

## Functional assignment to JEV proteins using SVM

Ganesh Chandra Sahoo<sup>1,\*</sup>, Manas Ranjan Dikhit<sup>1</sup> and Pradeep Das<sup>1</sup>

<sup>1</sup>BioMedical Informatics Division, Rajendra Memorial Research Institute of Medical Sciences, Agam Kuan, Patna-800007, India; Ganesh Chandra Sahoo\* - Email: ganeshitkgp@gmail.com; \* Corresponding author

received May 24, 2008; revised July 08, 2008; accepted July 20, 2008; published September 02, 2008

### Abstract:

Identification of different protein functions facilitates a mechanistic understanding of *Japanese encephalitis* virus (JEV) infection and opens novel means for drug development. Support vector machines (SVM), useful for predicting the functional class of distantly related proteins, is employed to ascribe a possible functional class to *Japanese encephalitis* virus protein. Our study from SVMProt and available JE virus sequences suggests that structural and nonstructural proteins of JEV genome possibly belong to diverse protein functions, are expected to occur in the life cycle of JE virus. Protein functions common to both structural and non-structural proteins are iron-binding, metal-binding, lipid-binding, copper-binding, transmembrane, outer membrane, channels/Pores - Pore-forming toxins (proteins and peptides) group of proteins. Non-structural proteins perform functions like actin binding, zinc-binding, calcium-binding, hydrolases, Carbon-Oxygen Lyases, P-type ATPase, proteins belonging to major facilitator family (MFS), secreting main terminal branch (MTB) family, phosphotransfer-driven group translocators and ATP-binding cassette (ABC) family group of proteins. Whereas structural proteins besides belonging to same structural group of proteins (capsid, structural, envelope), they also perform functions like nuclear receptor, antibiotic resistance, RNA-binding, DNA-binding, magnesium-binding, isomerase (intra-molecular), oxidoreductase and participate in type II (general) secretory pathway (IISP).

**Keywords:** *Japanese encephalitis* virus (JEV); protein function family; support vector machine; SVMProt

### Background:

*Japanese encephalitis* virus (JEV) is a mosquito borne neurotropic flavivirus, endemic-epidemic throughout Asia [1]. Severe JEV infection is marked by quick onset, headache, high fever, neck stiffness, occasional convulsions (especially in infants) and spastic (but rarely flaccid) paralysis, aseptic meningitis, encephalitis and even death [2]. It is transmitted by female mosquitoes of genera *Culex*, *Anopheles*, *Aedes monsonia* and *Armegeres* mosquitoes. JE virus infection is wide spread in southern states of India viz. Andhra Pradesh, TamilNadu, Uttar Pradesh and Karnataka.

JEV contains a single positive-sense RNA strand with about 11Kb nucleotides [3]. A single precursor polyprotein derived from JEV genome is subsequently processed by the host and viral protease to produce three structural proteins (Capsid (C), membrane(prM/M) and envelope (E)) and seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) [4]. These three structural proteins are synthesized in the order of C, M and E from the 5' half of a single long open reading frames of the flavivirus genome. The glycosylated prM (precursor of M protein) and E proteins appear to be released from the nascent polyprotein following cotranslational cleavage by signal peptidases. Late in virion maturation, prM is cleaved to M [5].

NS1 is a non-structural protein associated with viral RNA replication. Experiments have shown that small interfering

RNAs (siRNAs) could be used to inhibit JEV replication by silencing NS1 protein expression in Vero cells which was evaluated by fluorescence microscope, flow cytometry assay, Western blot and RT-PCR [6].

NS2A is cleaved from NS1 by a membrane bound host protease. NS2B is the cofactor of viral serine protease, correlating with stabilization and substrate recognition of NS3 protease. The NS2B residues Ser46 to Ile60 were the essential region required for both *cis* and *trans* activity of the NS3 protease which has been demonstrated by *cis*- and *trans*-cleavage assays of the deletions at the N-terminal of NS2B [7]. The non-structural protein 3(NS3) of JEV has been proposed to originate from rough endoplasmic reticulum (rER), golgi apparatus or the *trans*-Golgi network (TGN), and serves as a reservoir for viral proteins during virus assembly. Microtubules and TSG101 associate with NS3, which are incorporated into the JEV-induced structure during JEV replication [8].

High hydrophobicity of the NS4 protein supports the fact that this protein played a role as a membrane component and the poor nucleotide sequence conservativity among JEV strains suggested that this region might be important to adapt each viral growth environment [9]. NS5 is a key component of the viral RNA replicase complex that presumably includes other viral nonstructural and cellular proteins, carries both methyl

transferase and RNA-dependent RNA polymerase (RdRp) domains [10].

Functions of different JEV proteins important in different stages of its life cycle haven't yet been determined experimentally and it would take more time for experimental study. The purpose of the study is to find the novel functions of different structural and non-structural proteins of the JEV from its amino acid sequence. By using SVMProt we are able to assign certain functional properties to each protein. Novel vaccine candidates can be prepared targeting the functions of these proteins.

### Methodology:

All databases and softwares used in these studies are publicly available on the world-wide web. Retrieval of *Japanese encephalitis* amino acid sequence was carried out from different databases (NCBI, Swiss Protein Databank (<http://us.expasy.org/sprot/>) and other related databases). Prediction and analysis of different protein function family of different proteins of *Japanese encephalitis* were completed through online search at BIDD (<http://jing.cz3.nus.edu.sg/cgi-bin/svmprot.cgi>). Protein function prediction is of significance in studying biological processes. The web-based software, SVMProt, Support vector machine (SVM) classifies a protein into functional families from its primary sequence based on physico-chemical properties of amino acids. SVMProt shows a certain degree of capability for the classification of distantly related proteins and homologous proteins of different function and thus is used as a protein function prediction tool that complements sequence alignment methods [11].

Scoring of SVM classification of proteins has been estimated by a reliability index and its usefulness has been demonstrated by statistical analysis [12]. R-Value is a scoring function for

estimating the accuracy of support vector machine classification. It is defined as: where  $d$  is the distance between the position of the vector of a classified protein and the optimal separating hyperplane in the hyperspace. P-Value is expected classification accuracy (probability of correct classification). It is derived from the statistical relationship between the R-value and actual classification accuracy based on the analysis of 9,932 positive and 45,999 negative samples of proteins.

### Discussion:

No computational functional analysis of different proteins of JE virus is available till date. All these *in silico* functional analysis give us an idea concerning the role of different proteins of JEV in replication, survival and spread of JEV in the host. With reference to protein function family detection by SVMProt, *Japanese encephalitis* virus proteins belong to different function families which are given in Table 1 (see supplementary material).

From SVMProt it is known that core (capsid) protein and membrane protein precursor does multiple functions e.g. lipid-binding (62.2%), iron binding, metal-binding, calcium-binding, copper-binding, rRNA, mRNA and RNA binding (78-56%). It may also act as channels/pores - pore-forming toxins (proteins and peptides), participate in formation of outer membrane. Function for this protein from SVMProt has been shown in Figure 1. This protein may be responsible for binding of DNA in the host responsible for viral replication and for antibiotic resistance. Experiments have shown that core protein is released from ER membrane, bind to genomic RNA and form nucleocapsid [13]. Membrane protein precursor itself can perform similar functions along with functions like binding to magnesium ion and actin protein, act as enzyme e.g. Transferases-Acyltransferases and P-type ATPase (P-ATPase).

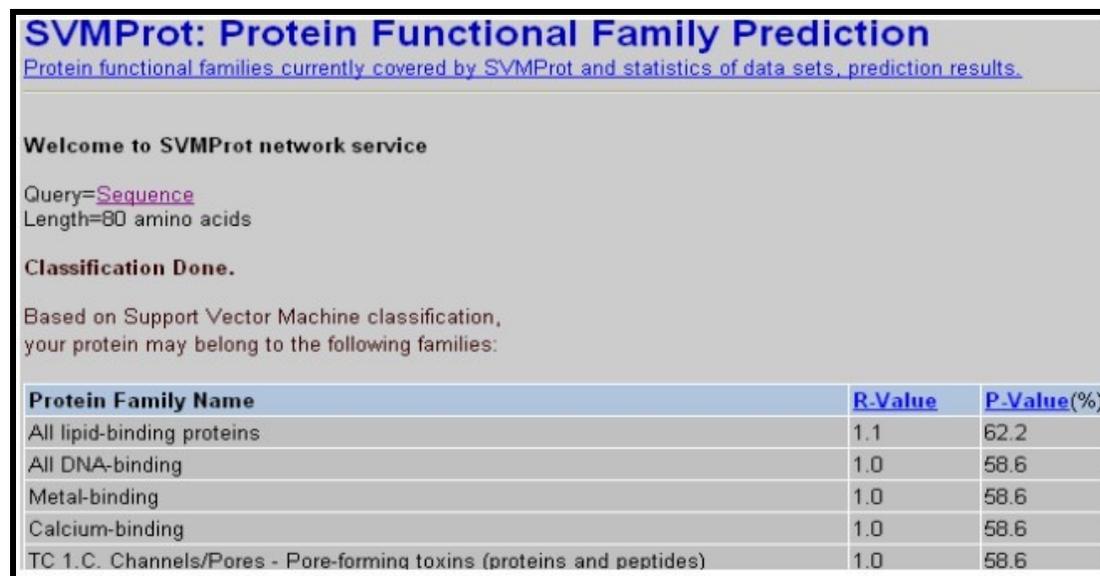


Figure 1: Showing the SVMProt results for core protein and membrane protein precursor of *Japanese encephalitis* virus.

It has been indicated from experiments that host B23 phosphoprotein and core protein interact during JEV replication [27], hence DNA binding and actin binding properties of core protein detected by SVMProt proves that core protein is involved in viral replication by binding to host DNA and actin protein. The electrophoretic mobility shift assay has shown that the purified recombinant capsid protein have DNA binding property [14]. In case of another capsid and PrM protein of the JEV strain 014173 isolated from blood clots collected during the acute phase of infection from an outbreak in the Lakhimpur area of Uttar Pradesh, India, SVMProt classifies it into same protein function families along with zinc-binding and lipid degradation proteins (65.4%).

Capsid protein C belongs to DNA-binding, RNA-binding, iron-binding, copper-binding metal-binding, lipid-binding, and ATP-binding cassette (ABC) family group of proteins. Sequence analysis, homology modeling using SYBYL (based upon a high resolution X-ray structure of ferredoxin (PDB code: 1awd)) and molecular dynamics simulations in aqueous medium using AMBER 6 of the NS4 region of *Japanese encephalitis virus* (JEV) suggests it is closely related to the iron-binding protein ferredoxin as it aligns with essential Cys residues of ferredoxin and might play a role in JEV infection and replication via TNF and other cellular stimuli mediated via redox mechanisms [15]. From SVMProt that NS4B has also been identified as an iron binding protein. Iron is an essential nutrient for the survival of most organisms and has played a central role in the virulence of many infectious disease pathogens [26]. Iron-binding proteins, in combination with antibodies, often have powerful bacteriostatic effects in vitro and are essential for protection against many infections [16]; it suggests that after the regression of JEV infection, antibodies produced with respect to iron binding proteins may protect the individual from bacterial infection. Lipid binding proteins play important roles in cell signaling and membrane trafficking [17], lipid metabolism and transport [18, 19], innate immune responses to bacterial infections [20], and the regulation of gene expression and cell growth [21].

In case of envelope protein comprising of 98 amino acids of the strain GP-14 of CSF of a viremic patient belonging to Gorakhpur, India (ncbi id: ABI94053) it can function as protein responsible for iron-binding, metal-binding, Incompletely Characterized Transport Systems - Recognized transporters of unknown biochemical mechanism, calcium-binding, outer membrane (the name 'envelope' itself indicates that it is an outer membrane protein). NCBI detects that this protein is a Flavivirus glycoprotein having immunoglobulin-like domain of pfam database (pfam02832). Another envelope protein (Chinese strain: GZ04-43) belongs to different protein function families e.g. transmembrane (97%), coat protein (88.1), aptamer-binding protein (88%), zinc-binding (86.8%), metal-binding (68.5%), Hydrolases - Acting on peptide bonds (Peptidases) (65.4%), outer membrane (58.6%), actin binding (58.6%). Envelope protein precursor consisting of 143 amino acids (Japanese strain Oki 589S/JPN03 and host is a swine not human) is possibly an

outer membrane protein (59%) having hydrolytic enzyme activity and acting on ester bonds (62%).

Analysis of glycoprotein M(matrix protein) by SVMProt shows that it may be a transmembrane region protein, participates in Type II (general) secretory pathway (IISP) and isomerase type intramolecular oxidoreductase activities. Analysis of JE virus strain (ncbi id: AAK15789) structural protein (total) proves that it not only belongs to envelope protein, structural protein (matrix protein, core protein, viral occlusion body, keratin), coat protein but also belongs to some other protein families like antibiotic resistance, transmembrane, Lyases - Carbon-Carbon Lyases, metal-binding, zinc-binding, iron-binding, aptamer-binding, lipid metabolism. Other structural protein sequence search at BIDD proves that along with these it can also act as enzyme e.g. Oxidoreductases - acting on CH-OH group of donors and Hydrolases - acting on acid anhydrides.

Analysis of non-structural protein NS1 by SVMProt shows that this protein belongs to zinc-binding, iron-binding, metal-binding, lipid-binding and copper-binding protein function families. Functions of NS2A and NS2B have not yet been determined experimentally. Analysis of non-structural protein NS2A by SVMProt shows that it is a transmembrane-like protein having lipid-binding, metal-binding, calcium-binding functions and belongs to major facilitator family (MFS) which includes transporter proteins. NS2B region of polyprotein shows calcium binding, metal binding, copper binding and iron binding nature but when polyprotein is converted into separate protein i.e. non-structural protein NS2B, it shows different protein functions e.g. transmembrane-like protein (85.4%), responsible for diversity of functions like Transferases-transferring phosphorus-containing groups (68.5%), copper-binding (58.6%) and antibiotic resistance (58.6%).

Analysis of non-structural protein NS3 by SVMProt reveals that it is a membrane-like lipoprotein having zinc-binding, iron-binding activities. It can also act as an enzyme i.e. ATP-binding cassette (ABC) family and Lyases - carbon-oxygen Lyases. Analysis of another NS3 protein having helicase and ATP binding domains whose crystal structure has been determined recently, discloses that it is having zinc-binding, iron-binding, lipid-binding, calcium-binding, magnesium-binding, metal-binding, DNA-binding activities and is responsible for lyses of carbon-carbon bond(carbon-carbon Lyases). Experimental work have established that NS3 have ATPase and helicase activity, motif I, II and VI were composed of an NTP- binding pockets [22]. NS3 protein has been claimed to have serine protease activity. Hence this in silico analysis is an additional evidence to support the experimental result that NS3 possess ATPase activity i.e. ATP-dependent DNA or RNA unwinding action (similar to ABC protein family) and belongs to serine protease where the Carbon-Oxygen bond of serine residue is broken (Lyases - Carbon-Oxygen Lyases protein family) [23]. Other protein functional families for NS3 (e.g. zinc-binding, lipoprotein) are the novel functions described only by SVMProt.

Analysis of non-structural protein NS4A by SVMProt reveals that it is a transmembrane-like protein (82.2%). It may be acting as an antigen (62.2%) in the host, participates in Type II (general) secretory pathway (IISP) protein (58.6%) function families. Analysis of NS4A region of a polyprotein by SVMProt shows that it is also a transmembrane protein, responsible for Group Translocators-Phosphotransfer-driven group translocators, actin binding, calcium-binding and zinc-binding (62.2%).

Analysis of non-structural protein NS4B by SVMProt reveals that it is a transmembrane-like protein (95.7%) and can function as enzyme i.e. Transferases - Transferring Phosphorus-Containing Groups (85.4%) , Hydrolases - Acting on Acid Anhydrides (85.4%), P-type ATPase (P-ATPase) family. It can also act as pore-forming toxins (proteins and peptides), participate in Type II (general) secretory pathway (IISP) family, and it belongs to Major facilitator family (MFS) protein. It may also act as magnesium-binding, iron-binding or metal-binding protein.

Analysis of non-structural protein NS5 by SVMProt shows that it may be involved in functions like lipid-binding, DNA-binding, metal-binding and actin binding. From experiments it is known that NS5 performs two functions e.g. RNA-dependent RNA polymerase and methyl transferase. It is predicted that binding of host DNA, metal and actin filament may be necessary for RdRp and methyl transferase activities of NS5. The GDD motif conserved in most RdRps of plus-strand RNA viruses is essential for metal binding and is considered as catalytic site of the enzyme [24]. Mutation of the first Asp in the GDD motif to Ala in NS5<sub>D668A</sub> results in loss of the RdRp activity.

Glycoprotein M (matrix) is a transmembrane-like protein having enzyme-like functions e.g. isomerases-intramolecular, oxidoreductases and participates in type II (general) secretory pathway (IISP). As this protein also belongs to nuclear receptor group of protein, it is predicted that this protein may have signals required for reaching at the destination site (host cell nucleus) of JEV genome for replication, consequently infection in the host. PrM is an outer membrane protein having lipid-binding, metal-binding, magnesium-binding and copper-binding properties, responsible for nullifying the action of antibiotics as this protein belongs to antibiotic resistance group protein function family. This protein may be helping in host DNA replication so that no error occurs in survival of host cell in which it replicates its own RNA as host cell survival is also important for virus replication.

Three proteins (Matrix, NS2B and Core) have been found to be responsible for antibiotic resistance i.e. prevents action of antibiotics. Three proteins i.e. NS1, NS2A and NS4B have been found to be belonging to MFS (major facilitator family) group of proteins. Core, envelope, NS4A and NS5 proteins have been found to be responsible for binding to actin protein of the host, required during JEV replication in the host.

Experiments based on envelope protein shows that the mutations at residues 364 (Ser-Phe), and 367 (Asn-Ile) of this protein affect early virus-cell interaction in Vero cells and virulence in 3-week-old mice e.g. reduced virulence in 3-week-old mice after peripheral inoculation but were virulent when inoculated intracranially [25]. Mutation in this protein did not make any difference in function still SVMProt correctly predicts that the envelope protein is a transmembrane-like coat protein and some other novel protein function families which add to the earlier experimental knowledge on envelope protein. Protein function families of different JE virus proteins like 6K which are less than fifty amino acids have not been predicted by SVMProt into any protein families.

### Conclusion:

Hence the protein function family predicted by SVMProt is different for each structural and non-structural protein of JE virus strain, some of which may be responsible for virulence or pathogenicity of the virus and others for replication of the virus in the host. Prediction of the functional roles of lipid binding proteins is important for facilitating the study of various biological processes and the search for new therapeutic targets.

From this analysis, it is predicted that there is presence of network of functions performed by JEV proteins which brings about severe complicated clinical manifestations e.g. toxin-like pore forming property of core, matrix and NS4B proteins of JEV is responsible for causing acute flaccid paralysis as pore formation in the host causes release of water, micronutrients and macronutrients which can also occur in nerve cells causing severe inflammation of nerve cells and hence the patient suffers from acute meningitis.

### References:

- [01] T. P. Monath *et al.*, *J Virol.*, 74: 1742 (2000) [PMID: 111650]
- [02] T. Solomona *et al.*, *J Neurol Neurosurg Psychiatry*, 68: 405 (2000)
- [03] H. Sumiyoshi *et al.*, *Virology*, 161: 497 (1987)
- [04] F. Zhang *et al.*, *Viol J.*, 4: 17 (2007) [PMID: 1817644]
- [05] X. Liu *et al.*, *Virus Genes.*, 33: 69 (2006) [PMID: 16791421]
- [06] C. W. Lina *et al.*, *Virus Research*, 127: 88 (2007) [PMID: 17467838]
- [07] C. T. Chiou *et al.*, *J. Gen. Virol.*, 84: 2795 (2003) [PMID: 13679614]
- [08] E. J. Kim *et al.*, *J. Korean Soc. Virol.*, 27: 197 (1997)
- [09] Y. G. Kim *et al.*, *BMC Mol Biol.*, 11: 8 (2007) [PMID: 1934914]
- [10] C. Z. Cai *et al.*, *Nucleic Acids Res.*, 31: 3692 (2003) [PMID: 169006]
- [11] Y. D. Cai *et al.*, *Comput. Chem.*, 26: 293 (2002)
- [12] A. Khromykh *et al.*, *J Virol Methods*, 61: 47 (1996)
- [13] H. N. Tseng *et al.*, *Virus Genes.*, 35: 483 (2007) [PMID: 17957476]
- [14] H. Zhong *et al.*, *Journal of Molecular Graphics*, 23: 223 (2004)

- 
- [15] J. J. Bullen, *Rev Infect Dis.*, 3: 1127 (1981) [PMID: 7043704]
- [16] P. Downes *et al.*, *Trends Cell Biol.*, 15: 259 (2005) [PMID: 15866030]
- [17] J. F. Glatz *et al.*, *Mol. Cell. Biochem.*, 239: 3 (2002) [PMID: 12479562]
- [18] N. H. Haunerland *et al.*, *Prog. Lipid Res.*, 43: 328 (2004) [PMID: 15234551]
- [19] D. Bingle *et al.*, *Trends Immunol.*, 25: 53 (2004) [PMID: 15106612]
- [20] A. Bernlohr *et al.*, *Annu. Rev. Nutr.*, 17: 277 (1997) [PMID: 9240929]
- [21] T. Tsutomu *et al.*, *Virus Genes*, 9: 105 (1995)
- [22] T. Yamashita *et al.*, *Virology*, 373: 426 (2008) [PMID: 18201743]
- [23] K. O'Reilly *et al.*, *Virology.*, 252: 287 (1998) [PMID: 9878607]
- [24] H. Hasegawa *et al.*, *Virology*, 191: 158 (1992 ) [PMID: 1329314]
- [25] C. Yukari *et al.*, *Proc Natl Acad Sci U S A.*, 96: 12844 (1999) [PMID: 23125]
- [26] Y. Tsuda *et al.*, *Med. Entomol. Zool.*, 57: 75 (2006)
- [27] <http://us.expasy.org/sprot/>
- [28] <http://jing.cz3.nus.edu.sg/cgi-bin/svmprot.cgi>

Edited by P. Kanguane

Citation: Sahoo *et al.*, *Bioinformatics* 3(1): 1-7 (2008)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

### Supplementary material

Protein	Functional assignment
Core (capsid) protein and membrane protein precursor. (C and prM)	TC 3.A.1 ATP-binding cassette (ABC) family Magnesium-binding Zinc-binding Lipid degradation proteins (65.4%) rRNA, mRNA and RNA binding Channels/pores - pore-forming toxins (proteins and peptides) Actin binding (62.2%) All lipid-binding proteins (62.2%) Iron-binding (58.6%) All DNA-binding (58.6%) Metal-binding (58.6%) Calcium-binding (58.6%) Lipid degradation (58.6%) Copper-binding (58.6%)
Envelope glycoprotein M (matrix)	Transmembrane TC 3.A.5 Type II (general) secretory pathway (IISP) family EC 5.3.-.-: Isomerases - Intramolecular Oxidoreductases
Envelope glycoprotein(E)	Iron-binding Calcium-binding All lipid-binding proteins TC 9.A. Incompletely Characterized Transport Systems - Recognized transporters of unknown biochemical mechanism Transmembrane (97% ), Coat protein (88.1%), Aptamer-binding protein (88%), Zinc-binding (86.8%), Metal-binding (68.5%), EC 3.4.-.-: Hydrolases - Acting on peptide bonds (Peptidases) (65.4%) Outer membrane (58.6%), Actin binding (58.6%). EC 3.4.-.-: Hydrolases - Acting on peptide bonds (Peptidases) (65.4%) Hydrolase enzyme activity -acting on ester bonds (62%)
Non-Structural Protein NS1	Zinc-binding Iron-binding Metal-binding All lipid-binding proteins Copper-binding
Non-Structural protein NS2A	Transmembrane (98%) All lipid-binding proteins (68.5%) TC 2.A.1 Major facilitator family (MFS) (58.6%) Metal-binding (58.6%) Calcium-binding (58.6%)
Non-Structural protein NS2B	Iron-binding Metal-binding Copper-binding Calcium-binding
NS3 serine protease	Zinc-binding (92%) Iron-binding (91%) Outer membrane (58.6%) TC 3.A.1 ATP-binding cassette (ABC) family EC 4.2.-.-: Lyases - Carbon-Oxygen Lyases(65%) Lipoprotein(85%)
Non-Structural protein NS4A	Transmembrane(82.2%) TC 4.A. Group Translocators - Phosphotransfer-driven group translocators TC 3.A.5 Type II (general) secretory pathway (IISP) family (58.6%)

Non-Structural protein <b>NS4B</b>	Antigen (62.2%) Transmembrane (95.7%) EC 2.7.-.-: Transferases - Transferring Phosphorus-Containing Groups (85.4%) TC 2.A. Electrochemical Potential-driven transporters - Porters (uniporters, symporters, antiporters) (85.4%) EC 3.6.-.-: Hydrolases - Acting on Acid Anhydrides (65.4%) Iron-binding (58.6%) TC 2.A.1 Major facilitator family (MFS)(58.6% ) TC 3.A.3 P-type ATPase (P-ATPase) family(58.6% ) Metal-binding (62.2%) TC 1.C. Channels/Pores - Pore-forming toxins (proteins and peptides) (58.6% )
Flavivirus RNA-directed RNA polymerase (NS5)	Metal-binding lipid-binding DNA-binding actin binding
Polyprotein	All lipid-binding proteins (62.2%) Iron-binding (58.6%) DNA replication (58.6%) Metal-binding (58.6%) All DNA-binding (58.6%) Lipid degradation (58.6%) Calcium-binding (58.6%)
NS4A of Polyprotein	Copper-binding (58.6%) Antibiotic resistance (58.6%) Magnesium-binding (58.6%) Calcium-binding (58.6%) TC 3.A.15 The Outer Membrane Protein Secreting Main Terminal Branch (MTB) family Actin binding Zinc-binding (62.2%)
polyprotein propeptide	Metal-binding Copper-binding Calcium-binding EC 1.2.-.-: Oxidoreductases - Acting on the aldehyde or oxo group of donors Magnesium-binding

**Table1:** SVMProt predicted protein function families for the different structural and non-structural proteins of JE virus.