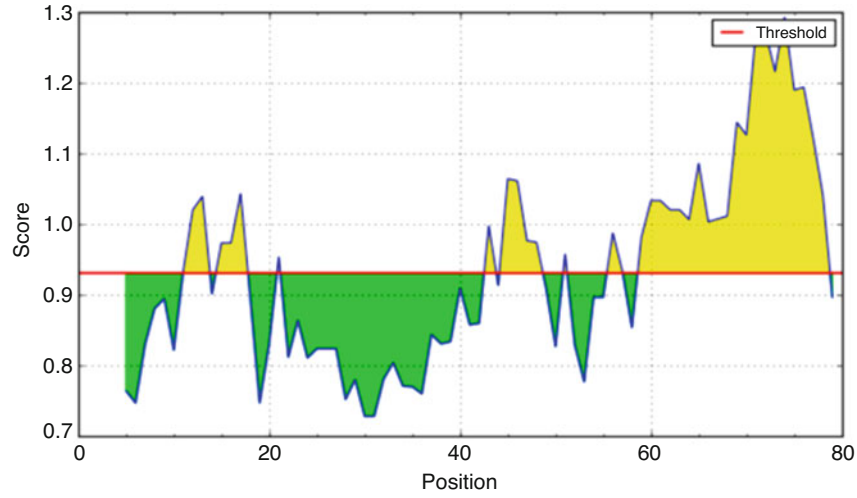

Fig. 23 Chou and Fasman beta turn prediction of E protein modified sequence. The desired epitope residue showed in yellow color, while green color below the red horizontal line indicates low beta turn threshold $\leq$ (1.009)

NetCTL for neural network-based prediction of proteasomal cleavage sites (NetChop), and T-cell epitopes (NetCTL and NetCTL-pan) with MHC-NP for prediction of peptides that's naturally processed by the MHC in IEDB software program.

3.2.1 MHC Class I Binding Predictions

Analysis of peptide sequence that's binding to MHC class I molecules by consensus method was assessed by the conserved epitopes that bind to alleles at score equal or less than 1.0 percentile. The

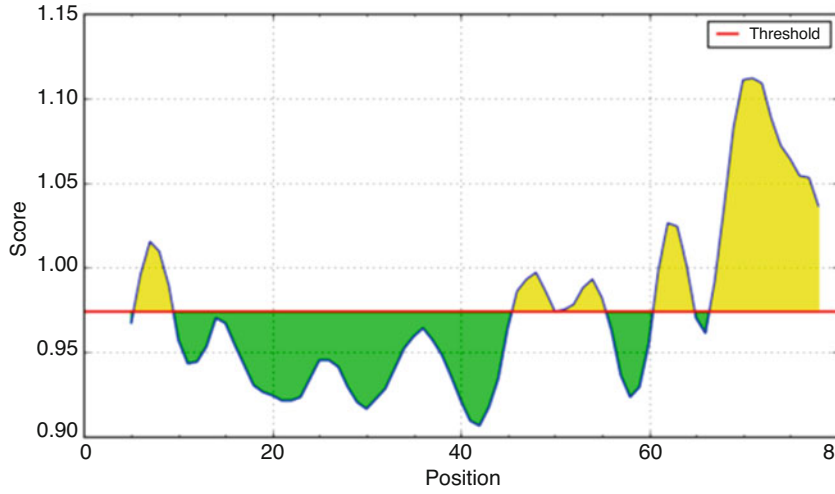

Fig. 24 Karplus and Schulz flexibility prediction of E protein modified sequence. The desired epitope residue showed in yellow color that illustrates flexibility threshold $\leq$ (0.992)

positive result numbers of selected peptide represent 602 out of 53,800 in S glycoprotein and 63 out of 3626 in E protein while in S and E modified sequence represents 612 out of 58,457 and 41 out of 3234 sequentially.

Seven alleles were not found in E protein modified sequence, including HLA-A∗03:01, HLA-A∗11:01, HLA-A∗31:01, HLA-A∗68:01, HLA-B∗14:02, HLA-B∗40:01, and HLA-B∗40:02, while in E protein four alleles were not found; they are HLA-B∗48:01, HLA-B∗58:02, HLA-C∗04:01, and HLA-E∗01:01; the ruminant of alleles are common between both of them; among them three peptide sequences are common such as CMTGFNTLL^{n}, MTGFNTLLV^{n}, and QCMTGFNTL^{n}, while HLCVQCMTG, KPPLPEDVW, LLVCTAFLT, LLVQPALSL, LTATHLCVQ, LVCTAFLTA, PALSLYMTG, PNFFDFTVV^{n}, SLYMTGRSV, VCTAFLTAT, VQERIGWFI, VQPALSLYM, VVCDITLLV, and WFIPNFFDF^{n} are only found in E modified sequence.

HLA-A∗02:01 allele showed higher frequency numbers six, followed by HLA-A∗23:01, HLA-A∗29:02, HLA-A∗68:02, and HLA-B∗46:01 that had four frequency numbers, and the same for the peptide sequences FIFTVVCAI, ITLLVCMAF, IVNFFIFTV^{n}, and LVQPALYLY in E protein while in modified E, I found HLA-C∗03:03 represents higher frequency numbers forty-three, but HLA-A∗02:01, HLA-A∗02:06, HLA-A∗29:02, and HLA-B∗38:01 had the same frequency numbers three.

For the peptide sequences, I found FIFTVVCAI had a higher frequency numbers five, followed by ITLLVCMAF, IVNFFIFTV^{n}, and LVQPALYLY in E protein; reverse E protein modified

sequence, LVQPALSLY had a higher frequency numbers five then followed by CMTGFNTLL[n], FLTATHLCV, FVQERIGWF, ITLLVCTAF, LYMTGRSVY, WFIPNFFDF[n], and YMTGRSVYV which had a frequency numbers four except QCMTGFNTL[n] that had three frequency numbers.

N.B: [n]indicate presence of asparagine (N) in peptide sequences, that's hiding epitope from recognition by immune system so we should deal with the common epitope with the caution; they are 11 peptide sequence numbers with asparagine in E and 13 in modified E, while they are 8 in S and 46 in modified S sequence.

HLA-A∗30:02 allele was not found in S glycoprotein modified sequence, while HLA-B∗38:01, HLA-B∗39:01, HLA-B∗40:01, HLA-B∗40:02, HLA-B∗44:02, HLA-B∗44:03, HLA-B∗46:01, HLA-B∗48:01, HLA-B∗51:01, and HLA-B∗53:01 were not found in S sequence, but they were found in S modified sequence; these means 15 peptide sequences were absent in S sequence (AGYKVLPPL, APQVTYQNI[n], CKLPLGQSL, CVFFILCCV, DVKQFDNGF[n], DYYVYSAGH, FKLSIPTNF[n], FLLTPTSSY, GEMRLASIA, GNYTYYHKW[n], GPASARDLI, GTDTNSVCI[n], HKWPWYIWL, HSKFLLMFL, IAPVNGYFI[n]) but presented in modified S sequence; besides this it also lakes a 34 peptide sequences like AGPISQFNY[n], CMGKLKCNR[n], DLSQLHCSY, DVKQFANGF[n], FATYHTPAT, FLLTPTESY, FQFATLPVY, FVYDAYQNL[n], GTNCMGKLK[n], GVRQQRFVY, HSVFLLMFL, ICAQYVAGY, etc.; the other peptide sequences were not shown here.

In S glycoprotein HLA-A∗29:02 allele showed higher frequency numbers (41) then followed by HLA-A∗30:02 (37), HLA-A∗01:01 (31), HLA-B∗15:01 (29), HLA-C∗14:02 (27), HLA-A∗25:01 (25), HLA-A∗23:01 (24), HLA-B∗58:01 (23), and HLA-C∗06:02 (22); modified S glycoprotein sequence partially shared the same alleles with higher frequency numbers like in S glycoprotein which they are HLA-A∗29:02 allele that represented the most higher frequency numbers (33), followed by HLA-C∗14:02 (27), HLA-A∗01:01 (25), HLA-B∗46:01 (22)/HLA-A∗23:01, HLA-B∗58:01, and HLA-C∗06:02 (21)/HLA-B∗15:01 (20). In S glycoprotein the following peptide sequences had higher frequency numbers such as 10 in FSFGVTQEY and ITYQGLFPY peptides, 8 in WSYTGSSFY, 7 in KAWAAFYVY, and 6 in FVYDAYQNL[n], and ITITYQGLF, QTAQGVHLF, while it represented 5 in FQFATLPVY, NSYTSFATY[n], SLILDYFSY, STVWEDGDY, VSVPVSVIY, and YTYYNKWPW[n], but in modified S glycoprotein, the frequencies were different, like 10 in FSFGVTQEY peptide, 4 in FLLTPTSSY, FSSRYVDLY, FVANYSQDV[n], FYVYKLQPL, and IAFNHPIQV[n], while it's 3 in

ASIAFNHPI[n], DEILEWFGI, DYFSYPLSM, EAAYTSSLL, FCSKINQAL[n], FFNHTLVLL[n], FQDELDEFF, FSDGKMGRF, FSNPTCLIL[n], GEMRLASIA, GRFFNHTLV[n], HISSTMSQY, and HKWPWYIWL peptides.

N.B: [n] indicate presence of asparagine (N) in peptide sequences, that's hiding epitope from recognition by immune system.

3.2.2 MHC Class II Binding Predictions

Analysis of peptide binding to MHC class II molecules was assessed by the conserved epitopes that bind to alleles at scores equal or less than 10 percentile rank; the positive result numbers of selected epitopes showed 212 out of 4819 epitopes in S glycoprotein, 685 out of 4148 in E protein, and 6896 out of 75,206 with 685 out of 4148 in both S and E modified proteins sequentially.

The following alleles are more common between S glycoprotein, E protein, and S and E modified sequences, and they are HLA-DPA1∗01:03/DPB1∗02:01, HLA-DPA1∗02:01/DPB1∗01:01, HLA-DRB1∗01:01, HLA-DRB1∗01:02, HLA-DRB1∗04:04, HLA-DRB1∗04:05, HLA-DRB1∗04:08, HLA-DRB1∗04:10, HLA-DRB1∗04:23, HLA-DRB1∗07:01, HLA-DRB1∗07:03, HLA-DRB1∗08:06, HLA-DRB1∗11:04, HLA-DRB1∗11:06, HLA-DRB1∗12:01, HLADRB1∗13:04, HLA-DRB1∗13:11, HLA-DRB1∗13:21, and HLA-DRB4∗01:01, but in S and modified S glycoprotein, both of them contain other 42 different alleles not shown here. In E and modified E protein, HLA-DRB1∗01:01 had higher frequency numbers of alleles which represented 20, followed by 17 in HLA-DRB1∗01:02, 11 in HLA-DRB1∗12:01, 10 in HLA-DRB1∗11:04, HLA-DRB1∗11:06, and HLA-DRB1∗13:11, and 9 in HLA-DRB1∗07:01, HLA-DRB1∗07:03 and HLA-DRB1∗13:21, while in S and modified S glycoprotein, those alleles below had higher frequency numbers, which represented (200/199) in HLA-DRB1∗04:08/(199/201) HLA-DRB1∗04:01, HLA-DRB1∗04:21, and HLA-DRB1∗04:26/(194/190) in HLA-DRB1∗09:01/(192/189) in HLA-DRB1∗04:05/(167/167) in HLA-DRB1∗07:01, HLA-DRB1∗07:03/(164/167) in HLA-DRB1∗15:02, (160/159) in HLA-DRB1∗13:02/(159/159) in HLA-DRB1∗11:14, HLA-DRB1∗11:20, and HLA-DRB1∗13:23, and (152/158) in HLA-DRB3∗01:01.

E and modified E protein had the same peptide sequences with same frequency numbers, but the higher frequency numbers only showed in peptides below; it represented 15 with GFNTLLVQPALSLYM[n], 14 with TGFNTLLVQPALSLY[n], 13 with FNTLLVQPALSLYMT, 12 with MTGFNTLLVQPALSL[n],

11 with NTLLVQPALSLYMTG[n], and 10 with ALSLYMTGRSVYVPQ, LSLYMTGRSVYVPQQ, PALSLYMTGRSVYVP, and QPALSLYMTGRSVYV peptides.

N.B:-

1. The alleles below are not available for S glycoprotein, E protein, and S and E modified sequence, and they are DPA1∗01-DPB1∗ 04:01, DRB1∗03:09, DRB1∗08:17, and DRB1∗13:28.
2. The same peptide sequence shared more than one allele gene or the same allele has a different peptide sequence.
3. Variation in frequency numbers among both alleles and peptide sequences has been shown when comparing reference sequence of S & E protein with the modified sequence of both of them.
4. [n] that is present in peptide sequences above indicates presence of arginine in the sequence.

3.2.3 Proteasomal Cleavage/TAP Transport/ MHC Class I Combined Predictor

In NetMHCpan high scores mean high efficiency due to prediction of a quantity proportional to the amount of peptide presented by MHC molecules on the cell surface; total score higher or equal to 0 were selected for S and modified S glycoprotein, while in E protein total score equal or higher than 0.3 was selected, but in modified E protein total score equal or higher than −2.82 was selected; *see* Tables 3 and 4.

3.2.4 Neural Network-Based Prediction of Proteasomal Cleavage Sites (NetChop) and T-Cell Epitopes (NetCTL and NetCTLpan)

The positive prediction thresholds are 0.5 and 0.75 (green color) for NetChop and NetCTL sequentially considered as proteasomal cleavage sites for T-cell epitopes; *see* Figs. 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, and 38 with Table 5.

NetChop prediction score equal or greater than 0.5 in S glycoprotein represented a positive result; more than 300 peptides out of 1353 showed positive results, while in modified S glycoprotein, 5 out of 66 showed positive results, in E protein 28 out of 82 were positive, and 28 out of 82 in modified E protein were positive.

Both E & modified E protein showed 28 amino acid that's crossed the threshold; 0.5 with same residue position like: F → 33; L → 58, 50, 39, 51, 28, 56, 2; Q → 70; R → 63; Y → 59 and 66; V → 67, 65, 41, 21, 22, 52, 29; except: V → 82 in E protein while it's at position 10 in modified E protein, L → 76 in E protein while at position 34 and 6 in modified E protein, F → 69 in E protein while it's at positions 17 and 19 in modified E protein, W → 81 in E while it's at position 11 in modified E protein, R → 38 in E, I → 18 in E, K → 68 and 73 in E while A → 32 in modified E protein with M → 60,Y → 57 in E protein.

Table 3
Illustrate the positive selected peptide sequences for both S and modified S glycoprotein sequence by NetMHCpan prediction tool

S	Modified S
AFYCILEPR[a]	AFYCILEPR[a]
ASLNSFKEY[a,b]	ASLNSFKEY[a,b]
ATDCSDGNY[a,b]	ATDCSDGNY[a,b]
AYQNLVGYY[a,b]	AYQNLVGYY[a,b]
ALALCVFFI[a]	AAIPFAQSI
CGTLLRAFY[a]	ALGAMQTGF
CTFMYTYNI[a,b]	AVNNNAQAL[b]
CYSSLILDY[a]	ALALCVFFI[a]
CMGKLKCNR[a,b]	CGTLLRAFY[a]
DAYQNLVGY[a,b]	CTFMYTYNI[a,b]
ESFDVESGV	CYSSLILDY[a]
EMRLASIAF[a]	CMGKLKCNR[a,b]
ETKTHATLF[a]	DLSQLHCSY
ESAALSAQL[a]	DAYQNLVGY[a,b]
FANGFVVRI [b]	ETKTHATLF[a]
FLLTPTESY[a]	EMRLASIAF[a]
FFNHTLVLL[a,b]	EAAYTSSLL
FSDGKMGRF[a]	ESAALSAQL[a]
FSSRYVDLY[a]	FLLTPTSSY[a]
FQFATLPVY	FFNHTLVLL[a,b]
FSVDGYIRR	FSDGKMGRF[a]
FYVYKLQPL[a]	FSSRYVDLY[a]
FSNPTCLIL[a,b]	FTNCNYNLT[b]
FQNCTAVGV[a,b]	FYVYKLQPL[a]
FSFGVTQEY[a]	FSNPTCLIL[a,b]
FVVNAPNGL [b]	FQNCTAVGV[a,b]
FQDELDEFF[a]	FVYDAYQNL[b]
GVHLFSSRY[a]	FSFGVTQEY[a]
GLVNSSLFV[a,b]	FAQSIFYRL
GYYSDDGNY[a,b]	FQDELDEFF[a]
GLYFMHVGY[a]	GVHLFSSRY[a]

(continued)

Table 3
(continued)

S	Modified S
GQGTHIVSF	GVRQQRFVY
GRLTTLNAF[a,b]	GYYSDDGNY[a,b]
HSVFLLMFL	GLVNSSLFV[a,b]
HISSTMSQY[a]	GWTAGLSSF
IEVDIQQTF[a]	GRLTTLNAF[a,b]
IIYPQGRTY[c]	GLYFMHVGY[a]
ITITYQGLF	HISSTMSQY[a]
ITYQGLFPY[a]	IEVDIQQTF[a]
ITEDEILEW[a]	IIYPQTRTY[c]
IASNCYSSL[a,b]	ITYQGLFPY[a]
ILATVPHNL[a,b]	ITEDEILEW[a]
ILDYFSYPL[a]	IASNCYSSL[a,b]
ITKPLKYSY[a]	ILATVPHNL[a]
IAFNHPIQV[a,b]	ILDYFSYPL[a]
IEVVSAYGL[a]	ITKPLKYSY[a]
IAGLVALAL[a]	IAFNHPIQV[a,b]
KQFANGFVV[a,b]	ICAQYVAGY
KAWAAFYVY[a]	IPFAQSIFY
KLQPLTFLL[c]	IANKFNQAL [b]
KETKTHATL[a]	IEVVSAYGL1
KVTIADPGY[a]	IPNFGSLTF [b]
KVTVDCKQY[a]	IAGLVALAL[a]
KELGNYTYY[a,b]	KQFDNGFVV[a,b]
KYVAPQVTY[a]	KAWAAFYVY[a]
LLRAFYCIL[a]	KLQPLTFLW[c]
LLDFSVDGY	KETKTHATL[a]
LPVYDTIKY[a]	KVTVDCKQY[a]
LYGGNMFQF[b]	KVTIADPGY[a]
LSGTPPQVY[a]	KYVAPQVTY[a]
LSLFSVNDF [b]	KELGNYTYY[a,b]
LSIPTNFSF[a,b]	LLRAFYCIL[a]

(continued)

Table 3
(continued)

S	Modified S
LQMGFGITV[a]	LPVYDTIKY[a]
LINGRLTTL[a,b]	LSGTPPQVY[a]
LVRSESAAL[a]	LTFLWDFSV
LYFMHVGYY[a]	LQMGFGITV[a]
LVALALCVF[a]	LSIPTNFSF[a,b]
MGRFFNHTL[a,b]	LGSIAGVGW
MLGSSVGNF[a,b]	LSSFAAIPF
MGFGITVQY[a]	LASELSNTF [b]
MTEQLQMGF[a]	LINGRLTTL[a,b]
MLKRRDSTY	LVRSESAAL[a]
MSQYSRSTR[a]	LTFINTTLL[b]
NLRNCTFMY[a,b]	LYFMHVGYY[a]
NSYTSFATY[a,b]	LVALALCVF[a]
NSVCPKLEF[a,b]	MGRFFNHTL[a,b]
NHIEVVSAY[a,b]	MLGSSVGNF[a,b]
NTTLLDLTY [b]	MGFGITVQY[a]
PVYDTIKYY	MSQYSRSTR[a]
QFANGFVVR [b]	MTEQLQMGF[a]
QTAQGVHLF[a]	MEAAYTSSL
QPLTFLLDF[c]	NLRNCTFMY[a,b]
QSFSNPTCL1 [b]	NSYTSFATY[a,b]
QALHGANLR [b]	NSVCIKLEF[a,b]
QSSPIIPGF[a]	NHIEVVSAY[a,b]
RFFNHTLVL[a,b]	QTAQGVHLF[a]
RNCTFMYTY[a]	QLHCSYESF
RLVFTNCNY[a,b]	QPLTFLWDF[c]
RSTRSMLKR[a]	QSFSNPTCL[a,b]
RSAIEDLLF[a]	QQRFVYDAY
SVFLLMFLL	QVDQLNSSY [b]
SFKEYFNLR[a,b]	QSSPIIPGF[a]
SLNSFKEYF[a,b]	RFFNHTLVL[a,b]

(continued)

Table 3
(continued)

S	Modified S
SFDVESGVY[a]	RNCTFMYTY[a,b]
SGVYSVSSF[a]	RLVFTNCNY[a,b]
SLILDYFSY[a]	RSTRSMLKR[a]
SQFNYKQSF[a,b]	RSAIEDLLF[a]
SSAGPISQF[a]	SFKEYFNLR[a,b]
SPLEGGGWL[a]	SLNSFKEYF[a,b]
SQLGNCVEY[a,b]	SFDVESGVY[a]
STVAMTEQL	SGVYSVSSF[a]
STVWEDGDY[a]	SLILDYFSY[a]
SYINKCSRL[a,b]	SPLEGGGWL[a]
SSTMSQYSR[a]	SQFNYKQSF[a,b]
STLTPRSVR[a]	SSAGPISQF[a]
STRSMLKRR[a]	STVWEDGDY[a]
SVRNLFASV[a,b]	SYINKCSRL[a,b]
TFFDKTWPR[a]	SSTMSQYSR[a]
TYSNITITY[a,b]	STRSMLKRR[a]
TAVGVRQQR[a]	SQLGNCVEY[a,b]
TVWEDGDYY[a]	STLTPRSVR[a]
TLLDLTYEM	SLLGSIAGV
TSIPNFGSL[a,b]	SVRNLFASV[a,b]
TYQNISTNL[a,b]	TFFDKTWPR[a]
TYYNKWPWY[a,b]	TYSNITITY[a,b]
VSKADGIIY[a]	TTITKPLKY
VYKLQPLTF[a]	TVWEDGDYY[a]
VECDFSPLL[a]	TAVGVRQQR[a]
VYNFKRLVF[a,b]	TTNEAFQKV[b]
VASGSTVAM	TSIPNFGSL[a,b]
VSIVPSTVW[a]	TYQNISTNL[a,b]
VSVPVSVIY[a]	TYYHKWPWY[a]
VNAPNGLYF[a,b]	VSKADGIIY[a]
VVNAPNGLY[a,b]	VECDFSPLL[a]

(continued)

Table 3
(continued)

S	Modified S
VALALCVFF[a]	VYKLQPLTF[a]
VVKALNESY[a,b]	VYNFKRLVF[a,b]
WPWYIWLGF[a]	VSIVPSTVW[a]
WAAFYVYKL[a]	VSVPVSVIY[a]
YQGDHGDMY[c]	VNAPNGLYF[a,b]
YFNLRNCTF[a,b]	VVNAPNGLY[a,b]
YYSIIPHSI[a]	VALALCVFF[a]
YSIIPHSIR[a]	VVKALNESY[a,b]
YNLTKLLSL[a,b]	WPWYIWLGF[a]
YPLSMKSDL[a]	WSYTGSSFY
YSSLILDYF[a]	WTAGLSSFA
YGVSGRGVF[a]	WAAFYVYKL[a]
YINKCSRLL[a]	YQGDHGDYY[c]
YSLYGVSGR[a]	YFNLRNCTF[a,b]
YSYINKCSR[a,b]	YNLTKLLSL[a,b]
YYRKQLSPL[a]	YSIIPHSIR[a]
YSRSTRSML[a]	YYSIIPHSI[a]
YYSDDGNYY[a,b]	YINKCSRLL[a,b]
YYPSNHIEV[a,b]	YPLSMKSDL[a]
YAPEPITSL[a]	YSSLILDYF[a]
YTYYNKWPW[b,c]	YSYINKCSR[a,b]
YYNKWPWYI[b,c]	YYRKQLSPL[a] YGVSGRGVF[a] YSLYGVSGR[a] YSRSTRSML[a] YYSDDGNYY[a,b] YAPEPITSL[a] YYPSNHIEV[a,b] YTYYHKWPW[c] YYHKWPWYI[c]

[a]Indicates a common peptide sequence
[b]Indicates presence of arginine in sequence
[c]Indicates a partial similarity between both reference sequence and modified sequence

Table 4
Illustrate the positive selected peptide sequences for both E and modified E protein by NetMHCpan prediction tool

E	Modified E
ALYLYNTGR [a]	KPPLPEDVW
CMAFLTATR FTVVCAITL FVQERIGLF ITLLVCMAF LFIVNFFIF [a] LVQPALYLY LYNTGRSVY [a] MAFLTATRL RIGLFIVNF [a] TLLVQPALY	

[a]Indicates presence of arginine in sequence

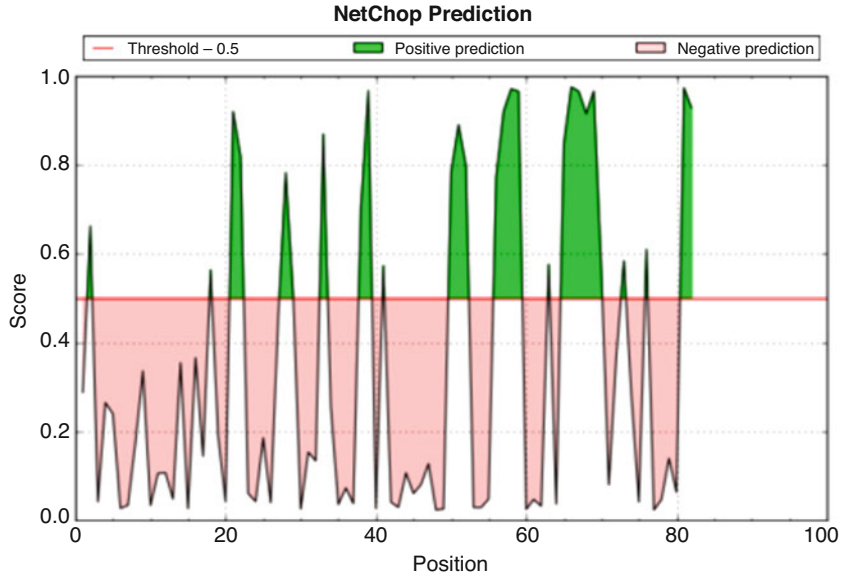

Fig. 25 Illustrate the NetChop positive prediction of E protein with threshold equal or greater than 0.5

N.B:-.

1. Peptide sequences of both E and modified E protein were different even if they had a similar residue position.
2. NetCTL was used for E and modified E protein just due to large amounts of data beside, time-consuming when it is used with S glycoprotein.
3. Modified E protein NetCTL charts were not shown here.

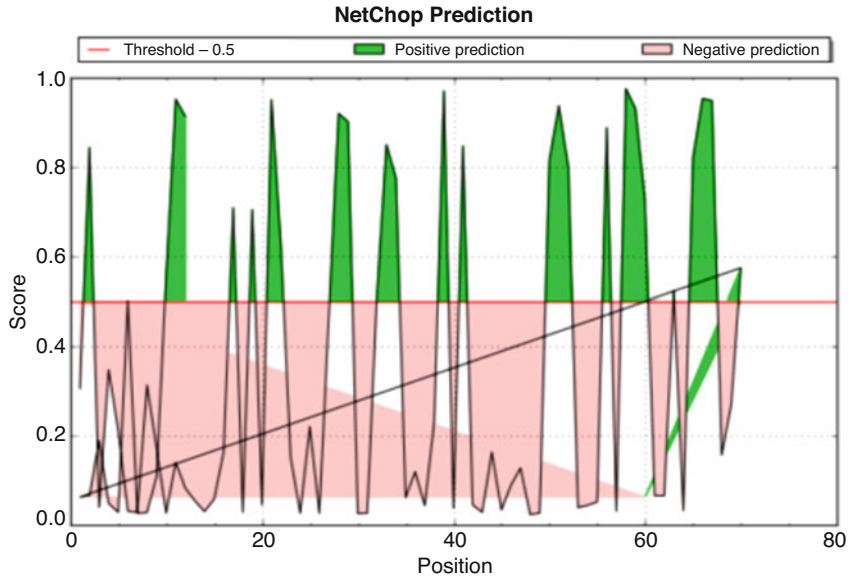

Fig. 26 Illustrate the NetChop positive prediction of modified E protein threshold equal or greater than 0.5

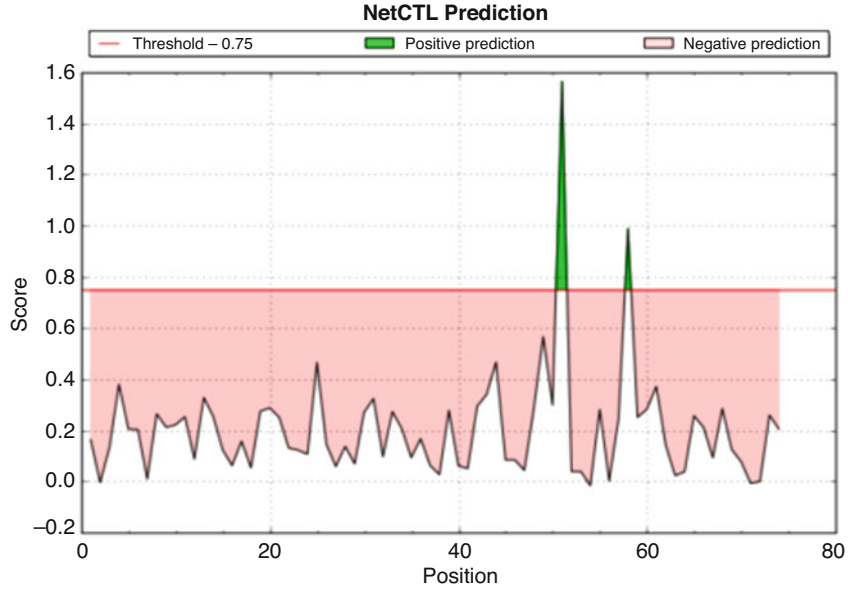

Fig. 27 Illustrate the NetCTL positive prediction of E protein supertype A1 that's indicated in a green color with threshold equal or greater than 0.75 above the red color

3.2.5 MHC-NP: Prediction of Peptides Naturally Processed by the MHC

The greater probe score was considered as naturally processing peptide; probe scores greater than 0 were considered as naturally processing peptides.

The total positive epitope number of naturally processing peptides represented 10,189 out of 10,760 in S glycoprotein and

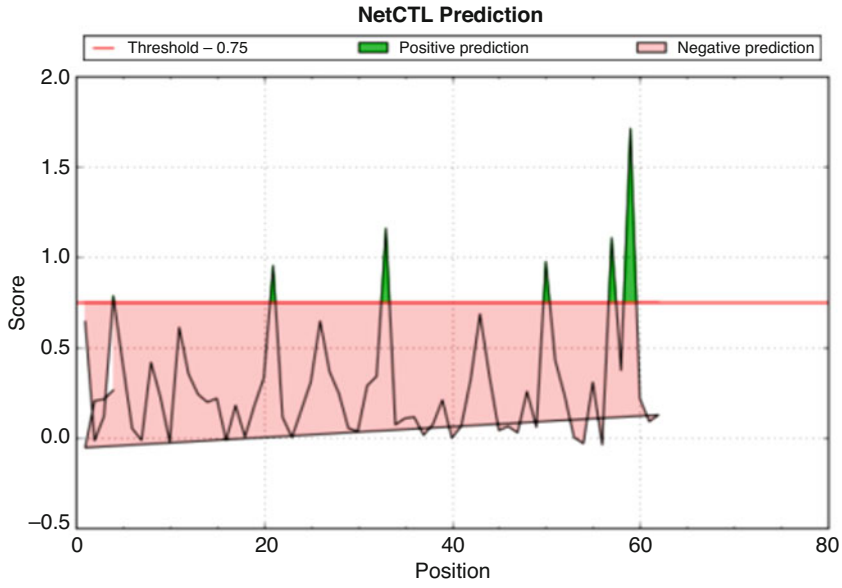

Fig. 28 Illustrate the NetCTL prediction of E protein supertype A2, the desired supertype A2 appeared in a green color with threshold equal or greater than 0.75 above the threshold red color

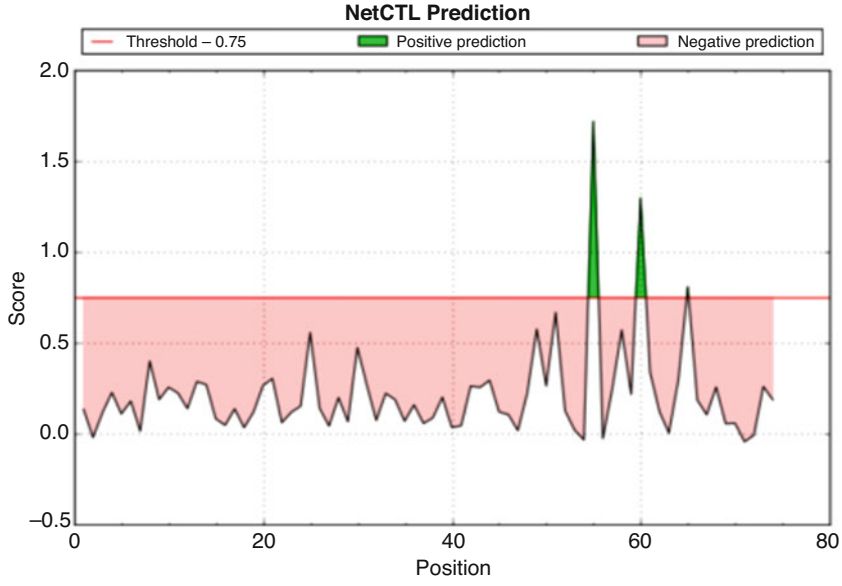

Fig. 29 Illustrate the NetCTL prediction of E protein supertype A3, the positive results appeared in a green color with threshold equal or greater than 0.75 above the red color

10,187 out of 10,760 in modified S glycoprotein, while it represents 568 out of 592 in E and 566 out of 592 in modified E protein.

E protein showed alleles frequencies: H-2-Db (74), H-2-Kb (74), HLA-A∗02:01 (68), HLA-B∗07:02 (66), HLA-B∗35:01

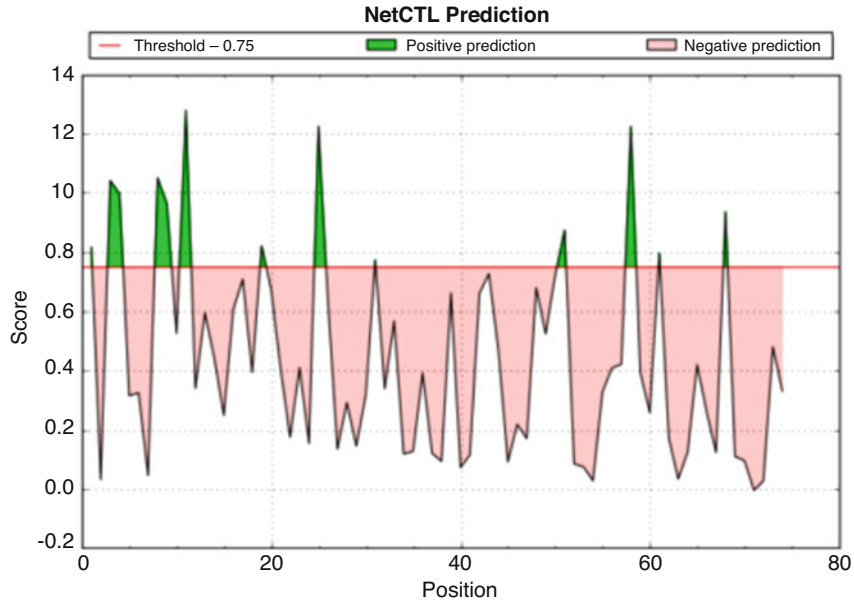

Fig. 30 Illustrate the NetCTL prediction of E protein supertype A24, positive results appeared in a green color with threshold equal or greater than 0.75 above the threshold red color

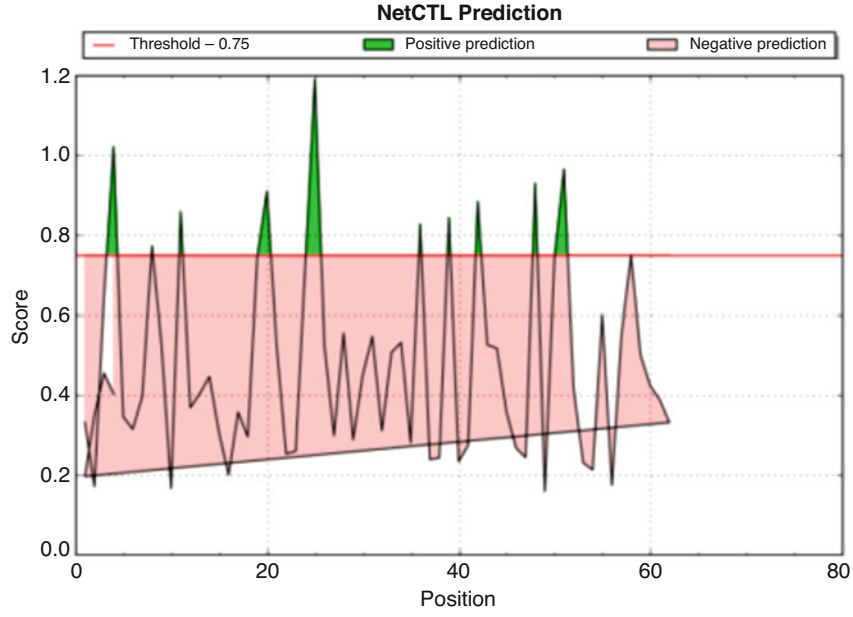

Fig. 31 Illustrate the NetCTL prediction of E protein supertype A26, positive results appeared in a green color with threshold equal or greater than 0.75 above the threshold red color

(74), HLA-B∗44:03 (74), HLA-B∗53:01 (73), HLA-B∗57:01 (62) while in modified E they are H-2-Db (28), H-2-Kb (16), HLA-A∗02:01 (5), HLA-B∗07:02 (2), HLA-B∗35:01 (6), HLA-B∗44:03 (28), HLA-B∗53:01 (60), and HLA-B∗57:01 (4).

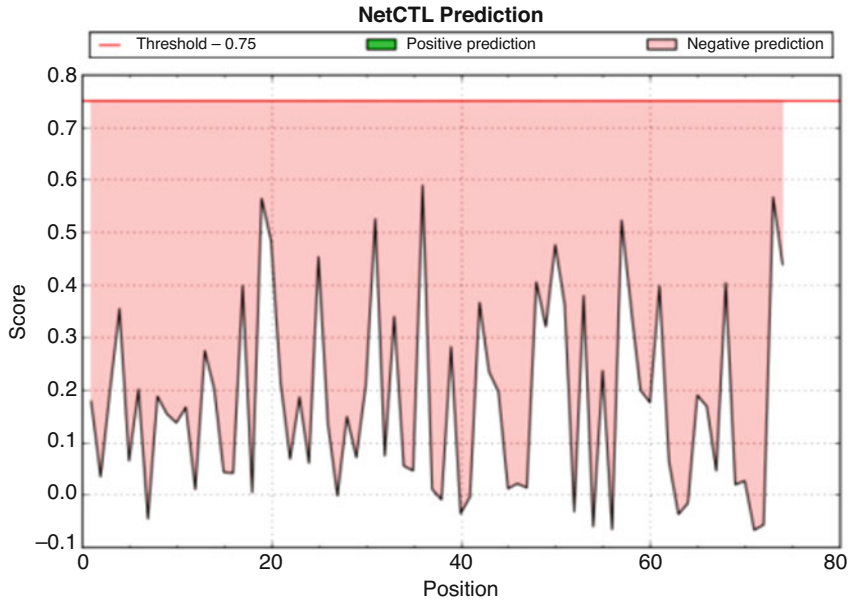

Fig. 32 Illustrate the NetCTL negative prediction of E protein supertype B7 with threshold below 0.75

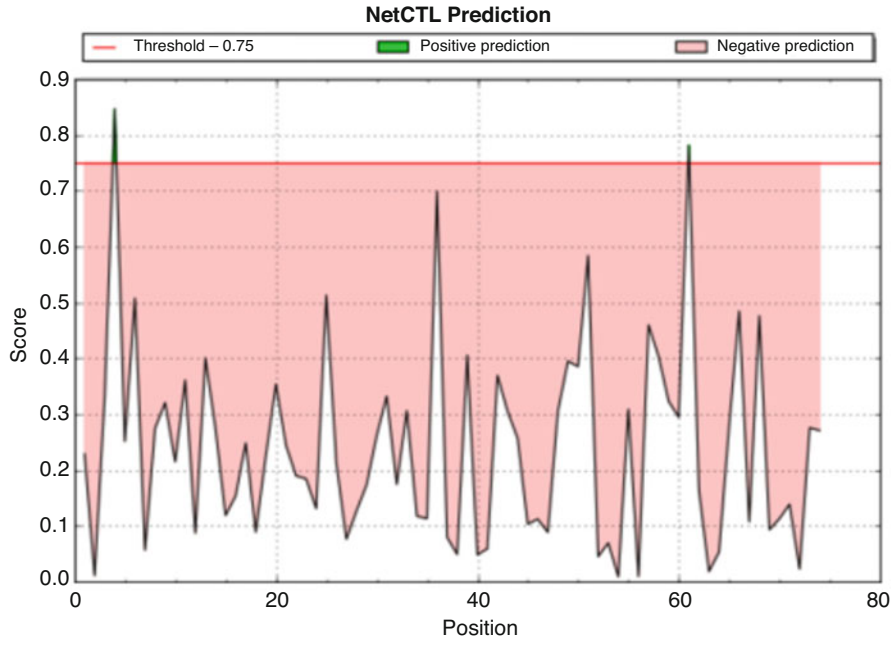

Fig. 33 Illustrate the NetCTL negative prediction of E protein supertype B8 with threshold below 0.75

N.B: modified E protein showed less allele frequency when compared with E protein in addition to some epitope differences even if at the same positions.

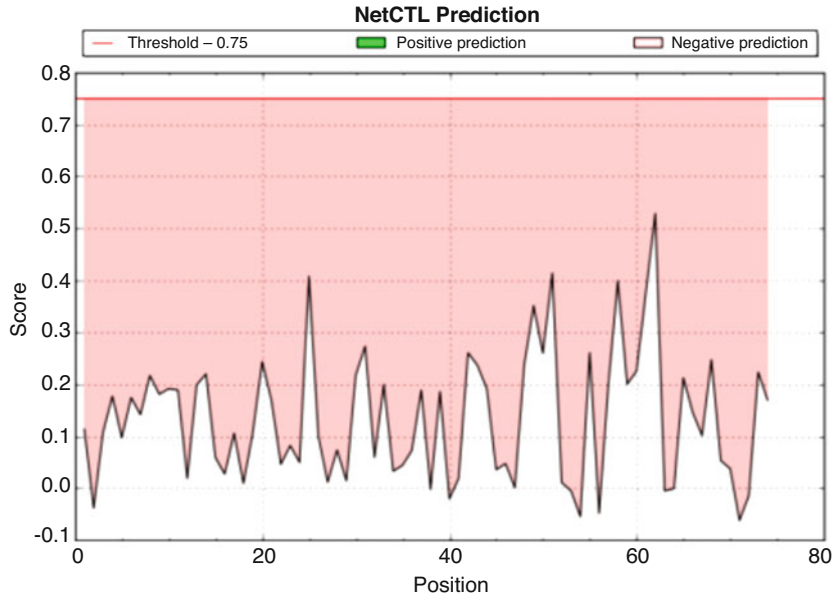

Fig. 34 Illustrate the NetCTL negative prediction of E protein supertype B27

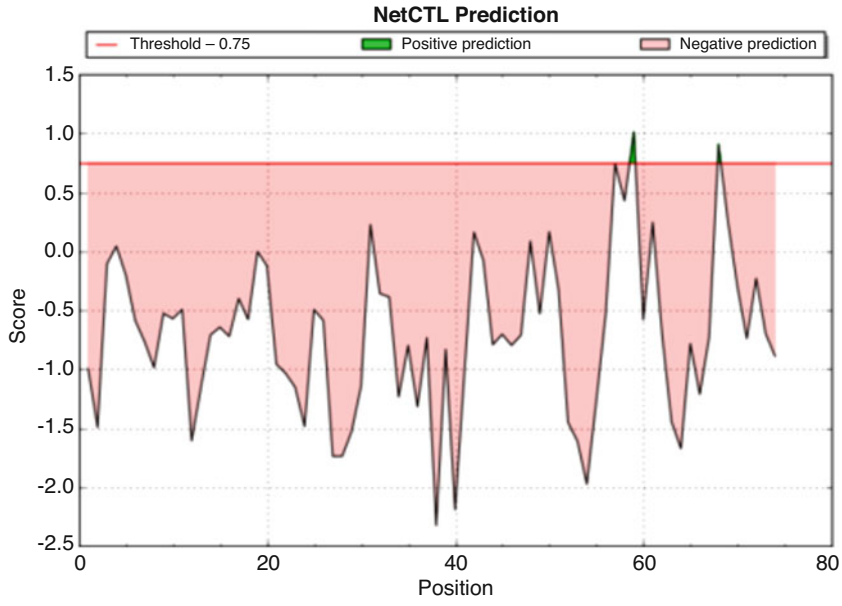

Fig. 35 Illustrate the NetCTL negative prediction of E protein supertype B39 with threshold below 0.75

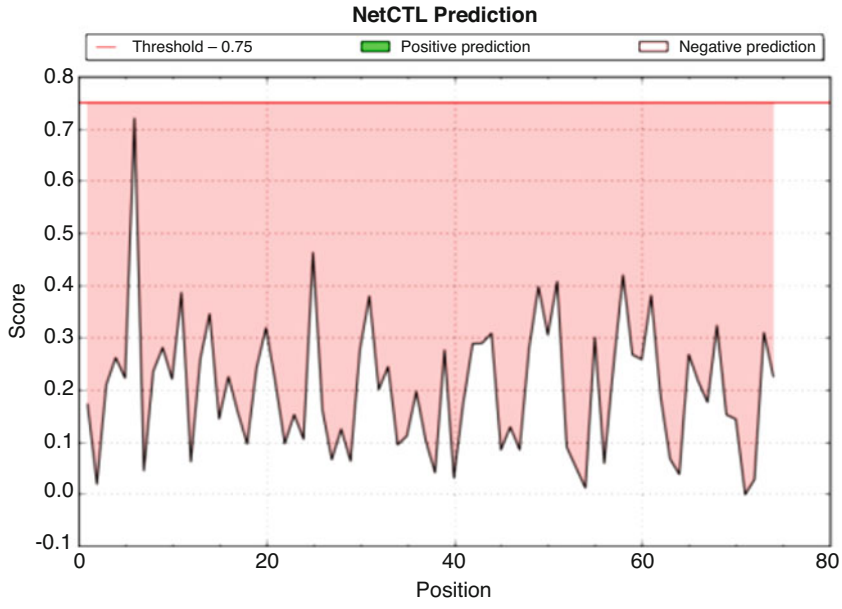

Fig. 36 Illustrate the NetCTL negative prediction of E protein supertype B44 with threshold below 0.75

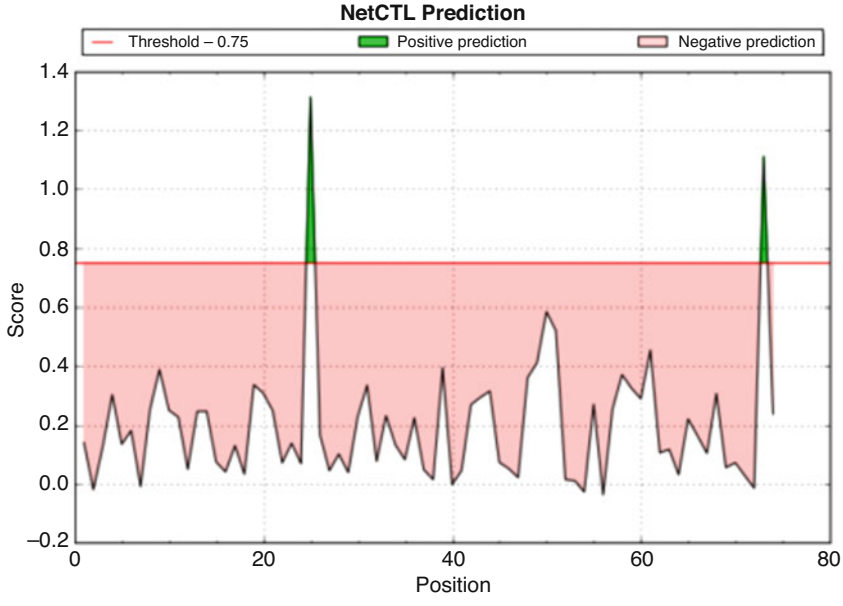

Fig. 37 Illustrate the NetCTL prediction of E protein supertype B58, positive results appeared in a green colored with threshold equal or greater than 0.75 above the threshold red color

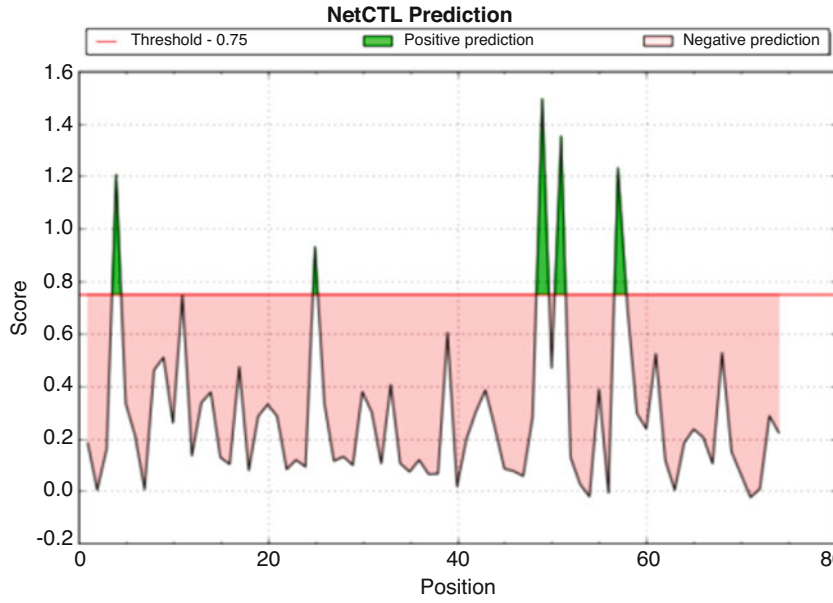

Fig. 38 Illustrate the NetCTL prediction of E protein supertype B62, positive results appeared in a green colored with threshold equal or greater than 0.75 above the threshold red color

3.3 Epitope Analysis Tools

3.3.1 Population Coverage Calculation

MHC-I and MHC-II interacted alleles by the IEDB population coverage calculation tool was computed by the average number of epitope hits/HLA combinations recognized by the population and a minimum number of epitope hits/HLA combinations recognized by 90% of the population (PC90); see tables below.

Those below represented a selected E protein epitopes for population coverage calculation:

PFVQER, VQERIG, QERIGL, FLTATR, LYLYNT, YLYNTG, LYNTGR, YNTGRS, NTGRSV, TGRSVY, RSVYVK, YVKFQD, VKFQDS, KFQDSK, FQDSKP, QDSKPP, DSKPPL, SKPPLP, KPPLPP, PPLPPD, PLPPDE, LPPDEW, PPDEWV, MLPFVQE, LPFVQER, PFVQERI, VQERIGL, RIGLFIV, IGLFIVN, GLFIVNF, LFIVNFF, FIVNFFI, IVNFFIF, and VNFFIFT.

There are differences between MHC-I and MHC-II population coverage percentage.

There are similarities between MHC-I between both E and modified E protein, but still there are differences between them at MHC-II.

Those below represented a selected modified E protein epitopes for population coverage calculation:

RSVYVP, LYMTGR, VYVPQQ, PLPEDV, QERIGW, TGRSVY, YMTGRS, QFVQER, VPQQDS, SKPPLP, PPLPED, DSKPPL, YVPQQD, KPPLPE, QDSKPP, PQQDSK, QQDSKP, PLPEDVW, QFVQERI, AFLTATH, MLQFVQE, ALSLYMT,

Table 5
Illustrate NetCTL +ve results in E and modified E protein with indications of similarities and differences in the peptide sequences between them, beside the totals numbers of them

Supertype	Peptide sequence for E protein	Peptide sequence for modified E protein	Residue position for E/modified E protein
A1	LVQPALYLY LYNTGRSVY	LVQPALSLY	51/51 58/58
A2	FVQERIGWF VVCDITLLV FLTATHLCV LLVQPALSL SLYMTGRSV YMTGRSVYV	FVQERIGWF VVCDITLLV FLTATHLCV LLVQPALSL SLYMTGRSV YMTGRSVYV	4/4 21/21 33/33 50/50 57/57 59/59
A3	ALYLYNTGR NTGRSVYVK VYVKFQDSK	ALSLYMTGR	55/55 60/– 65/–
A24	MLPFVQERI PFVQERIGL FVQERIGLF RIGLFIVNF IGLFIVNFF LFIVNFFIF FTVVCAITL ITLLVCMAF MAFLTATRL LVQPALYLY LYNTGRSVY TGRSVYVKF KFQDSKPPL	MLQFVQERI FVQERIGWF RIGWFIPNF WFIPNFFDF FTVVCDITL ITLLVCTAF LVQPALSLY LYMTGRSVY	1/1 3/4 4/8 8/11 9/19 11/25 19/51 25/58 31/– 51/– 58/– 61/– 68/–
A26	FVQERIGWF RIGWFIPNF WFIPNFFDF TVVCDITLL ITLLVCTAF ATHLCVQCM LCVQCMTGF QCMTGFNTL NTLLVQPAL LVQPALSLY	FVQERIGWF RIGWFIPNF WFIPNFFDF TVVCDITLL ITLLVCTAF ATHLCVQCM LCVQCMTGF QCMTGFNTL NTLLVQPAL LVQPALSLY	4/4 8/8 11/11 20/20 25/25 36/36 39/39 42/42 48/48 51/51
B7	–	LLVQPALSL QPALSLYMT KPPLPEDVW	–/50 –/53 –/3
B8	FVQERIGLF TGRSVYVKF	FVQERIGWF WFIPNFFDF	4/4 61/11
B27	–	–	–
B39	YNTGRSVYV KFQDSKPPL	YMTGRSVYV	59/59 68

(continued)

Table 5 (continued)

Supertype	Peptide sequence for E protein	Peptide sequence for modified E protein	Residue position for E/modified E protein
B44	–	–	–
B58	ITLLVCMAF KPPLPPDEW	IGWFIPNFF ITLLVCTAF KPPLPEDVW	25/9 73/25 −/3
B62	FVQERIGLF ITLLVCMAF TLLVQPALY LVQPALYLY YLYNTGRSV	FVQERIGWF WFIPNFFDF ITLLVCTAF LVQPALSLY LYMTGRSVY	4/4 25/11 49/25 51/51 57/58

LQFVQER, VQCMTGF, YVPQQDS, GFNTLLV, PPLPEDV, FLTATHL, TGRSVYV, PALSLYM, NTLLVQP, FNTLLVQ, LPEDVWV, and CTAFLTA.

The percentage of a coverage population was similar among both S glycoprotein reference sequence and modified S glycoprotein; it represented 95.60% of the world by MHC-I; 118 countries showed a higher percentage especially Chile Amerindian (100%), 69 other countries showed 0% while in East Asia (94.80%), South Korea and South Oriental Korea (92.84%), China (88.77%), Iran and Iran Persian (91.53%) but Iran Kurd (0.00%), Jordan and Jordan Arab (76.80%),Oman and Oman Arab (95.82%), Saudi Arabia and Saudi Arabia Arab (96.38%), United Arab Emirates and United Arab Emirates Arab (0.00%), Sudan (86.43%), Sudan Arab (49.41%), Sudan Black (0.00%), and Sudan Mixed (87.06%); please *see* Table 6.

According to the percentage of a coverage population that was similar between S glycoprotein reference sequence and modified S glycoprotein, the world MHC-II represent 81.81%; 64 countries showed a higher percentage especially Norway and Norway Caucasoid (94.71%), 59 other countries (0%) while in East Asia represents (94.80%), South Korea and South Oriental Korea (85.32%), China (59.99%), Iran (64.22%), Iran Persian (55.78%), Iran Kurd (65.72%), Jordan and Jordan Arab (52.88%), Oman and Oman Arab (0.00%), Saudi Arabia and Saudi Arabia Arab (80.14%), United Arab Emirates and United Arab Emirates Arab (32.92%), Sudan (60.56%), Sudan Arab (0.00%), Sudan Black (0.00%), and Sudan Mixed (60.56%), as in Table 7.

According to the percentage of MHC-I E protein coverage, the world MHC-I represents 95.60%; 116 countries showed a higher percentage especially Chile Amerindian (100%), 23 other countries showed more than 4% but less than 50% while in East Asia it

Table 6
MHC-I coverage population for S and modified S glycoprotein

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
World	95.60%	10.57	4.38
East Asia	94.80%	10.93	2.58
Japan	96.19%	11.44	3.12
Japan Oriental	96.19%	11.44	3.12
Korea, South	92.84%	10.41	2.16
Korea, South Oriental	92.84%	10.41	2.16
Mongolia	94.37%	10.07	3.12
Mongolia Oriental	94.37%	10.07	3.12
Northeast Asia	88.80%	9.38	0.89
China	88.77%	9.33	0.89
China Oriental	88.77%	9.33	0.89
Hong Kong	90.85%	10.01	1.91
Hong Kong Oriental	90.85%	10.01	1.91
South Asia	86.54%	8.03	0.74
India	82.00%	7.21	0.56
India Asian	82.00%	7.21	0.56
Pakistan	88.63%	8.74	1.76
Pakistan Asian	87.30%	8.38	1.58
Pakistan Mixed	91.12%	9.42	3.23
Sri Lanka	52.39%	3.74	0.84
Sri Lanka Asian	52.39%	3.74	0.84
Southeast Asia	87.81%	9.99	0.82
Borneo	0.00%	0	?
Borneo Austronesian	0.00%	0	?
Indonesia	76.44%	7.8	0.42
Indonesia Austronesian	76.44%	7.8	0.42
Malaysia	76.30%	7.64	0.42
Malaysia Austronesian	40.59%	3.17	0.34
Malaysia Oriental	84.44%	9.02	0.64
Philippines	92.86%	11.56	8.01

(continued)

Table 6
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Philippines Austronesian	92.86%	11.56	8.01
Singapore	85.74%	9.04	0.7
Singapore Austronesian	82.82%	8.55	0.58
Singapore Oriental	88.96%	9.64	0.91
Taiwan	92.58%	11.31	6.08
Taiwan Oriental	92.58%	11.31	6.08
Thailand	82.85%	7.46	0.58
Thailand Oriental	82.85%	7.46	0.58
Vietnam	84.58%	8.55	0.65
Vietnam Oriental	84.58%	8.55	0.65
Southwest Asia	85.77%	7.59	0.7
Iran	91.53%	8.6	1.33
Iran Kurd	0.00%	0	?
Iran Persian	91.53%	8.6	1.33
Israel	82.14%	7.29	0.56
Israel Arab	89.15%	9.13	0.92
Israel Jew	87.17%	7.84	0.78
Jordan	76.80%	6.52	0.43
Jordan Arab	76.80%	6.52	0.43
Lebanon	0.00%	0	0
Lebanon Arab	0.00%	0	?
Lebanon Mixed	0.00%	0	0
Oman	95.82%	9.96	3.04
Oman Arab	95.82%	9.96	3.04
Saudi Arabia	96.38%	9.87	3.65
Saudi Arabia Arab	96.38%	9.87	3.65
United Arab Emirates	0.00%	0	0
United Arab Emirates Arab	0.00%	0	0
Europe	97.81%	11.07	5.29
Austria	98.78%	11.29	6

(continued)

Table 6
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Austria Caucasoid	98.78%	11.29	6
Belarus	0.00%	0	?
Belarus Caucasoid	0.00%	0	?
Belgium	98.75%	10.62	6.02
Belgium Caucasoid	98.75%	10.62	6.02
Bulgaria	96.59%	11.08	4.52
Bulgaria Caucasoid	96.56%	11.25	4.57
Bulgaria Other	97.43%	10.02	4.35
Croatia	97.76%	11.79	6.12
Croatia Caucasoid	97.76%	11.79	6.12
Czech Republic	96.20%	9.39	4.33
Czech Republic Caucasoid	96.20%	9.39	4.33
Czech Republic Other	0.00%	0	?
Denmark	0.00%	0	0
Denmark Caucasoid	0.00%	0	0
England	99.29%	11.43	6.21
England Caucasoid	99.29%	11.43	6.21
England Jew	0.00%	0	0
England Mixed	0.00%	0	?
Finland	99.80%	12.56	7.8
Finland Caucasoid	99.80%	12.56	7.8
France	98.05%	10.72	4.75
France Caucasoid	98.05%	10.72	4.75
Georgia	95.62%	10.98	4.48
Georgia Caucasoid	97.22%	11.66	6.21
Georgia Kurd	89.99%	9.26	1
Germany	99.07%	11.71	6.4
Germany Caucasoid	99.07%	11.71	6.4
Greece	0.00%	0	?
Greece Caucasoid	0.00%	0	?

(continued)

Table 6
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Ireland Northern	99.40%	11.43	6.27
Ireland Northern Caucasoid	99.40%	11.43	6.27
Ireland South	98.83%	10.82	4.85
Ireland South Caucasoid	98.83%	10.82	4.85
Italy	96.52%	9.83	4.16
Italy Caucasoid	96.52%	9.83	4.16
Macedonia	11.83%	0.86	0.45
Macedonia Caucasoid	11.83%	0.86	0.45
Netherlands	0.00%	0	?
Netherlands Caucasoid	0.00%	0	?
Norway	0.00%	0	?
Norway Caucasoid	0.00%	0	?
Poland	97.99%	11.25	6.02
Poland Caucasoid	97.99%	11.25	6.02
Portugal	97.11%	10.98	4.73
Portugal Caucasoid	97.11%	10.98	4.73
Romania	97.94%	11.56	5.94
Romania Caucasoid	97.94%	11.56	5.94
Russia	96.71%	11.38	4.59
Russia Caucasoid	0.00%	0	0
Russia Mixed	0.00%	0	0
Russia Other	98.34%	12.46	6.71
Russia Siberian	97.30%	11.52	4.53
Scotland	15.91%	0.81	0.24
Scotland Caucasoid	15.91%	0.81	0.24
Serbia	43.75%	0.78	0.18
Serbia Caucasoid	43.75%	0.78	0.18
Slovakia	0.00%	0	?
Slovakia Caucasoid	0.00%	0	?
Slovenia	0.00%	0	?

(continued)

Table 6
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Slovenia Caucasoid	0.00%	0	?
Spain	71.85%	5.51	0.36
Spain Caucasoid	71.85%	5.51	0.36
Spain Jew	0.00%	0	?
Spain Other	0.00%	0	?
Sweden	99.69%	12.61	6.84
Sweden Caucasoid	99.69%	12.61	6.84
Switzerland	0.00%	0	0
Switzerland Caucasoid	0.00%	0	0
Turkey	44.80%	3.58	1.45
Turkey Caucasoid	44.80%	3.58	1.45
Ukraine	0.00%	0	?
Ukraine Caucasoid	0.00%	0	?
United Kingdom	0.00%	0	0
United Kingdom Caucasoid	0.00%	0	0
Wales	0.00%	0	0
Wales Caucasoid	0.00%	0	0
East Africa	86.99%	6.96	0.77
Kenya	85.86%	6.62	0.71
Kenya Black	85.86%	6.62	0.71
Uganda	91.04%	8.19	1.48
Uganda Black	91.04%	8.19	1.48
Zambia	95.32%	7.98	4.01
Zambia Black	95.32%	7.98	4.01
Zimbabwe	91.57%	7.69	1.71
Zimbabwe Black	91.57%	7.69	1.71
West Africa	92.60%	8.71	1.67
Burkina Faso	58.50%	3.24	0.24
Burkina Faso Black	58.50%	3.24	0.24
Cape Verde	96.69%	10.09	4.14

(continued)

Table 6
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Cape Verde Black	96.69%	10.09	4.14
Gambia	0.00%	0	?
Gambia Black	0.00%	0	?
Ghana	0.00%	0	0
Ghana Black	0.00%	0	0
Guinea-Bissau	92.66%	8.7	1.49
Guinea-Bissau Black	92.66%	8.7	1.49
Ivory Coast	58.05%	0.78	0.24
Ivory Coast Black	58.05%	0.78	0.24
Liberia	0.00%	0	?
Liberia Black	0.00%	0	?
Nigeria	0.00%	0	?
Nigeria Black	0.00%	0	?
Senegal	95.03%	9.11	4
Senegal Black	95.03%	9.11	4
Central Africa	84.98%	6.7	0.67
Cameroon	88.67%	7.35	0.88
Cameroon Black	88.67%	7.35	0.88
Central African Republic	10.75%	0.27	0.11
Central African Republic Black	10.75%	0.27	0.11
Congo	0.00%	0	?
Congo Black	0.00%	0	?
Equatorial Guinea	0.00%	0	0
Equatorial Guinea Black	0.00%	0	0
Gabon	0.00%	0	?
Gabon Black	0.00%	0	?
Rwanda	23.09%	1.33	0.13
Rwanda Black	23.09%	1.33	0.13
Sao Tome and Principe	95.54%	8.72	2.29
Sao Tome and Principe Black	95.54%	8.72	2.29

(continued)

Table 6
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
North Africa	91.87%	8.61	1.86
Algeria	0.00%	0	?
Algeria Arab	0.00%	0	?
Ethiopia	0.00%	0	?
Ethiopia Black	0.00%	0	?
Mali	94.28%	8.82	1.74
Mali Black	94.28%	8.82	1.74
Morocco	95.95%	9.47	4.19
Morocco Arab	97.89%	10.2	4.47
Morocco Caucasoid	94.32%	8.96	4.02
Sudan	86.43%	7.53	0.74
Sudan Arab	49.41%	4.62	0.59
Sudan Black	0.00%	0	0
Sudan Mixed	87.06%	7.56	0.77
Tunisia	96.04%	9.85	4.19
Tunisia Arab	96.04%	9.85	4.19
Tunisia Berber	0.00%	0	?
South Africa	91.05%	8	2.1
South Africa	91.05%	8	2.1
South Africa Black	86.71%	6.67	0.75
South Africa Other	93.82%	9.59	2.73
West Indies	97.34%	10.78	4.6
Cuba	97.20%	10.65	4.53
Cuba Caucasoid	97.64%	11.2	4.77
Cuba Mixed	0.00%	0	?
Cuba Mulatto	96.58%	9.66	4.09
Jamaica	0.00%	0	?
Jamaica Black	0.00%	0	?
Martinique	22.56%	2.03	1.16
Martinique Black	22.56%	2.03	1.16

(continued)

Table 6
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Trinidad and Tobago	0.00%	0	0
Trinidad and Tobago Asian	0.00%	0	0
North America	96.88%	10.98	4.65
Canada	0.00%	0	?
Canada Amerindian	0.00%	0	?
Mexico	97.10%	11	6.02
Mexico Amerindian	99.86%	13	7.84
Mexico Mestizo	96.78%	10.7	4.46
United States	96.93%	10.98	4.66
United States Amerindian	99.44%	13.15	8.19
United States Asian	92.39%	10.32	2.29
United States Austronesian	0.00%	0	?
United States Black	94.18%	8.83	2.54
United States Caucasoid	98.65%	11.4	6.08
United States Hispanic	97.46%	11.01	4.77
United States Mestizo	98.09%	11.2	4.97
United States Polynesian	97.53%	11.57	3.62
Central America	5.10%	0.16	0.11
Costa Rica	0.00%	0	?
Costa Rica Mestizo	0.00%	0	?
Guatemala	5.10%	0.16	0.11
Guatemala Amerindian	5.10%	0.16	0.11
South America	86.24%	8.01	0.73
Argentina	98.02%	8.76	2.61
Argentina Amerindian	98.02%	8.76	2.61
Argentina Caucasoid	0.00%	0	?
Bolivia	0.00%	0	?
Bolivia Amerindian	0.00%	0	?
Brazil	93.72%	9.43	2.69
Brazil Amerindian	92.35%	8.37	2.16

(continued)

Table 6
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Brazil Caucasoid	97.68%	11.33	5.35
Brazil Mixed	95.06%	9.85	3.75
Brazil Mulatto	0.00%	0	?
Brazil Other	0.00%	0	0
Chile	94.93%	10.63	4.37
Chile Amerindian	100.00%	14.31	9.11
Chile Hispanic	0.00%	0	?
Chile Mixed	87.43%	8.16	0.8
Colombia	9.86%	0.76	0.67
Colombia Amerindian	0.00%	0	0
Colombia Black	5.79%	0.42	0.64
Colombia Mestizo	14.81%	1.17	0.7
Ecuador	76.97%	8.77	1.74
Ecuador Amerindian	76.97%	8.77	1.74
Ecuador Black	0.00%	0	?
Paraguay	0.00%	0	?
Paraguay Amerindian	0.00%	0	?
Peru	99.98%	13.69	8.37
Peru Amerindian	99.98%	13.69	8.37
Peru Mestizo	0.00%	0	0
Venezuela	88.37%	9.05	0.86
Venezuela Amerindian	88.88%	8.98	0.9
Venezuela Caucasoid	9.18%	0.83	0.99
Venezuela Mestizo	7.84%	0.71	0.98
Venezuela Mixed	0.00%	0	?
Oceania	91.82%	10.92	4.06
American Samoa	95.26%	12.14	7.15
American Samoa Polynesian	95.26%	12.14	7.15
Australia	89.30%	9.93	0.93
Australia Australian Aborigines	82.36%	9.31	0.57

(continued)

Table 6
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Australia Caucasoid	99.06%	11.46	6.16
Chile	94.93%	10.63	4.37
Chile Amerindian	100.00%	14.31	9.11
Cook Islands	0.00%	0	?
Cook Islands Polynesian	0.00%	0	?
Fiji	0.00%	0	?
Fiji Melanesian	0.00%	0	?
Kiribati	0.00%	0	?
Kiribati Micronesian	0.00%	0	?
Nauru	0.00%	0	?
Nauru Micronesian	0.00%	0	?
New Caledonia	96.70%	12.14	8.63
New Caledonia Melanesian	96.70%	12.14	8.63
New Zealand	0.00%	0	?
New Zealand Polynesian	0.00%	0	?
Niue	0.00%	0	?
Niue Polynesian	0.00%	0	?
Papua New Guinea	97.26%	12.58	8.57
Papua New Guinea Melanesian	97.26%	12.58	8.57
Samoa	0.00%	0	?
Samoa Polynesian	0.00%	0	?
Tokelau	0.00%	0	?
Tokelau Polynesian	0.00%	0	?
Tonga	0.00%	0	?
Tonga Polynesian	0.00%	0	?
Average	55.31%	5.73	?
(Standard deviation)	−44.16%	−4.92	(?)

[a]Projected population coverage
[b]Average number of epitope hits/HLA combinations recognized by the population
[c]Minimum number of epitope hits/HLA combinations recognized by 90% of the population

Table 7
The MHC-II coverage population for S and modified S glycoprotein

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
World	81.81%	8.16	1.1
East Asia	81.82%	8.83	1.1
Japan	74.83%	7.85	0.79
Japan Oriental	74.83%	7.85	0.79
Korea, South	85.32%	9.56	1.36
Korea, South Oriental	85.32%	9.56	1.36
Mongolia	81.85%	7.79	1.1
Mongolia Oriental	81.85%	7.79	1.1
Northeast Asia	59.99%	5.33	0.5
China	59.99%	5.33	0.5
China Oriental	59.99%	5.33	0.5
Hong Kong	0.00%	0	?
Hong Kong Oriental	0.00%	0	?
South Asia	75.38%	7.4	0.81
India	74.99%	7.35	0.8
India Asian	74.99%	7.35	0.8
Pakistan	1.18%	0.09	0.81
Pakistan Asian	1.45%	0.12	0.81
Pakistan Mixed	0.00%	0	0
Sri Lanka	0.00%	0	?
Sri Lanka Asian	0.00%	0	?
Southeast Asia	56.98%	4.98	0.46
Borneo	49.02%	4.03	0.39
Borneo Austronesian	49.02%	4.03	0.39
Indonesia	47.84%	4.4	0.38
Indonesia Austronesian	47.84%	4.4	0.38
Malaysia	57.99%	5.34	0.48
Malaysia Austronesian	55.38%	5.12	0.45
Malaysia Oriental	70.35%	6.57	0.67
Philippines	28.56%	2.52	0.28

(continued)

Table 7
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Philippines Austronesian	28.56%	2.52	0.28
Singapore	65.78%	6.04	0.58
Singapore Austronesian	65.78%	6.04	0.58
Singapore Oriental	0.00%	0	?
Taiwan	67.88%	6.13	0.62
Taiwan Oriental	67.88%	6.13	0.62
Thailand	63.90%	5.92	0.55
Thailand Oriental	63.90%	5.92	0.55
Vietnam	54.44%	4.43	0.44
Vietnam Oriental	54.44%	4.43	0.44
Southwest Asia	43.93%	3.65	0.36
Iran	64.22%	5.65	0.56
Iran Kurd	55.78%	4.74	0.45
Iran Persian	65.72%	5.83	0.58
Israel	68.79%	6.4	0.64
Israel Arab	67.51%	6.2	0.62
Israel Jew	69.65%	6.51	0.66
Jordan	52.88%	4.56	0.42
Jordan Arab	52.88%	4.56	0.42
Lebanon	70.46%	6.48	0.68
Lebanon Arab	70.46%	6.48	0.68
Lebanon Mixed	0.00%	0	?
Oman	0.00%	0	?
Oman Arab	0.00%	0	?
Saudi Arabia	80.14%	8.31	1.01
Saudi Arabia Arab	80.14%	8.31	1.01
United Arab Emirates	32.92%	0.66	0.3
United Arab Emirates Arab	32.92%	0.66	0.3
Europe	85.83%	8.88	1.41
Austria	93.34%	10.8	2.82

(continued)

Table 7
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Austria Caucasoid	93.34%	10.8	2.82
Belarus	43.81%	3.55	1.25
Belarus Caucasoid	43.81%	3.55	1.25
Belgium	79.39%	7.16	0.97
Belgium Caucasoid	79.39%	7.16	0.97
Bulgaria	57.23%	4.95	0.47
Bulgaria Caucasoid	57.23%	4.95	0.47
Bulgaria Other	0.00%	0	?
Croatia	66.71%	5.89	0.6
Croatia Caucasoid	66.71%	5.89	0.6
Czech Republic	86.21%	9.23	1.45
Czech Republic Caucasoid	88.76%	9.66	1.78
Czech Republic Other	64.14%	6.4	0.56
Denmark	88.98%	9.04	1.81
Denmark Caucasoid	88.98%	9.04	1.81
England	93.48%	10.49	2.74
England Caucasoid	93.48%	10.49	2.74
England Jew	0.00%	0	?
England Mixed	0.00%	0	0
Finland	51.14%	4.24	0.41
Finland Caucasoid	51.14%	4.24	0.41
France	88.54%	9.29	1.74
France Caucasoid	88.54%	9.29	1.74
Georgia	75.05%	7.09	0.8
Georgia Caucasoid	75.05%	7.09	0.8
Georgia Kurd	0.00%	0	?
Germany	91.14%	10.14	2.26
Germany Caucasoid	91.14%	10.14	2.26
Greece	66.92%	6.29	0.6
Greece Caucasoid	66.92%	6.29	0.6

(continued)

Table 7
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Ireland Northern	94.65%	10.58	2.89
Ireland Northern Caucasoid	94.65%	10.58	2.89
Ireland South	93.15%	10	2.51
Ireland South Caucasoid	93.15%	10	2.51
Italy	85.90%	5.93	1.42
Italy Caucasoid	85.90%	5.93	1.42
Macedonia	66.53%	6.2	0.6
Macedonia Caucasoid	66.53%	6.2	0.6
Netherlands	83.44%	8.33	1.21
Netherlands Caucasoid	83.44%	8.33	1.21
Norway	94.71%	10.56	3.01
Norway Caucasoid	94.71%	10.56	3.01
Poland	84.46%	8.85	1.29
Poland Caucasoid	84.46%	8.85	1.29
Portugal	78.00%	7.74	0.91
Portugal Caucasoid	78.00%	7.74	0.91
Romania	0.00%	0	?
Romania Caucasoid	0.00%	0	?
Russia	77.62%	7.24	0.89
Russia Caucasoid	88.52%	9.81	1.74
Russia Mixed	0.00%	0	0
Russia Other	85.01%	9.2	1.33
Russia Siberian	78.83%	7.14	0.94
Scotland	90.82%	10.1	2.2
Scotland Caucasoid	90.82%	10.1	2.2
Serbia	0.00%	0	?
Serbia Caucasoid	0.00%	0	?
Slovakia	18.28%	0.37	0.24
Slovakia Caucasoid	18.28%	0.37	0.24
Slovenia	84.85%	8.74	1.32

(continued)

Table 7
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Slovenia Caucasoid	84.85%	8.74	1.32
Spain	80.51%	8.28	1.03
Spain Caucasoid	80.84%	8.34	1.04
Spain Jew	0.00%	0	?
Spain Other	6.30%	0.57	0.96
Sweden	88.07%	9.13	1.68
Sweden Caucasoid	88.07%	9.13	1.68
Switzerland	0.00%	0	?
Switzerland Caucasoid	0.00%	0	?
Turkey	76.19%	7.3	0.84
Turkey Caucasoid	76.19%	7.3	0.84
Ukraine	50.64%	4.17	1.42
Ukraine Caucasoid	50.64%	4.17	1.42
United Kingdom	0.00%	0	0
United Kingdom Caucasoid	0.00%	0	0
Wales	0.00%	0	0
Wales Caucasoid	0.00%	0	0
East Africa	68.30%	5.65	0.63
Kenya	0.00%	0	0
Kenya Black	0.00%	0	0
Uganda	0.00%	0	0
Uganda Black	0.00%	0	0
Zambia	0.00%	0	?
Zambia Black	0.00%	0	?
Zimbabwe	68.30%	5.65	0.63
Zimbabwe Black	68.30%	5.65	0.63
West Africa	65.23%	6.13	0.58
Burkina Faso	0.00%	0	?
Burkina Faso Black	0.00%	0	?
Cape Verde	80.38%	8.1	1.02

(continued)

Table 7
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Cape Verde Black	80.38%	8.1	1.02
Gambia	0.00%	0	0
Gambia Black	0.00%	0	0
Ghana	0.00%	0	?
Ghana Black	0.00%	0	?
Guinea-Bissau	71.16%	7.04	0.69
Guinea-Bissau Black	71.16%	7.04	0.69
Ivory Coast	0.00%	0	?
Ivory Coast Black	0.00%	0	?
Liberia	0.00%	0	0
Liberia Black	0.00%	0	0
Nigeria	0.00%	0	0
Nigeria Black	0.00%	0	0
Senegal	30.28%	2.32	0.29
Senegal Black	30.28%	2.32	0.29
Central Africa	62.71%	5.17	0.54
Cameroon	49.87%	3.31	0.4
Cameroon Black	49.87%	3.31	0.4
Central African Republic	82.69%	6.47	1.16
Central African Republic Black	82.69%	6.47	1.16
Congo	68.66%	5.93	0.64
Congo Black	68.66%	5.93	0.64
Equatorial Guinea	47.58%	3.55	0.38
Equatorial Guinea Black	47.58%	3.55	0.38
Gabon	41.78%	3.84	1.2
Gabon Black	41.78%	3.84	1.2
Rwanda	62.79%	5.38	0.54
Rwanda Black	62.79%	5.38	0.54
Sao Tome and Principe	66.50%	4.89	0.6
Sao Tome and Principe Black	66.50%	4.89	0.6

(continued)

Table 7
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
North Africa	75.06%	7	0.8
Algeria	77.15%	7.25	0.88
Algeria Arab	77.15%	7.25	0.88
Ethiopia	83.00%	8.71	1.18
Ethiopia Black	83.00%	8.71	1.18
Mali	0.00%	0	?
Mali Black	0.00%	0	?
Morocco	83.44%	8.14	1.21
Morocco Arab	85.07%	8.25	1.34
Morocco Caucasoid	79.75%	8.07	0.99
Sudan	60.56%	4.52	0.51
Sudan Arab	0.00%	0	?
Sudan Black	0.00%	0	0
Sudan Mixed	60.56%	4.52	0.51
Tunisia	74.26%	6.82	0.78
Tunisia Arab	74.97%	6.78	0.8
Tunisia Berber	74.47%	7.43	0.78
South Africa	32.10%	1.11	0.29
South Africa	32.10%	1.11	0.29
South Africa Black	32.10%	1.11	0.29
South Africa Other	0.00%	0	?
West Indies	69.22%	6.67	0.65
Cuba	85.48%	9.66	1.38
Cuba Caucasoid	0.00%	0	?
Cuba Mixed	85.48%	9.66	1.38
Cuba Mulatto	0.00%	0	?
Jamaica	27.41%	2.28	0.28
Jamaica Black	27.41%	2.28	0.28
Martinique	74.51%	7.17	0.78
Martinique Black	74.51%	7.17	0.78

(continued)

Table 7
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Trinidad and Tobago	0.00%	0	?
Trinidad and Tobago Asian	0.00%	0	?
North America	87.89%	9.12	1.65
Canada	38.41%	2.21	0.32
Canada Amerindian	38.41%	2.21	0.32
Mexico	55.04%	4.3	0.44
Mexico Amerindian	42.59%	3.09	0.35
Mexico Mestizo	68.51%	5.97	0.64
United States	88.10%	9.17	1.68
United States Amerindian	42.79%	3.31	0.35
United States Asian	78.84%	8.03	0.95
United States Austronesian	58.09%	5.47	0.48
United States Black	71.50%	6.44	0.7
United States Caucasoid	90.15%	9.68	2.03
United States Hispanic	72.95%	6.9	0.74
United States Mestizo	72.23%	6.78	0.72
United States Polynesian	73.18%	5.87	0.75
Central America	49.91%	4.06	0.4
Costa Rica	24.31%	2.21	0.26
Costa Rica Mestizo	24.31%	2.21	0.26
Guatemala	49.16%	3.37	0.39
Guatemala Amerindian	49.16%	3.37	0.39
South America	58.59%	4.77	0.48
Argentina	62.67%	5.36	0.54
Argentina Amerindian	45.78%	3.4	0.37
Argentina Caucasoid	80.65%	7.85	1.03
Bolivia	77.82%	5.97	0.9
Bolivia Amerindian	77.82%	5.97	0.9
Brazil	63.80%	5.16	0.55
Brazil Amerindian	48.60%	3.23	0.39

(continued)

Table 7
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Brazil Caucasoid	84.39%	8.81	1.28
Brazil Mixed	77.50%	6.94	0.89
Brazil Mulatto	74.09%	6.89	0.77
Brazil Other	0.00%	0	?
Chile	67.08%	5.82	0.61
Chile Amerindian	72.65%	6.09	0.73
Chile Hispanic	0.00%	0	0
Chile Mixed	52.65%	4.39	0.42
Colombia	54.02%	4.34	0.43
Colombia Amerindian	47.40%	3.65	0.38
Colombia Black	65.25%	5.28	0.58
Colombia Mestizo	56.31%	4.8	0.46
Ecuador	52.17%	3.75	1.25
Ecuador Amerindian	52.17%	3.75	1.25
Ecuador Black	0.00%	0	0
Paraguay	4.90%	0.29	0.63
Paraguay Amerindian	4.90%	0.29	0.63
Peru	49.87%	3.47	0.4
Peru Amerindian	49.87%	3.47	0.4
Peru Mestizo	0.00%	0	0
Venezuela	3.01%	0.06	0.21
Venezuela Amerindian	0.00%	0	0
Venezuela Caucasoid	0.00%	0	?
Venezuela Mestizo	0.00%	0	?
Venezuela Mixed	3.17%	0.06	0.21
Oceania	59.87%	5.38	0.5
American Samoa	0.00%	0	?
American Samoa Polynesian	0.00%	0	?
Australia	33.15%	2.21	0.3
Australia Australian Aborigines	33.15%	2.21	0.3

(continued)

Table 7
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Australia Caucasoid	0.00%	0	?
Chile	67.08%	5.82	0.61
Chile Amerindian	72.65%	6.09	0.73
Cook Islands	78.59%	6.44	0.93
Cook Islands Polynesian	78.59%	6.44	0.93
Fiji	79.87%	7.5	0.99
Fiji Melanesian	79.87%	7.5	0.99
Kiribati	10.89%	0.85	0.22
Kiribati Micronesian	10.89%	0.85	0.22
Nauru	38.66%	3.4	0.33
Nauru Micronesian	38.66%	3.4	0.33
New Caledonia	81.41%	8.44	3.77
New Caledonia Melanesian	81.41%	8.44	3.77
New Zealand	84.46%	6.76	1.29
New Zealand Polynesian	84.46%	6.76	1.29
Niue	77.82%	4.27	0.9
Niue Polynesian	77.82%	4.27	0.9
Papua New Guinea	69.15%	7.16	0.65
Papua New Guinea Melanesian	69.15%	7.16	0.65
Samoa	80.86%	7.29	1.04
Samoa Polynesian	80.86%	7.29	1.04
Tokelau	55.11%	2.82	0.45
Tokelau Polynesian	55.11%	2.82	0.45
Tonga	71.91%	6.12	0.71
Tonga Polynesian	71.91%	6.12	0.71
Average	51.14%	4.7	?
(Standard deviation)	−32.55%	−3.35	(?)

[a]Projected population coverage
[b]Average number of epitope hits/HLA combinations recognized by the population
[c]Minimum number of epitope hits/HLA combinations recognized by 90% of the population

represents 94.80%, South Korea and South Oriental Korea (92.84%), China (88.77%), Iran and Iran Persian (91.53%%), Jordan and Jordan Arab (76.80%), Oman and Oman Arab (95.82%), Saudi Arabia and Saudi Arabia Arab (96.38%), Sudan (86.43%), Sudan Arab (49.41%), Sudan Black (0.00%), and Sudan Mixed (87.06%); *see* Table 8. Iran Kurd, United Arab Emirates, and United Arab Emirates Arab were not mentioned and showed results in this tool.

According to the percentage of MHC-I modified E protein coverage population that represented 95.60% of the world population, 112 countries showed a higher percentile rate especially Chile Amerindian which represents 100.00%, 96 other countries showed 0% while in East Asia represents 94.80%, South Korea and South Oriental Korea (92.84%), China (88.77%), Iran (91.53%), Iran Persian (91.53%), Iran Kurd (0.00%), Jordan and Jordan Arab (76.80%), Oman and Oman Arab (95.82%), Saudi Arabia and Saudi Arabia Arab (96.38%), United Arab Emirates and United Arab Emirates Arab (0.0%), Sudan (60.56%), Sudan Arab (0.00%), Sudan Black (0.00%), and Sudan Mixed (60.56%); *see* Table 9.

According to the percentile rates of MHC-II E protein coverage population that represented 81.81% of the world population, 63 countries showed a higher percentage especially Norway and Norway Caucasoid (94.71%), 45 other countries showed from 0% to less than 50% while in East Asia represents 94.80%, South Korea and South Oriental Korea (85.32%), China (59.99%), Iran (64.22%), Iran Persian (65.72%), Iran Kurd (55.78%), Saudi Arabia and Saudi Arabia Arab (80.14%), United Arab Emirates and United Arab Emirates Arab (32.92%), and Sudan and Sudan Mixed (60.56%); *see* Table 10. Oman, Jordan, Sudan Black, and Arab were not mentioned and showed results in this tool.

According to the percentage of MHC-II modified E protein coverage population that represented 81.81% of the world population, 62 countries showed a higher percentage especially Norway and Norway Caucasoid (94.71%), 59 other countries showed 0% while in East Asia represents 94.80%, South Korea and South Oriental Korea (85.32%), China (59.99%), Iran (64.22%), Iran Persian (65.72%), Iran Kurd (55.78%), Jordan and Jordan Arab (52.88%), Oman and Oman Arab (0.00%), Saudi Arabia and Saudi Arabia Arab (80.14%), United Arab Emirates and United Arab Emirates Arab (32.92%), Sudan and Sudan Mixed (60.56%), and Sudan Arab and Sudan Black (0.00%); *see* Table 11.

3.4 Homology Modeling

The results of homology modeling were not shown here because they are not necessary.

Table 8
MHC-I coverage population for E protein

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
World	95.60%	10.57	4.38
East Asia	94.80%	10.93	2.58
Japan	96.19%	11.44	3.12
Japan Oriental	96.19%	11.44	3.12
Korea, South	92.84%	10.41	2.16
Korea, South Oriental	92.84%	10.41	2.16
Mongolia	94.37%	10.07	3.12
Mongolia Oriental	94.37%	10.07	3.12
Northeast Asia	88.80%	9.38	0.89
China	88.77%	9.33	0.89
China Oriental	88.77%	9.33	0.89
Hong Kong	90.85%	10.01	1.91
Hong Kong Oriental	90.85%	10.01	1.91
South Asia	86.54%	8.03	0.74
India	82.00%	7.21	0.56
India Asian	82.00%	7.21	0.56
Pakistan	88.63%	8.74	1.76
Pakistan Asian	87.30%	8.38	1.58
Pakistan Mixed	91.12%	9.42	3.23
Sri Lanka	52.39%	3.74	0.84
Sri Lanka Asian	52.39%	3.74	0.84
Southeast Asia	87.81%	9.99	0.82
Indonesia	76.44%	7.8	0.42
Indonesia Austronesian	76.44%	7.8	0.42
Malaysia	76.30%	7.64	0.42
Malaysia Austronesian	40.59%	3.17	0.34
Malaysia Oriental	84.44%	9.02	0.64
Philippines	92.86%	11.56	8.01
Philippines Austronesian	92.86%	11.56	8.01
Singapore	85.74%	9.04	0.7

(continued)

Table 8
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Singapore Austronesian	82.82%	8.55	0.58
Singapore Oriental	88.96%	9.64	0.91
Taiwan	92.58%	11.31	6.08
Taiwan Oriental	92.58%	11.31	6.08
Thailand	82.85%	7.46	0.58
Thailand Oriental	82.85%	7.46	0.58
Vietnam	84.58%	8.55	0.65
Vietnam Oriental	84.58%	8.55	0.65
Southwest Asia	85.77%	7.59	0.7
Iran	91.53%	8.6	1.33
Iran Persian	91.53%	8.6	1.33
Israel	82.14%	7.29	0.56
Israel Arab	89.15%	9.13	0.92
Israel Jew	87.17%	7.84	0.78
Jordan	76.80%	6.52	0.43
Jordan Arab	76.80%	6.52	0.43
Oman	95.82%	9.96	3.04
Oman Arab	95.82%	9.96	3.04
Saudi Arabia	96.38%	9.87	3.65
Saudi Arabia Arab	96.38%	9.87	3.65
Europe	97.81%	11.07	5.29
Austria	98.78%	11.29	6
Austria Caucasoid	98.78%	11.29	6
Belgium	98.75%	10.62	6.02
Belgium Caucasoid	98.75%	10.62	6.02
Bulgaria	96.59%	11.08	4.52
Bulgaria Caucasoid	96.56%	11.25	4.57
Bulgaria Other	97.43%	10.02	4.35
Croatia	97.76%	11.79	6.12
Croatia Caucasoid	97.76%	11.79	6.12

(continued)

Table 8
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Czech Republic	96.20%	9.39	4.33
Czech Republic Caucasoid	96.20%	9.39	4.33
England	99.29%	11.43	6.21
England Caucasoid	99.29%	11.43	6.21
Finland	99.80%	12.56	7.8
Finland Caucasoid	99.80%	12.56	7.8
France	98.05%	10.72	4.75
France Caucasoid	98.05%	10.72	4.75
Georgia	95.62%	10.98	4.48
Georgia Caucasoid	97.22%	11.66	6.21
Georgia Kurd	89.99%	9.26	1
Germany	99.07%	11.71	6.4
Germany Caucasoid	99.07%	11.71	6.4
Ireland Northern	99.40%	11.43	6.27
Ireland Northern Caucasoid	99.40%	11.43	6.27
Ireland South	98.83%	10.82	4.85
Ireland South Caucasoid	98.83%	10.82	4.85
Italy	96.52%	9.83	4.16
Italy Caucasoid	96.52%	9.83	4.16
Macedonia	11.83%	0.86	0.45
Macedonia Caucasoid	11.83%	0.86	0.45
Poland	97.99%	11.25	6.02
Poland Caucasoid	97.99%	11.25	6.02
Portugal	97.11%	10.98	4.73
Portugal Caucasoid	97.11%	10.98	4.73
Romania	97.94%	11.56	5.94
Romania Caucasoid	97.94%	11.56	5.94
Russia	96.71%	11.38	4.59
Russia Other	98.34%	12.46	6.71
Russia Siberian	97.30%	11.52	4.53

(continued)

Table 8
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Scotland	15.91%	0.81	0.24
Scotland Caucasoid	15.91%	0.81	0.24
Serbia	43.75%	0.78	0.18
Serbia Caucasoid	43.75%	0.78	0.18
Spain	71.85%	5.51	0.36
Spain Caucasoid	71.85%	5.51	0.36
Sweden	99.69%	12.61	6.84
Sweden Caucasoid	99.69%	12.61	6.84
Turkey	44.80%	3.58	1.45
Turkey Caucasoid	44.80%	3.58	1.45
East Africa	86.99%	6.96	0.77
Kenya	85.86%	6.62	0.71
Kenya Black	85.86%	6.62	0.71
Uganda	91.04%	8.19	1.48
Uganda Black	91.04%	8.19	1.48
Zambia	95.32%	7.98	4.01
Zambia Black	95.32%	7.98	4.01
Zimbabwe	91.57%	7.69	1.71
Zimbabwe Black	91.57%	7.69	1.71
West Africa	92.60%	8.71	1.67
Burkina Faso	58.50%	3.24	0.24
Burkina Faso Black	58.50%	3.24	0.24
Cape Verde	96.69%	10.09	4.14
Cape Verde Black	96.69%	10.09	4.14
Guinea-Bissau	92.66%	8.7	1.49
Guinea-Bissau Black	92.66%	8.7	1.49
Ivory Coast	58.05%	0.78	0.24
Ivory Coast Black	58.05%	0.78	0.24
Senegal	95.03%	9.11	4
Senegal Black	95.03%	9.11	4

(continued)

Table 8
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Central Africa	84.98%	6.7	0.67
Cameroon	88.67%	7.35	0.88
Cameroon Black	88.67%	7.35	0.88
Central African Republic	10.75%	0.27	0.11
Central African Republic Black	10.75%	0.27	0.11
Rwanda	23.09%	1.33	0.13
Rwanda Black	23.09%	1.33	0.13
Sao Tome and Principe	95.54%	8.72	2.29
Sao Tome and Principe Black	95.54%	8.72	2.29
North Africa	91.87%	8.61	1.86
Mali	94.28%	8.82	1.74
Mali Black	94.28%	8.82	1.74
Morocco	95.95%	9.47	4.19
Morocco Arab	97.89%	10.2	4.47
Morocco Caucasoid	94.32%	8.96	4.02
Sudan	86.43%	7.53	0.74
Sudan Arab	49.41%	4.62	0.59
Sudan Black	0.00%	0	0
Sudan Mixed	87.06%	7.56	0.77
Tunisia	96.04%	9.85	4.19
Tunisia Arab	96.04%	9.85	4.19
South Africa	91.05%	8	2.1
South Africa	91.05%	8	2.1
South Africa Black	86.71%	6.67	0.75
South Africa Other	93.82%	9.59	2.73
West Indies	97.34%	10.78	4.6
Cuba	97.20%	10.65	4.53
Cuba Caucasoid	97.64%	11.2	4.77
Cuba Mulatto	96.58%	9.66	4.09
Martinique	22.56%	2.03	1.16

(continued)

Table 8
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Martinique Black	22.56%	2.03	1.16
North America	96.88%	10.98	4.65
Mexico	97.10%	11	6.02
Mexico Amerindian	99.86%	13	7.84
Mexico Mestizo	96.78%	10.7	4.46
United States	96.93%	10.98	4.66
United States Amerindian	99.44%	13.15	8.19
United States Asian	92.39%	10.32	2.29
United States Black	94.18%	8.83	2.54
United States Caucasoid	98.65%	11.4	6.08
United States Hispanic	97.46%	11.01	4.77
United States Mestizo	98.09%	11.2	4.97
United States Polynesian	97.53%	11.57	3.62
Central America	5.10%	0.16	0.11
Guatemala	5.10%	0.16	0.11
Guatemala Amerindian	5.10%	0.16	0.11
South America	86.24%	8.01	0.73
Argentina	98.02%	8.76	2.61
Argentina Amerindian	98.02%	8.76	2.61
Brazil	93.72%	9.43	2.69
Brazil Amerindian	92.35%	8.37	2.16
Brazil Caucasoid	97.68%	11.33	5.35
Brazil Mixed	95.06%	9.85	3.75
Chile	94.93%	10.63	4.37
Chile Amerindian	100.00%	14.31	9.11
Chile Mixed	87.43%	8.16	0.8
Colombia	9.86%	0.76	0.67
Colombia Black	5.79%	0.42	0.64
Colombia Mestizo	14.81%	1.17	0.7
Ecuador	76.97%	8.77	1.74

(continued)

Table 8
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Ecuador Amerindian	76.97%	8.77	1.74
Peru	99.98%	13.69	8.37
Peru Amerindian	99.98%	13.69	8.37
Venezuela	88.37%	9.05	0.86
Venezuela Amerindian	88.88%	8.98	0.9
Venezuela Caucasoid	9.18%	0.83	0.99
Venezuela Mestizo	7.84%	0.71	0.98
Oceania	91.82%	10.92	4.06
American Samoa	95.26%	12.14	7.15
American Samoa Polynesian	95.26%	12.14	7.15
Australia	89.30%	9.93	0.93
Australia Australian Aborigines	82.36%	9.31	0.57
Australia Caucasoid	99.06%	11.46	6.16
Chile	94.93%	10.63	4.37
Chile Amerindian	100.00%	14.31	9.11
New Caledonia	96.70%	12.14	8.63
New Caledonia Melanesian	96.70%	12.14	8.63
Papua New Guinea	97.26%	12.58	8.57
Papua New Guinea Melanesian	97.26%	12.58	8.57
Average	55.31%	5.73	?
(Standard deviation)	−44.16%	−4.92	(?)

[a]Projected population coverage
[b]Average number of epitope hits/HLA combinations recognized by the population
[c]Minimum number of epitope hits/HLA combinations recognized by 90% of the population

3.5 Confirmation of Amino Acid Change in Spike Glycoprotein (S) and Envelope Protein (E) Sequence

The results of confirmatory amino acid change were not shown here because they are not necessary.

3.6 Peptide Search Tool

The results of peptide search tool showed presence of selected peptide sequence in another organisms such as *Leishmania donovani*, *Drosophila sechellia* (fruit fly), *Leishmania infantum*,

Table 9
MHC-I coverage population for modified E protein

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
World	95.60%	10.57	4.38
East Asia	94.80%	10.93	2.58
Japan	96.19%	11.44	3.12
Japan Oriental	96.19%	11.44	3.12
Korea, South	92.84%	10.41	2.16
Korea, South Oriental	92.84%	10.41	2.16
Mongolia	94.37%	10.07	3.12
Mongolia Oriental	94.37%	10.07	3.12
Northeast Asia	88.80%	9.38	0.89
China	88.77%	9.33	0.89
China Oriental	88.77%	9.33	0.89
Hong Kong	90.85%	10.01	1.91
Hong Kong Oriental	90.85%	10.01	1.91
South Asia	86.54%	8.03	0.74
India	82.00%	7.21	0.56
India Asian	82.00%	7.21	0.56
Pakistan	88.63%	8.74	1.76
Pakistan Asian	87.30%	8.38	1.58
Pakistan Mixed	91.12%	9.42	3.23
Sri Lanka	52.39%	3.74	0.84
Sri Lanka Asian	52.39%	3.74	0.84
Southeast Asia	87.81%	9.99	0.82
Borneo	0.00%	0	?
Borneo Austronesian	0.00%	0	?
Indonesia	76.44%	7.8	0.42
Indonesia Austronesian	76.44%	7.8	0.42
Malaysia	76.30%	7.64	0.42
Malaysia Austronesian	40.59%	3.17	0.34
Malaysia Oriental	84.44%	9.02	0.64
Philippines	92.86%	11.56	8.01

(continued)

Table 9
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Philippines Austronesian	92.86%	11.56	8.01
Singapore	85.74%	9.04	0.7
Singapore Austronesian	82.82%	8.55	0.58
Singapore Oriental	88.96%	9.64	0.91
Taiwan	92.58%	11.31	6.08
Taiwan Oriental	92.58%	11.31	6.08
Thailand	82.85%	7.46	0.58
Thailand Oriental	82.85%	7.46	0.58
Vietnam	84.58%	8.55	0.65
Vietnam Oriental	84.58%	8.55	0.65
Southwest Asia	85.77%	7.59	0.7
Iran	91.53%	8.6	1.33
Iran Kurd	0.00%	0	?
Iran Persian	91.53%	8.6	1.33
Israel	82.14%	7.29	0.56
Israel Arab	89.15%	9.13	0.92
Israel Jew	87.17%	7.84	0.78
Jordan	76.80%	6.52	0.43
Jordan Arab	76.80%	6.52	0.43
Lebanon	0.00%	0	0
Lebanon Arab	0.00%	0	?
Lebanon Mixed	0.00%	0	0
Oman	95.82%	9.96	3.04
Oman Arab	95.82%	9.96	3.04
Saudi Arabia	96.38%	9.87	3.65
Saudi Arabia Arab	96.38%	9.87	3.65
United Arab Emirates	0.00%	0	0
United Arab Emirates Arab	0.00%	0	0
Europe	97.81%	11.07	5.29
Austria	98.78%	11.29	6

(continued)

Table 9
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Austria Caucasoid	98.78%	11.29	6
Belarus	0.00%	0	?
Belarus Caucasoid	0.00%	0	?
Belgium	98.75%	10.62	6.02
Belgium Caucasoid	98.75%	10.62	6.02
Bulgaria	96.59%	11.08	4.52
Bulgaria Caucasoid	96.56%	11.25	4.57
Bulgaria Other	97.43%	10.02	4.35
Croatia	97.76%	11.79	6.12
Croatia Caucasoid	97.76%	11.79	6.12
Czech Republic	96.20%	9.39	4.33
Czech Republic Caucasoid	96.20%	9.39	4.33
Czech Republic Other	0.00%	0	?
Denmark	0.00%	0	0
Denmark Caucasoid	0.00%	0	0
England	99.29%	11.43	6.21
England Caucasoid	99.29%	11.43	6.21
England Jew	0.00%	0	0
England Mixed	0.00%	0	?
Finland	99.80%	12.56	7.8
Finland Caucasoid	99.80%	12.56	7.8
France	98.05%	10.72	4.75
France Caucasoid	98.05%	10.72	4.75
Georgia	95.62%	10.98	4.48
Georgia Caucasoid	97.22%	11.66	6.21
Georgia Kurd	89.99%	9.26	1
Germany	99.07%	11.71	6.4
Germany Caucasoid	99.07%	11.71	6.4
Greece	0.00%	0	?
Greece Caucasoid	0.00%	0	?

(continued)

Table 9
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Ireland Northern	99.40%	11.43	6.27
Ireland Northern Caucasoid	99.40%	11.43	6.27
Ireland South	98.83%	10.82	4.85
Ireland South Caucasoid	98.83%	10.82	4.85
Italy	96.52%	9.83	4.16
Italy Caucasoid	96.52%	9.83	4.16
Macedonia	11.83%	0.86	0.45
Macedonia Caucasoid	11.83%	0.86	0.45
Netherlands	0.00%	0	?
Netherlands Caucasoid	0.00%	0	?
Norway	0.00%	0	?
Norway Caucasoid	0.00%	0	?
Poland	97.99%	11.25	6.02
Poland Caucasoid	97.99%	11.25	6.02
Portugal	97.11%	10.98	4.73
Portugal Caucasoid	97.11%	10.98	4.73
Romania	97.94%	11.56	5.94
Romania Caucasoid	97.94%	11.56	5.94
Russia	96.71%	11.38	4.59
Russia Caucasoid	0.00%	0	0
Russia Mixed	0.00%	0	0
Russia Other	98.34%	12.46	6.71
Russia Siberian	97.30%	11.52	4.53
Scotland	15.91%	0.81	0.24
Scotland Caucasoid	15.91%	0.81	0.24
Serbia	43.75%	0.78	0.18
Serbia Caucasoid	43.75%	0.78	0.18
Slovakia	0.00%	0	?
Slovakia Caucasoid	0.00%	0	?
Slovenia	0.00%	0	?

(continued)

Table 9
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Slovenia Caucasoid	0.00%	0	?
Spain	71.85%	5.51	0.36
Spain Caucasoid	71.85%	5.51	0.36
Spain Jew	0.00%	0	?
Spain Other	0.00%	0	?
Sweden	99.69%	12.61	6.84
Sweden Caucasoid	99.69%	12.61	6.84
Switzerland	0.00%	0	0
Switzerland Caucasoid	0.00%	0	0
Turkey	44.80%	3.58	1.45
Turkey Caucasoid	44.80%	3.58	1.45
Ukraine	0.00%	0	?
Ukraine Caucasoid	0.00%	0	?
United Kingdom	0.00%	0	0
United Kingdom Caucasoid	0.00%	0	0
Wales	0.00%	0	0
Wales Caucasoid	0.00%	0	0
East Africa	86.99%	6.96	0.77
Kenya	85.86%	6.62	0.71
Kenya Black	85.86%	6.62	0.71
Uganda	91.04%	8.19	1.48
Uganda Black	91.04%	8.19	1.48
Zambia	95.32%	7.98	4.01
Zambia Black	95.32%	7.98	4.01
Zimbabwe	91.57%	7.69	1.71
Zimbabwe Black	91.57%	7.69	1.71
West Africa	92.60%	8.71	1.67
Burkina Faso	58.50%	3.24	0.24
Burkina Faso Black	58.50%	3.24	0.24
Cape Verde	96.69%	10.09	4.14

(continued)

Table 9
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Cape Verde Black	96.69%	10.09	4.14
Gambia	0.00%	0	?
Gambia Black	0.00%	0	?
Ghana	0.00%	0	0
Ghana Black	0.00%	0	0
Guinea-Bissau	92.66%	8.7	1.49
Guinea-Bissau Black	92.66%	8.7	1.49
Ivory Coast	58.05%	0.78	0.24
Ivory Coast Black	58.05%	0.78	0.24
Liberia	0.00%	0	?
Liberia Black	0.00%	0	?
Nigeria	0.00%	0	?
Nigeria Black	0.00%	0	?
Senegal	95.03%	9.11	4
Senegal Black	95.03%	9.11	4
Central Africa	84.98%	6.7	0.67
Cameroon	88.67%	7.35	0.88
Cameroon Black	88.67%	7.35	0.88
Central African Republic	10.75%	0.27	0.11
Central African Republic Black	10.75%	0.27	0.11
Congo	0.00%	0	?
Congo Black	0.00%	0	?
Equatorial Guinea	0.00%	0	0
Equatorial Guinea Black	0.00%	0	0
Gabon	0.00%	0	?
Gabon Black	0.00%	0	?
Rwanda	23.09%	1.33	0.13
Rwanda Black	23.09%	1.33	0.13
Sao Tome and Principe	95.54%	8.72	2.29
Sao Tome and Principe Black	95.54%	8.72	2.29

(continued)

Table 9
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
North Africa	91.87%	8.61	1.86
Algeria	0.00%	0	?
Algeria Arab	0.00%	0	?
Ethiopia	0.00%	0	?
Ethiopia Black	0.00%	0	?
Mali	94.28%	8.82	1.74
Mali Black	94.28%	8.82	1.74
Morocco	95.95%	9.47	4.19
Morocco Arab	97.89%	10.2	4.47
Morocco Caucasoid	94.32%	8.96	4.02
Sudan	86.43%	7.53	0.74
Sudan Arab	49.41%	4.62	0.59
Sudan Black	0.00%	0	0
Sudan Mixed	87.06%	7.56	0.77
Tunisia	96.04%	9.85	4.19
Tunisia Arab	96.04%	9.85	4.19
Tunisia Berber	0.00%	0	?
South Africa	91.05%	8	2.1
South Africa	91.05%	8	2.1
South Africa Black	86.71%	6.67	0.75
South Africa Other	93.82%	9.59	2.73
West Indies	97.34%	10.78	4.6
Cuba	97.20%	10.65	4.53
Cuba Caucasoid	97.64%	11.2	4.77
Cuba Mixed	0.00%	0	?
Cuba Mulatto	96.58%	9.66	4.09
Jamaica	0.00%	0	?
Jamaica Black	0.00%	0	?
Martinique	22.56%	2.03	1.16
Martinique Black	22.56%	2.03	1.16

(continued)

Table 9
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Trinidad and Tobago	0.00%	0	0
Trinidad and Tobago Asian	0.00%	0	0
North America	96.88%	10.98	4.65
Canada	0.00%	0	?
Canada Amerindian	0.00%	0	?
Mexico	97.10%	11	6.02
Mexico Amerindian	99.86%	13	7.84
Mexico Mestizo	96.78%	10.7	4.46
United States	96.93%	10.98	4.66
United States Amerindian	99.44%	13.15	8.19
United States Asian	92.39%	10.32	2.29
United States Austronesian	0.00%	0	?
United States Black	94.18%	8.83	2.54
United States Caucasoid	98.65%	11.4	6.08
United States Hispanic	97.46%	11.01	4.77
United States Mestizo	98.09%	11.2	4.97
United States Polynesian	97.53%	11.57	3.62
Central America	5.10%	0.16	0.11
Costa Rica	0.00%	0	?
Costa Rica Mestizo	0.00%	0	?
Guatemala	5.10%	0.16	0.11
Guatemala Amerindian	5.10%	0.16	0.11
South America	86.24%	8.01	0.73
Argentina	98.02%	8.76	2.61
Argentina Amerindian	98.02%	8.76	2.61
Argentina Caucasoid	0.00%	0	?
Bolivia	0.00%	0	?
Bolivia Amerindian	0.00%	0	?
Brazil	93.72%	9.43	2.69
Brazil Amerindian	92.35%	8.37	2.16

(continued)

Table 9
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Brazil Caucasoid	97.68%	11.33	5.35
Brazil Mixed	95.06%	9.85	3.75
Brazil Mulatto	0.00%	0	?
Brazil Other	0.00%	0	0
Chile	94.93%	10.63	4.37
Chile Amerindian	100.00%	14.31	9.11
Chile Hispanic	0.00%	0	?
Chile Mixed	87.43%	8.16	0.8
Colombia	9.86%	0.76	0.67
Colombia Amerindian	0.00%	0	0
Colombia Black	5.79%	0.42	0.64
Colombia Mestizo	14.81%	1.17	0.7
Ecuador	76.97%	8.77	1.74
Ecuador Amerindian	76.97%	8.77	1.74
Ecuador Black	0.00%	0	?
Paraguay	0.00%	0	?
Paraguay Amerindian	0.00%	0	?
Peru	99.98%	13.69	8.37
Peru Amerindian	99.98%	13.69	8.37
Peru Mestizo	0.00%	0	0
Venezuela	88.37%	9.05	0.86
Venezuela Amerindian	88.88%	8.98	0.9
Venezuela Caucasoid	9.18%	0.83	0.99
Venezuela Mestizo	7.84%	0.71	0.98
Venezuela Mixed	0.00%	0	?
Oceania	91.82%	10.92	4.06
American Samoa	95.26%	12.14	7.15
American Samoa Polynesian	95.26%	12.14	7.15
Australia	89.30%	9.93	0.93
Australia Australian Aborigines	82.36%	9.31	0.57

(continued)

Table 9
(continued)

	Class I		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Australia Caucasoid	99.06%	11.46	6.16
Chile	94.93%	10.63	4.37
Chile Amerindian	100.00%	14.31	9.11
Cook Islands	0.00%	0	?
Cook Islands Polynesian	0.00%	0	?
Fiji	0.00%	0	?
Fiji Melanesian	0.00%	0	?
Kiribati	0.00%	0	?
Kiribati Micronesian	0.00%	0	?
Nauru	0.00%	0	?
Nauru Micronesian	0.00%	0	?
New Caledonia	96.70%	12.14	8.63
New Caledonia Melanesian	96.70%	12.14	8.63
New Zealand	0.00%	0	?
New Zealand Polynesian	0.00%	0	?
Niue	0.00%	0	?
Niue Polynesian	0.00%	0	?
Papua New Guinea	97.26%	12.58	8.57
Papua New Guinea Melanesian	97.26%	12.58	8.57
Samoa	0.00%	0	?
Samoa Polynesian	0.00%	0	?
Tokelau	0.00%	0	?
Tokelau Polynesian	0.00%	0	?
Tonga	0.00%	0	?
Tonga Polynesian	0.00%	0	?
Average	55.31%	5.73	?
(Standard deviation)	−44.16%	−4.92	(?)

[a]Projected population coverage
[b]Average number of epitope hits/HLA combinations recognized by the population
[c]Minimum number of epitope hits/HLA combinations recognized by 90% of the population

Table 10
The MHC-II coverage population for E protein

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
World	81.81%	8.16	1.1
East Asia	81.82%	8.83	1.1
Japan	74.83%	7.85	0.79
Japan Oriental	74.83%	7.85	0.79
Korea, South	85.32%	9.56	1.36
Korea, South Oriental	85.32%	9.56	1.36
Mongolia	81.85%	7.79	1.1
Mongolia Oriental	81.85%	7.79	1.1
Northeast Asia	59.99%	5.33	0.5
China	59.99%	5.33	0.5
China Oriental	59.99%	5.33	0.5
South Asia	75.38%	7.4	0.81
India	74.99%	7.35	0.8
India Asian	74.99%	7.35	0.8
Pakistan	1.18%	0.09	0.81
Pakistan Asian	1.45%	0.12	0.81
Southeast Asia	56.98%	4.98	0.46
Borneo	49.02%	4.03	0.39
Borneo Austronesian	49.02%	4.03	0.39
Indonesia	47.84%	4.4	0.38
Indonesia Austronesian	47.84%	4.4	0.38
Malaysia	57.99%	5.34	0.48
Malaysia Austronesian	55.38%	5.12	0.45
Malaysia Oriental	70.35%	6.57	0.67
Philippines	28.56%	2.52	0.28
Philippines Austronesian	28.56%	2.52	0.28
Singapore	65.78%	6.04	0.58
Singapore Austronesian	65.78%	6.04	0.58
Singapore Oriental	0.00%	0	?
Taiwan	67.88%	6.13	0.62

(continued)

Table 10
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Taiwan Oriental	67.88%	6.13	0.62
Thailand	63.90%	5.92	0.55
Thailand Oriental	63.90%	5.92	0.55
Vietnam	54.44%	4.43	0.44
Vietnam Oriental	54.44%	4.43	0.44
Southwest Asia	43.93%	3.65	0.36
Iran	64.22%	5.65	0.56
Iran Kurd	55.78%	4.74	0.45
Iran Persian	65.72%	5.83	0.58
Israel	68.79%	6.4	0.64
Israel Arab	67.51%	6.2	0.62
Israel Jew	69.65%	6.51	0.66
Jordan	52.88%	4.56	0.42
Jordan Arab	52.88%	4.56	0.42
Lebanon	70.46%	6.48	0.68
Lebanon Arab	70.46%	6.48	0.68
Saudi Arabia	80.14%	8.31	1.01
Saudi Arabia Arab	80.14%	8.31	1.01
United Arab Emirates	32.92%	0.66	0.3
United Arab Emirates Arab	32.92%	0.66	0.3
Europe	85.83%	8.88	1.41
Austria	93.34%	10.8	2.82
Austria Caucasoid	93.34%	10.8	2.82
Belarus	43.81%	3.55	1.25
Belarus Caucasoid	43.81%	3.55	1.25
Belgium	79.39%	7.16	0.97
Belgium Caucasoid	79.39%	7.16	0.97
Bulgaria	57.23%	4.95	0.47
Bulgaria Caucasoid	57.23%	4.95	0.47
Croatia	66.71%	5.89	0.6

(continued)

Table 10
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Croatia Caucasoid	66.71%	5.89	0.6
Czech Republic	86.21%	9.23	1.45
Czech Republic Caucasoid	88.76%	9.66	1.78
Czech Republic Other	64.14%	6.4	0.56
Denmark	88.98%	9.04	1.81
Denmark Caucasoid	88.98%	9.04	1.81
England	93.48%	10.49	2.74
England Caucasoid	93.48%	10.49	2.74
Finland	51.14%	4.24	0.41
Finland Caucasoid	51.14%	4.24	0.41
France	88.54%	9.29	1.74
France Caucasoid	88.54%	9.29	1.74
Georgia	75.05%	7.09	0.8
Georgia Caucasoid	75.05%	7.09	0.8
Germany	91.14%	10.14	2.26
Germany Caucasoid	91.14%	10.14	2.26
Greece	66.92%	6.29	0.6
Greece Caucasoid	66.92%	6.29	0.6
Ireland Northern	94.65%	10.58	2.89
Ireland Northern Caucasoid	94.65%	10.58	2.89
Ireland South	93.15%	10	2.51
Ireland South Caucasoid	93.15%	10	2.51
Italy	85.90%	5.93	1.42
Italy Caucasoid	85.90%	5.93	1.42
Macedonia	66.53%	6.2	0.6
Macedonia Caucasoid	66.53%	6.2	0.6
Netherlands	83.44%	8.33	1.21
Netherlands Caucasoid	83.44%	8.33	1.21
Norway	94.71%	10.56	3.01
Norway Caucasoid	94.71%	10.56	3.01

(continued)

Table 10
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Poland	84.46%	8.85	1.29
Poland Caucasoid	84.46%	8.85	1.29
Portugal	78.00%	7.74	0.91
Portugal Caucasoid	78.00%	7.74	0.91
Russia	77.62%	7.24	0.89
Russia Caucasoid	88.52%	9.81	1.74
Russia Other	85.01%	9.2	1.33
Russia Siberian	78.83%	7.14	0.94
Scotland	90.82%	10.1	2.2
Scotland Caucasoid	90.82%	10.1	2.2
Slovakia	18.28%	0.37	0.24
Slovakia Caucasoid	18.28%	0.37	0.24
Slovenia	84.85%	8.74	1.32
Slovenia Caucasoid	84.85%	8.74	1.32
Spain	80.51%	8.28	1.03
Spain Caucasoid	80.84%	8.34	1.04
Spain Other	6.30%	0.57	0.96
Sweden	88.07%	9.13	1.68
Sweden Caucasoid	88.07%	9.13	1.68
Turkey	76.19%	7.3	0.84
Turkey Caucasoid	76.19%	7.3	0.84
Ukraine	50.64%	4.17	1.42
Ukraine Caucasoid	50.64%	4.17	1.42
East Africa	68.30%	5.65	0.63
Zimbabwe	68.30%	5.65	0.63
Zimbabwe Black	68.30%	5.65	0.63
West Africa	65.23%	6.13	0.58
Cape Verde	80.38%	8.1	1.02
Cape Verde Black	80.38%	8.1	1.02
Guinea-Bissau	71.16%	7.04	0.69

(continued)

Table 10
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Guinea-Bissau Black	71.16%	7.04	0.69
Senegal	30.28%	2.32	0.29
Senegal Black	30.28%	2.32	0.29
Central Africa	62.71%	5.17	0.54
Cameroon	49.87%	3.31	0.4
Cameroon Black	49.87%	3.31	0.4
Central African Republic	82.69%	6.47	1.16
Central African Republic Black	82.69%	6.47	1.16
Congo	68.66%	5.93	0.64
Congo Black	68.66%	5.93	0.64
Equatorial Guinea	47.58%	3.55	0.38
Equatorial Guinea Black	47.58%	3.55	0.38
Gabon	41.78%	3.84	1.2
Gabon Black	41.78%	3.84	1.2
Rwanda	62.79%	5.38	0.54
Rwanda Black	62.79%	5.38	0.54
Sao Tome and Principe	66.50%	4.89	0.6
Sao Tome and Principe Black	66.50%	4.89	0.6
North Africa	75.06%	7	0.8
Algeria	77.15%	7.25	0.88
Algeria Arab	77.15%	7.25	0.88
Ethiopia	83.00%	8.71	1.18
Ethiopia Black	83.00%	8.71	1.18
Morocco	83.44%	8.14	1.21
Morocco Arab	85.07%	8.25	1.34
Morocco Caucasoid	79.75%	8.07	0.99
Sudan	60.56%	4.52	0.51
Sudan Mixed	60.56%	4.52	0.51
Tunisia	74.26%	6.82	0.78
Tunisia Arab	74.97%	6.78	0.8

(continued)

Table 10
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Tunisia Berber	74.47%	7.43	0.78
South Africa	32.10%	1.11	0.29
South Africa	32.10%	1.11	0.29
South Africa Black	32.10%	1.11	0.29
West Indies	69.22%	6.67	0.65
Cuba	85.48%	9.66	1.38
Cuba Mixed	85.48%	9.66	1.38
Jamaica	27.41%	2.28	0.28
Jamaica Black	27.41%	2.28	0.28
Martinique	74.51%	7.17	0.78
Martinique Black	74.51%	7.17	0.78
North America	87.89%	9.12	1.65
Canada	38.41%	2.21	0.32
Canada Amerindian	38.41%	2.21	0.32
Mexico	55.04%	4.3	0.44
Mexico Amerindian	42.59%	3.09	0.35
Mexico Mestizo	68.51%	5.97	0.64
United States	88.10%	9.17	1.68
United States Amerindian	42.79%	3.31	0.35
United States Asian	78.84%	8.03	0.95
United States Austronesian	58.09%	5.47	0.48
United States Black	71.50%	6.44	0.7
United States Caucasoid	90.15%	9.68	2.03
United States Hispanic	72.95%	6.9	0.74
United States Mestizo	72.23%	6.78	0.72
United States Polynesian	73.18%	5.87	0.75
Central America	49.91%	4.06	0.4
Costa Rica	24.31%	2.21	0.26
Costa Rica Mestizo	24.31%	2.21	0.26
Guatemala	49.16%	3.37	0.39

(continued)

Table 10
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Guatemala Amerindian	49.16%	3.37	0.39
South America	58.59%	4.77	0.48
Argentina	62.67%	5.36	0.54
Argentina Amerindian	45.78%	3.4	0.37
Argentina Caucasoid	80.65%	7.85	1.03
Bolivia	77.82%	5.97	0.9
Bolivia Amerindian	77.82%	5.97	0.9
Brazil	63.80%	5.16	0.55
Brazil Amerindian	48.60%	3.23	0.39
Brazil Caucasoid	84.39%	8.81	1.28
Brazil Mixed	77.50%	6.94	0.89
Brazil Mulatto	74.09%	6.89	0.77
Chile	67.08%	5.82	0.61
Chile Amerindian	72.65%	6.09	0.73
Chile Mixed	52.65%	4.39	0.42
Colombia	54.02%	4.34	0.43
Colombia Amerindian	47.40%	3.65	0.38
Colombia Black	65.25%	5.28	0.58
Colombia Mestizo	56.31%	4.8	0.46
Ecuador	52.17%	3.75	1.25
Ecuador Amerindian	52.17%	3.75	1.25
Paraguay	4.90%	0.29	0.63
Paraguay Amerindian	4.90%	0.29	0.63
Peru	49.87%	3.47	0.4
Peru Amerindian	49.87%	3.47	0.4
Venezuela	3.01%	0.06	0.21
Venezuela Mixed	3.17%	0.06	0.21
Oceania	59.87%	5.38	0.5
Australia	33.15%	2.21	0.3
Australia Australian Aborigines	33.15%	2.21	0.3

(continued)

Table 10
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Chile	67.08%	5.82	0.61
Chile Amerindian	72.65%	6.09	0.73
Cook Islands	78.59%	6.44	0.93
Cook Islands Polynesian	78.59%	6.44	0.93
Fiji	79.87%	7.5	0.99
Fiji Melanesian	79.87%	7.5	0.99
Kiribati	10.89%	0.85	0.22
Kiribati Micronesian	10.89%	0.85	0.22
Nauru	38.66%	3.4	0.33
Nauru Micronesian	38.66%	3.4	0.33
New Caledonia	81.41%	8.44	3.77
New Caledonia Melanesian	81.41%	8.44	3.77
New Zealand	84.46%	6.76	1.29
New Zealand Polynesian	84.46%	6.76	1.29
Niue	77.82%	4.27	0.9
Niue Polynesian	77.82%	4.27	0.9
Papua New Guinea	69.15%	7.16	0.65
Papua New Guinea Melanesian	69.15%	7.16	0.65
Samoa	80.86%	7.29	1.04
Samoa Polynesian	80.86%	7.29	1.04
Tokelau	55.11%	2.82	0.45
Tokelau Polynesian	55.11%	2.82	0.45
Tonga	71.91%	6.12	0.71
Tonga Polynesian	71.91%	6.12	0.71
Average	51.14%	4.7	?
(Standard deviation)	−32.55%	−3.35	(?)

[a]Projected population coverage
[b]Average number of epitope hits/HLA combinations recognized by the population
[c]Minimum number of epitope hits/HLA combinations recognized by 90% of the population

Table 11
The MHC-II coverage population for modified E protein

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
World	81.81%	8.16	1.1
East Asia	81.82%	8.83	1.1
Japan	74.83%	7.85	0.79
Japan Oriental	74.83%	7.85	0.79
Korea, South	85.32%	9.56	1.36
Korea, South Oriental	85.32%	9.56	1.36
Mongolia	81.85%	7.79	1.1
Mongolia Oriental	81.85%	7.79	1.1
Northeast Asia	59.99%	5.33	0.5
China	59.99%	5.33	0.5
China Oriental	59.99%	5.33	0.5
Hong Kong	0.00%	0	?
Hong Kong Oriental	0.00%	0	?
South Asia	75.38%	7.4	0.81
India	74.99%	7.35	0.8
India Asian	74.99%	7.35	0.8
Pakistan	1.18%	0.09	0.81
Pakistan Asian	1.45%	0.12	0.81
Pakistan Mixed	0.00%	0	0
Sri Lanka	0.00%	0	?
Sri Lanka Asian	0.00%	0	?
Southeast Asia	56.98%	4.98	0.46
Borneo	49.02%	4.03	0.39
Borneo Austronesian	49.02%	4.03	0.39
Indonesia	47.84%	4.4	0.38
Indonesia Austronesian	47.84%	4.4	0.38
Malaysia	57.99%	5.34	0.48
Malaysia Austronesian	55.38%	5.12	0.45
Malaysia Oriental	70.35%	6.57	0.67
Philippines	28.56%	2.52	0.28

(continued)

Table 11
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Philippines Austronesian	28.56%	2.52	0.28
Singapore	65.78%	6.04	0.58
Singapore Austronesian	65.78%	6.04	0.58
Singapore Oriental	0.00%	0	?
Taiwan	67.88%	6.13	0.62
Taiwan Oriental	67.88%	6.13	0.62
Thailand	63.90%	5.92	0.55
Thailand Oriental	63.90%	5.92	0.55
Vietnam	54.44%	4.43	0.44
Vietnam Oriental	54.44%	4.43	0.44
Southwest Asia	43.93%	3.65	0.36
Iran	64.22%	5.65	0.56
Iran Kurd	55.78%	4.74	0.45
Iran Persian	65.72%	5.83	0.58
Israel	68.79%	6.4	0.64
Israel Arab	67.51%	6.2	0.62
Israel Jew	69.65%	6.51	0.66
Jordan	52.88%	4.56	0.42
Jordan Arab	52.88%	4.56	0.42
Lebanon	70.46%	6.48	0.68
Lebanon Arab	70.46%	6.48	0.68
Lebanon Mixed	0.00%	0	?
Oman	0.00%	0	?
Oman Arab	0.00%	0	?
Saudi Arabia	80.14%	8.31	1.01
Saudi Arabia Arab	80.14%	8.31	1.01
United Arab Emirates	32.92%	0.66	0.3
United Arab Emirates Arab	32.92%	0.66	0.3
Europe	85.83%	8.88	1.41
Austria	93.34%	10.8	2.82

(continued)

Table 11
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Austria Caucasoid	93.34%	10.8	2.82
Belarus	43.81%	3.55	1.25
Belarus Caucasoid	43.81%	3.55	1.25
Belgium	79.39%	7.16	0.97
Belgium Caucasoid	79.39%	7.16	0.97
Bulgaria	57.23%	4.95	0.47
Bulgaria Caucasoid	57.23%	4.95	0.47
Bulgaria Other	0.00%	0	?
Croatia	66.71%	5.89	0.6
Croatia Caucasoid	66.71%	5.89	0.6
Czech Republic	86.21%	9.23	1.45
Czech Republic Caucasoid	88.76%	9.66	1.78
Czech Republic Other	64.14%	6.4	0.56
Denmark	88.98%	9.04	1.81
Denmark Caucasoid	88.98%	9.04	1.81
England	93.48%	10.49	2.74
England Caucasoid	93.48%	10.49	2.74
England Jew	0.00%	0	?
England Mixed	0.00%	0	0
Finland	51.14%	4.24	0.41
Finland Caucasoid	51.14%	4.24	0.41
France	88.54%	9.29	1.74
France Caucasoid	88.54%	9.29	1.74
Georgia	75.05%	7.09	0.8
Georgia Caucasoid	75.05%	7.09	0.8
Georgia Kurd	0.00%	0	?
Germany	91.14%	10.14	2.26
Germany Caucasoid	91.14%	10.14	2.26
Greece	66.92%	6.29	0.6
Greece Caucasoid	66.92%	6.29	0.6

(continued)

Table 11
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Ireland Northern	94.65%	10.58	2.89
Ireland Northern Caucasoid	94.65%	10.58	2.89
Ireland South	93.15%	10	2.51
Ireland South Caucasoid	93.15%	10	2.51
Italy	85.90%	5.93	1.42
Italy Caucasoid	85.90%	5.93	1.42
Macedonia	66.53%	6.2	0.6
Macedonia Caucasoid	66.53%	6.2	0.6
Netherlands	83.44%	8.33	1.21
Netherlands Caucasoid	83.44%	8.33	1.21
Norway	94.71%	10.56	3.01
Norway Caucasoid	94.71%	10.56	3.01
Poland	84.46%	8.85	1.29
Poland Caucasoid	84.46%	8.85	1.29
Portugal	78.00%	7.74	0.91
Portugal Caucasoid	78.00%	7.74	0.91
Romania	0.00%	0	?
Romania Caucasoid	0.00%	0	?
Russia	77.62%	7.24	0.89
Russia Caucasoid	88.52%	9.81	1.74
Russia Mixed	0.00%	0	0
Russia Other	85.01%	9.2	1.33
Russia Siberian	78.83%	7.14	0.94
Scotland	90.82%	10.1	2.2
Scotland Caucasoid	90.82%	10.1	2.2
Serbia	0.00%	0	?
Serbia Caucasoid	0.00%	0	?
Slovakia	18.28%	0.37	0.24
Slovakia Caucasoid	18.28%	0.37	0.24
Slovenia	84.85%	8.74	1.32

(continued)

Table 11
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Slovenia Caucasoid	84.85%	8.74	1.32
Spain	80.51%	8.28	1.03
Spain Caucasoid	80.84%	8.34	1.04
Spain Jew	0.00%	0	?
Spain Other	6.30%	0.57	0.96
Sweden	88.07%	9.13	1.68
Sweden Caucasoid	88.07%	9.13	1.68
Switzerland	0.00%	0	?
Switzerland Caucasoid	0.00%	0	?
Turkey	76.19%	7.3	0.84
Turkey Caucasoid	76.19%	7.3	0.84
Ukraine	50.64%	4.17	1.42
Ukraine Caucasoid	50.64%	4.17	1.42
United Kingdom	0.00%	0	0
United Kingdom Caucasoid	0.00%	0	0
Wales	0.00%	0	0
Wales Caucasoid	0.00%	0	0
East Africa	68.30%	5.65	0.63
Kenya	0.00%	0	0
Kenya Black	0.00%	0	0
Uganda	0.00%	0	0
Uganda Black	0.00%	0	0
Zambia	0.00%	0	?
Zambia Black	0.00%	0	?
Zimbabwe	68.30%	5.65	0.63
Zimbabwe Black	68.30%	5.65	0.63
West Africa	65.23%	6.13	0.58
Burkina Faso	0.00%	0	?
Burkina Faso Black	0.00%	0	?
Cape Verde	80.38%	8.1	1.02

(continued)

Table 11
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Cape Verde Black	80.38%	8.1	1.02
Gambia	0.00%	0	0
Gambia Black	0.00%	0	0
Ghana	0.00%	0	?
Ghana Black	0.00%	0	?
Guinea-Bissau	71.16%	7.04	0.69
Guinea-Bissau Black	71.16%	7.04	0.69
Ivory Coast	0.00%	0	?
Ivory Coast Black	0.00%	0	?
Liberia	0.00%	0	0
Liberia Black	0.00%	0	0
Nigeria	0.00%	0	0
Nigeria Black	0.00%	0	0
Senegal	30.28%	2.32	0.29
Senegal Black	30.28%	2.32	0.29
Central Africa	62.71%	5.17	0.54
Cameroon	49.87%	3.31	0.4
Cameroon Black	49.87%	3.31	0.4
Central African Republic	82.69%	6.47	1.16
Central African Republic Black	82.69%	6.47	1.16
Congo	68.66%	5.93	0.64
Congo Black	68.66%	5.93	0.64
Equatorial Guinea	47.58%	3.55	0.38
Equatorial Guinea Black	47.58%	3.55	0.38
Gabon	41.78%	3.84	1.2
Gabon Black	41.78%	3.84	1.2
Rwanda	62.79%	5.38	0.54
Rwanda Black	62.79%	5.38	0.54
Sao Tome and Principe	66.50%	4.89	0.6
Sao Tome and Principe Black	66.50%	4.89	0.6

(continued)

Table 11
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
North Africa	75.06%	7	0.8
Algeria	77.15%	7.25	0.88
Algeria Arab	77.15%	7.25	0.88
Ethiopia	83.00%	8.71	1.18
Ethiopia Black	83.00%	8.71	1.18
Mali	0.00%	0	?
Mali Black	0.00%	0	?
Morocco	83.44%	8.14	1.21
Morocco Arab	85.07%	8.25	1.34
Morocco Caucasoid	79.75%	8.07	0.99
Sudan	60.56%	4.52	0.51
Sudan Arab	0.00%	0	?
Sudan Black	0.00%	0	0
Sudan Mixed	60.56%	4.52	0.51
Tunisia	74.26%	6.82	0.78
Tunisia Arab	74.97%	6.78	0.8
Tunisia Berber	74.47%	7.43	0.78
South Africa	32.10%	1.11	0.29
South Africa	32.10%	1.11	0.29
South Africa Black	32.10%	1.11	0.29
South Africa Other	0.00%	0	?
West Indies	69.22%	6.67	0.65
Cuba	85.48%	9.66	1.38
Cuba Caucasoid	0.00%	0	?
Cuba Mixed	85.48%	9.66	1.38
Cuba Mulatto	0.00%	0	?
Jamaica	27.41%	2.28	0.28
Jamaica Black	27.41%	2.28	0.28
Martinique	74.51%	7.17	0.78
Martinique Black	74.51%	7.17	0.78

(continued)

Table 11
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Trinidad and Tobago	0.00%	0	?
Trinidad and Tobago Asian	0.00%	0	?
North America	87.89%	9.12	1.65
Canada	38.41%	2.21	0.32
Canada Amerindian	38.41%	2.21	0.32
Mexico	55.04%	4.3	0.44
Mexico Amerindian	42.59%	3.09	0.35
Mexico Mestizo	68.51%	5.97	0.64
United States	88.10%	9.17	1.68
United States Amerindian	42.79%	3.31	0.35
United States Asian	78.84%	8.03	0.95
United States Austronesian	58.09%	5.47	0.48
United States Black	71.50%	6.44	0.7
United States Caucasoid	90.15%	9.68	2.03
United States Hispanic	72.95%	6.9	0.74
United States Mestizo	72.23%	6.78	0.72
United States Polynesian	73.18%	5.87	0.75
Central America	49.91%	4.06	0.4
Costa Rica	24.31%	2.21	0.26
Costa Rica Mestizo	24.31%	2.21	0.26
Guatemala	49.16%	3.37	0.39
Guatemala Amerindian	49.16%	3.37	0.39
South America	58.59%	4.77	0.48
Argentina	62.67%	5.36	0.54
Argentina Amerindian	45.78%	3.4	0.37
Argentina Caucasoid	80.65%	7.85	1.03
Bolivia	77.82%	5.97	0.9
Bolivia Amerindian	77.82%	5.97	0.9
Brazil	63.80%	5.16	0.55
Brazil Amerindian	48.60%	3.23	0.39

(continued)

Table 11
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Brazil Caucasoid	84.39%	8.81	1.28
Brazil Mixed	77.50%	6.94	0.89
Brazil Mulatto	74.09%	6.89	0.77
Brazil Other	0.00%	0	?
Chile	67.08%	5.82	0.61
Chile Amerindian	72.65%	6.09	0.73
Chile Hispanic	0.00%	0	0
Chile Mixed	52.65%	4.39	0.42
Colombia	54.02%	4.34	0.43
Colombia Amerindian	47.40%	3.65	0.38
Colombia Black	65.25%	5.28	0.58
Colombia Mestizo	56.31%	4.8	0.46
Ecuador	52.17%	3.75	1.25
Ecuador Amerindian	52.17%	3.75	1.25
Ecuador Black	0.00%	0	0
Paraguay	4.90%	0.29	0.63
Paraguay Amerindian	4.90%	0.29	0.63
Peru	49.87%	3.47	0.4
Peru Amerindian	49.87%	3.47	0.4
Peru Mestizo	0.00%	0	0
Venezuela	3.01%	0.06	0.21
Venezuela Amerindian	0.00%	0	0
Venezuela Caucasoid	0.00%	0	?
Venezuela Mestizo	0.00%	0	?
Venezuela Mixed	3.17%	0.06	0.21
Oceania	59.87%	5.38	0.5
American Samoa	0.00%	0	?
American Samoa Polynesian	0.00%	0	?
Australia	33.15%	2.21	0.3
Australia Australian Aborigines	33.15%	2.21	0.3

(continued)

Table 11
(continued)

	Class II		
Population/Area	**Coverage[a]**	**Average hit[b]**	**PC90[c]**
Australia Caucasoid	0.00%	0	?
Chile	67.08%	5.82	0.61
Chile Amerindian	72.65%	6.09	0.73
Cook Islands	78.59%	6.44	0.93
Cook Islands Polynesian	78.59%	6.44	0.93
Fiji	79.87%	7.5	0.99
Fiji Melanesian	79.87%	7.5	0.99
Kiribati	10.89%	0.85	0.22
Kiribati Micronesian	10.89%	0.85	0.22
Nauru	38.66%	3.4	0.33
Nauru Micronesian	38.66%	3.4	0.33
New Caledonia	81.41%	8.44	3.77
New Caledonia Melanesian	81.41%	8.44	3.77
New Zealand	84.46%	6.76	1.29
New Zealand Polynesian	84.46%	6.76	1.29
Niue	77.82%	4.27	0.9
Niue Polynesian	77.82%	4.27	0.9
Papua New Guinea	69.15%	7.16	0.65
Papua New Guinea Melanesian	69.15%	7.16	0.65
Samoa	80.86%	7.29	1.04
Samoa Polynesian	80.86%	7.29	1.04
Tokelau	55.11%	2.82	0.45
Tokelau Polynesian	55.11%	2.82	0.45
Tonga	71.91%	6.12	0.71
Tonga Polynesian	71.91%	6.12	0.71
Average	51.14%	4.7	?
(Standard deviation)	−32.55%	−3.35	(?)

[a]Projected population coverage
[b]Average number of epitope hits/HLA combinations recognized by the population
[c]Minimum number of epitope hits/HLA combinations recognized by 90% of the population

Trypanosoma cruzi Dm28c, *Strigamia maritime*, and *Nocardioides dokdonensis*; besides some species of *Mycobacteria*, *Salmonella*, *Streptococcus*, these may mean the presence of these peptides in those organisms had a relationship with respiratory disease but still needs to go deeper to confirm this suggestion, other things we can easily synthesis the desired peptides in laboratory by using one of these organisms (cloning techniques) because it is easy and no risk from acquired a very dangers infections beside determination of the peptide sequences impact on immune system via injected laboratory animals with those selected peptide sequences from any organisms.

3.7 AllerHunter: Cross-Reactive Allergen Prediction Program

Any sequence can be considered as a cross-reactive allergen if its probability is ≧0.06. The results considered that envelope (E) protein, spike (S) glycoprotein, and modified S glycoprotein are potential non-allergens with scores of 0.01, 0.0, and 0.0, respectively, while modified E protein sequence was too short for prediction (AllerHunter predicted the query sequence as a potential allergen with score of 0.07). According to the FAO/WHO, E and modified E protein sequences are classified as a non-allergen because they do not meet the criteria set by the FAO/WHO evaluation scheme for cross-reactive allergen prediction, but in S and modified S glycoprotein, they are classified as a potential allergen based on the FAO/WHO evaluation scheme because query sequence matches at least one sequence in the AllerHunter data set with at least 35 percent identity over 80 amino acids.

3.8 AlgPred: Prediction of Allergenic Proteins and Mapping of IgE Epitopes

AlgPred showed non-allergen for all four sequences (S, E, modified S and E proteins) as follows:

1. Prediction by mapping of IgE epitope: The protein sequence does not contain experimentally proven IgE epitope.
2. MAST RESULT: No Hits found; NON ALLERGEN.
3. BLAST results of ARPS: No hits found, NON-ALLERGEN.
4. Prediction by hybrid approach: NON-ALLERGEN/ALLERGEN.

There were slightly differences between the four sequences in SVM prediction methods according to amino acid composition/dipeptide composition as in Tables 12 and 13.

3.9 VaxiJen v2.0

VaxJen servers showed three protein sequences out of two, considered as probable antigens, as illustrated below:

S glycoprotein: threshold for this model, 0.4; overall antigen prediction, 0.4827 (probable ANTIGEN).

Modified S glycoprotein: threshold for this model, 0.4; overall antigen prediction, 0.4907 (probable ANTIGEN).

Table 12
SVM prediction methods based on amino acid composition for the four protein sequences

Types of protein sequence	SVM prediction based on amino acid composition	Score	Threshold	Positive predictive value	Negative predictive value
S glycoprotein	Allergen	0.014762929	−0.4	70.05%	80.74%
Modified S glycoprotein	Allergen	0.0065929692	−0.4	70.05%	80.74%
E protein	Allergen	−0.3638541	−0.4	47.13%/	89.71%
Modified E protein	Non-allergen	−1.08932	−0.4	15.19%	94.18%.

Table 13
Illustrates SVM prediction methods based on dipeptide composition for the four protein sequences

Types of protein sequence	SVM prediction based on amino acid composition	Score	Threshold	Positive predictive value	Negative predictive value
S glycoprotein	Allergen	−0.04096577	−0.2	63.1%	85.56%
Modified S glycoprotein	Allergen	−0.059498832	−0.2	63.1%	85.56%
E protein	Non-allergen	−0.7511982	−0.2	13.26%	74.19%
Modified E protein	Non-allergen	−0.65278098	−0.2	13.26%	74.19%

E protein: threshold for this model, 0.4; overall antigen prediction, 0.3811 (probable NON-ANTIGEN).

Modified E protein: threshold for this model, 0.4; overall antigen prediction, 0.4417 (probable ANTIGEN).

4 Discussions

Today, there are so many different ways to develop MERS-CoV vaccine; some of them partially succeed but the others failed while the remaining nor succeed neither failed because it depends on software program for different reasons and still need to go under vaccine protocols processing, in those studies that consist with S1 protein subunit especially RBD (the most mutable region that containing mutation sites which define antibody escape variants) was considered the basis for several MERS-CoV vaccine candidates in many studies such as using RBD with aluminum salt or oil-in-water adjuvants; can elicited neutralizing antibodies of high potency across multiple viral strains by Modjarrad [4] and Wang

et al. [6] said that the full-length S DNA and a truncated S1 subunit glycoprotein can elicit a higher titer of neutralizing antibodies; this kind of immunization protected non-human primates (NHPs) from severe lung disease after intratracheal challenge with MERS-CoV injection; in another study that was done in Iran by Poorinmohammad et al. [15] [NetCTL 1.2 (Larsen et al., 2007), EpiJen (Doytchinova et al, 2006), and NHLApred (Bhasin and Raghava, 2007), they were selected computational prediction tools with PEPstr server for modeling (Kaur et al., 2007)] to identify cytotoxic T-lymphocyte epitopes presented by the human leukocyte antigen (HLA)-A∗0201; as this is the most frequent HLA class I allele among Middle Eastern populations with this selected RBD for their study, they showed LLSGTPPQV, ILDYFSYPL ILATVPHNL, NLTTITKPL, LQMGFGITV, and FSNPTCLIL as selected epitopes but LLSGTPPQV and FSNPTCLIL were considered as real epitope due to the following: peptides with binding orientations closer to the native structure and lower binding free energy scores are ranked higher in having the potential to be real epitopes reverse another study were done by Shi J et al. [19] by using the Immune Epitope Database, that said: 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 when compared with spike (S) protein; in addition 71 peptides were identified as helper T-cell epitopes, 34 peptides were identified as CTL epitopes; just top 10 helper T-cell epitopes and CTL epitopes based on maximum HLA binding alleles, can elicit protective cellular immune responses against MERS-CoV were considered as MERS vaccine candidates and they are covering 15 geographic regions [19].

In this study that consists of two parts reference and modified sequence of both S glycoprotein and E protein, I found that the most common B-cell epitope that passed all B-cell prediction methods [IEDB prediction tool] for E protein is YVKFQDS in position 69 and for modified E they are VYVPQQD, YVPQQDS, and PPLPED/PPLPEDV epitopes at positions 68, 69, and 77 sequentially; while for S and modified S, they are DVGPDSV, PDSVKSA, DSVKSAC, PRPIDVS, HTPATDC, AKPSGSV, KPSGSVV, SGTPPQV, GTPPQVY, TPPQVYN, QLSPLEG, YGPLQTP, PRSVRSV, RSVRSVP, SVKSSQS, VKSSQSS, SQSSPII, and SLNTKYV at positions 23, 26, 27, 48, 211, 371, 372, 393, 394, 395, 547, 707, 750, 751, 856, 859 (857 in modified S glycoprotein), and 1202 sequentially, but QVDQLNS and VDQLNSS epitopes at positions 772 and 773 are only found in S glycoprotein, while LTPTSSY, TPTSSYV, PTSSYVD, TSSYVDV, DHGDYYV, YSQDVKQ, ANQYSPC, NQYSPCV, and YYRKQLS epitopes at positions 15, 16, 17, 18, 83, 108, 523, 524, and 543 are only found in modified S glycoprotein; according to my study, I found that the results of S and modified S glycoprotein they are

partially agree with the study that was done in Africa city of Technology-Khartoum, Sudan by Badawi et al, [16] in those epitopes GTPPQVY in position 391–397 and LTPRSVRSVP in position 745–754, may be do you to different numbers of selected MERS-CoV protein sequence.

Prediction of cytotoxic T-lymphocyte epitopes and their interaction with MHC Class I, the results showed ILDYFSYPL was similar according my study, Badwai et al [16] and Poorinmohammad and Mohabatkar [15] studies; partially similarity with Iranian study [15] in LLSGTPPQV, ILATVPHNL, LQMGFGITV, and FSNPTCLIL epitopes were noticed except NLTTITKPL epitope that was absent from my study in S and modified S sequence; FSNPTCLIL represents the only epitope that is found in my study in S and modified S sequence; FSFGVTQEY have a high affinity to bind to many alleles and these findings agree with Badawi et al. [16] in addition to ITYQGLFPY in my study through S glycoprotein sequence, but still there are differences in the numbers of selected epitopes that reacted with MHC-I which were higher than that in Badawi et al. [16], while in E protein FIFTVVCAI epitope has a higher allele affinity followed by ITLLVCMAF, IVNFFIFTV, and LVQPALYLY reverse modified E protein; LVQPALSLY epitope has shown high affinity and then followed by LYMTGRSVY, WFIPNFFDF, YMTGRSVYV, ITLLVCTAF, FVQERIGWF, FLTATHLCV, and CMTGFNTLL, the last epitope which is common between E and modified E protein sequences.

Prediction of T-helper cell epitopes and their interactions with MHC Class II showed FNLTLLEPVSISTGS epitope that was considered as the most suitable epitope with a high affinity to 26 alleles in Badawi et al. [16]; this epitope was actually found in S and modified S sequence of my study, but the difference is that it cannot considered that the most suitable epitope with a high binding affinity to different alleles like in in Badawi et al, [16] study.

There is no research results related to E protein and modified E and S glycoprotein epitope vaccine instead of partial similarity that I found between S and modified S glycoprotein.

No previous study illustrates S glycoprotein and E protein allergic reactions except the study that were done by Shi J et al. [19] for N protein, but in this study, S and E protein showed no allergic reaction according to AllerHunter services. Furthermore Shi J et al. [19] said that, for N protein, 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 (363KKEKKQ368), but the minimum value of surface probability was 0.074 for 205GIGAVG210 peptides, while in the analysis of the flexibility of the predicted peptides, they showed that the maximum flexibility value was 1.160 at amino acid position from 170 to 176 (167GNSQSSS173) with the minimum value

0.903 for peptides 97RWYFYYT103; in MHC-II the epitope 329LRYSGAIKL337 interacting with 357 HLA-DR alleles was considered the epitope that possesses the maximum number of binding HLA-DR alleles, while 230VKQSQPKVI238 interacting with 94 HLA-DR alleles is the epitope that possesses the minimum number of binding HLA-DR alleles, and also the same occurred with MHC-I; KQLAPRWYF100 had the highest number of binding HLA-A alleles in MHC-I and then followed by 343NYNKWLELL351,72AQNAGYWRR80, and 387RVQGSITQR395 (*see* [19]) paper for coverage population); in addition to the above, the studies that were done by Sharmin and Islam [20] showed that WDYPKCDRA was considered as a highly conserved epitope in the RNA directed RNA polymerase of human coronaviruses after applying multiple sequence alignment (MSA) approach for spike (S), membrane (M), enveloped (E), and nucleocapsid (N) protein and replicase polyprotein 1ab to identify which one is highly conserved in all coronavirus strains, followed by using various in silico tools to predict consensus immunogenic and conserved peptide.

Furthermore information that were not shown here are that I used the software below to confirm MHC-II results, and their results partially agree with IEDB MHC-I results and I do not know why. EpiDOCK: Molecular docking—based tool for MHC class II binding prediction (http://epidock.ddg-pharmfac.net/), EpiTOP1.0 (http://www.pharmfac.net/EpiTOP/index.php), other things that I do not agree with Shi J et al. [19] when he did alignments for S, E, M. . . .., with all human coronavirus & said he just found the most common peptide was N protein alone, because when I trying to made alignment for S, M, ORFA1,.., I found some alignments between those proteins and different coronavirus strains and this may be means presence of some common peptide but it still needs more studies.

4.1 Conclusions

As I mentioned before, software vaccine and drug design became very important in the first and third world countries to avoid wasting resources, time, and efforts; for MERS-CoV vaccine, it is important to design effective vaccine that cannot be protected against MERS-CoV but also the emergence of new strain besides the other human coronavirus especially when MERS-CoV vaccines they are not passed all vaccine design protocols.

In this study I found the following points: Emergence of a new strains may had a minor change in peptide sequence vaccine especially when the selected viruses parts nor longer neither smaller in their length.

In B-cell prediction; mutations can lead to increased numbers of selected epitopes with very few sequence changes noticed, in addition to a large number of shared epitopes between reference and modified sequence; this means mutated sequence has the ability to elicit the same immune response (IR) (response to virus by the same antibodies as in first infections).

Mutations of the virus sequence can change the frequency of allele and peptide numbers eithers through increased or decreased these numbers, beside presences or absences of some new/old alleles or peptides; same alleles had a different peptide sequences and vice versa.

For MHC-II there were not changed in E & modified E protein alleles & their frequencies & also in peptide sequences & their frequencies were noticed, these may be due to short E protein sequence, while for S & modified S glycoprotein there are minor difference in some peptide frequency numbers either by adding/lowering one or two numbers just & same for alleles.

There is an allele similarity between E, S, and modified E and S proteins in MHC-II, besides presence of a tiny difference in S and modified S peptide sequences in MHC-II due to the modification that I was introduced before in S reference sequence.

The absence of very few numbers of peptide sequences from S reference sequence in modified S sequence leads to the presence of a new peptide sequences.

In MHC-I a lot of selected peptide sequences that are represented in S glycoprotein reference sequence are missing from the modified one reverse E protein reference sequence due to presence of additional epitopes in E protein modified sequence.

The presence of arginine in some selected peptide sequence vaccine makes it ineffective, so we need to solve this problem either by replacing it with other amino acid from the same group or by finding another ways that make those epitopes visible for immune system (IS).

The presence of mutated sequence can effect on the coverage population in MHC-II by presence/absence of some countries, with the percentage changes, reverse MHC-I no changes were noticed.

Acknowledgments

The author would like to thank Allah, her family, for always supporting her, and the National Ribat University members.

References

1. Coronavirus-Vaccine-a-6110.html, 2013
2. http://en.wikipedia.org/wiki/Coronavirus, 2014
3. Khan G (2013) A novel coronavirus capable of lethal human infections: an emerging picture. Virol J 10:66. http://virologyj.biomedcentral.com/articles/10.1186/1743-422X-10-66
4. Modjarrad K (2016) MERS-CoV vaccine candidates in development: the current landscape. Vaccine 34(26):2982–2987
5. Ithete NL, Stoffberg S, Corman VM, Cottontail VM, Richards LR, Schoeman MC, Drosten C, Drexler JF, Preiser W (2013) Close relative of human middle east respiratory syndrome coronavirus in Bat, South Africa. Emerg Infect Dis 19(10):1697–1699
6. Wang L, Shi W, Joyce GM, Modjarrad K, Zhang Y, Leung K, Lees RC, Zhou T, Yassine MH et al (2015) Evaluation of candidate vaccine approaches for MERS-CoV. Nat Commun

6:7712. http://www.nature.com/articles/ncomms8712

7. Kim Y, Ponomarenko J, Zhu Z, Tamang D, Wang P, Greenbaum J, Lundegaard C, Sette A, Lund O, Bourne PE, Nielsen M, Peters B (2012) Immune epitope database analysis resource. Nucleic Acids Res 40: W525–W530
8. Sidney J, Assarsson E, Moore C, Ngo S, Pinilla C, Sette A, Peters B (2008) Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. Immunome Res 4:2
9. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, Buus S, Nielsen M (2009) NetMHCpan, a method for MHC class I binding prediction beyond humans. Immunogenetics 61:1–13
10. Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lamberth K, Buus S, Brunak S, Lund O (2003) Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci 12:1007–1017
11. Peters B, Sette A (2005) Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BMC Bioinformatics 6:132
12. Karosiene E, Rasmussen M, Blicher T, Lund O, Buus S, Nielsen M (2013) 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 65(10):711
13. Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, Roder G, Peters B, Sette A, Lund O, Buus S (2007) NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. PLoS One 2:e796
14. Nielsen M, Lundegaard C, Blicher T, Peters B, Sette A, Justesen S, Buus S, Lund O (2008) Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. PLoS Comput Biol 4(7): e1000107
15. Poorinmohammad N, Mohabatkar H (2014) Identification of HLA-A*0201-restricted CTL epitopes from the receptor-binding domain of MERS-CoV spike protein using a combinatorial in silico approach. *Turk J Biol* 38:628–632. http://journals.tubitak.gov.tr/biology/issues/biy-14-38-5/biy-38-5-10-1401-21.pdf
16. Badawi MM, Salaheldin AM, Suliman MM, AbduRahim AS, Mohammed AEA, SidAhmed SAA, Othman MM, Salih AM (2016) In silico prediction of a novel universal multi-epitope peptide vaccine in the whole spike glycoprotein of MERS CoV. Am. J. Microbiol. Res 4 (4):101–121
17. Du L, Zhao G, Kou Z (2013) 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 87 (17):9939–9942
18. Mohamed HA, Mohamed YO, Salam AB, Yousif AH, Hassan MM, Kaheel HH, Hassan AM (2014) In silico analysis of single nucleotide polymorphisms (SNPs) in human FANCA gene. Int J Comput Bioinform In Silico Model 3(5):502–513
19. Shi J, Zhang J, Li S, Sun J, Teng Y, Wu M, Li J, Li Y, Hu N, Wang H, Hu Y (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
20. Sharmin R, Islam AB (2014) 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 15:161
21. Saha S, Raghava GPS (2006) AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic Acids Res 34:W202–W209
22. Doytchinova AI, Flower RD (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics 8:4
23. Doytchinova AI, Flower RD (2007) Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. Vaccine 25:856–866
24. Doytchinova AI, Flower RD (2008) Bioinformatic approach for identifying parasite and fungal candidate subunit vaccines. Open Vaccines J 1:22–26