Microbial Pathogenesis 125 (2018) 129–143

Contents lists available at ScienceDirect

Microbial Pathogenesis

journal homepage: www.elsevier.com/locate/micpath

In silico identification of immunodominant B-cell and T-cell epitopes of non-structural proteins of *Usutu Virus*

Rohit Satyam[a], Essam Mohammed Janahi[b], Tulika Bhardwaj[c], Pallavi Somvanshi[c,*], Shafiul Haque[d,**], Mohammad Zeeshan Najm[a]

[a] Department of Biotechnology, Noida Institute of Engineering and Technology, 19, Knowledge Park-II, Greater Noida, 201308, Uttar Pradesh, India
[b] Department of Biology, College of Science, University of Bahrain, Sakhir, Bahrain
[c] Department of Biotechnology, TERI School of Advanced Studies, 10, Institutional Area, Vasant Kunj, New Delhi, 110070, India
[d] Research and Scientific Studies Unit, College of Nursing & Allied Health Sciences, University of Jazan, Jazan, 45142, Saudi Arabia

ARTICLE INFO

Keywords:
Usutu virus
Physicochemical characterization
Phylogenetic analysis
Epitope prediction
Homology modelling
Toxicity prediction
Population coverage analysis

ABSTRACT

Usutu Virus (USUV; flavivirus) is a re-emerging pathogen invading the territories of European countries, Asia, and Africa. It is a mosquito-borne zoonotic virus with a bi-directional transmission route from animal to human and vice versa, and causes neurological disorders such as meningoencephalitis in bats, *Homo sapiens*, birds and horses. Due to limited availability of information about USUV and its deleterious effects on neural cells causing neurologic impairments, it becomes imperative to study this virus in detail to equip ourselves with a solution beforehand. The current study aims to identify immunodominant peptides that could be exploited in future for designing global peptide vaccine for combating the infections caused by USUV. In this study, an immunoinformatics approach was applied to evaluate the immunogenicity of 7 non-structural proteins and determined 64 continuous B-cell epitopes, numerous probable discontinuous B-cell epitopes, 64 MHC Class-I binders, 126 MHC class-II binders and 52 promiscuous binders with a maximum population coverage of 98.55%(MHC Class-I binder ofYP_164815.1 NS4a) and 81.81% (MHC Class-II binders of YP_164812.1 NS2a, YP_164813.1 NS2b, YP_164814.1 NS3, YP_164817.1 NS4b, YP_164818.1 NS5). Further, studies involving experimental validation of these predicted epitopes is warranted to ensure the potential of B-cells and T-cells stimulation for their effective use as vaccine candidates, and as diagnostic agents against USUV.

1. Introduction

Usutu Virus (USUV) is a pathogenic member of family *Flaviviridae*. USUV is a re-emerging pathogen conquering the broad regions of European countries, Asia, and Africa. It is a mosquito-borne flavivirus of Japanese encephalitis virus (JEV) group considering *Culex pipiens* mosquito being the most important vector [1]. It is the causative agent of neurological disorders such as meningoencephalitis or meningitis in bats, *Homo sapiens*, birds and horses [2,3]. The clinical manifestations of USUV infection include neurological signs, jaundice, rash and fever [3]. A detailed target surveillance demarcates decline in the common blackbird (*Turdus merula*) population due to meningoencephalitis to 15.7% in USUV-associated areas, which raised a serious concern to combat this virus [4,5]. Initially, USUV was isolated from *Culex neavei* mosquito and named as SouthAfrica-1959 strain which serves as a reference strain [6]. The relatively high mortality and morbidity rates in birds and the rapid invasion and spread of the virus in human population settings have raised serious concerns about its possible dire consequences on public health.

The open reading frames (ORF)of USUV encodes a polyprotein precursor (3434 amino acid residues long) cleaved by both viral and cellular proteolytic enzyme proteases producing three structural and eight non-structural (NS) proteins post-translation. The non-structural proteins of USUV are coded by the genomic RNA that acts as mRNA of the virus and are expressed in the host cells. The non-structural proteins of USUV include a soluble complement-fixing antigen [NS1 (Nucleotide 2476–3531)], serine protease/RNA helicase [NS2a (Nucleotide 3532–4212), NS2b(Nucleotide 4213–4605), NS3(Nucleotide 4604–6462)]; NS4a(Nucleotide 6463–6840), NS4b(Nucleotide 6910–7683), 2K protein (Nucleotide 6841–6909)and RNA-dependent RNA polymerase/methyltransferase [NS5 (Nucleotide 7684–10398)] [7]. The viral non-structural proteins are regulatory proteins, they do

* Corresponding author.
** Corresponding author.
E-mail addresses: psomvanshi@gmail.com (P. Somvanshi), shafiul.haque@hotmail.com (S. Haque).

https://doi.org/10.1016/j.micpath.2018.09.019
Received 28 June 2018; Received in revised form 10 September 2018; Accepted 11 September 2018
Available online 11 September 2018
0882-4010/

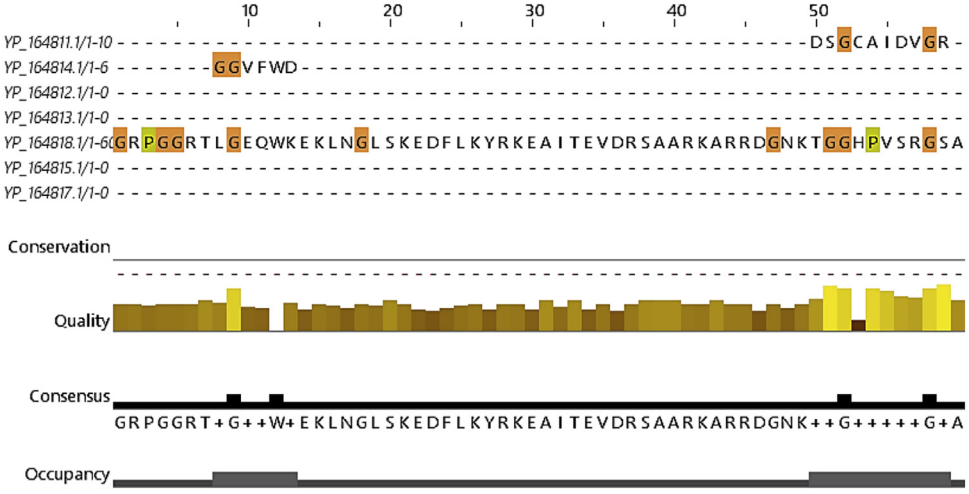


Fig. 1. Multiple sequence alignment of non-structural proteins using JalView.

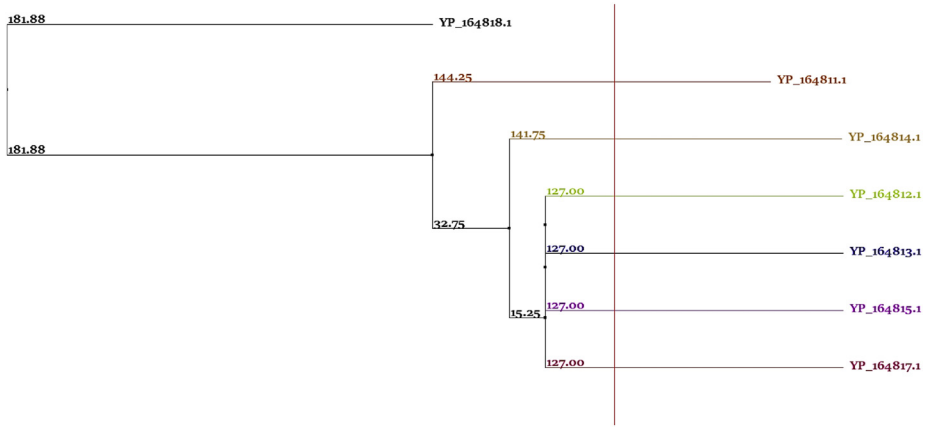


Fig. 2. Generated phylogenetic tree of non-structural proteins of USUV.

Table 1
Immunogenic and physicochemical characterization of 7 non-structural proteins of *Usutu Virus*.

Protein	Length	Molecular Weight	Isoelectric point	Extinction Coefficient	-R (Asp + Glu)	+R (Arg + Lys)	Instability Index	Aliphatic Index	GRAVY	Vaxijen score
YP_164811.1 NS1	352	39723.02	5.98	82680	50	44	49.02 unstable	75.82	−0.435	0.5161 (Probable antigen)
YP_164812.1 NS2a	227	24388.23	10.15	22585	10	21	35.11 Stable	124.36	0.685	0.6431 (Probable antigen)
YP_164813.1 NS2b	131	14284.52	4.48	48470	15	9	18.46 Stable	105.80	0.384	0.7578 (Probable antigen)
YP_164814.1 NS3	619	68664.02	7.75	113010	75	76	33.78 Stable	79.60	−0.398	0.4702 (Probable antigen)
YP_164815.1 NS4a	126	13541.29	6.70	6990	12	12	27.12 Stable	121.59	0.664	0.4144 (Probableantigen).
YP_164817.1 NS4b	258	27442.06	9.14	46200	15	20	29.09 Stable	112.40	0.401	0.484 (Probable antigen).
YP_164818.1 NS5	905	103658.09	8.88	224095	118	131	35.13 stable	70.96	−0.603	0.4421 (Probable antigen)

not get assembled in the virion capsid, membrane or envelope. Rather, they carry out essential functions that affect the replication process [8]. Their functional annotation reveals their involvement in major transcriptional machinery.

NS1 is involved in several functions like viral replication (replicon formation), immune evasion (inhibiting signal transduction originating from Toll-like receptor 3 (TLR3) [9], and pathogenesis (binding to the host macrophages and dendritic cells). After the proteolytic cleavage post-translation, it is targeted to three destinations: the viral replication cycle, the plasma membrane and the extracellular compartment. NS2a forms a part of replication complex and aids in virion assembly [10]. It also antagonizes the host α/β interferon antiviral immune response. NS2b is a cofactor required for serine protease activity of NS3 [11] and might have membrane-destabilizing activity and viroporins formation activity. 2K peptide acts as a signal peptide for NS4B, which in turn has the interferon antagonism activity. NS4a allows NS3 helicase to conserve energy during the unwinding by regulating its ATPase activity [12]. NS4binduces the ER-derived membrane vesicles formation, where

Table 2
Secondary structure prediction of 7 non-structural proteins of *Usutu Virus*.

Protein	Alpha helix (Hh)	Random coil (Cc)	Beta-turn (Tt)	Extended strand (Ee)
YP_164811.1 NS1	26.70%	39.49%	8.24%	25.57
YP_164812.1 NS2a	54.19%	21.15%	8.81%	15.86%
YP_164813.1 NS2b	35.11%	31.30%	11.45%	22.14%
YP_164814.1 NS3	26.82%	35.54%	13.57%	24.07%
YP_164815.1 NS4a	54.76%	20.63%	8.73%	15.87%
YP_164817.1 NS4b	48.06%	25.97%	10.08%	15.89%
YP_164818.1 NS5	44.42%	30.94%	8.40%	16.24%

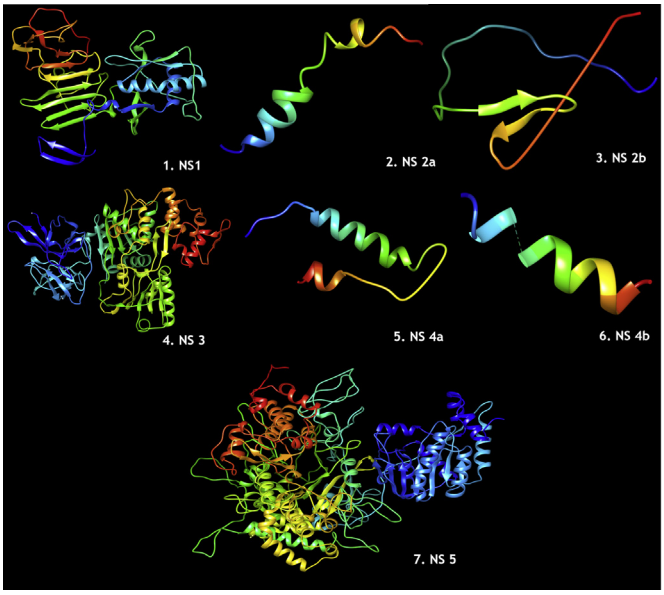


Fig. 3. Visualization of three-dimensional models of seven non-structural proteins using Chimera 1.12.

the viral replication process takes place. It inhibits interferon (IFN)-induced host STAT1 (signal transducer and activator of transcription 1) phosphorylation and nuclear translocation and blocks the IFN-α/β pathway, thereby averting the establishment of the cellular antiviral state. It also inhibits STAT2 translocation in the nucleus [13]. NS5 is RNA-dependent RNA polymerase that reproduces the viral (+) and (-) RNA genome, performs the capping of genome in the cytoplasm [14], methylates viral RNA cap at guanine N-7 and ribose 2′-O positions, regulate and prevent the establishment of cellular antiviral state by blocking the IFN-α/β signalling pathway [15]. Also, it inhibits host TYK2 (tyrosine kinase 2) and STAT2 phosphorylation, thereby preventing the activation of Janus kinase (JAK), signal transducer of activation (STAT)signalling pathway [16].

In the present study, an *in-silico* approach was employed to introduce the metabolic glitch in the replication machinery of USUV by targeting the non-structural proteins to prevent the viral infection to the healthy host cell. Therefore, epitope identification *viz*. B-cell and T-cell epitopes (HLA class-I $CD8^+$ T-cells and HLA class-II $CD4^+$ T-cells) and characterization was performed to provide the platform for future researchers to develop potential vaccines against USUV [17–19]. Also, to overcome the limitations of extreme polymorphism among maximal population setting, population coverage analysis was performed, which aids in the novel vaccine discovery against USUV [20,21].

2. Materials and methodology

2.1. Dataset collection

The complete protein sequences of 7 non-structural proteins of USUV were retrieved from the Viral Genome database of NCBI (https://www.ncbi.nlm.nih.gov/genome/viruses/) in FASTA format. The sequence similarity search was performed using BLASTp to trace the presence of already determined protein structures, if any. Further, the selected proteins were subjected to the prediction of immunodominant epitopes.

2.2. Phylogenetic analysis

The multiple sequence alignment (MSA) of 7 non-structural proteins was performed by using Clustal Omega of EMBL-EBI (https://www.ebi.ac.uk/Tools/msa/clustalo/). The obtained alignment was then visualized using Jalview (http://www.jalview.org/Help/Getting-Started), a JAVA based MSA visualization and editing tool from University of Dundee. The phylogenetic tree was then prepared by using MEGA 6.0 employing neighbor-joining method fixed at default parameters (BLOSUM 62 matrix and similarity percentage of 70%) [22].

2.3. Physicochemical characterization

Various physicochemical properties of the proteins such as instability index, molecular weight, aliphatic index, extinction coefficient, theoretical pI (isoelectric point), and GRAVY (grand average of hydropathicity) index, amino acid compositions were determined by using PROTPARAM (online ExPASy proteomics server) [23].

2.4. Structure prediction

The secondary structure prediction of the selected 7 proteins was performed by using SOPMA (https://npsa-prabi.ibcp.fr/NPSA/npsa_sopma.html) and PSIPRED for four conformational states. The tertiary structures of the proteins were then determined via comparative protein structure modelling using MODELLER (https://salilab.org/modeller/) and Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index).

2.5. Immunogenicity prediction

The immunogenicity of the proteins was evaluated by using VaxiJen v2.0, an alignment-independent server (http://www.ddg-pharmfac.net/vaxijen/Vaxi-Jen/VaxiJen.html)that enables differentiation of protective antigens from non-antigens; a classification explicitly based on physicochemical properties of the protein with the precision level of 70–89% by employing Auto Cross-Covariance (ACC) algorithm. The protective antigens are imperative in vaccine development and biological markers prediction for the diagnosis of infectious and non-infectious disease and analysis of fundamental host immune response against them [24].

2.6. B-cell epitope prediction

The prediction of both linear continuous and discontinuous epitopes was performed to identify short peptides that can be directly used or can mimic the immunogenic properties and structure of an epitope for

Table 3a
List of predicted continuous B-cell epitopes 7 non-structural proteins of *Usutu Virus*.

Protein Accession No.	Start	End	B cell Continuous Epitopes	Number of residues	Score
YP_164811.1 NS1	1	23	DSGCAIDVGRRELRCGQGIFIHN	23	0.86
	101	132	PQRLALTSEEFEIGWKAWGKSLVFAPELANHT	32	0.768
	334	348	EIRPMKHDETTLVKS	15	0.742
	73	87	DELNTLLRENAVDLS	15	0.711
	278	320	DYCPGTTVTITEACGKRGPSIRTTTSSGRLVTDWCCRSCTLPP	43	0.71
	206	210	KNTTW	5	0.703
	47	56	EQAHAKGICG	10	0.644
	156	166	EDFGFGIMSTR	11	0.61
	139	148	ETKECPDAKR	10	0.571
	226	239	PETHTLWSDGVVES	14	0.557
YP_164812.1 NS2a	89	96	FLAMTFLR	8	0.67
	69	77	NSGGDVVHL	9	0.544
YP_164813.1 NS2b	88	95	INDPGVPW	8	0.677
	50	69	TDLWLERAADITWETDAAIT	20	0.563
YP_164814.1 NS3	550	598	KVASNGIQYTDRKWCFDGPRSNIILEDNNEVEIVTRTGERKMLKPRWLD	49	0.836
	1	38	GGVFWDTPAPRTYPKGDTSPGVYRIMSRYILGTYQAGV	38	0.79
	370	384	NEIAQCLQRAGKKVI	15	0.789
	336	359	IQAEVPDRAWSSGFEWITEYTGKT	24	0.776
	51	66	HTTRGAAIRSGEGRLT	16	0.747
	233	259	AEALKGLPVRYLTPAVNREHSGTEIVD	27	0.739
	150	163	YGNGVILGNGSYVS	14	0.722
	608	619	LKWFKDFAAGKR	12	0.7
	206	218	PQIIKDAIQRRLR	13	0.687
	91	118	GLDDVQLIIVAPGKAAINIQTKPGIFKT	28	0.684
	396	405	PKCKNGDWDF	10	0.68
	125	136	AVSLDYPEGTSG	12	0.67
	435	441	LEEGEGR	7	0.645
	519	533	TMDGEYRLRGEERKT	15	0.641
	272	280	SPLRAPNYN	9	0.633
	499	506	IHLPNGLV	8	0.623
	464	472	RNPSQIGDE	9	0.552
	307	310	LGEA	4	0.541
YP_164815.1 NS4a	9	15	LGRMPEH	7	0.71
	31	38	ATAEKGGK	8	0.703
YP_164817.1 NS4b	2	8	EYGMLR	6	0.567

(*continued on next page*)

Table 3a (*continued*)

Protein Accession No.	Start	End	B cell Continuous Epitopes	Number of residues	Score
YP_164818.1 NS5	1	42	GRPGGRTLGEQWKEKLNGLSKEDFLKYRKEAITEVDRSAARK	42	0.866
	1135	1185	LIGTRTRATWAENIYAAINQVRAIIGQEKYRDYMLSLRRYEEVNVQEDRVL	51	0.821
	147	260	IGESSSSAEVEEQRTLRILEMVSDWLQRGPREFCIKVLCPYMPRVMERLEVLQRRYGGGLVRVPLSRNSNHEMYWVSGAAGNIVHAVNMTSQVLIGRMEKRTWHGPKYEEDVNL	114	0.791
	522	545	GRMEKRTWHGPKYEEDVNLGSGTR	24	0.768
	1000	1009	ELIMKDGRTL	10	0.762
	548	568	GKPQPHTNQEKIKARIQRLKE	21	0.759
	719	735	EERENHLKGECHTCIYN	17	0.759
	905	942	LMEAEGVIGQEHLESLPRKTKYAVRTWLFENGEERVTR	38	0.757
	617	642	ILNVTTMAMTDTTPFGQQRVFKEKVD	26	0.7
	650	688	SGVREVMDETTNWLWAFLAREKKPRLCTREEFKRKVNSN	39	0.692
	1102	1130	WIQDNEWMLDKTPVQSWTDIPYTGKREDI	29	0.691
	953	958	KPLDDR	6	0.67
	362	375	CGRGGWSYYAATLK	14	0.666
	598	608	SASSLVNGVVR	11	0.657
	380	399	VRGYTKGGPGHEEPMLMQSY	20	0.653
	325	333	RDGNKTGGH	9	0.649
	491	503	LSRNSNHEMYWVS	13	0.648
	284	309	GGRTLGEQWKEKLNGLSKEDFLKYRK	26	0.638
	60	75	AKLRWMVERQFVKPIG	16	0.635
	802	812	REMSHHSGGKM	11	0.627
	1062	1073	NAICSAVPSNWV	12	0.626
	773	783	FLNEDHWLGRK	11	0.616
	1090	1100	TTDDMLEVWNK	11	0.592
	505	515	AAGNIVHAVNM	11	0.589
	579	584	DHPYRT	6	0.555
	272	282	PHTNQEKIKGR	11	0.534
	451	455	WLQRG	5	0.534
	478	481	LQRR	4	0.509

Table 3b
List of predicted discontinuous B-cell epitopes of seven non-structural proteins of *Usutu Virus*.

Accession No. of proteins	B-cell Discontinuous epitopes residues	Number of residues	Score
YP_164811.1 NS1	_:D1, _:S2, _:G3, _:C4, _:A5, _:I6, _:D7, _:V8, _:G9, _:R10, _:R11, _:E12, _:L13, _:R14, _:C15, _:G16, _:Q17, _:G18, _:I19	19	0.941
	_:K40, _:A43, _:K44, _:E47, _:Q48, _:A49, _:H50, _:A51, _:K52, _:G53, _:I54, _:C55, _:G56, _:D73, _:E74, _:N76, _:T77, _:L78, _:L79, _:R80,	75	0.683
	_:E81, _:N82, _:A83, _:V84, _:D85, _:L86, _:S87, _:A100, _:P101, _:Q102, _:R103, _:L104, _:A105, _:L106, _:T107, _:S108, _:E109, _:E110, _:		
	F111, _:E112, _:I113, _:G114, _:W115, _:K116, _:A117, _:W118, _:G119, _:K120, _:S121, _:L122, _:V123, _:F124, _:A125, _:P126, _:E127, _:L128, _:A129, _:N130, _:H131, _:T132, _:E139, _:T140, _:K141, _:E142, _:C143, _:P144, _:A146, _:K147, _:R148, _:A149, _:R172, _:E173, _:H174, _:N175, _:T176		
	_:R31, _:G159, _:F160, _:G161, _:I162, _:M163, _:S164, _:R166	8	0.667
	_:A187, _:V188, _:K189, _:G190, _:D191, _:I192, _:K206, _:N207, _:T208, _:T209, _:W210, _:P226, _:E227, _:T228, _:H229, _:T230, _:L231, _:W232, _:S233, _:D234, _:G235, _:V236, _:V237, _:S239, _:K251, _:S252, _:N253, _:H254, _:R256, _:R257, _:E258, _:G259, _:Y260, _:K261, _:V262, _:Q265, _:D278, _:Y279, _:C280, _:P281, _:G282, _:T283, _:T284, _:V285, _:T286, _:I287, _:T288, _:E289, _:A290, _:C291, _:G292, _:K293, _:R294, _:G295, _:P296, _:S297, _:I298, _:R299, _:T300, _:T301, _:T302, _:S303, _:S304, _:G305, _:R306, _:L307, _:V308, _:T309, _:D310, _:W311, _:C312, _:C313, _:R314, _:S315, _:C316, _:T317, _:L318, _:P319, _:P320, _:N327, _:G328, _:G332, _:M333, _:E334, _:I335, _:R336, _:P337, _:M338, _:K339, _:H340, _:D341, _:E342, _:T343, _:T344, _:L345, _:V346, _:K347, _:S348, _:S349	99	0.641
	_:F20, _:I21, _:H22	3	0.547
YP_164812.1 NS2a	_:F89, _:L90, _:T93, _:F94	4	0.652
	_:N69, _:S70, _:G71, _:G72, _:D73, _:V74, _:V75, _:H76, _:I80	9	0.536
YP_164813.1 NS2b	_:G92, _:V93, _:P94, _:W95	4	0.717
	_:T50, _:D51, _:L52, _:W53, _:L54	5	0.665
	_:D76, _:I88, _:N89	3	0.587
	_:R56, _:A57, _:A58, _:D59, _:I60, _:T61, _:W62, _:E63, _:T64, _:D65, _:A66, _:A67, _:I68, _:T69	14	0.516
YP_164814.1 NS3	_:L435, _:E436, _:E437, _:G438, _:E439, _:G440, _:R441, _:N498, _:T519, _:M520, _:D521, _:G522, _:E523, _:Y524, _:R525, _:L526, _:R527, _:G528, _:E529, _:E530, _:R531, _:K532, _:T533, _:E536, _:T540, _:A541, _:K550, _:V551, _:A552, _:S553, _:N554, _:G555, _:I556, _:Q557, _:Y558, _:T559, _:D560, _:R561, _:K562, _:W563, _:C564, _:F565, _:D566, _:G567, _:P568, _:R569, _:S570, _:N571, _:I572, _:I573, _:L574, _:E575, _:D576, _:N577, _:N578, _:E579, _:V580, _:E581, _:I582, _:V583, _:T584, _:R585, _:T586, _:G587, _:E588, _:R589, _:K590, _:M591, _:L592, _:K593, _:P594, _:R595, _:W596, _:L597, _:D598, _:A599, _:V601, _:Y602, _:A603, _:H605, _:L608, _:K609, _:F611, _:K612, _:D613, _:F614, _:A615, _:A616, _:G617, _:K618, _:R619	91	0.74
	_:G1, _:G2, _:V3, _:F4, _:W5, _:D6, _:T7, _:P8, _:A9, _:P10, _:R11, _:T12, _:Y13, _:P14, _:K15, _:G16, _:D17, _:T18, _:S19, _:P20, _:G21, _:V22, _:Y23, _:R24, _:I25, _:M26, _:S27, _:R28, _:Y29, _:I30, _:L31, _:G32, _:T33, _:Y34, _:Q35, _:A36, _:G37, _:V38, _:L49, _:T52, _:T53, _:R54, _:G55, _:A56, _:A57, _:I58, _:R59, _:S60, _:G61, _:E62, _:G63, _:R64, _:L65, _:T66, _:N90, _:G91, _:D93, _:D94, _:V95, _:Q96, _:L97, _:I98, _:I99, _:V100, _:A101, _:P102, _:G103, _:K104, _:A105, _:A106, _:I107, _:N108, _:I109, _:Q110, _:T111, _:K112, _:P113, _:G114, _:I115, _:F116, _:A125, _:V126, _:S127, _:L128, _:D129, _:Y130, _:P131, _:E132, _:G133, _:T134, _:S135, _:G136, _:S137, _:G144, _:Y150, _:N152, _:G153, _:V154, _:I155, _:L156, _:G157, _:N158, _:G159, _:S160, _:Y161, _:V162, _:S163	107	0.73
	_:N330, _:A331, _:P332, _:I336, _:Q337, _:A338, _:E339, _:V340, _:P341, _:D342, _:R343, _:A344, _:W345, _:S346, _:S347, _:G348, _:F349, _:E350, _:W351, _:I352, _:T353, _:E354, _:Y355, _:T356, _:G357, _:K358, _:T359, _:N370, _:E371, _:I372, _:A373, _:Q374, _:C375, _:L376, _:Q377, _:R378, _:A379, _:G380, _:K381, _:K382, _:V383, _:I384, _:P396, _:K397, _:K399, _:N400, _:G401, _:D402, _:W403, _:D404, _:F405, _:G419, _:A420, _:S421, _:R422, _:N465, _:P466, _:S467, _:Q468, _:I469, _:G470, _:E472, _:G476	63	0.709
	_:R185, _:K186, _:K187, _:Q188, _:T201, _:P206, _:Q207, _:I209, _:K210, _:D211, _:A212, _:I213, _:Q214, _:R215, _:R216, _:L217, _:R218, _:A233, _:E234, _:A235, _:L236, _:K237, _:G238, _:L239, _:P240, _:V241, _:R242, _:Y243, _:L244, _:T245, _:P246, _:A247, _:V248, _:N249, _:R250, _:E251, _:H252, _:S253, _:G254, _:T255, _:E256, _:I257, _:V258, _:D259, _:S272, _:P273, _:L274, _:R275, _:A276, _:P277, _:N278, _:Y279, _:N280, _:L307, _:G308, _:E309, _:A310	57	0.676
	_:H500, _:L501, _:P502, _:N503, _:G504, _:L505, _:V506	7	0.659
YP_164815.1 NS4a	_:L9, _:G10, _:R11, _:M12, _:P13, _:E14, _:H15	7	0.71
	_:G36, _:G37, _:K38	3	0.676
	_:V30, _:A31, _:T32, _:A33, _:E34	5	0.573
YP_164817.1 NS4b	_:Y3, _:G4, _:M5, _:L6, _:R8	5	0.507

(*continued on next page*)

Table 3b (*continued*)

Accession No. of proteins	B-cell Discontinuous epitopes residues	Number of residues	Score
YP_164818.1 NS5	_:G1, _:R2, _:P3, _:G4, _:G5, _:R6, _:T7, _:L8, _:G9, _:E10, _:Q11, _:W12, _:K13, _:K15, _:L16, _:N17, _:G18, _:L19, _:S20, _:K21, _:E22, _:D23, _:F24, _:L25, _:K26, _:Y27, _:R28, _:K29, _:E30, _:A31, _:I32, _:T33, _:E34, _:V35, _:D36, _:R37, _:S38, _:A39, _:R41, _:K42, _:G51, _:G52, _:H53, _:P54, _:S56, _:R57, _:G58, _:A60, _:K61, _:R63, _:W64, _:M65, _:V66, _:E67, _:R68, _:F70, _:V71, _:K72, _:P73, _:I74, _:G75, _:L94, _:G96, _:V97, _:S137, _:P139, _:C140, _:D141, _:D146, _:I147, _:G148, _:E149, _:S150, _:S151, _:S152, _:S153, _:A154, _:E155, _:V156, _:E157, _:E158, _:Q159, _:R160, _:T161, _:L162, _:I164, _:L165, _:E166, _:V168, _:S169, _:D170, _:L172, _:Q173, _:R174, _:G175, _:P176, _:R177, _:E178, _:F179, _:C180, _:I181, _:K182, _:V183, _:L184, _:C185, _:P186, _:Y187, _:M188, _:P189, _:R190, _:V191, _:M192, _:E193, _:R194, _:L195, _:E196, _:V197, _:L198, _:Q199, _:R200, _:R201, _:Y202, _:G203, _:G204, _:G205, _:L206, _:V207, _:R208, _:V209, _:P210, _:L211, _:S212, _:R213, _:N214, _:S215, _:N216, _:H217, _:E218, _:M219, _:Y220, _:W221, _:V222, _:S223, _:G224, _:A225, _:A226, _:G227, _:N228, _:I229, _:V230, _:H231, _:A232, _:V233, _:N234, _:M235, _:T236, _:S237, _:Q238, _:V239, _:L240, _:I241, _:G242, _:R243, _:M244, _:E245, _:K246, _:T248, _:W249, _:H250, _:G251, _:P252, _:K253, _:Y254, _:E255, _:E256, _:D257, _:V258, _:N259, _:L260, _:G261	180	0.767
	_:M499, _:Y500, _:W501	3	0.732
	_:T384, _:K385, _:G386, _:G387, _:P388, _:G389, _:H390, _:E391, _:E392, _:P393, _:M394	11	0.708
	_:C362, _:G363, _:R364, _:G365, _:G366, _:W367, _:S368, _:Y369, _:A371, _:A372, _:T373, _:L374, _:K375, _:V377, _:Q378, _:V380, _:R381, _:G382, _:Y383, _:L395, _:M396, _:Q397, _:S398, _:Y399, _:K578, _:H580, _:P581, _:Y582, _:R583, _:T584, _:G597, _:S598, _:A599, _:S600, _:S601, _:L602, _:V603, _:N604, _:G605, _:V606, _:V607, _:R608, _:E1000, _:L1001, _:I1002, _:M1003, _:K1004, _:D1005, _:G1006, _:R1007, _:T1008, _:L1009, _:N1032, _:V1033, _:R1034, _:F1052, _:H1053, _:R1055, _:R1058, _:L1059, _:N1062, _:A1063, _:C1065, _:S1066, _:A1067, _:V1068, _:P1069, _:S1070, _:N1071, _:W1072, _:V1073, _:T1090, _:T1091, _:D1092, _:D1093, _:M1094, _:L1095, _:E1096, _:W1098, _:N1099, _:K1100, _:W1102, _:I1103, _:Q1104, _:D1105, _:N1106, _:E1107, _:W1108, _:M1109, _:L1110, _:D1111, _:K1112, _:T1113, _:P1114, _:V1115, _:Q1116, _:S1117, _:W1118, _:T1119, _:D1120, _:I1121, _:P1122, _:Y1123, _:T1124, _:G1125, _:K1126, _:R1127, _:E1128, _:I1130, _:G1133, _:S1134, _:L1135, _:I1136, _:G1137, _:T1138, _:R1139, _:T1140, _:R1141, _:A1142, _:T1143, _:W1144, _:A1145, _:E1146, _:N1147, _:I1148, _:Y1149, _:A1150, _:A1151, _:I1152, _:N1153, _:Q1154, _:V1155, _:R1156, _:A1157, _:I1158, _:I1159, _:G1160, _:Q1161, _:E1162, _:K1163, _:Y1164, _:R1165, _:D1166, _:Y1167, _:M1168, _:L1169, _:S1170, _:L1171, _:R1172, _:R1173, _:Y1174, _:E1176, _:V1177, _:N1178, _:V1179, _:Q1180, _:E1181	161	0.694
	_:P272, _:H273, _:T274, _:N275, _:Q276, _:E277, _:K278, _:K280, _:G281, _:R282, _:P283, _:G284, _:G285, _:R286, _:T287, _:L288, _:G289, _:E290, _:Q291, _:W292, _:K293, _:E294, _:K295, _:N297, _:G298, _:L299, _:S300, _:K301, _:E302, _:D303, _:F304, _:L305, _:K306, _:Y307, _:R324, _:D326, _:G327, _:N328, _:K329, _:T330, _:G331, _:G332, _:H333, _:P334, _:W451, _:L452, _:Q453, _:A505, _:A506, _:G507, _:N508, _:V510, _:H511, _:A512, _:N514, _:M515, _:Q518, _:V519, _:G522, _:R523, _:E525, _:K526, _:R527, _:T528, _:W529, _:H530, _:G531, _:P532, _:Y534, _:E535, _:E536, _:D537, _:V538, _:N539, _:L540, _:G541, _:S542, _:G543, _:T544, _:G548, _:K549, _:P550, _:Q551, _:P552, _:H553, _:T554, _:N555, _:Q556, _:E557, _:K558, _:I559, _:K560, _:A561, _:R562, _:I563, _:Q564, _:R565, _:L566, _:K567, _:E568, _:I617, _:L618, _:N619, _:V620, _:T621, _:T622, _:M623, _:A624, _:M625, _:T626, _:D627, _:T628, _:T629, _:P630, _:F631, _:G632, _:Q633, _:Q634, _:R635, _:V636, _:F637, _:K638, _:E639, _:K640, _:V641, _:D642, _:S650, _:G651, _:V652, _:R653, _:E654, _:V655, _:M656, _:D657, _:E658, _:T659, _:T660, _:N661, _:W662, _:L663, _:W664, _:A665, _:F666, _:A668, _:R669, _:E670, _:K671, _:K672, _:P673, _:L675, _:C676, _:T677, _:R678, _:E680, _:F681, _:K682, _:R683, _:K684, _:V685, _:N686, _:S687, _:N688, _:E719, _:E720, _:R721, _:E722, _:N723, _:H724, _:L725, _:K726, _:G727, _:E728, _:C729, _:H730, _:T731, _:C732, _:I733, _:Y734, _:A755, _:I756, _:W757, _:F758, _:W760, _:A763, _:R764, _:F765, _:L766, _:E767, _:E769, _:A770, _:G772, _:F773, _:L774, _:N775, _:E776, _:D777, _:H778, _:L780, _:G781, _:R782, _:K783, _:N784, _:E803, _:S805, _:H806, _:H807, _:S808, _:G809, _:G810, _:K811, _:M812, _:Y813, _:L836, _:E837, _:L838, _:M839, _:E840, _:G841, _:E842, _:H843, _:R844, _:Q845, _:L846, _:R848, _:L853, _:H857, _:L905, _:M906, _:E907, _:A908, _:E909, _:G910, _:V911, _:I912, _:G913, _:Q914, _:E915, _:H916, _:L917, _:E918, _:S919, _:L920, _:P921, _:R922, _:K923, _:T924, _:K925, _:Y926, _:A927, _:V928, _:R929, _:T930, _:W931, _:L932, _:F933, _:E934, _:N935, _:G936, _:E937, _:E938, _:R939, _:T941, _:R942, _:K953, _:P954, _:L955, _:D956, _:D957, _:R958, _:N961, _:A962, _:L963, _:H964, _:F965, _:W979, _:G984, _:H986	277	0.675
	_:L491, _:S492, _:R493, _:N494, _:S495, _:N496, _:H497	7	0.646
	_:L478, _:Q479, _:R480, _:Y482	4	0.532

vaccine designing and immunodiagnostic purpose [25]. ElliPro suite of Immune Epitope Database and Analysis Resource (IEDB) was used for the prediction of continuous and discontinuous B-cell epitopes from the tertiary structures of the proteins. This server enables B-cell epitope prediction via subsequence Kernel-based (Support Vector Machine) SVM classifier with improved AUC (the area under the receiver operating characteristic curve) performance (0.758) outperforming AAP method (AUC 0.7) [26]. The structure-based discontinuous epitopes prediction was carried out using ElliPro suite of IEDB and identified 15.5% of residues present in discontinuous epitopes with a specificity of 95% [27,28].

2.7. T-cell epitope prediction

The prediction of immunogenic peptides binding to MHC class-I and class-II molecules was carried out using a standalone tool of IEDB i.e., TepiTool (http://tools.iedb.org/tepitool/). Optimum peptides were obtained using IEDB recommended settings and were screened out based on percentile score [29]. Additionally, MHC class-I epitopes were further screened by using MHC I immunogenicity tool (http://tools.iedb.org/immunogenicity/) and the scores were analyzed. The promiscuous binders were also predicted by using TepiTool employing 7 allele methods to tap down top immunodominant epitopes which could be used to further screen MHC Class-II epitopes for most potent vaccine candidates.

2.8. Population coverage analysis

IEDB population coverage tool (http://tools.immuneepitope.org/tools/population/iedb_input) was used to assess the population coverage of predicted T-cell epitopes (MHC class-I and class-II). Population coverage analysis is crucial to identify the putative vaccine candidates for global vaccine development [30].

2.9. Toxicity assessment

The epitopes procured from the above stated screening procedure were further analyzed for peptide toxicity using Protein Scanning module of ToxinPred server (http://crdd.osdd.net/raghava/toxinpred/). This server employs machine learning SVM technique for the discrimination of non-toxic and toxic peptides [31].

Table 4
Predicted T-cell epitopes (MHC Class I) of7 non-structural proteins of *Usutu Virus*.

Accession No.	Peptide start	Peptide end	MHC Class I Peptide	Allele	Length	MHC Class I Immunogenicity Score	Consensus Percentile rank	%age Population Coverage
YP_164811.1 NS1	127	135	ELANHTFVV	HLA-A* 02:01	9	0.19715	0.3	39.08%
	105	113	ALTSEEFEI	HLA-A*02:01	9	0.19518	1	39.08%
	209	217	TWRLERAVF	HLA- A*23:01	9	0.209	0.6	21.38%
	224	232	TWPETHTLW	HLA-A*24:02	9	0.19168	0.3	21.38%
	279	287	YCPGTTVTI	HLA-A*24:02	9	0.16236	1	21.38%
	24	32	DVEAWVDRY	HLA-A*26:01	9	0.37539	0.35	17.34%
YP_164812.1 NS2a	166	174	YLDTYRITL	HLA-A*02:01	9	0.22638	0.5	39.08%
	164	172	LLYLDTYRI	HLA-A*02:01	9	0.0729	0.4	39.08%
	15	23	VMFLATQEV	HLA-A*02:01	9	0.06222	0.5	39.08%
	169	177	TYRITLIII	HLA-A*24:02	9	0.36816	0.55	21.38%
	75	83	VHLALIAAF	HLA-A*24:02	9	0.20613	0.95	21.38%
	166	174	YLDTYRITL	HLA-A*01:01	9	0.22638	0.8	17.34%
	28	36	WTARLTVPA	HLA-A*01:01	9	0.11888	0.8	17.34%
	99	107	WTNQENILL	HLA-A*01:01	9	0.07859	0.7	17.34%
YP_164813.1 NS2b	13	21	LMFAIVGGL	HLA-A*02:01	9	0.29423	1	39.08%
	52	60	LWLERAADI	HLA-A*24:02	9	0.23036	0.65	21.38%
	49	57	STDLWLERA	HLA-A*01:01	9	0.31604	0.35	17.34%
	119	127	GIGYWLTVK	HLA-A*11:01	9	0.26942	0.85	15.53%
	116	124	IPAGIGYWL	HLA-B*07:02	9	0.3346	0.5	12.78%
YP_164814.1 NS3	282	290	FVMDEAHFT	HLA-A*02:01	9	0.19535	0.6	39.08%
	355	363	YTGKTVWFV	HLA-A*68:02	9	0.126	0.7	39.08%
	355	363	YTGKTVWFV	HLA-A*02:01	9	0.126	0.8	39.08%
	41	49	MYEGVLHTL	HLA-A*24:02	9	0.14634	0.5	21.38%
	344	352	AWSSGFEWI	HLA-A*24:02	9	0.13695	0.7	21.38%
	77	85	ITYGGPWKF	HLA-A*24:02	9	0.1164	0.45	21.38%
	557	565	QYTDRKWCF	HLA-A*24:02	9	0.03776	0.25	21.38%
	354	362	EYTGKTVWF	HLA-A*24:02	9	0.0375	0.45	21.38%
	28	36	RYILGTYQA	HLA-A*24:02	9	0.03078	0.45	21.38%
	479	487	TSEDDTIAA	HLA-A*01:01	9	0.24814	0.9	17.34%
	26	34	MSRYILGTY	HLA-A*01:01	9	0.18352	0.5	17.34%
	558	566	YTDRKWCFD	HLA-A*01:01	9	0.08069	0.5	17.34%
	34	42	YQAGVGVMY	HLA-A*01:01	9	0.05114	1	17.34%
YP_164815.1 NS4a	86	94	VLGLATFFL	HLA-A*02:01	9	0.24168	0.9	39.08%
	62	70	VMTAGVFLL	HLA-A*02:01	9	0.21615	0.5	39.08%
	87	95	LGLATFFLW	HLA-A*24:02	9	0.27609	0.7	21.38%
	63	71	MTAGVFLLL	HLA-A*01:01	9	0.18136	0.45	17.34%
	20	28	TREAFDTMY	HLA-A*01:01	9	0.13691	0.9	17.34%
	11	19	RMPEHFAGK	HLA-A*03:01	9	0.29167	0.3	16.81%
YP_164817.1 NS4b	128	136	MLPGWQAEA	HLA-A*02:01	9	0.22868	1	39.08%
	91	99	FTELDFTVV	HLA-A*01:01	9	0.21002	0.75	17.34%
	117	125	AAALATLHY	HLA-A*01:01	9	0.08572	0.7	17.34%
	241	249	ASIAWTLIK	HLA-A*03:01	9	0.40321	0.5	16.81%
	241	249	ASIAWTLIK	HLA-A*11:01	9	0.40321	0.15	15.53%
	52	60	TVVLTPLIK	HLA-A*11:01	9	0.098	0.45	15.53%

(*continued on next page*)

Table 4 (*continued*)

Accession No.	Peptide start	Peptide end	MHC Class I Peptide	Allele	Length	MHC Class I Immunogenicity Score	Consensus Percentile rank	%age Population Coverage
YP_164818.1 NS5	478	486	FMWLGARFL	HLA-A*02:01	9	0.24265	0.4	39.08%
	373	381	REVMDETTN	HLA-B*40:01	9	0.00799	0.8	39.08%
	289	297	EYAATWHHD	HLA-A*24:02	9	0.34458	1	21.38%
	645	653	KYAVRTWLF	HLA-A*24:02	9	0.32164	0.1	21.38%
	865	873	AENIYAAIN	HLA-B*44:03	9	0.27583	0.9	21.38%
	477	485	WFMWLGARF	HLA-A*24:02	9	0.25025	0.3	21.38%
	609	617	TYALNTFTN	HLA-A*24:02	9	0.15326	0.3	21.38%
	532	540	MYADDTAGW	HLA-A*24:02	9	0.14598	0.3	21.38%
	764	772	AQMWLLLYF	HLA-A*24:02	9	0.13313	0.6	21.38%
	201	209	RYGGGLVRV	HLA-A*24:02	9	0.13274	0.7	21.38%
	678	686	RFANALHFL	HLA-A*24:02	9	0.12955	0.6	21.38%
	838	846	WTDIPYTGK	HLA-A*01:01	9	0.1794	0.8	17.34%
	714	722	CSNHFQELI	HLA-A*01:01	9	0.11343	0.8	17.34%
	615	623	FTNIAVQLI	HLA-A*01:01	9	0.10454	0.85	17.34%
	446	454	KGECHTCIY	HLA-A*01:01	9	0.07855	0.9	17.34%
	303	311	RTWTYHGSY	HLA-A*01:01	9	0.06975	0.5	17.34%
	779	787	LMANAICSA	HLA-A*02:03	9	0.02741	0.3	17.34%
	136	144	PSEPCDTLF	HLA-A*01:01	9	0.016	0.65	17.34%
	394	402	RLCTREEFK	HLA-A*03:01	9	0.31911	0.3	16.81%
	582	590	VMRPGTDGK	HLA-A*03:01	9	0.1137	0.65	16.81%

Table 5
List of promiscuous binders of seven non-structural proteins of *Usutu Virus*.

Accession Number	Peptide start	Peptide end	Promiscuous Binders	Median consensus percentile
YP_164811.1	91	105	EKPKGMYKSAPQRLA	7.17
NS1	96	110	MYKSAPQRLALTSEE	7.43
	121	135	SLVFAPELANHTFVV	8.86
YP_164812.1	71	85	GGDVVHLALIAAFKI	1.29
NS2a	161	175	GMRLLYLDTYRITLI	2.44
	51	65	YTDLLRYVLLVGAAF	2.79
	151	165	VMPLLCLLAPGMRLL	3.35
	111	125	AAFFQMAATDLNFSL	3.76
	76	90	HLALIAAFKIQPGFL	4.29
	11	25	LGLLVMFLATQEVLR	4.34
	56	70	RYVLLVGAAFAEANS	5.52
	101	115	NQENILLALGAAFFQ	5.74
	171	185	RITLIIIGICSLIGE	6.69
	136	150	WMLLRAATQPSTSAI	7
	21	35	QEVLRKRWTARLTVP	7.6
	176	190	IIGICSLIGERRRAA	7.65
	116	130	MAATDLNFSLPGILN	8.75
	16	30	MFLATQEVLRKRWTA	9.36
YP_164813.1	96	110	KIWVIRMTALGFAAW	4.94
NS2b	6	20	EVLTAVGLMFAIVGG	12.05
YP_164814.1	206	220	PQIIKDAIQRRLRTA	3.95
NS3	21	35	GVYRIMSRYILGTYQ	4.54
	436	450	EEGEGRVILSNPSPI	5.12
	96	110	QLIIVAPGKAAINIQ	6.84
	441	455	RVILSNPSPITSASA	7.59
	605	619	HQSLKWFKDFAAGKR	7.64
	531	545	RKTFLELLRTADLPV	7.72
	356	370	TGKTVWFVASVKMGN	8.69
	261	275	MCHATLTHRLMSPLR	9.12
YP_164815.1	51	65	LETITLIVALAVMTA	1.77
NS4a	66	80	GVFLLLVQRRGIGKL	5.39
	106	120	GTLLLALLMMIVLIP	6.03
	56	70	LIVALAVMTAGVFLL	8.29
YP_164817.1	176	190	VGQILLIGVSAAALL	3.39
NS4b	56	70	TPLIKHLVTSEYITT	5.13
	181	195	LIGVSAAALLVNPCV	8.94
	66	80	EYITTSLASISAQAG	8.98
YP_164818.1	766	780	MWLLLYFHRRDLRLM	3.39
NS5	771	785	YFHRRDLRLMANAIC	3.81
	681	695	NALHFLNSMSKVRKD	5.13
	476	490	IWFMWLGARFLEFEA	5.47
	611	625	ALNTFTNIAVQLIRL	6.2
	866	880	ENIYAAINQVRAIIG	6.54
	621	635	QLIRLMEAEGVIGQE	6.63
	606	620	QVVTYALNTFTNIAV	7.03
	516	530	KLGYILREMSHHSGG	7.47
	226	240	AGNIVHAVNMTSQVL	7.95
	61	75	KLRWMVERQFVKPIG	8.34
	111	125	EEPMLMQSYGWNLVT	8.42
	511	525	GLGVQKLGYILREMS	9.72
	616	630	TNIAVQLIRLMEAEG	9.87

3. Results& discussion

The non-structural proteins are the proteins coded by the viral genome that never assembled into a virion. These non-structural proteins are expressed in the host cell and assist in replication process (replicon formation), assembly process, immunomodulation etc [32]. In order to evaluate the role of cell-mediated or humoral immunity against USUV infection, a computational pipeline was designed for the immunogenicity prediction of7 viral proteins (YP_164811.1 NS1, YP_164812.1 NS2a, YP_164813.1 NS2b, YP_164813.1 NS2b, YP_164815.1 NS4a, YP_164817.1 NS4b, YP_164818.1 NS5) and identification of immunodominant B-cell and T-cell epitopes [33]. Viral Genome database of NCBI was mined for the retrieval of protein sequences in FASTA format. Further, phylogenetic analysis revealed the percentage of similarity shared between non-structural proteins of USUV (Fig. 1 and Fig. 2). To trace the ancestral path of evolution,

Table 6
Predicted T-cell epitopes (MHC Class II) of7 non-structural proteins of *Usutu Virus*.

Accession No.	Peptide start	Peptide end	MHC Class II binders	Allele	Consensus percentile rank	%age Population coverage
YP_164811.1 NS1	188	202	VKGDIAVHSDLSYWI	HLA-DRB1*15:01	9.17	18.41%
	157	171	DFGFGIMSTRVWLKV	HLA-DRB1*15:01	9.6	18.41%
	122	136	LVFAPELANHTFVVD	HLA-DRB1*15:01	9.65	18.41%
	91	105	EKPKGMYKSAPQRLA	HLA-DRB1*07:01	2.32	18.23%
	157	171	DFGFGIMSTRVWLKV	HLA-DRB1*07:01	4.42	18.23%
	273	287	IVLDFDYCPGTTVTI	HLA-DRB1*07:01	4.59	18.23%
	41	55	QLAKVIEQAHAKGIC	HLA-DRB1*07:01	5.84	18.23%
	183	197	IIGTAVKGDIAVHSD	HLA-DQA1*05:01/DQB1*03:01	5.92	18.23%
	296	310	PSIRTTTSSGRLVTD	HLA-DRB1*07:01	8.5	18.23%
	198	212	LSYWIESHKNTTWRL	HLA-DRB1*07:01	9.34	18.23%
	162	176	IMSTRVWLKVREHNT	HLA-DRB1*07:01	9.65	18.23%
	1	15	DSGCAIDVGRRELRC	HLA-DRB1*03:01	0.13	17.84%
	335	349	IRPMKHDETTLVKSS	HLA-DRB1*03:01	0.5	17.84%
	233	247	SDGVVESDLVVPVTL	HLA-DRB1*03:01	0.67	17.84%
	127	141	ELANHTFVVDGPETK	HLA-DRB1*03:01	0.98	17.84%
	132	146	TFVVDGPETKECPDA	HLA-DRB1*03:01	0.98	17.84%
	188	202	VKGDIAVHSDLSYWI	HLA-DRB1*03:01	1.4	17.84%
	228	242	THTLWSDGVVESDLV	HLA-DRB1*03:01	1.62	17.84%
	157	171	DFGFGIMSTRVWLKV	HLA-DPA1*01:03/DPB1*02:01	1.86	17.84%
YP_164812.1 NS2a	37	51	IVGALLVLILGGITY	HLA-DRB1*15:01	1.25	18.41%
	75	89	VHLALIAAFKIQPGF	HLA-DRB1*15:01	1.92	18.41%
	42	56	LVLILGGITYTDLLR	HLA-DRB1*15:01	2.03	18.41%
	173	187	TLIIIGICSLIGERR	HLA-DRB1*15:01	2.17	18.41%
	163	177	RLLYLDTYRITLIII	HLA-DRB1*15:01	2.24	18.41%
	54	68	LLRYVLLVGAAFAEA	HLA-DRB1*15:01	2.3	18.41%
	104	118	NILLALGAAFFQMAA	HLA-DRB1*15:01	3.14	18.41%
	150	164	IVMPLLCLLAPGMRL	HLA-DRB1*15:01	4.46	18.41%
	20	34	TQEVLRKRWTARLTV	HLA-DRB1*15:01	4.57	18.41%
	80	94	IAAFKIQPGFLAMTF	HLA-DRB1*15:01	4.75	18.41%
	32	46	LTVPAIVGALLVLIL	HLA-DRB1*15:01	5.07	18.41%
	13	27	LLVMFLATQEVLRKR	HLA-DRB1*15:01	5.74	18.41%
	6	20	IDPFQLGLLVMFLAT	HLA-DRB1*15:01	6.04	18.41%
	155	169	LCLLAPGMRLLYLDT	HLA-DRB1*15:01	6.49	18.41%
	168	182	DTYRITLIIIGICSL	HLA-DRB1*15:01	8.33	18.41%
	109	123	LGAAFFQMAATDLNF	HLA-DRB1*07:01	1.29	18.23%
	54	68	LLRYVLLVGAAFAEA	HLA-DRB1*07:01	2.03	18.23%
	25	39	RKRWTARLTVPAIVG	HLA-DRB1*07:01	2.81	18.23%
	104	118	NILLALGAAFFQMAA	HLA-DRB1*07:01	3.3	18.23%
	150	164	IVMPLLCLLAPGMRL	HLA-DRB1*07:01	3.81	18.23%
YP_164813.1 NS2b	96	110	KIWVIRMTALGFAAW	HLA-DRB1*15:01	4.94	18.41%
	9	23	TAVGLMFAIVGGLAE	HLA-DRB1*15:01	8.26	18.41%
	27	41	DSMSIPFVLAGLMAV	HLA-DRB1*15:01	8.99	18.41%
	91	105	PGVPWKIWVIRMTAL	HLA-DRB1*15:01	9.3	18.41%
	32	46	PFVLAGLMAVSYTIS	HLA-DRB1*15:01	10	18.41%
	96	110	KIWVIRMTALGFAAW	HLA-DRB1*07:01	3.04	18.23%
	91	105	PGVPWKIWVIRMTAL	HLA-DRB1*07:01	3.68	18.23%
	103	117	TALGFAAWTPWAIIP	HLA-DRB1*07:01	3.82	18.23%
	74	88	RLDVKLDDDGDFHLI	HLA-DRB1*03:01	2.71	17.84%
	27	41	DSMSIPFVLAGLMAV	HLA-DRB1*03:01	2.93	17.84%
YP_164814.1 NS3	18	32	TSPGVYRIMSRYILG	HLA-DRB1*15:01	0.12	18.41%
	144	158	GDIVGLYGNGVILGN	HLA-DRB1*15:01	0.3	18.41%
	152	166	NGVILGNGSYVSAIV	HLA-DRB1*15:01	3.11	18.41%
	235	249	ALKGLPVRYLTPAVN	HLA-DRB1*15:01	3.29	18.41%
	265	279	TLTHRLMSPLRAPNY	HLA-DRB1*15:01	4.69	18.41%
	91	105	GLDDVQLIIVAPGKA	HLA-DRB1*15:01	5.12	18.41%
	436	450	EEGEGRVILSNPSPI	HLA-DRB1*15:01	5.12	18.41%
	96	110	QLIIVAPGKAAINIQ	HLA-DRB1*15:01	5.15	18.41%
	531	545	RKTFLELLRTADLPV	HLA-DRB1*15:01	6.27	18.41%
	493	507	KIMLDNIHLPNGLVA	HLA-DRB1*15:01	7.03	18.41%
	24	38	RIMSRYILGTYQAGV	HLA-DRB1*15:01	7.16	18.41%
	441	455	RVILSNPSPITSASA	HLA-DRB1*15:01	7.55	18.41%
	212	226	AIQRRLRTAVLAPTR	HLA-DRB1*15:01	9.11	18.41%
	18	32	TSPGVYRIMSRYILG	HLA-DRB1*07:01	0.56	18.23%
	531	545	RKTFLELLRTADLPV	HLA-DRB1*07:01	0.81	18.23%
	48	62	TLWHTTRGAAIRSGE	HLA-DRB1*07:01	1.31	18.23%
	400	414	NGDWDFVITTDISEM	HLA-DRB1*07:01	1.57	18.23%
	536	550	ELLRTADLPVWLAYK	HLA-DRB1*07:01	2.63	18.23%
	436	450	EEGEGRVILSNPSPI	HLA-DRB1*07:01	2.89	18.23%
	358	372	KTVWFVASVKMGNEI	HLA-DRB1*07:01	3.88	18.23%

(*continued on next page*)

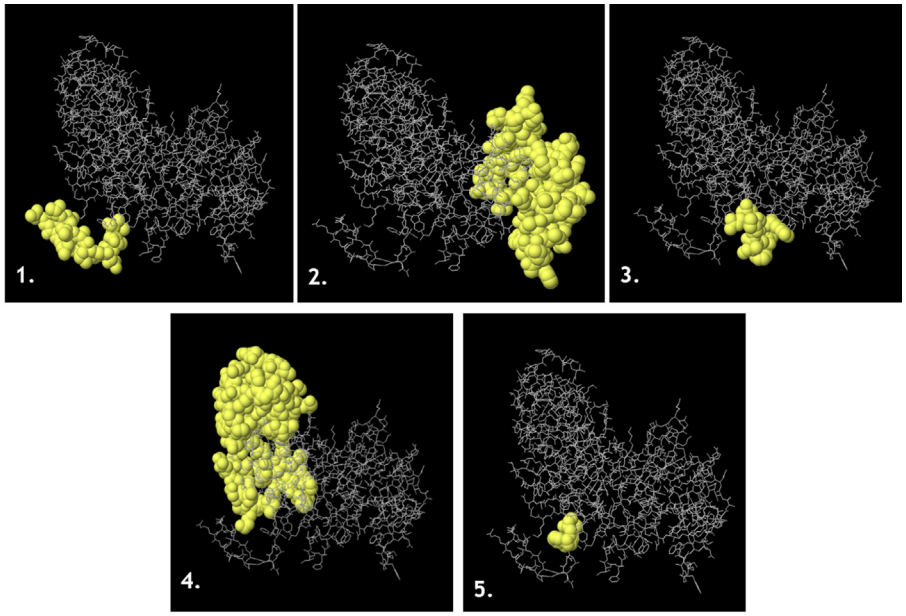
Table 6 (*continued*)

Accession No.	Peptide start	Peptide end	MHC Class II binders	Allele	Consensus percentile rank	%age Population coverage
YP_164815.1 NS4a	106	120	GTLLLALLMMIVLIP	HLA-DRB1*15:01	0.78	18.41%
	48	62	PDALETITLIVALAV	HLA-DRB1*15:01	1.74	18.41%
	82	96	LGGMVLGLATFFLWM	HLA-DRB1*15:01	2.15	18.41%
	57	71	IVALAVMTAGVFLLL	HLA-DRB1*15:01	5.84	18.41%
	87	101	LGLATFFLWMADVSG	HLA-DRB1*15:01	7.22	18.41%
	62	76	VMTAGVFLLLVQRRG	HLA-DRB1*15:01	9.53	18.41%
	68	82	FLLLVQRRGIGKLGL	HLA-DRB1*07:01	7.6	18.23%
	106	120	GTLLLALLMMIVLIP	HLA-DRB1*03:01	1.01	17.84%
	57	71	IVALAVMTAGVFLLL	HLA-DRB1*03:01	1.16	17.84%
	62	76	VMTAGVFLLLVQRRG	HLA-DRB1*03:01	1.95	17.84%
	82	96	LGGMVLGLATFFLWM	HLA-DRB1*03:01	2.17	17.84%
	111	125	ALLMMIVLIPEPEKQ	HLA-DRB1*03:01	4.5	17.84%
	68	82	FLLLVQRRGIGKLGL	HLA-DRB1*03:01	5.89	17.84%
	92	106	FFLWMADVSGTKIAG	HLA-DRB1*03:01	6.07	17.84%
	48	62	PDALETITLIVALAV	HLA-DRB1*03:01	7.08	17.84%
	48	62	PDALETITLIVALAV	HLA-DRB1*01:01	1.81	11.53%
	57	71	IVALAVMTAGVFLLL	HLA-DRB1*01:01	2.28	11.53%
YP_164817.1 NS4b	198	212	VREAGILISAALLTL	HLA-DRB1*15:01	1.56	18.41%
	178	192	QILLIGVSAAALLVN	HLA-DRB1*15:01	3.39	18.41%
	56	70	TPLIKHLVTSEYITT	HLA-DRB1*15:01	4.41	18.41%
	39	53	LRPATAWALYGGSTV	HLA-DRB1*15:01	5.09	18.41%
	203	217	ILISAALLTLWDNGA	HLA-DRB1*15:01	6.96	18.41%
	97	111	TVVLVFLGCWGQVSL	HLA-DRB1*15:01	7.69	18.41%
	218	232	IAVWNSTTATGLCHV	HLA-DRB1*07:01	0.11	18.23%
	111	125	LTTLITAAALATLHY	HLA-DRB1*07:01	1.06	18.23%
	77	91	AQAGSLFNLPRGLPF	HLA-DRB1*07:01	1.41	18.23%
	34	48	ALALDLRPATAWALY	HLA-DRB1*07:01	2.49	18.23%
	61	75	HLVTSEYITTSLASI	HLA-DRB1*07:01	2.49	18.23%
	56	70	TPLIKHLVTSEYITT	HLA-DRB1*07:01	2.78	18.23%
	82	96	LFNLPRGLPFTELDF	HLA-DRB1*07:01	2.81	18.23%
	178	192	QILLIGVSAAALLVN	HLA-DRB1*07:01	3.57	18.23%
	160	174	TDVPELERTTPLMQK	HLA-DRB1*07:01	6.88	18.23%
	39	53	LRPATAWALYGGSTV	HLA-DRB1*07:01	6.99	18.23%
	66	80	EYITTSLASISAQAG	HLA-DRB1*07:01	8.98	18.23%
	34	48	ALALDLRPATAWALY	HLA-DRB1*03:01	0.58	17.84%
	178	192	QILLIGVSAAALLVN	HLA-DRB1*03:01	1.26	17.84%
	198	212	VREAGILISAALLTL	HLA-DRB1*03:01	2.93	17.84%
YP_164818.1 NS5	766	780	MWLLLYFHRRDLRLM	HLA-DRB1*15:01	0.44	18.41%
	112	126	EPMLMQSYGWNLVTM	HLA-DRB1*15:01	1.58	18.41%
	609	623	TYALNTFTNIAVQLI	HLA-DRB1*15:01	1.81	18.41%
	680	694	ANALHFLNSMSKVRK	HLA-DRB1*15:01	3.25	18.41%
	195	209	LEVLQRRYGGGLVRV	HLA-DRB1*15:01	3.39	18.41%
	326	340	VVRLMSKPWDAILNV	HLA-DRB1*15:01	3.49	18.41%
	619	633	AVQLIRLMEAEGVIG	HLA-DRB1*15:01	3.74	18.41%
	771	785	YFHRRDLRLMANAIC	HLA-DRB1*15:01	4.69	18.41%
	761	775	KAYAQMWLLLYFHRR	HLA-DRB1*15:01	6.95	18.41%
	885	899	RDYMLSLRRYEEVNV	HLA-DRB1*15:01	7.05	18.41%
	706	720	HDWQQVPFCSNHFQE	HLA-DRB1*15:01	7.07	18.41%
	21	35	KEDFLKYRKEAITEV	HLA-DRB1*15:01	7.3	18.41%
	158	172	EQRTLRILEMVSDWL	HLA-DRB1*15:01	7.41	18.41%
	517	531	LGYILREMSHHSGGK	HLA-DRB1*15:01	7.47	18.41%
	880	894	GQEKYRDYMLSLRRY	HLA-DRB1*15:01	8.01	18.41%
	117	131	QSYGWNLVTMKSGVD	HLA-DRB1*15:01	8.57	18.41%
	473	487	SRAIWFMWLGARFLE	HLA-DRB1*15:01	9.12	18.41%
	717	731	HFQELIMKDGRTLVV	HLA-DRB1*15:01	9.24	18.41%
	177	191	REFCIKVLCPYMPRV	HLA-DRB1*15:01	9.3	18.41%
	868	882	IYAAINQVRAIIGQE	HLA-DRB1*15:01	9.66	18.41%

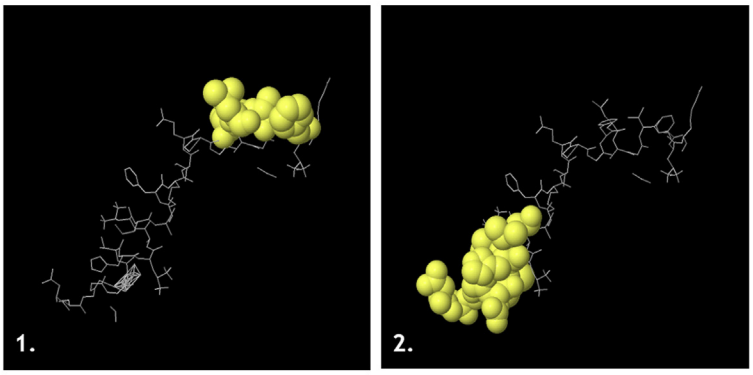
nonstructural proteins were subjected to similarity search discerning diversity and conservation pattern. The physicochemical characterization of the proteins was carried out using PROTPARAM to determine the various parameters enlisted in Table 1.

The characterization of physicochemical attributes of antigenic proteins was a major step discerning the information about the biological activity of viral protein sequences such as instability index, extinction coefficient, GRAVY, aliphatic index, theoretical pI of the protein sequences of USUV [23]. As per the tabulated results, YP_164813.1 NS2b protein and YP_164815.1 NS4a protein showed least instability index; hence conferring for more stability under *in vitro* conditions. Among all viral non-structural sequences, the instability index of YP_164811.1 NS1 was found to be > 40 suggesting that it could be unstable in the test tube environment. Physicochemical characterization enabled the computation of the strength of protein absorbance at the given wavelength per molar concentration, which was reflected by the total amino acid composition [34]. Aliphatic index, i.e., the volume occupied by aliphatic side chains in a viral protein sequence revealed thermo-stability of the protein sequences [35]. The amount of positively charged hydrophobic amino acid residues ranging between -2 and $+2$ estimated the solubility of the protein sequences (Table 1).

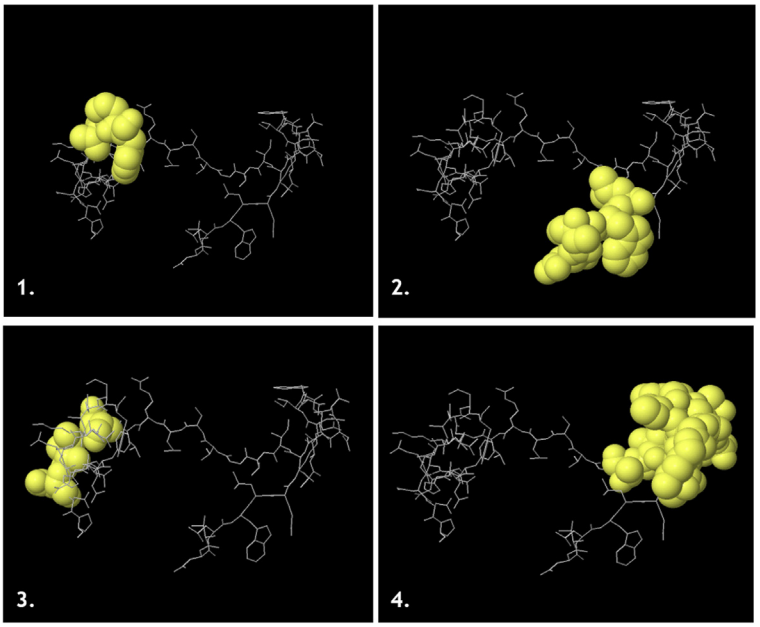
The sequence similarity search BLAST was performed against Protein Data Bank (PDB) using BLASTp using E threshold cut-off value of 0.01 to trace any predetermined tertiary structure of any of the 7 proteins. The absence of predetermined structures of the proteins in the existing PDB necessitated the *in-silico* secondary and the tertiary structure prediction. The secondary structures of the proteins were



(A)

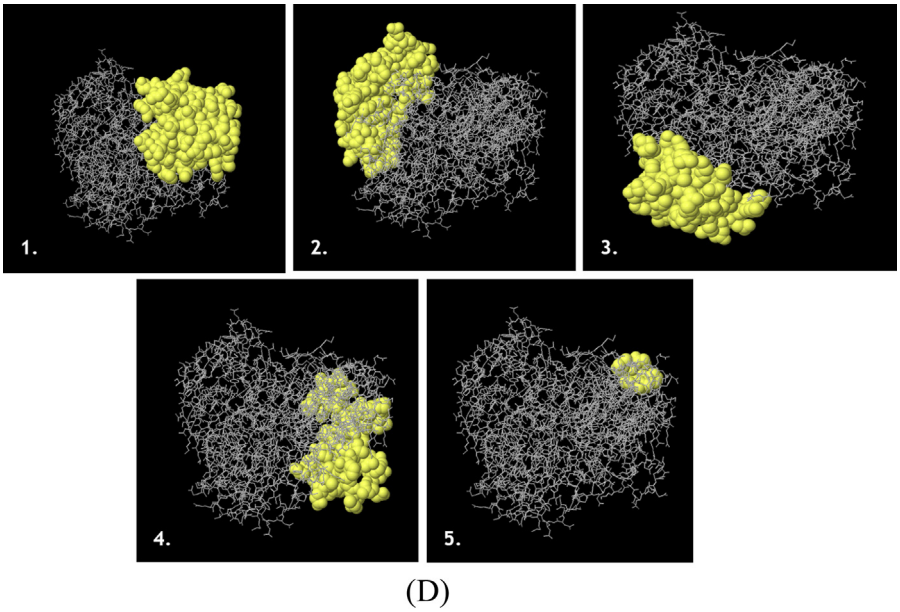


(B)

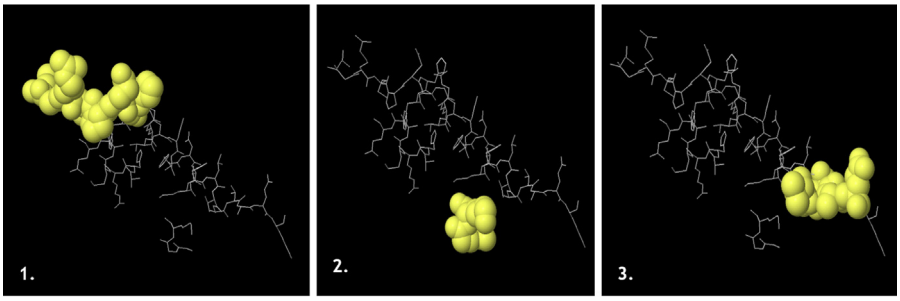


(C)

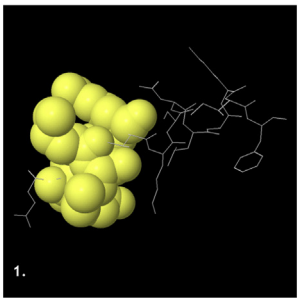
Fig. 4. (A-G)Visualization of (A) NS1 (B) NS 2a (C) NS 2b (D) NS3 (E) NS 4a (F) NS 4b (G) NS 5 discontinuous B cell epitope.



(D)



(E)



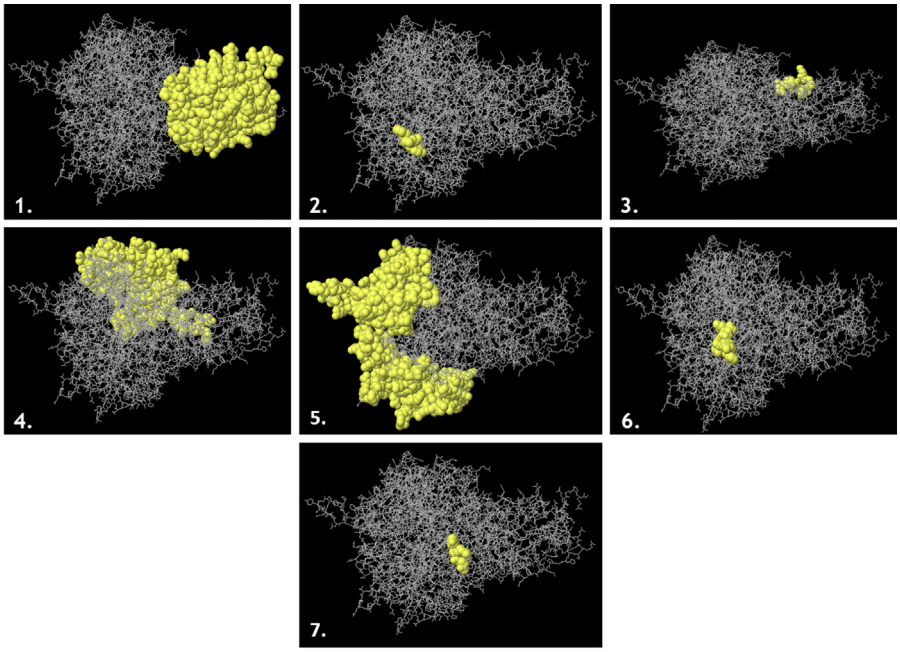
(F)

Fig. 4. (*continued*)

determined using SOPMA server and validated by PSIPRED. The primary protein sequences were evaluated for four conformational states (alpha-helix, extended strand, beta turn and random coil) as given in Table 2. The presence of low percentage of alpha helices in YP_164811.1 NS1 further confers the low stability of the protein molecule. The tertiary structures of all the seven proteins were then computationally determined (Fig. 3)using homology modelling based tool, MODELLER and the structures were validated with PROCHECK (seven generated models with > 90% [36] and QMEAN with maximum Z-scores [37].

The antigenicity of the viral proteins was confirmed by using VaxiJen V2.0 server at the constant threshold of 0.4 (Table 1). Epitope or antigenic determinants are short antigenic peptides that reproducibly elicit B-cell and T-cell response [38]. Both continuous and discontinuous epitopes were predicted by ElliPro [28] suite of IEDB. As ElliPro considers residues that protrude from the surface of globular protein as epitopes, therefore performs two-step procedural epitope identification by (a) performing protein shape approximation as an ellipsoid (b) calculating residue protrusion index (PI) and the other clustering the neighbouring residues based on PI values. The former algorithm defines the continuous epitopes (Table 3a) and the later one defines discontinuous epitopes (Table 3b). The threshold values of protrusion index were taken as 0.5 and 6 as the distance between the centre of mass of each residue [39].

Further, in order to maximize the immune coverage, the identification of T-cell epitopes present on antigen presenting cells (MHC class-I and MHC class-II) was performed using Tepitool [29] of IEDB (Table 4). MHC class-I epitopeswere identified using a total number of 27 most frequent alleles covering > 97% of the population [34] (A*01:01, A*02:01, A*02:03, A*02:06, A*03:01, A*11:01, A*23:01, A*24:02, A*26:01, A*30:01, A*30:02, A*31:01, A*32:01, A*33:01, A*68:01,



(G)

Fig. 4. (*continued*)

Table 7
Population coverage analysis (%) of MHC Class-I and MHC Class-II.

Proteins	Population coverage (World)		Average number of epitope hits/HLA combinations recognized by the population		PC90 (minimum number of epitope hits/HLA combinations recognized by 90% of the population)	
	Class-I	Class-II	Class-I	Class-II	Class-I	Class-II
YP_164811.1 NS1	97.6%	79.83%	5.35	11.36	2.16	1.49
YP_164812.1 NS2a	98.18%	81.81%	6.67	17.2	3.11	1.1
YP_164813.1 NS2b	95.13%	81.81%	3.33	6.27	1.24	1.1
YP_164814.1 NS3	98.55%	81.81%	10.53	21.62	4.5	1.1
YP_164815.1 NS4a	96.83%	78.79%	3.86	6.12	1.76	0.47
YP_164817.1 NS4b	93.85%	81.81%	3.65	13.03	1.23	3.3
YP_164818.1 NS5	98.09%	81.81%	14.48	26	4.31	3.3

A*68:02,B*07:02,B*08:01,B*15:01,B*35:01,B*40:01,B*44:02,-B*44:03,B*51:01,B*53:01, B*57:01, B*58:01) [30]and the window size of 9-mer peptides was considered for prediction; since, MHC class-I binding groove is small, and can accommodate comparatively smaller peptides. IEDB recommended prediction method was employed with predicted consensus percentile rank ≤1 to cover most of the immune response [29]. Further, the peptides were subjected to MHC class-I immunogenicity evaluation using default parameters that masks first, second, and C-terminal amino acids. The scores obtained are sums of propensity score at all unmasked regions. Higher the score and positive, the more immunogenic will be the peptide [40]. Similarly, MHC class-II epitopes were identified using pre-selected panel of 26 most frequent alleles (DRB1*01:01, DRB1*03:01, DRB1*04:01, DRB1*04:05, DRB1*07:01, DRB1*08:02, DRB1*09:01, DRB1*11:01, DRB1*12:01, DRB1*13:02, DRB1*15:01, DRB3*01:01, DRB3*02:02, DRB4*01:01, DRB5*01:01, DPA1*01/DPB1*04:01, DPA1*01:03/DPB1*02:01, DPA1*02:01/DPB1*01:01, DPA1*02:01/DPB1*05:01, DPA1*03:01/DPB1*04:02, DQA1*01:01/DQB1*05:01, DQA1*01:02/DQB1*06:02, DQA1*03:01/DQB1*03:02, DQA1*04:01/DQB1*04:02, DQA1*05:01/DQB1*02:01, DQA1*05:01/DQB1*03:01) [41] and the window size of 15-mer peptides were predicted using IEDB recommended method with predicted consensus percentile rank ≤ 10; since, the MHC class-II has open groove, it can accommodate comparatively longer peptides.

The IEDB recommended method utilizes consensus method that employs a combination of SMM-align, NN-align, and CombLib/Sturniolo for the alleles prediction and works on NetMHCIIpan (NetMHCpan for class-I epitopes) algorithm by default [42]. The protein sequences were first broken down into a set of all possible 15-mers followed by binding affinity prediction of peptides based on the above stated methods. The predicted affinity was then compared to a large set of randomly selected peptides and a percentile rank was then assigned on the individual predicted affinity. The consensus percentile rank score was then calculated as the median of ranks/scores of three different methods. Lower the percentile rank, higher will be the affinity [43]. Additionally, the MHC class-II promiscuous binders were predicted using TepiTool [29] using seven allele method (DRB1*03:01, DRB1*07:01, DRB1*15:01, DRB3*01:01, DRB3*02:02, DRB4*01:01, DRB5*01:01) and the peptides with median consensus percentile rank threshold ≤ 20 were selected(Table 6). This method selects the immunodominant epitopes based on consensus percentile rank of seven DR alleles and covers 50% of the immune response. Promiscuous binders are peptides that can bind to multiple MHC molecules (Fig. 4A–G). Therefore, it has higher antigenicity, covering larger population [44]. The potential immunodominant epitopes have been tabulated in Table 5.

The identified T-cell epitopes were also evaluated for the population coverage to gauge the suitability of the potential vaccine at the global scale. The population dataset for the world was chosen with a motive to

validate the worldwide acceptability of the future vaccine candidates. Class-I and class-II coverage analysis was separately covered (Table 7). The MHC class-I binders were screened primarily based on percentile rank (≤1). Further screening of the peptides was done based on MHC class-I Immunogenicity Scores and the peptides with positive scores were used for population coverage analysis. This was adopted to reduce the redundancy and unnecessary bulk of epitopes generated by the tool. The top 10% of the total peptides contributing significantly to the total percentage of the population coverage for each non-structural protein were selected and have been provided in Table 7. Likewise, the MHC class-II peptides with consensus rank threshold ≤20 were selected and top 10% major contributors to the total population coverage percentile are enlisted in Table 7.

4. Conclusion

The proposed work deals with the application of bioinformatics tools and provides a deep insight into the underlying antigenicity related to USUV. The prediction of B-cell and T-cell MHC class-I & MHC class-II epitopes of non-structural proteins of USUVand their respective toxic potential enables the cost-effective diagnosis and paves the way for the development of vaccine soon to combat USUV induced infections. The suggested *in silico* immunoinformatics strategy might provide a solid platform for the experimental biologists in rapid screening and identification of probable vaccine candidates of USUV. The identified non–toxic epitopes of non-structural proteins of USUV would be a relevant representative of a large proportion of the human population. However, the future in-depth analysis involving experimental validation of the screened epitopes is warranted as a first step towards long term goal of vaccine development against *Usutu Virus*.

Acknowledgments

The authors are thankful to the bioinformatics laboratory facility provided by the Department of Biotechnology, TERI School of Advanced Studies (New Delhi, India). The author (Shafiul Haque) is grateful to the Deanship of Scientific Research, Jazan University, Saudi Arabia for providing the access to the Saudi Digital Library for this research work.

References

[1] J.M. Rijks, M.L. Kik, R. Slaterus, R.P.B. Foppen, A. Stroo, J. IJzer, C.B.E.M. Reusken, Widespread *Usutu virus* outbreak in birds in The Netherlands, Euro Surveill. 21 (45) (2016) 30391.
[2] L. Barbic, T. Vilibic-Cavlek, E. Listes, V. Stevanovic, I. Gjenero-Margan, S. Ljubin-Sternak, G. Savini, Demonstration of *Usutu virus* antibodies in horses, Croatia, Vector Borne Zoonotic Dis. 13 (10) (2013) 772–774.
[3] T. Bakonyi, C. Jungbauer, S.W. Aberle, J. Kolodziejek, K. Dimmel, K. Stiasny, N. Nowotny, *Usutu Virus* infections among blood donors, Austria, July and August 2017–Raising awareness for diagnostic challenges, Euro Surveill. 22 (41) (2017).
[4] U. Ashraf, J. Ye, X. Ruan, S. Wan, B. Zhu, S. Cao, *Usutu virus*: an emerging flavivirus in europe, Viruses 7 (1) (2015) 219–238.
[5] R. Luhken, H. Jöst, D. Cadar, S.M. Thomas, S. Bosch, E. Tannich, J. Schmidt-Chanasit, Distribution of *Usutu virus* in Germany and its effect on breeding bird populations, Emerg. Infect. Dis. 23 (12) (2017) 1994–2001.
[6] J.P. Woodall, The viruses isolated from arthropods at the east african virus research Institute in the 26 years ending december 1963, Proc E Afr Acad 2 (1964) 141–146.
[7] T. Bakonyi, E.A. Gould, J. Kolodziejek, H. Weissenböck, N. Nowotny, 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 328 (2) (2004) 301–310.
[8] J.D. Zeidler, L.O. Fernandes-Siqueira, G.M. Barbosa, A.T. Da Poian, Non-canonical roles of dengue virus non-structural proteins, Viruses 9 (3) (2017) 42.
[9] G. Pauli, U. Bauerfeind, J. Blümel, et al., Usutu Virus, Transfusion Medicine and Hemotherapy, (2014), pp. 73–82 41(1).
[10] D. Vlachakis, Theoretical study of the Usutu virus helicase 3D structure, by means of computer-aided homology modeling, Theor. Biol. Med. Model. 6 (2009) 9.
[11] K.D. Leung, H. Schroder, N.X. White, M.J. Fang, G. Stoermer, J.L. Abbenante, P.R. Martin, D.P. Young, Fairlie Activity of recombinant dengue 2 virusNS3 protease in the presence of a truncated NS2B co-factor, small peptide substrates, and inhibitors, J. Biol. Chem. 276 (2001) 45762–45771.
[12] P. Gaibani, F. Cavrini, E.A. Gould, G. Rossini, A. Pierro, M.P. Landini, V. Sambri, Comparative genomic and phylogenetic analysis of the first usutu virus isolate from a human patient presenting with neurological symptoms. Kapoor A, ed, PLoS One 8 (5) (2013) e64761.
[13] M.L. Yarbrough, M.A. Mata, R. Sakthivel, B.M.A. Fontoura, Viral subversion of nucleocytoplasmic trafficking, Traffic (Copenhagen, Denmark) 15 (2) (2014) 127–140.
[14] H. Malet, M.P. Egloff, B. Selisko, R.E. Butcher, P.J. Wright, M. Roberts, A. Gruez, G. Suzlenbacher, C. Vonrhein, G. Bricogne, J.M. Mackenzie, A.A. Khromykh, A.D. Davidson, B. Canard, Crystal structure of the RNA polymerase domain of the West Nile virus non-structural protein 5, J. Biol. Chem. 282 (14) (2007) 10678–10689.
[15] D. Engel, H. Jöst, M. Wink, Reconstruction of the evolutionary history and dispersal of *Usutu virus*, a neglected emerging arbovirus in europe and africa, mBio 7 (1) (2016) e01938-15.
[16] F. Seif, M. Khoshmirsafa, H. Aazami, M. Mohsenzadegan, G. Sedighi, M. Bahar, The role of JAK-STAT signaling pathway and its regulators in the fate of T helper cells, Cell Commun. Signal.: CCS 15 (2017) 23.
[17] V. Singh, I. Mani, D.K. Chaudhary, P. Somvanshi, Molecular detection and cloning of thermostable hemolysin gene from *Aeromonas hydrophila*, Mol. Biol. 45 (2011) 551–560.
[18] P. Somvanshi, V. Singh, P.K. Seth, Prediction of epitopes in hemagglutinin and neuraminidase proteins of Influenza A virus H5N1 strain: a clue for diagnostic and vaccine development, OMICS A J. Integr. Biol. 12 (2008) 61–69.
[19] P. Somvanshi, V. Singh, P.K. Seth, In silico prediction of epitopes in virulence proteins of *Mycobacterium tuberculosis* H37Rv for diagnostic and subunit vaccine design, J. Proteonomics Bioinf. 1 (2008) 143–153.
[20] P. Somvanshi, P.K. Seth, Analysis of antigenic diversity of T cell epitopes in the prediction of distinct variants of dengue virus by *in silico* methods, Indian J. Biotechnol. (2009) 193–198.
[21] J. Shu, X. Fan, J. Ping, X. Jin, P. Hao, Designing peptide-based HIV vaccine for Chinese, BioMed Res. Int. 2014 (2014) 272950.
[22] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: molecular evolutionary genetics analysis version 6.0, Mol. Biol. Evol. 30 (12) (2013) 2725–2729.
[23] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein identification and analysis tools on the ExPASy server, in: John M. Walker (Ed.), The Proteomics Protocols Handbook, Humana Press, 2005, pp. 571–607.
[24] I.A. Doytchinova, D.R. Flower, VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, BMC Bioinf. 8 (1) (2007) 4.
[25] H. Singh, H.R. Ansari, G.P. Raghava, Improved method for linear B-cell epitope prediction using antigen's primary sequence, PLoS One 8 (5) (2013) e62216.
[26] L. Potocnakova, M. Bhide, L.B. Pulzova, An introduction to B-cell epitope mapping and in silico epitope prediction, J. Immunol. Res. (2016) 6760830.
[27] P. Haste Andersen, M. Nielsen, O. Lund, Prediction of residues in discontinuous B-cell epitopes using protein 3D structures, Protein Sci. 15 (11) (2006) 2558–2567.
[28] J. Ponomarenko, H.H. Bui, W. Li, et al., ElliPro: a new structure-based tool for the prediction of antibody epitopes, BMC Bioinf. 9 (2008) 514.
[29] S. Paul, J. Sidney, A. Sette, B. Peters, TepiTool: a pipeline for computational prediction of T cell epitope candidates, Curr. Protoc. Im. 114 (2016) 18.19.1-18.19.24.
[30] H.H. Bui, J. Sidney, K. Dinh, S. Southwood, M.J. Newman, A. Sette, Predicting population coverage of T-cell epitope-based diagnostics and vaccines, BMC Bioinf. 7 (1) (2006) 153.
[31] S. Gupta, P. Kapoor, K. Chaudhary, A. Gautam, R. Kumar, Open Source Drug Discovery Consortium and G.P.S. Raghava, *In- silico* approach for predicting the toxicity of peptides and proteins, PLoS One 8 (9) (2013) e73957.
[32] V. Uversky, S. Longhi, Flexible Viruses: Structural Disorder in Viral Proteins vol 11, John Wiley & Sons, 2011.
[33] S. Saha, G.P.S. Raghava, Prediction of continuous B-cell epitopes in an antigen using the recurrent neural network, Proteins 65 (1) (2006) 40–48.
[34] S.C. Gill, P.H. von Hippel, Calculation of protein extinction coefficients from amino acid sequence data, Anal. Biochem. 182 (1989) 319–326.
[35] A. Ikai, Thermostability and aliphatic index of globular proteins, J. Biochem. 88 (6) (1980) 1895–1898.
[36] R.A. Laskowski, M.W. MasArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291.
[37] P. Benkert, S.C.E. Tosatto, D. Schomburg, QMEAN: a comprehensive scoring function for model quality assessment, Proteins Struct Funct Bioinforma 71 (2008) 261e77.
[38] P.I. Delves, I.M.C.E. Roitt, Encyclopedia of Immunology, Academic Press, 1998 pp. 3-000.
[39] A.K. Dey, P. Malyala P, M. Singh, Physicochemical and functional characterization of vaccine antigens and adjuvants, Expet Rev. Vaccine 13 (5) (2014) 671–685.
[40] J. Greenbaum, J. Sidney, J. Chung, C. Brander, B. Peters, A. Sette, Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes, Immunogenetics 63 (6) (2011) 325–335.
[41] M. Moutaftsi, B. Peters, V. Pasquetto, D.C. Tscharke, J. Sidney, H.H. Bui, et al., A consensus epitope prediction approach identifies the breadth of murine T CD8+-cell responses to vaccinia virus, Nat. Biotechnol. 24 (7) (2006) 817.
[42] J.J. Calis, M. Maybeno, J.A. Greenbaum, D. Weiskopf, A.D. De Silva, A. Sette, et al., Properties of MHC class I presented peptides that enhance immunogenicity, PLoS Comput. Biol. 9 (10) (2013) e1003266.
[43] M. Nielsen, M. Andreatta, NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptors and peptide length datasets, Genome Med. 8 (1) (2016) 33.
[44] S. Paul, C.S.L. Arlehamn, T.J. Scriba, M.B. Dillon, C. Oseroff, D. Hinz, et al., Development and validation of a broad scheme for prediction of HLA class II-restricted T cell epitopes, J. Immunol. Meth. 422 (2015) 28–34.