doi: 10.1111/sji.12544

# Computational Prediction of Usutu Virus E Protein B Cell and T Cell Epitopes for Potential Vaccine Development

N. Palanisamy[1,2] & J. Lennerstrand[1]

[1]Section of Clinical Virology, Department of Medical Sciences, Uppsala University, Uppsala, Sweden

*Received 2 November 2016; Accepted in revised form 26 February 2017*

*Correspondence to*: N. Palanisamy, Molecular and Cellular Engineering Group, Im Neuenheimer Feld 267 (BioQuant), 69120 Heidelberg, Germany. E-mail: navaneethan.palanisamy@bioquant.uni-heidelberg.de

[2]Present address: Molecular and Cellular Engineering Group, BioQuant, University of Heidelberg, Heidelberg, Germany; The Hartmut Hoffmann-Berling International Graduate School of Molecular and Cellular Biology, University of Heidelberg, Heidelberg, Germany

## Abstract

Usutu virus (family *Flaviviridae*), once confined to Africa, has emerged in Europe a decade ago. The virus has been spreading throughout Europe at a greater pace mostly affecting avian species. While most bird species remain asymptomatic carriers of this virus, few bird species are highly susceptible. Lately, Usutu virus (USUV) infections in humans were reported sporadically with severe neuroinvasive symptoms like meningoencephalitis. As so much is unknown about this virus, which potentially may cause severe diseases in humans, there is a need for more studies of this virus. In this study, we have used computational tools to predict potential B cell and T cell epitopes of USUV envelope (E) protein. We found that amino acids between positions 68 and 84 could be a potential B cell epitope, while amino acids between positions 53 and 69 could be a potential major histocompatibility complex (MHC) class I- and class II-restricted T cell epitope. By homology 3D modeling of USUV E protein, we found that the predicted B cell epitope was predominantly located in the coil region, while T cell epitope was located in the beta-strand region of the E protein. Additionally, the potential MHC class I T cell epitope (LAEVRSYCYL) was predicted to bind to nearly 24 human leucocyte antigens (HLAs) ($IC_{50}$ ≤5000 nM) covering nearly 86.44% of the Black population and 96.90% of the Caucasoid population. Further *in vivo* studies are needed to validate the predicted epitopes.

## Introduction

Usutu virus (USUV), a mosquito-borne virus, belongs to the Japanese encephalitis virus serogroup of the *Flaviviridae* family [1]. The other well-known, human pathogenic members of *Flaviviridae* family include hepatitis C virus, dengue virus, Zika virus, yellow fever virus, Japanese encephalitis virus, tick-borne encephalitis virus and West Nile virus. USUV was first discovered in 1959 in South Africa from a mosquito species *Culex univittatus*, which bites humans and animals, followed by in 1962 in Uganda from another mosquito species *Mansonia aurites*, which usually bites birds [2]. Like other members of the *Flaviviridae* family, the USUV genome consists of a single-stranded positive-sense RNA of approximately 11 kb, which encodes a polyprotein [3]. The polyprotein consists of 10 individual proteins, from N-terminal to C-terminal: capsid (C), the precursor of M (pre-M), envelope (E), NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5 [3]. This polyprotein is cleaved off into individual proteins by the host cellular protease and viral protease (NS3). The 5′ of the RNA has cap structure, but the 3′ lacks poly(A) tail [3]. The 5′ and 3′ UTR sequences form terminal stem-loop structures which are vital for the replication of the virus [4, 5]. A study has found that these sequences also play a part in flaviviral cytopathicity and pathogenicity [6].

For nearly a decade after its discovery, USUV was neglected as the virus was confined to Africa and only two human cases with milder symptoms such as fever and rash were reported [7]. In 2001, USUV was identified in Austria causing deaths among Eurasian blackbirds (*Turdus merula*) and has since been enzootic [8, 9]. This was the first identification of USUV in Europe, although later, another study reported the isolation of USUV from preserved tissue samples of birds that have died in the Tuscany region of Italy in 1996 [10]. The virus has gradually spread across Hungary [11], Switzerland [12], Czech Republic [13], Italy [14], Germany [15], Croatia [16], Poland [17], Spain [18] and the UK [19, 20] mainly causing deaths of blackbirds and chicken. It is still not clear how the virus got introduced into Europe. Even if serological studies suggest that many avian species encountered USUV in Europe,





SCANDINAVIAN JOURNAL OF Immunology AN INTERNATIONAL MONTHLY JOURNAL

only certain bird species are susceptible, while others are asymptomatic carriers. The first human case of USUV infection in Europe was identified in 2009 in Italy from an immunocompromised patient with meningoencephalitis and other severe neuroinvasive symptoms [21]. In 2013, three more patients with neuroinvasive symptoms were diagnosed with USUV infection in Croatia [22]. Additionally, studies conducted in Germany from 4200 human blood donors showed that there is a low prevalence of USUV infection [23]. For more information, there is a review on Usutu virus by Ashraf *et al.*, 2015 [7].

Four assays are currently available to diagnose USUV in blood and cerebrospinal fluid. Two of these assays are based on real-time RT-PCR principle [24, 25]. One assay is based on ELISA principle, where IgG antibodies are used to capture USUV proteins [23]. Finally, there is plaque reduction neutralization test which is the standard assay for discriminating USUV from other flaviviruses such as West Nile virus and Zika virus [14, 20]. There is no drug or vaccine available to treat USUV infections [26]. As much is not known about this virus and human patients with USUV infections are being reported periodically, there is a need for further studies of the virus to find ways to limit its spread. Vaccination is obviously one of the main strategies to prevent the spread of a virus. In this study, we have used available computer tools and servers to predict potential B cell and T cell epitopes of USUV E protein which could be used for vaccine development. Fig. 1 provides an overview of this study. The envelope (E) protein of USUV, such as E protein of other viruses, aids the virus in binding to specific receptors, thereby initiating an infection. The E protein also plays a role in cellular tropism [27]. This study is similar to a small number of earlier studies where epitopes for human pathogenic viruses such as Zika virus [28, 29], dengue virus [30], Saint Louis encephalitis virus [31] and Chikungunya virus [32] have been predicted.

## Materials and methods

*USUV sequences retrieval*. USUV sequences (both nucleotide and protein) were retrieved from NCBI database. For accession numbers, see Fig. 2.

*Construction of phylogenetic tree*. Multiple sequence alignment of USUV nucleotide sequences (complete genome) was carried out using ClustalW function in MEGA 6.0 [33] with default parameters. Using the same software, a phylogenetic tree was constructed using an un-weighted pair of group method with arithmetic mean (UPGMA) algorithm, with default parameters.

*Multiple sequence alignment of protein sequences*. Multiple sequence alignment of USUV E protein sequences was carried out using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). Alignment figure was generated using ESPript 3.0 (http://espript.ibcp.fr/ESPript/ESPript/), a free online tool [34].

*Analysis of protein properties*. Protein properties such as amino acid composition, isoelectric point (pI), molecular weight, grand average of hydropathicity (GRAVY), estimated half-life and stability index were determined using ProtParam, a freely accessed online server (http://web.expasy.org/protparam/) [35]. The secondary structure of the protein was predicted from primary protein sequence using SOPMA online server

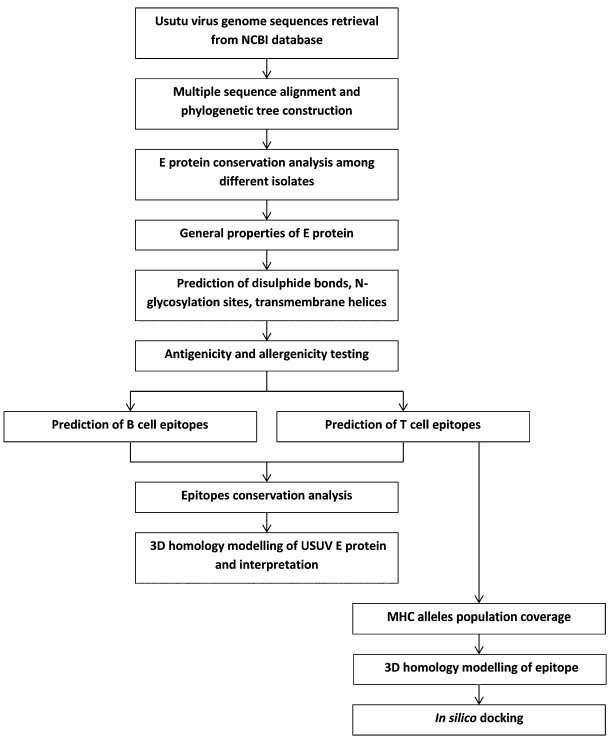


**Figure 1** Overview of the current study.

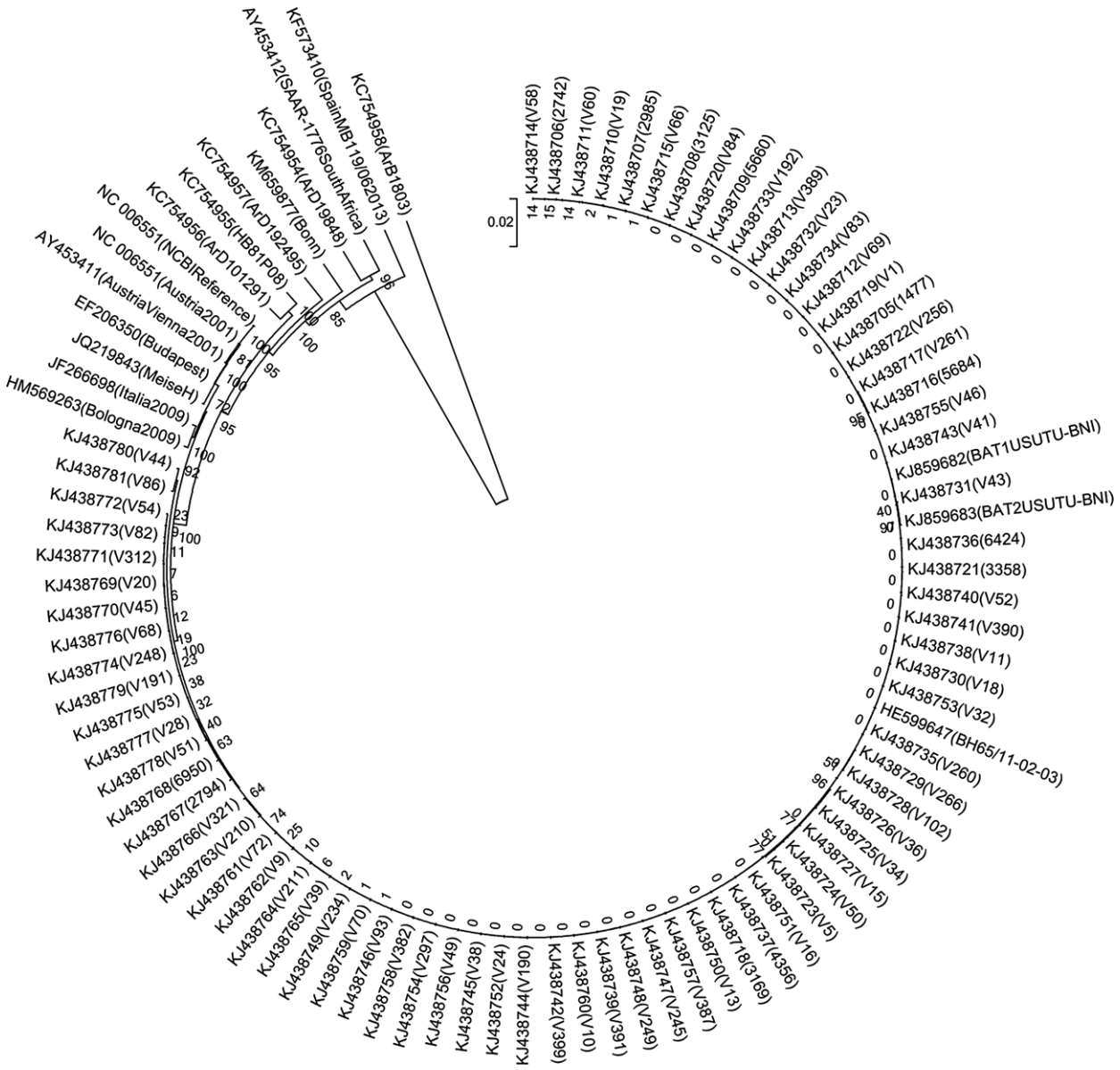


**Figure 2** Phylogenetic analysis of 95 Usutu virus isolates by an un-weighted pair of group method with arithmetic mean algorithm. Isolates in the figure are represented by their accession number and name.

(https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html), with default parameters.

*Prediction of transmembrane topology, disulphide bonds and N-glycosylation sites*. Transmembrane helices in protein were predicted using TMHMM server 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) [36] and MEMSAT-SVM [37] (a default inbuilt program in Phyre2). Disulphide bonds in protein were predicted using DiANNA 1.1 web server (http://clavius.bc.edu/~clotelab/DiANNA/) [38]. N-glycosylation sites in protein were predicted using NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/).

*Antigenicity prediction*. Antigenicity of protein was determined using VaxiJen 2.0 server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) [39]. Apart from protein sequence as input, a threshold of 0.5 was given as an additional input.

*Allergenicity prediction*. Allergenicity of protein was determined using AllerHunter online tool (http://tiger.dbs.nus.edu.sg/AllerHunter/) [40] and PREAL online tool (http://gmobl.sjtu.edu.cn/PREAL/index.php) [41].

*Prediction of B cell and T cell epitopes*. Linear B cell epitopes were predicted using ABCpred (http://www.imtech.res.in/raghava/abcpred/) [42] and BCPREDS (http://ailab.ist.psu.edu/bcpreds/predict.html) [43] servers. The predicted B cell epitopes were further confirmed using the following tools: (1) Bepipred linear epitope prediction [44],

(2) Chou and Fasman beta-turn prediction [45], (3) Emini surface accessibility prediction [46], (4) Karplus and Schulz flexibility prediction [47], (5) Kolaskar and Tongaonkar antigenicity [48] and (6) Parker hydrophilicity prediction [49]. All these tools are available at IEDB analysis resource (http://tools.immuneepitope.org/bcell/). Major histocompatibility complex (MHC) class I T cell epitopes were predicted using NetCTL (http://tools.iedb.org/stools/netchop/netchop.do?app=netchop) [50] and NetMHC (http://tools.iedb.org/mhci/) [51] servers available at IEDB. MHC class II T cell epitopes were predicted using ProPred server (http://www.imtech.res.in/raghava/propred/) [52].

*Modeling 3D structure of USUV E protein*. The 3D structure of USUV E protein was modeled using three different freely accessible online servers namely SWISSMODEL (https://swissmodel.expasy.org/) [53], Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) [54] and I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) [55]. While using SWISSMODEL and I-TASSER, the cryoelectron structure of Zika virus E protein (PDB ID: 5ire) was used as a template. Phyre2, on the other hand, used a combination of 3D structures that are already available in PDB, including the Zika virus E protein. The modeled structures were refined using ModRefiner (http://zhanglab.ccmb.med.umich.edu/ModRefiner/) [56]. The quality of the modeled structure was verified by Ramachandran plot, based on phi and psi angles, using the PROCHECK [57] function in PDBSUM tool (https://www.ebi.ac.uk/thorntonsrv/databases/pdbsum/Generate.html).

*Epitope conservation analysis*. The conservation of predicted epitopes among different isolates was analysed using IEDB epitope conservancy analysis tool (http://tools.immuneepitope.org/tools/conservancy/iedb_input) [58].

*Population coverage*. IEDB population coverage tool (http://tools.immuneepitope.org/tools/population/iedb_input) was used to predict the percentage of people that might respond to a given epitope based on HLA alleles data [59].

*Docking*. Crystal structure of HLA-B*35:01 was retrieved from protein data bank (PDB ID: 1zhk) [60]. A three-dimensional model of the peptide (epitope) was generated using a PEP-FOLD3 tool (http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/) [61]. Phyre2-generated complete protein model was used as a template in generating the epitope 3D model. *In silico* docking was performed using AutoDock Vina (http://vina.scripps.edu/) [62].

## Results

### Evolutionary relationship between different USUV isolates

Complete genome sequence of 95 USUV isolates (including the NCBI reference isolate) was retrieved from the NCBI database. A multiple sequence alignment of nucleotides (complete genome) was carried out using the ClustalW function in MEGA 6.0 using default parameters. A phylogenetic tree was constructed with 1000 bootstrap replications using the UPGMA statistical method in MEGA 6.0 (Fig. 2). From the phylogenetic analysis, we found that ArB1803 isolate (accession number: KC754958) was the farthest neighbour to rest of the isolates separated by an evolutionary distance of 0.12. This isolate, also called CAR_1969, was isolated in the Central African Republic. USUV infects both avian species and humans. The evolutionary relationship study gives an idea about how far the non-human and human USUV isolates differ genetically. This, in turn, will aid in the rational development of a vaccine that can be effective against most of the isolates.

Based on the evolutionary relationship, we chose seven isolates for the protein consensus studies. Isolate AustriaVienna2001 (accession number: AY453411) is also identical to the NCBI reference sequence. USUV E protein sequence from the seven isolates was retrieved, and a multiple sequence alignment was performed using the Clustal Omega online tool. The BLOSUM62 matrix was used to check for amino acid conservation. We found that the sequence had 93.4% identity (Fig. 3). We used the USUV E protein of AustriaVienna2001 isolate for all the analysis in this study. The GenBank ID was AAS59402.1.

### USUV E protein properties

USUV E protein was analysed using ExPASy ProtParam tool. The primary amino acid sequence of the protein was given as input. USUV E protein consisted of 500 amino acids with 9% of negatively charged amino acids (aspartate and glutamate) and 8.4% of positively charged amino acids (arginine and lysine). The molecular weight of the protein was about 53571 Da. The theoretical isoelectric point (pI) was estimated to be about 6.56, implying that protein is negatively charged at neutral pH. The extinction coefficient of the protein was predicted assuming water as solvent at 280 nm. The extinction coefficients were predicted to be 68130 and 67380 $M^{-1}$ $cm^{-1}$ when assuming all pairs of cysteines form cystines and when assuming all pairs of cysteines are reduced, respectively. The estimated half-lives of the protein were 1.1 h, 3 and 2 min in mammalian reticulocytes (*in vitro*), yeast (*in vivo*) and *Escherichia coli* (*in vivo*), respectively. The instability index was computed to be 24.79 classifying it to be a stable protein. The aliphatic index was estimated to be 80.16, and GRAVY was estimated to be −0.028.

The secondary structure of USUV E protein was analysed using SOPMA secondary structure prediction method. The primary sequence of the protein was given as input. Other parameters were left with default values. The protein was predicted to be made of 26.40% of alpha-helix (Hh), 29.20% of extended strand (Ee), 10.00% of beta-turn (Tt) and 34.40% of random coil (Cc).

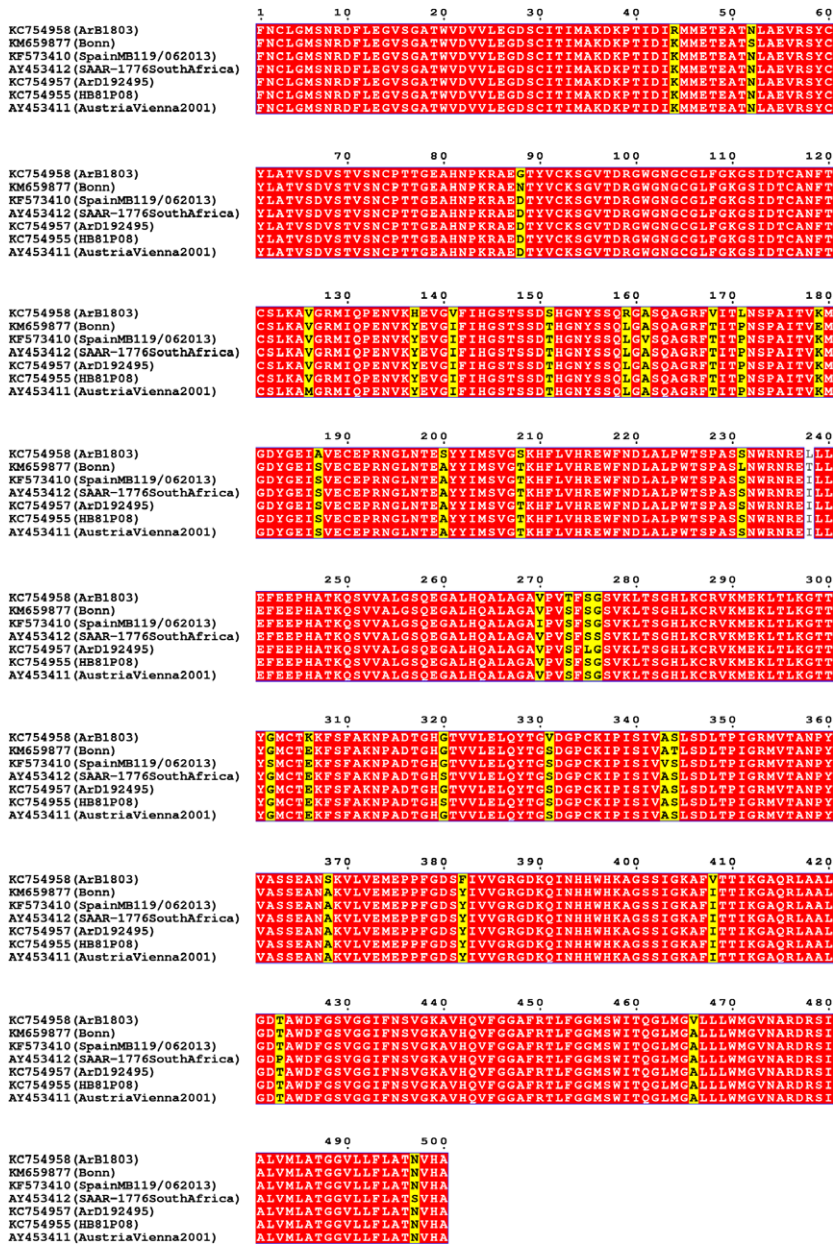


**Figure 3** Multiple sequence alignment of E protein from seven different Usutu virus isolates.

## Predicted disulphide bonds in USUV E protein

The disulphide bonds in protein were predicted using DiANNA 1.0 web server. With primary amino acid sequence as input, the server predicted six disulphide bonds. The server algorithm consists of five steps. In step 1, the submitted input sequence was run in PSI-BLAST. In step 2, the secondary structure of the protein was predicted using PSIPRED. In step 3, disulphide oxidation state was predicted. In step 4, disulphide bonds were predicted using a diresidue neural network. Finally, the predicted disulphide bonds were weighed by an Ed Rothberg's implementation of the Edmonds-Gabow maximum weight matching algorithm. The disulphide bonds were predicted to be formed between positions 3–335, 30–116, 60–92, 74–105, 121–287 and 190–304.

## Transmembrane helices and N-glycosylation sites in USUV E protein

We used TMHMM server to predict the transmembrane helices. This server predicted that amino acids from 451 to 473 and from 480 to 499 form transmembrane helices. The Phyre2 online tool has Memsat SVM as an inbuilt to predict the transmembrane helices (see 3D homology modeling part). Amino acids from 442 to 472 and from 481 to 496 were predicted to form transmembrane helices. We considered the amino acids predicted by both the tools as transmembrane forming.

Asparagine in N-X-S/T sequence, where N is asparagine, X is any amino acid, S is serine and T is threonine, is usually glycosylated in proteins produced by eukaryotes, archaea and very rarely bacteria. We used NetNGlyc 1.0 server to predict the N-glycosylation sites in USUV E protein. By default, 0.5 was considered as threshold potential. The server predicted two potential glycosylation sites. Asparagine at positions 118 (with sequence NFTC; N = asparagine, F = phenylalanine, T = threonine and C = cysteine) and at 154 (with sequence NYSS; N = asparagine, Y = tyrosine and S = serine) were predicted, with position 154 having the highest agreement among the two positions.

### Antigenicity and allergenicity of USUV E whole protein

Antigenicity of USUV E protein was determined using VaxiJen v2.0 tool. The primary sequence of the protein was submitted, and a threshold of 0.5 was given as an additional input. We also selected virus as the model. The tool predicted that USUV E protein can be an antigen with an overall antigen prediction score of 0.6329.

Allergenicity of USUV E protein was determined using AllerHunter and PREAL online tools. For both the tools, the primary sequence of the protein in FASTA was given as input. AllerHunter tool predicted the protein to be a non-allergen with a score of 0.0 (SE = 91.6%, SP = 89.3%). PREAL tool predicted the protein to be non-allergen as well with a probability of 0.986392.

Antigenicity and allergenicity information are helpful if one plans to deliver the complete protein as a vaccine using the unilamellar liposomes.

### Predicted linear B cell epitopes of USUV E protein

Linear B cell epitopes of USUV E protein were predicted using ABCpred and BCPREDS online servers. For ABCpred, the primary sequence of the protein in plain format was given as input. Additionally, we gave 0.51 as a threshold and 16 amino acids in window length, and the overlapping filter was kept 'ON'. We considered only amino acids till 450 for epitope prediction as amino acids between 451 and 499 were found to form transmembrane helices. ABCpred predicted several epitopes and ranked them using trained recurrent neural network. Based on the ABCpred ranking, epitope with the sequence YGEISVECEPRNGLNT had the highest score (Table 1). Seven of 53 epitopes predicted by ABCpred were from the transmembrane region. For BCPREDS, similar to ABCpred, the primary sequence of the protein in plain format was given as input. Additionally, we restricted the server to perform fixed length epitope prediction of length 16 amino acids. We also provided 75% as specificity and asked the server to report only non-overlapping epitopes. The server predicted 11 potential epitopes. Based on BCPREDS score, epitope with the sequence GASQAGRFTITPNSPA had the highest (Table 2). Of the 11 predicted epitopes, one is from the transmembrane region.

An ensemble of factors determines the potentiality for B cell epitopes (e.g. surface accessibility, hydrophilicity, secondary structures, flexibility and antigenicity). The predicted B cell epitope(s) should have good surface accessibility, antigenicity, flexibility and hydrophilicity, and also should have beta-turn motif forming amino acid residues. We used (1) Kolaskar and Tongaonkar antigenicity, (2) Chou and Fasman beta-turn prediction, (3) Emini surface accessibility prediction, (4) Karplus and Schulz flexibility prediction, (5) Parker hydrophilicity prediction and (6) Bepipred linear epitope prediction tools available at IEDB analysis resource to predict and validate (against ABCpred and BCPRED) the epitopes. For all the aforementioned tools, the primary amino acid sequence of USUV E protein was submitted as input. The window length was adjusted to 16 (except in Karplus and Schulz flexibility prediction where the window length was 7). An ideal B cell epitope should (1) be antigenic, (2) have beta-turn, (3) be accessible, (4) be flexible and (5) be hydrophilic. Combining the results of ABCpred, BCPRED, Bepipred (Figure S1F), antigenicity (Figure S1A), beta-turn (Figure S1B), surface accessibility (Figure S1C), flexibility (Figure S1D) and hydrophilicity (Figure S1E), we found that amino acids between 68 and 84 (i.e. VSTVSNCPTTGEAHNP) could be a potential B cell epitope.

The predicted peptide VSTVSNCPTTGEAHNP was found to be conserved in 94/94 (i.e. in all) isolates.

### Predicted T cell epitopes of USUV E protein

For predicting the MHC class I T cell epitopes, we first used the NetCTL server at IEDB analysis resource. The primary protein sequence was given as input. Weights for C-terminal cleavage and TAP transport efficiency were left with default values, that is 0.15 and 0.05, respectively. We chose all the supertypes (i.e. 12 supertypes) for the study. For threshold, a value of 0.5 was given as input because at that threshold epitopes can be screened with 89% sensitivity and 94% specificity. Of the several epitopes predicted by the NetCTL server, we considered only the top 15 epitopes based on the NetCTL score. Table 3 lists the predicted MHC class I T cell epitopes by NetCTL. The epitopes are ordered based on the descending order of the NetCTL score. The NetCTL score is based on the integration of MHC class I binding, proteasomal cleavage and TAP transport efficiency scores. MHC class I binding and proteasomal cleavage were predicted by the artificial neural network, while TAP transport efficiency was predicted using a weight matrix. To filter antigenic epitopes from non-antigenic epitopes, we again used Kolaskar and Tongaonkar antigenicity prediction tool at IEDB analysis resource, but this time

**Table 1** Linear B cell epitopes of USUV E protein predicted using ABCpred.

| S.No. | ABCpred Rank | Sequence | Start position | ABCpred score |
|---|---|---|---|---|
| 1 | 1 | YGEISVECEPRNGLNT | 183 | 0.99 |
| 2 | *2* | *GMSWITQGLMGALLLW* | *455* | *0.97* |
| 3 | 3 | PKRAEDTYVCKSGVTD | 83 | 0.92 |
| 4 | 4 | VLEGDSCITIMAKDKP | 24 | 0.91 |
| 5 | 5 | CKSGVTDRGWGNGCGL | 92 | 0.89 |
| 6 | 5 | TTYGMCTEKFSFAKNP | 299 | 0.89 |
| 7 | 6 | SNCPTTGEAHNPKRAE | 72 | 0.88 |
| 8 | 6 | CITIMAKDKPTIDIKM | 30 | 0.88 |
| 9 | 6 | GIFIHGSTSSDTHGNY | 140 | 0.88 |
| 10 | 6 | KGSIDTCANFTCSLKA | 110 | 0.88 |
| 11 | 7 | LTPIGRMVTANPYVAS | 348 | 0.87 |
| 12 | 8 | RGWGNGCGLFGKGSID | 99 | 0.84 |
| 13 | 8 | ALPWTSPASSNWRNRE | 222 | 0.84 |
| 14 | 8 | TVKMGDYGEISVECEP | 177 | 0.84 |
| 15 | 9 | CYLATVSDVSTVSNCP | 60 | 0.83 |
| 16 | 9 | SVGTKHFLVHREWFND | 205 | 0.83 |
| 17 | 10 | AALGDTAWDFGSVGGI | 418 | 0.82 |
| 18 | 11 | GMSNRDFLEGVSGATW | 5 | 0.81 |
| 19 | 11 | SSIGKAFITTIKGAQR | 401 | 0.81 |
| 20 | 11 | NGLNTEAYYIMSVGTK | 194 | 0.81 |
| 21 | *12* | *VGKAVHQVFGGAFRTL* | *437* | *0.80* |
| 22 | 13 | PPFGDSYIVVGRGDKQ | 376 | 0.79 |
| 23 | 13 | HGTVVLELQYTGSDGP | 319 | 0.79 |
| 24 | 13 | TCSLKAMGRMIQPENV | 120 | 0.79 |
| 25 | 14 | SQAGRFTITPNSPAIT | 162 | 0.78 |
| 26 | 14 | SSQLGASQAGRFTITP | 156 | 0.78 |
| 27 | 14 | STSSDTHGNYSSQLGA | 146 | 0.78 |
| 28 | 15 | TIDIKMMETEATNLAE | 40 | 0.77 |
| 29 | 15 | KVLVEMEPPFGDSYIV | 369 | 0.77 |
| 30 | 15 | REILLEFEEPHATKQS | 236 | 0.77 |
| 31 | 16 | ITTIKGAQRLAALGDT | 408 | 0.76 |
| 32 | 16 | EKFSFAKNPADTGHGT | 306 | 0.76 |
| 33 | 17 | SFSGSVKLTSGHLKCR | 273 | 0.75 |
| 34 | 18 | GDKQINHHWHKAGSSI | 388 | 0.74 |
| 35 | 18 | YIVVGRGDKQINHHWH | 382 | 0.74 |
| 36 | 18 | SGHLKCRVKMEKLTLK | 282 | 0.74 |
| 37 | *19* | *VFGGAFRTLFGGMSWI* | *444* | *0.73* |
| 38 | 19 | DGPCKIPISIVASLSD | 332 | 0.73 |
| 39 | 20 | ANPYVASSEANAKVLV | 357 | 0.72 |
| 40 | 20 | GSQEGALHQALAGAVP | 256 | 0.72 |
| 41 | *21* | *GLMGALLLWMGVNARD* | *462* | *0.71* |
| 42 | 22 | ETEATNLAEVRSYCYL | 47 | 0.70 |
| 43 | 22 | GATWVDVVLEGDSCIT | 17 | 0.70 |
| 44 | 23 | NPADTGHGTVVLELQY | 313 | 0.69 |
| 45 | 23 | AGAVPVSFSGSVKLTS | 267 | 0.69 |
| 46 | *24* | *LVMLATGGVLLFLATN* | *482* | *0.67* |
| 47 | *25* | *LWMGVNARDRSIALVM* | *469* | *0.66* |
| 48 | *26* | *SVGGIFNSVGKAVHQV* | *429* | *0.64* |
| 49 | 27 | EPHATKQSVVALGSQE | 244 | 0.61 |
| 50 | 28 | ENVKYEVGIFIHGSTS | 133 | 0.60 |
| 51 | 29 | HREWFNDLALPWTSPA | 214 | 0.56 |
| 52 | 30 | SSEANAKVLVEMEPPF | 363 | 0.54 |
| 53 | 30 | PASSNWRNREILLEFE | 228 | 0.54 |

Epitopes overlapping with the transmembrane regions are represented in italics.

the window length was set to 9. From Kolaskar and Tongaonkar antigenicity prediction (Figure S2), we could see that 10 of top 15 epitopes (based on NetCTL score) were non-antigenic. We considered the five antigenic peptides for further analysis.

**Table 2** Linear B cell epitopes of USUV E protein predicted using BCPREDS.

| S.No. | Amino acid position | Epitope sequence | BCPred score |
|---|---|---|---|
| 1 | 160 | GASQAGRFTITPNSPA | 0.999 |
| 2 | 143 | IHGSTSSDTHGNYSSQ | 0.997 |
| 3 | 68 | VSTVSNCPTTGEAHNP | 0.997 |
| 4 | 222 | ALPWTSPASSNWRNRE | 0.995 |
| *5* | *437* | *VGKAVHQVFGGAFRTL* | *0.99* |
| 6 | 97 | TDRGWGNGCGLFGKGS | 0.99 |
| 7 | 325 | ELQYTGSDGPCKIPIS | 0.99 |
| 8 | 370 | VLVEMEPPFGDSYIVV | 0.988 |
| 9 | 350 | PIGRMVTANPYVASSE | 0.964 |
| 10 | 306 | EKFSFAKNPADTGHGT | 0.96 |
| 11 | 398 | KAGSSIGKAFITTIKG | 0.925 |

Epitope overlapping with the transmembrane region is represented in italics.

Although NetCTL integrates NetMHC algorithm, we found that it was difficult to interpret epitopes and predicted binding affinities. NetMHC output data were, on the other hand, easy to interpret. So, we used NetMHC server as well at IEDB analysis resource. The primary amino acid sequence of the protein was given as input in FASTA format. SMM was used as the prediction method. Human MHC class I alleles were chosen for the binding prediction. We restricted the MHC class I alleles with $IC_{50}$ less than or equal to 5000 nM. Table 4 lists the predicted MHC class I alleles for the potential epitopes. The epitope with the sequence ATKQSVVAL (spanning between 247 and 255) was conserved in all isolates and was predicted to bind to 18 MHC class I alleles. Epitopes GTVVLELQY and GHGTVVLEL spanning between positions 318 and 328 were overlapping, and together, they were found to bind to 24 MHC class I alleles. But, the amino acid at position 320 was not conserved in a few isolates. Epitopes LAEVRSYCY and AEVRSYCYL spanning between positions 53 and 62 were conserved in all isolates and together were found to bind to 24 MHC class I alleles. Considering the conservation, antigenicity, proteasomal cleavage, TAP transport efficiency and a number of MHC class I alleles binding, epitopes with the sequences LAEVRSYCY and AEVRSYCYL were considered as the potential MHC class I T cell candidates. The 24 MHC class I alleles that were predicted to bind to LAEVRSYCY and AEVRSYCYL epitopes were further studied for population coverage using IEDB population coverage tool. As USUV is currently a threat in Europe and Africa, we considered Black and Caucasoid populations for the population coverage study. The predicted 24 MHC class I alleles were given as input. We found that nearly 86.44% of the Black population and 96.90% of the Caucasoid population would respond to the predicted epitope.

**Table 3** T cell epitopes of USUV E protein predicted using NetCTL.

| S.No. | Position | Peptide | NetCTL score | Proteasomal cleavage score | TAP transport efficiency |
|---|---|---|---|---|---|
| 1 | 147-155 | TSSDTHGNY | 4.1276 | 0.9990 | 2.9940 |
| 2 | 352-360 | GRMVTANPY | 2.7162 | 0.9957 | 3.0260 |
| 3 | 93-101 | KSGVTDRGW | 2.7116 | 0.0382 | 0.8620 |
| *4* | *450-458* | *RTLFGGMSW* | *2.6839* | *0.8809* | *1.0200* |
| 5 | 225-233 | WTSPASSNW | 2.6698 | 0.9568 | 0.6870 |
| 6 | 299-307 | TTYGMCTEK | 2.2148 | 0.9318 | 0.5790 |
| 7 | 247-255 | ATKQSVVAL | 2.0828 | 0.9891 | 1.0520 |
| 8 | 76-84 | TTGEAHNPK | 2.0481 | 0.9921 | 0.1140 |
| 9 | 320-328 | GTVVLELQY | 1.8758 | 0.9983 | 2.7790 |
| 10 | 318-326 | GHGTVVLEL | 1.8619 | 0.9991 | 0.6300 |
| 11 | 53-61 | LAEVRSYCY | 1.8566 | 0.9474 | 2.9780 |
| 12 | 54-62 | AEVRSYCYL | 1.8422 | 0.9741 | 1.0890 |
| 13 | 151-159 | THGNYSSQL | 1.8355 | 0.9990 | 0.6390 |
| 14 | 148-156 | SSDTHGNYS | 1.7770 | 0.0195 | −2.3180 |
| 15 | 12-20 | LEGVSGATW | 1.7119 | 0.0472 | 0.4850 |

Epitope overlapping with the transmembrane region is represented in italics.

**Table 4** MHC class I alleles predicted using NetMHC to bind to the potential epitopes of USUV E protein.

| | | | MHC class I alleles (based on binding affinity) | | |
|---|---|---|---|---|---|
| S.No. | Position | Peptide | <50 nM | 51–500 nM | 501–5000 nM |
| 1 | 247-255 | ATKQSVVAL | HLA-C*12:03, HLA-C*03:03, HLA-B*15:02 | HLA-A*30:01, HLA-C*14:02 | HLA-B*15:01, HLA-C*07:02, HLA-A*32:01, HLA-B*58:02, HLA-B*07:02, HLA-B*08:01, HLA-C*08:02, HLA-C*15:02, HLA-A*02:06, HLA-A*68:02, HLA-B*35:03, HLA-B*14:02, HLA-C*06:02 |
| 2 | 320-328 | GTVVLELQY | HLA-C*12:03 | HLA-A*29:02, HLA-C*03:03, HLA-C*14:02, HLA-B*15:02, HLA-B*58:01, HLA-A*30:02 | HLA-A*01:01, HLA-C*07:02, HLA-B*57:01, HLA-B*15:01, HLA-A*11:01, HLA-C*05:01, HLA-C*15:02, HLA-A*26:01, HLA-B*58:02, HLA-B*35:01, HLA-A*68:01, HLA-B*44:03 |
| 3 | 318-326 | GHGTVVLEL | - | HLA-C*12:03, HLA-C*03:03, HLA-B*15:02, HLA-C*07:02, HLA-C*14:02 | HLA-B*38:01, HLA-B*39:01, HLA-C*15:02, HLA-C*05:01, HLA-C*07:01, HLA-C*08:02, HLA-B*14:02, HLA-B*58:02 |
| 4 | 53-61 | LAEVRSYCY | HLA-C*12:03, HLA-C*03:03, HLA-C*05:01, HLA-B*35:01 | HLA-B*15:02, HLA-A*30:02, HLA-C*07:01, HLA-C*14:02 | HLA-C*15:02, HLA-A*01:01, HLA-A*29:02, HLA-B*08:01, HLA-C*06:02, HLA-C*07:02, HLA-B*15:01, HLA-B*53:01, HLA-B*14:02 |
| 5 | 54-62 | AEVRSYCYL | HLA-C*03:03, HLA-B*15:02, HLA-B*40:01 | HLA-C*12:03, HLA-B*40:02, HLA-B*44:03, HLA-C*07:02 | HLA-B*44:02, HLA-C*14:02, HLA-C*15:02, HLA-A*02:06, HLA-B*18:01, HLA-C*08:02 |

HLA, human leucocyte antigen; MHC, major histocompatibility complex.

For predicting the epitopes binding to MHC class II alleles, we used ProPred online server. The server predicts the MHC class II binding regions from the primary amino acid sequence using quantitative matrices. The primary amino acid sequence of USUV E protein was given as input in FASTA format. Additionally, the threshold was set to 4%. Fig. 4 shows the predicted epitopes with their potential MHC class II binding partners. The epitope with the sequence LVMLATGGV was restricted by all the MHC class II alleles found at the ProPred server. But, this epitope was located in the transmembrane region of the E protein. The epitope YYIMSVGTK was restricted by 33 MHC class II alleles. This epitope was found to be conserved in 93/94 isolates. The epitope with amino acids

between positions 56 and 69 was also restricted by nearly 20 MHC class II alleles.

### 3D model of USUV E protein

We wanted to study the locations of potential epitopes in the 3D structure of USUV E protein. A 3D crystal or cryoelectron microscopy structure of USUV E protein is not yet available. We used three different web servers namely SWISSMODEL, Phyre2 and I-TASSER to model the 3D protein structure. We used three different servers simply because each of these servers uses different algorithms. By progressive alignment ordering, we found that USUV E protein closely resembles other flaviviruses in the following order: Murray Valley encephalitis virus > Saint Louis encephalitis virus > Zika virus > West Nile virus > Japanese encephalitis virus > yellow fever virus > dengue virus type 2 > tick-borne encephalitis

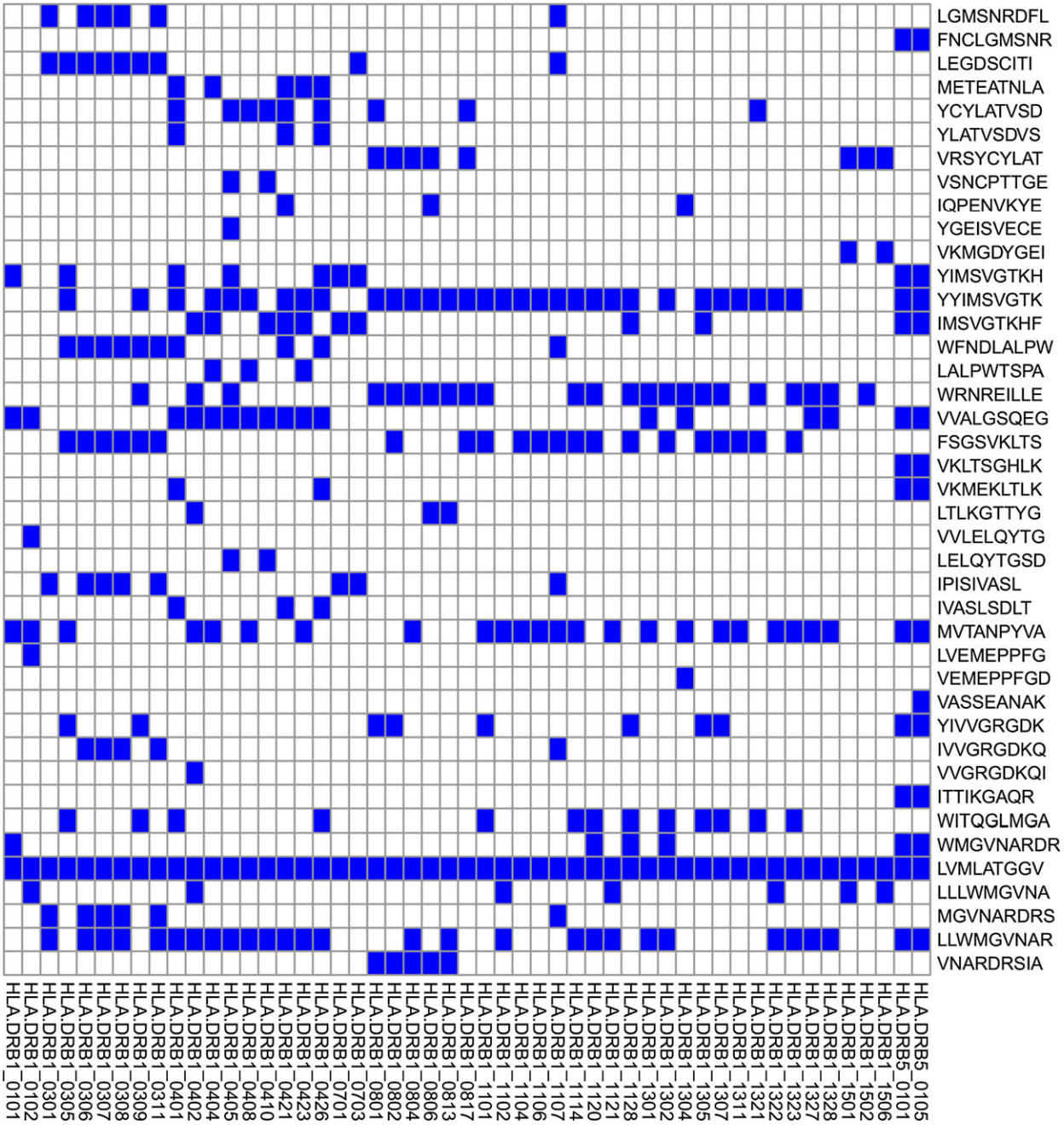


**Figure** 4 Prediction of major histocompatibility complex (MHC) class II-restricted T cell epitopes of Usutu virus E protein using ProPred online server (http://www.imtech.res.in/raghava/propred/). The epitope and its MHC class II binding partner(s) are represented by a blue square.

virus. Additionally, considering the availability of structures, identity (%) and coverage, we found that Zika virus E protein could act as a template in modeling the USUV E protein structure. So, while modeling using SWISSMODEL and I-TASSER, Zika virus E protein (PDB: 5ire) electron microscopy structure was used as a template (Fig. 5A). Phyre2, on the other hand, has HHpred 1.51 inbuilt script, which automatically detected different available PDB structures including Zika virus E protein as a template. The modeled structures were refined using ModRefiner online server. We compared the quality of the generated models by Ramachandran plot. For this, the generated 3D models were submitted in PDBSUM tool. Three-dimensional model by all the aforementioned servers was quite identical. Based on the phi and psi angles, the model generated by Phyre2 was better among the three with 88.7% of residues in most favoured regions, 8.7% of the residues in additional allowed regions, 1.2% of residues in generously allowed regions and 1.4% of residues in disallowed regions (Fig. 5C). We further validated the model using Verify_3D web server (http://services.mbi.ucla.edu/Verify_3D/). We found that 84.20% of the residues had an averaged 3D-1D score $\geq$ 0.2 (i.e. pass). Our model can be accessed from the protein model database (PMDB, Link: https://bioinformatics.cineca.it/PMDB/) with ID PM0080679. We found that the predicted B cell epitope was predominantly found in the coil region, while the predicted T cell epitope was found in the beta-strand ($\beta$-strand) region.

## Docking

The peptide sequence LAEVRSYCY was considered for docking studies. Three-dimensional structure of the peptide was generated using PEP-FOLD3 via RPBS Mobyle portal. The sequence of the peptide was given as input. Additionally, we gave the Phyre2-generated protein model (complete protein) as a reference input. Other parameters were left with default values. The structure prediction program defines conformation of four consecutive residues with structural alphabets. These structural alphabets series are coupled to a greedy algorithm and a coarse-grained force field. By this procedure, several viable models are generated with model 1 being the best model in most cases.

From T cell MHC class I prediction, LAEVRSYCY peptide was predicted to bind strongly to HLA-B*35:01 allele. Of all the strong binders (i.e. affinity <50 nM), we chose HLA-B*35:01 allele because the crystal structure with higher resolution was readily available and an earlier study had also used the same allele for docking [32]. Three-dimensional structure of this MHC molecule (PDB

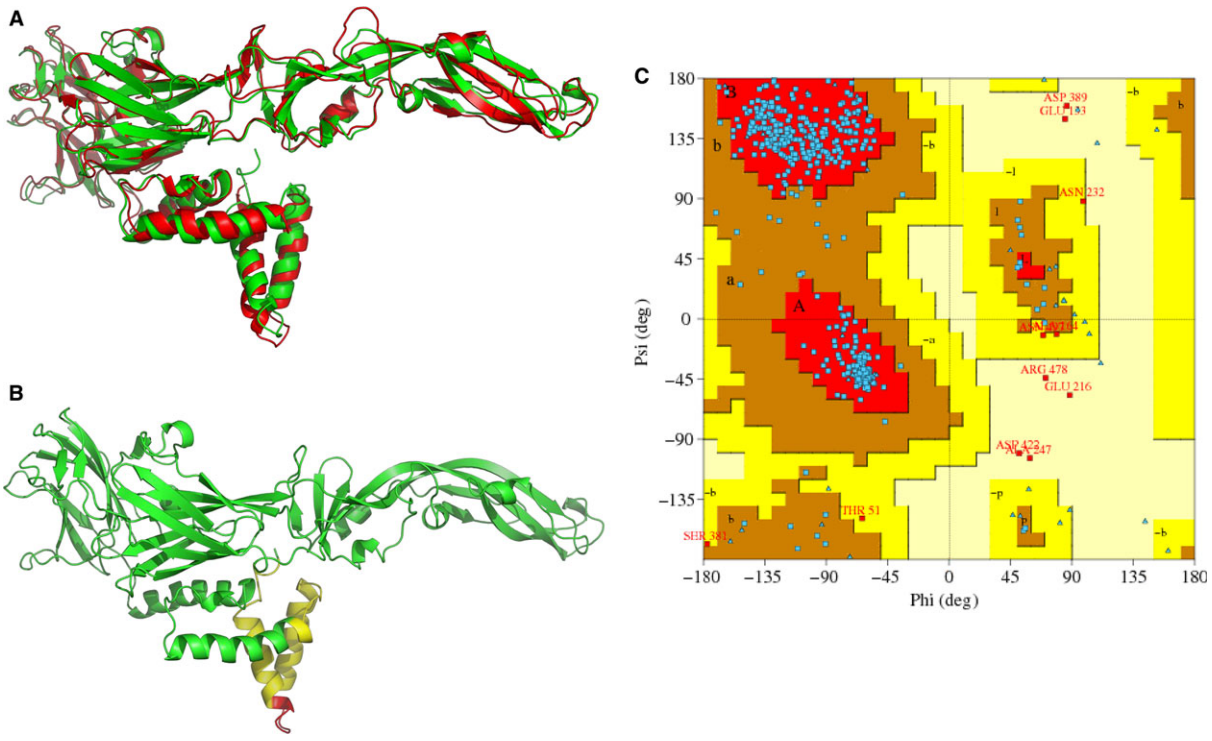


**Figure 5** (A) Three-dimensional structure comparison between cryoelectron microscopic Zika virus E protein and *in silico* Phyre2-modeled Usutu virus (USUV) E protein. (B) Three-dimensional model of USUV E protein with residues traversing through the membrane is coloured in yellow. (C) Ramachandran plot for USUV E protein model. The plot was generated using PROCHECK in PDBSUM (https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html).

ID: 1zhk) was retrieved from protein data bank. The structure was resolved at 1.6 Å. This molecule was co-crystallized with a 13-mer epitope from lytic antigen BZLF1 of Epstein–Barr virus. Prior to docking, this 13-mer epitope was removed. Using AutoDock tools, hydrogen atoms were added to polar groups in the protein macromolecule. Additionally, the grid size was visualized. The peptide was dock-prepped using UCSF Chimera. AutoDock Vina was used for docking. We gave the following search space x = 50, y = 40 and z = 34 (in Å), and centre x = 20.651, y = 11.818 and z = 21.814. As the search space was greater than 30*30*30 $Å^3$, we increased the exhaustiveness to 24. The binding affinity of LAEVRSYCY peptide with HLA-B*35:01 was found to be −7.8 kcal/mol. Fig. 6 shows LAEVRSYCY peptide docked with HLA-B*35:01. As a control, we also docked the Epstein–Barr virus peptide with HLA-B*35:01. For this, the binding affinity was found to be slightly higher (i.e. −8.5 kcal/mol).

## Discussion

The spread of USUV is increasing steadily among the avian species in Europe. The virus is spreading between birds by mosquitos. Due to global warming, mosquito populations have started to thrive in new places previously not supportive for their growth. Controlling mosquitos help to prevent the spread of this virus. Humans are assumed to be the dead-end host for this virus [63]. USUV in many cases co-exists with other members of *Flaviviridae* family like West Nile virus and is therefore poorly diagnosed [64]. As much is not known about this virus, there is a need for further studies to stop the spread of USUV infection.

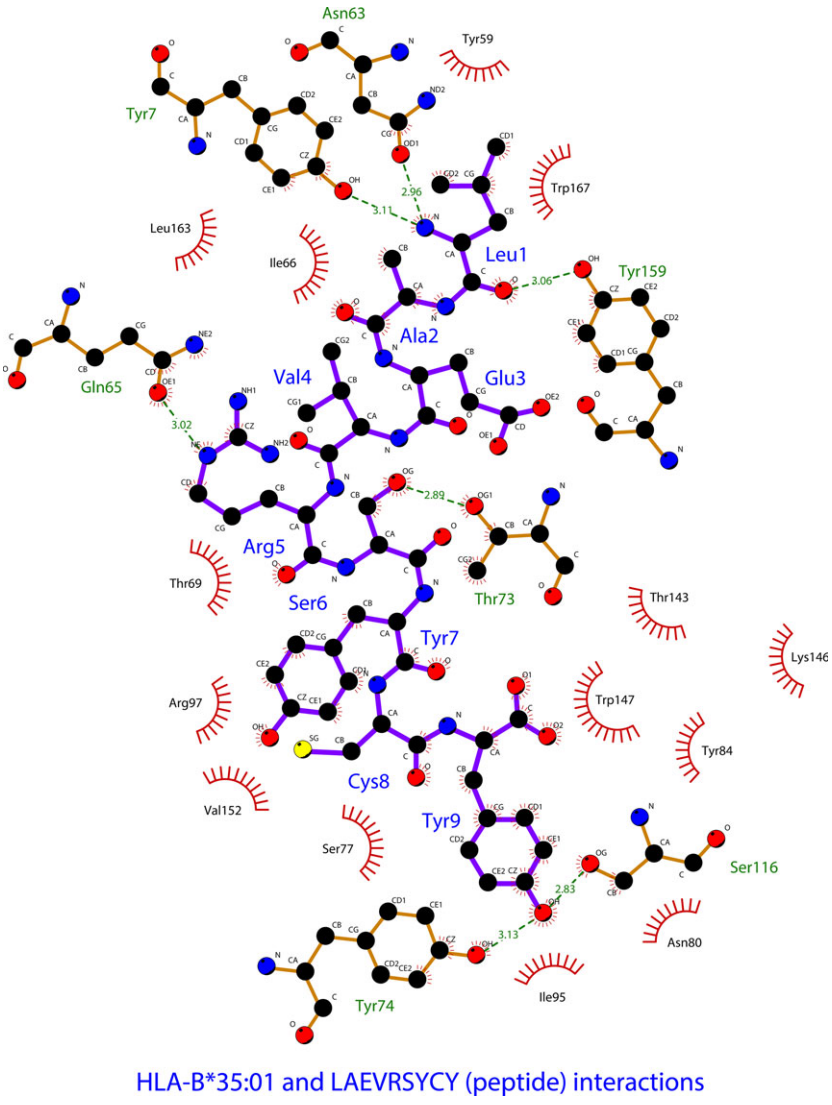


**Figure 6** Docking of LAEVRSYCY peptide with human leucocyte antigen (HLA)-B*35:01. The *in silico* docking was performed using AutoDock Vina (http://vina.scripps.edu/). The hydrogen bonds are represented by green dotted lines, and the hydrophobic interactions are represented by arcs with lines. The image was generated using LigPlot+ (https://www.ebi.ac.uk/thornton-srv/software/LigPlus/download.html). The residues in chain A of HLA-B*35:01 interact with the peptide (epitope).

In the present study, we have used available computational tools to predict potential B cell and T cell epitopes for USUV E protein. This is the first ever study for USUV in this endeavour. We focused on the envelope protein (E) of the virus in the present study because it is exposed on the surface of the virus and can be intercepted by the host immune system. We have predicted that amino acids between positions 68 and 84 could be a potential B cell epitope, while amino acids between positions 53 and 69 could be a potential T cell MHC class I- and class II-restricted epitope. Both the B and T cell epitopes were conserved in all isolates. Recently, the 3D structure of Zika virus envelope (E) protein bound to broadly protective antibody of Flavivirus has been published [65]. The B cell epitope of USUV E protein predicted in our study matches to the antibody-binding region in Zika virus E protein. This further validates our predicted B cell epitope. Population coverage using the MHC class I-restricted peptide showed that nearly 86.44% of the Black population and 96.90% of the Caucasoid population respond to the predicted epitope.

We used DiANNA 1.0 web server to predict the disulphide bonds between cysteines in the USUV E protein. We found that there was a disulphide bond between cysteines 74 and 105 in the protein. As the potential B cell epitope spans amino acids between 68 and 84, further studies are needed regarding the role of this particular disulphide bond in the protein structure.

We used the E protein sequence from the AustriaVienna2001 isolate (accession number: AY453411) for the current prediction study. The isolate Bologna2009 (accession number: HM569263) was known to cause neuroinvasive infection in humans. The E protein sequence of these two isolates differs by a single amino acid at position 302, where glycine was mutated to serine in the Bologna2009 isolate.

After a mosquito bite, the first cell to get infected with USUV is still unknown. It is hypothesized that the resident dendritic cells of the skin (i.e. the Langerhans cells) are the first to get infected [66]. A study has reported that USUV can infect immature dendritic cells and can induce interferon-alpha [66]. However, the replication kinetics of the virus was poor compared to West Nile virus [66]. A few research groups are currently working on developing laboratory animal models to study USUV virus. Adult mice are generally resistant to USUV infection but not the suckling mice [67]. A recent study has shown that adult mice lacking interferon receptor alpha and beta are susceptible to USUV infections [67]. The group also tested a DNA vaccine expressing the USUV envelope and pre-M proteins and found that interferon receptor-deficient adult mice could be protected against USUV infections [67].

So far, USUV infections with neuroinvasive symptoms are reported only in immunocompromised patients. To treat such patients, antibodies generated against USUV from other animals or humans can be artificially transfused; thereby, they acquire passive immunity. For controlling USUV infections, for example domestic birds or birds protected in sanctuaries, vaccination would be a better option than treatment with antiviral drugs. Avian species also possess B cells and T cells. The epitopes predicted in our study can also be tested in birds to check for their protection capabilities.

There have been several claims regarding cross-protection against antigenically similar viruses [68]. However, in some antigenically similar flaviviral infections, it can rather be a bane than a boon. An example is the dengue virus with four subtypes. Infection in a human with one subtype gives lifelong immunity against that particular subtype. On the other hand, it makes the person more susceptible to dengue virus infections to other subtypes due to cross-reactivity of the antibodies resulting in antibody-dependent enhancement of the viral uptake by the dendritic cells [68]. The E protein of West Nile virus and USUV have a sequence identity of approx. 77.26% and have a sequence similarity of approx. 85.02% (alignment not shown in this study). It has been shown that USUV-infected mice have cross-protection against heterologous neurovirulent West Nile virus, but not vice versa [26]. Moreover, West Nile virus and USUV interact differently with the host immune response [26]. The predicted B cell and T cell epitopes of USUV E protein have 69% and 82.4% sequence identities, respectively, with E protein of West Nile virus. Hence, the cross-protection of predicted epitopes can be also studied.

Several factors come into play to make a vaccination project successful. One of the major contributing factors is the accessibility. The inactivated viral vaccines and recombinant protein vaccines need to be stored at lower temperatures. Many remote places cannot afford cold storage units. The cost of the vaccine is majorly determined by the transportation costs. Peptide vaccines, on the other hand, are easy to synthesize and could be stored at room temperature or refrigerator. These advantages will greatly reduce the cost of the vaccine and will enhance the accessibility. The peptide vaccines also have a negative side. Epitope scanning of complete protein experimentally is cumbersome and expensive. *In silico* epitope prediction studies, like ours, aid in reducing the cost and time in identifying the potential epitopes. Using *in silico* prediction methods, several potential epitopes have been tested and verified (both *in vitro* and *in vivo*) and are in clinical trials now. For more information, there is a review by Soria-Guerra [69] on computational epitope prediction and implications on vaccine development.

## Acknowledgment

N.P. would like to thank Rabea Binte Akram for her moral support. N.P. would like to extend his gratitude to Mr. N. Sezhian (Senior General Manager) and EDP section of

Sakthi Sugars Limited, Sakthi Nagar, India, for providing the computational facility.

## Funding source

None.

## Conflict of interest

None.

## Author's contributions

N.P. conceived the idea, designed the study, performed the computational analysis and generated the results. N.P. and J.L. wrote the manuscript together. Both the authors read the manuscript and approved it.

## References

1 Kuno G, Chang GJ, Tsuchiya KR, Karabatsos N, Cropp CB. Phylogeny of the genus flavivirus. *J Virol* 1998;72:73–83.
2 Williams MC, Simpson DI, Haddow AJ, Knight EM. The isolation of West Nile virus from man and of Usutu virus from the bird-biting mosquito *Mansonia aurites* (Theobald) in the Entebbe area of Uganda. *Ann Trop Med Parasitol* 1964;58:367–74.
3 Bakonyi T, Gould EA, Kolodziejek J, Weissenböck H, Nowotny N. Complete genome analysis and molecular characterization of Usutu virus that emerged in Austria in 2001: comparison with the South African strain SAAR-1776 and other flaviviruses. *Virology* 2004; 328:301–10.
4 Brinton MA, Fernandez AV, Dispoto JH. The 3′-nucleotides of flavivirus genomic RNA form a conserved secondary structure. *Virology* 1986;153:113–21.
5 Brinton MA, Dispoto JH. Sequence and secondary structure analysis of the 5′-terminal region of flavivirus genome RNA. *Virology* 1988; 162:290–9.
6 Pijlman GP, Funk A, Kondratieva N *et al.* A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity. *Cell Host Microbe* 2008;4:579–91.
7 Ashraf U, Ye J, Ruan X, Wan S, Zhu B, Cao S. Usutu virus: an emerging flavivirus in Europe. *Viruses* 2015;7:219–38.
8 Weissenböck H, Kolodziejek J, Url A, Lussy H, Rebel-Bauder B, Nowotny N. Emergence of Usutu virus, an African mosquito-borne flavivirus of the Japanese encephalitis virus group, central Europe. *Emerging Infect Dis* 2002;8:652–6.
9 Chvala S, Bakonyi T, Bukovsky C *et al.* Monitoring of Usutu virus activity and spread by using dead bird surveillance in Austria, 2003-2005. *Vet Microbiol* 2007;122:237–45.
10 Mani P, Rossi G, Perrucci S, Bertini S. Mortality of *Turdus merula* in Tuscany. *Sel Vet* 1998;8:749–53.
11 Bakonyi T, Erdélyi K, Ursu K *et al.* Emergence of Usutu virus in Hungary. *J Clin Microbiol* 2007;45:3870–4.
12 Steinmetz HW, Bakonyi T, Weissenböck H *et al.* Emergence and establishment of Usutu virus infection in wild and captive avian species in and around Zurich, Switzerland–genomic and pathologic comparison to other central European outbreaks. *Vet Microbiol* 2011;148:207–12.
13 Hubálek Z, Rudolf I, Čapek M, Bakonyi T, Betášová L, Nowotny N. Usutu virus in blackbirds (*Turdus merula*), Czech Republic, 2011–2012. *Transbound Emerg Dis* 2014;61:273–6.
14 Lelli R, Savini G, Teodori L *et al.* Serological evidence of Usutu virus occurrence in north-eastern Italy. *Zoonoses Public Health* 2008;55:361–7.
15 Jöst H, Bialonski A, Maus D *et al.* Isolation of Usutu virus in Germany. *Am J Trop Med Hyg* 2011;85:551–3.
16 Barbic L, Vilibic-Cavlek T, Listes E *et al.* Demonstration of Usutu virus antibodies in horses. *Croatia. Vector Borne Zoonotic Dis* 2013;13:772–4.
17 Hubálek Z, Wegner E, Halouzka J *et al.* Serologic survey of potential vertebrate hosts for West Nile virus in Poland. *Viral Immunol* 2008;21:247–53.
18 Vázquez A, Ruiz S, Herrero L *et al.* West Nile and Usutu viruses in mosquitoes in Spain, 2008–2009. *Am J Trop Med Hyg* 2011;85:178–81.
19 Buckley A, Dawson A, Moss SR, Hinsley SA, Bellamy PE, Gould EA. Serological evidence of West Nile virus, Usutu virus and Sindbis virus infection of birds in the UK. *J Gen Virol* 2003;84:2807–17.
20 Buckley A, Dawson A, Gould EA. Detection of seroconversion to West Nile virus, Usutu virus and Sindbis virus in UK sentinel chickens. *Virol J* 2006;3:71.
21 Pecorari M, Longo G, Gennari W *et al.* First human case of Usutu virus neuroinvasive infection, Italy, August-September 2009. *Euro Surveill* 2009;14:e19446.
22 Vilibic-Cavlek T, Kaic B, Barbic L *et al.* First evidence of simultaneous occurrence of West Nile virus and Usutu virus neuroinvasive disease in humans in Croatia during the 2013 outbreak. *Infection* 2014;42:689–95.
23 Allering L, Jöst H, Emmerich P *et al.* Detection of Usutu virus infection in a healthy blood donor from south-west Germany, 2012. *Euro Surveill* 2012;17:e20341.
24 Cavrini F, Della Pepa ME, Gaibani P *et al.* A rapid and specific real-time RT-PCR assay to identify Usutu virus in human plasma, serum, and cerebrospinal fluid. *J Clin Virol* 2011;50:221–3.
25 Nikolay B, Weidmann M, Dupressoir A *et al.* Development of a Usutu virus specific real-time reverse transcription PCR assay based on sequenced strains from Africa and Europe. *J Virol Methods* 2014;197:51–4.
26 Blázquez AB, Escribano-Romero E, Martín-Acebes MA, Petrovic T, Saiz JC. Limited susceptibility of mice to Usutu virus (USUV) infection and induction of flavivirus cross-protective immunity. *Virology* 2015;482:67–71.
27 Beasley DW, Barrett AD. Identification of neutralizing epitopes within structural domain III of the West Nile virus envelope protein. *J Virol* 2002;76:13097–100.
28 Ashfaq UA, Ahmed B. *De novo* structural modeling and conserved epitopes prediction of Zika virus envelop protein for vaccine development. *Viral Immunol* 2016;29:1–8.
29 Alam A, Ali S, Ahamad S, Malik MZ, Ishrat R. From ZikV genome to vaccine: *in silico* approach for the epitope-based peptide vaccine against Zika virus envelope glycoprotein. *Immunology* 2016;149:386–399. doi: 10.1111/imm.12656. [Epub ahead of print].
30 Chakraborty S, Chakravorty R, Ahmed M *et al.* A computational approach for identification of epitopes in dengue virus envelope protein: a step towards designing a universal dengue vaccine targeting endemic regions. *In Silico Biol* 2010;10:235–46.
31 Hasan MA, Hossain M, Alam MJ. A computational assay to design an epitope-based peptide vaccine against Saint Louis Encephalitis virus. *Bioinform Biol Insights* 2013;7:347–55.
32 Hasan MA, Khan MA, Datta A, Mazumder MH, Hossain MU. A comprehensive immunoinformatics and target site study revealed the corner-stone toward Chikungunya virus treatment. *Mol Immunol* 2015;65:189–204.

33 Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 2013;30:2725–9.
34 Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 2014;42: W320–4.
35 Gasteiger E, Hoogland C, Gattiker A *et al.* Protein identification and analysis tools on the ExPASy server. In: Walker JM, ed. *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press 2005: 571–607.
36 Sonnhammer EL, von Heijne G, Krogh A. A hidden markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* 1998;6:175–82.
37 Jones DT. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics* 2007;23:538–44.
38 Ferrè F, Clote P. DiANNA 1.1: an extension of the DiANNA web server for ternary cysteine classification. *Nucleic Acids Res* 2006;34: W182–5.
39 Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* 2007;8:4.
40 Muh HC, Tong JC, Tammi MT. AllerHunter: a SVM-pairwise system for assessment of allergenicity and allergic cross-reactivity in proteins. *PLoS ONE* 2009;4:e5861.
41 Wang J, Zhang D, Li J. PREAL: prediction of allergenic protein by maximum Relevance Minimum Redundancy (mRMR) feature selection. *BMC Syst Biol* 2013;7:S5–9.
42 Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* 2006;65:40–8.
43 El-Manzalawy Y, Dobbs D, Honavar V. Predicting linear B-cell epitopes using string kernels. *J Mol Recognit* 2008;21:243–55.
44 Larsen JEP, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Res* 2006;2:2.
45 Chou PY, Fasman GD. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol Relat Areas Mol Biol* 1978;47:45–148.
46 Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol* 1985;55:836–9.
47 Karplus PA, Schulz GE. Prediction of chain flexibility in proteins – a tool for the selection of peptide antigens. *Naturwissenschaften* 1985;72:212–3.
48 Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* 1990;276:172–4.
49 Parker JM, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* 1986;25:5425–32.
50 Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-Scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics* 2007;8:424.
51 Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics* 2005;6:132.
52 Singh H, Raghava GP. ProPred: prediction of HLA-DR binding sites. *Bioinformatics* 2001;17:1236–7.
53 Biasini M, Bienert S, Waterhouse A *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 2014;42:W252–8.
54 Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 2015;10:845–58.
55 Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods* 2015;12:7–8.
56 Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys J* 2011;101:2525–34.
57 Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK – a program to check the stereochemical quality of protein structures. *J App Cryst* 1993;26:283–91.
58 Bui HH, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics* 2007;8:361.
59 Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics* 2006;7:153.
60 Tynan FE, Borg NA, Miles JJ *et al.* High resolution structures of highly bulged viral epitopes bound to major histocompatibility complex class I. Implications for T-cell receptor engagement and T-cell immunodominance. *J Biol Chem* 2005;280:23900–9.
61 Lamiable A, Thévenet P, Rey J, Vavrusa M, Derreumaux P, Tufféry P. PEP-FOLD3: faster *de novo* structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res* 2016;44:W449–54.
62 Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comput Chem* 2010;31:455–61.
63 Hubálek Z, Rudolf I, Nowotny N. Arboviruses pathogenic for domestic and wild animals. *Adv Virus Res* 2014;89:201–75.
64 Nikolay B. A review of West Nile and Usutu virus co-circulation in Europe: how much do transmission cycles overlap? *Trans R Soc Trop Med Hyg* 2015;109:609–18.
65 Dai L, Song J, Lu X *et al.* Structures of the Zika virus envelope protein and its complex with a Flavivirus broadly protective antibody. *Cell Host Microbe* 2016;19:696–704.
66 Cacciotti G, Caputo B, Selvaggi C *et al.* Variation in interferon sensitivity and induction between Usutu and West Nile (lineages 1 and 2) viruses. *Virology* 2015;485:189–98.
67 Martín-Acebes MA, Blázquez AB, Cañas-Arranz R *et al.* A recombinant DNA vaccine protects mice deficient in the alpha/beta interferon receptor against lethal challenge with Usutu virus. *Vaccine* 2016;34:2066–73.
68 Lobigs M, Diamond MS. Feasibility of cross-protective vaccination against flaviviruses of the Japanese encephalitis serocomplex. *Expert Rev Vaccines* 2012;11:177–87.
69 Soria-Guerra RE, Nieto-Gomez R, Govea-Alonso DO, Rosales-Mendoza S. An overview of bioinformatics tools for epitope prediction: implications on vaccine development. *J Biomed Inform* 2015;53:405–14.

## Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

**Figure S1.** Prediction and validation of potential B-cell epitopes based on biophysical properties. (A) Kolaskar and Tongaonkar antigenicity prediction (Threshold = 1.027), (B) Chou and Fasman beta-turn prediction (Threshold = 0.993), (C) Emini surface accessibility prediction

(Threshold = 1.0), (D) Karplus and Schulz flexibility prediction (Threshold = 0.995), (E) Parker hydrophilicity prediction (Threshold = 1.469) and F) Bepipred linear epitope prediction (Threshold = 0.062). Regions with values above the threshold are represented in yellow while the regions with values below the threshold are represented in green.

**Figure S2.** Filtering of predicted T-cell MHC class-I epitopes based on antigenicity. Kolaskar and Tongaonkar antigenicity prediction tool was used (window length = 9). The antigenic regions (i.e. values above the threshold = 1.027) are represented in yellow and the non-antigenic regions are represented in green.