DOI: 10.21276/ijabpt.2016.7.4.3

INTERNATIONAL JOURNAL OF APPLIED BIOLOGY AND PHARMACEUTICAL TECHNOLOGY

www.ijabpt.com ISSN : 0976-4550

Volume-7, Issue-4, Oct-Dec-2016 CodenIJABFP-CAS-USA

Received: 18th June 2016 *Revised: 2nd July 2016* *Accepted: 2nd July 2016*

DOI: 10.21276/Ijabpt, http://dx.doi.org/10.21276/ijabptResearch article

SEQUENTIAL PHYLOGENETIC CHARACTERIZATION OF ZAIRE EBOLA VIRUS MATRIX PROTEIN VP40: AN *IN SILICO* ANALYSIS

Prashant Ranjan[1], Tanmoy Sarkar[1] and Parimal Das[1*]

[1]Centre for Genetic Disorders, Institute of Science, Banaras Hindu University, Varanasi-221005, India

ABSTRACT:*In silico* analysis of EBOV VP40 matrix protein revealed that VP40 contains hydrophilic segment with three potential phosphorylation sites and two sites for N-Glycosylation. Nine best active sites (Ligand binding site), 13 antigenic sites and 10 different catalytic domains were detected in that.Along with those several RNA binding residues viz. Lysine at position 225, 221, 275, 274, 279 and 274; Asparagines at 200 and 280; Serine at 222; Threonine at 277; Prolines at 50, 165 and 196; Leucine at 51 and Aspartic acid at 296 were detected during the analysis. MSA (Multiple Sequence Alignment) of all five species of EBOV matrix protein including Marburg VP40 shows that the length of peptide and VP40 domain are quite similar in all species than Marburg VP40. Conserve motif sequence analysis indicates that VP40 of Marburg virus has only two conserved motifs sequence while other species of EBOV possess 7 conserved motifs. The Phylogenetic analysis showed that VP40 EBOV out of 16 other negative sense RNA virus matrix proteins is orthologous to Marburg virus VP40 and seven best conserve motifs sequence found in all 17 negative strand RNA virus matrix protein. Novel observations thus made in this study could rationally be used for better and effective drug designing and help study protein function, their interaction and regulatory pathways.

Key words: EBOV, VP40, *Insilico* study, Phylogenetic analysis

*Corresponding author: ParimalDas, Centre for Genetic Disorders, Institute of Science, Banaras Hindu University, Varanasi-221005, IndiaEmail: parimal@bhu.ac.in

Abbreviations

EBOV, Ebola virus; ZEBOV, Zaire Ebolavirus; SEBOV, Sudan Ebolavirus; REBOV, Reston Ebolavirus; BEBOV, BundibugyoEbolavirus; VLPs, Virus Like Particles; NCBI, National Centre for Biotechnology Information; GRAVY, Grand Average of Hydropathicity; MEGA, Molecular Evolutionary Genetic Analysis; NLS, Nuclear Localization Signals; MEME, Multiple EM for Motif Elicitation; NJ,Neighbor-Joining method; CASTp, Computed Atlas of Surface Topography of Protein.

INTRODUCTION

Patients are dying by EBOV due to hyper secretion of cytokines, chemokines and growth factors which influenced the breakdown of vascular epithelium and multiorgan failure (Silva. L.P, et.al., 2012) Still now, there is no standard vaccine and remedy for treatment of EBOV. The EBOV genome contains seven genes viz., NP, VP35, V30, VP40, VP24, GP and L, which encodes seven viral structural proteins, those are NP (major nucleoprotein), VP35 (phosphoprotein), VP30 (minor nucleoprotein), VP40 (matrix protein), VP24 (secondary matrix protein), GP (glycoprotein) and L (RNA-dependent RNA polymerase (Nanbo. A, et. al., 2013). The matrix protein (VP40) is *located just beneath the viral membrane, which is most abundant protein in virions.*

Main role of matrix protein is in virus assembly and budding of virus like particles (VLPs), which takes place at the plasma membrane (Noda. T, et al., 2002) and also plays a key role in viral and host RNA metabolism for the duration of its replications (Xavier. F, et. al., 2003). VP40 protein expressed in mammalian cells, which induces the formation of VLPs. VP30, VP35 and VP40 have ability to suppress RNA silencing (Fabozzi. G, et. al., 2011). VP40 has Late-budding domains (L-domains) that mediate interactions with host proteins to release of VLPs and virus. ISG15 of human inhibits budding of Ebola VP40 VLPs by blocking the ligase activity of Nedd4 (Okumura. A, et. al., 2008).

Here we report the findings of active sites, RNA binding sites, antigenic determinants, Catalytic domains, motifs distribution, phosphorylation and N-Glycosylation sites of EBOV VP40 matrix protein using bioinformatics tools. Active sites, RNA binding sites and antigenic sites is well thought-out as effective drug target sites for molecular based drug designing. Catalytic domains, motifs distribution, phosphorylation and N-Glycosylation sites helps to understand the molecular, cellular and evolutionary role of VP40

MATERIALS AND METHODS

Sequence retrieval

A sequence of VP40 matrix protein of Zaire EBOV (accession ID: AAG40166) was retrieved in Fasta format from NCBI (http://www.ncbi.nlm.nih.gov/) database and crystal structure of VP40 was downloaded from Protein Data Bank (PDB ID- 4LDD).

Physicochemical parameter analysis

The total number of amino acid (a.a) Molecular weight (Mw), theoretical isoelectric point (pI), GRAVY, number of positively charge (+R) and negatively charge (-R) residues, aliphatic index (AI) and instability index (II) was calculated by Protparam tool (http://web.expasy.org/protparam/) (Gasteiger. E, et. al., 2005).

Prediction of Phosphorylation, N-Glycosylation site

Netphos 2.0 server (www.cbs.dtu.dk/services/NetPhos/) (Gupta. R, et. al. 2004) was used to predict the serine, threonine and tyrosinephosphorylation sites whereas N-Glycosylation 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/) (Blom. N, et. al., 1999) was used to predict the N-Glycosylation site.

Functional annotation

Functional enrichment of VP40 protein at molecular, biological and cellular level were analyzed with high confidence function prediction by PFP (Protein Function Prediction) tool (http://kiharalab.org/web/pfp.php)

Identifications of Active site, RNA binding site and Antigenic site

Active site prediction was done with the help of CASTp from the crystal structure (Dundas. J, et. al., 2006). Active site is the important analysis for docking studies. RNA binding interface prediction was done from crystal structure of protein by KYG (http://cib.cf.ocha.ac.jp/KYG/)(Kim. O.T.P, et. al., 2006) (http://cib.cf.ocha.ac.jp/KYG/). Antigenic site of protein sequence that is likely to be antigenic by eliciting an antibody response. Antigenic sites are determined using the method of Kolaskar and Tongaonkar. Predictions are based on a table that reflects the occurrence of amino acid residues in experimentally known segmental epitopes (http:/ /imed.med.ucm.es /Tools/antigenic.pl).

Prediction of catalytic domains and motifs analysis

Catalytic domains of Zaire EBOV VP40 were identified using Motif scan (myhits.isb-sib.ch/cgi-bin/ motif scan) (Sigrist. C. J, et. al., 2010). MEME (Multiple Expectation-Maximization for Motif Elicitation) 4.6.0 was useful to show presence ofdifferent motifs ofconservedsequences in EBOV and Marburg VP40 (Bailey. T.L, et. al., 2006).

Identification of nuclear and subcellular localization

Subcellular localization was done by WoLF PSORT program (Horton. P, et. al., 2007) (Jia. F, et al., 2013) and nuclear localization signal (NLS) was done by NLStradamus software (Nguyen. B.A.N, et. al., 2009).

Secondary structure prediction

Secondary structural properties of protein such as alpha helix, 310 helix, Pi helix, beta bridge, extended strand, beta turns, bend region, random coil, ambiguous states and other states of the protein was predicted through SOPMA program (Self-Optimized Prediction Method), the program which determined individual role of amino acid for building the secondary structure with their positions (http:// npsapbil.ibcp.fr/cgibin/npsa_automat.pl?page =/NPSA /npsa_sopma.html) (Geourjon. C, and Deléage. G, et. al.,1995).

Multiple Sequence Alignment.

The Clustal X offline software (Larkin. M.A, et. al., 2007) were used for multiple sequence alignment of VP40 matrix protein of all group of EBOV including Marburg virus.

Phylogenetic Analysis

VP40 is activated as octameric and hexameric form which is liable for virus replication. So phylogenetic analysis of VP40 may gave better evidence for better drug designing. VP40 protein was analyzed to study the evolutionary relationship with other 16 viral matrix proteins of negative-sense RNA. Along with, evolutionary analysis of all proteins (*NP, VP35, V30, VP40, VP24, GP and L*, proteins) of five types of EBOV and Marburg virus proteins was carried out based on multiple sequence alignments (MSA) and neighbor joining method using Molecular evolutionary Genetic analysis (MEGA5.1) software (Tamura. K, et. al., 2011). Bootstrap analysis of 1,000 replicates was performed to provide confidence estimates for the tree topologies.

RESULTS AND DISCUSSION

Physicochemical parameter analysis.

The physiochemical parameter showed the stability of the protein about its functions. The isoelectric point of VP 40 is 8.76 at this point; it is least soluble in nature. The extinction coefficient indicates that how much light absorbed by a protein at 280 nm in water. II of vp40 is 40.39, which indicates instability of protein. II estimate the stability and instability of protein in test tube and it should be less than 40 for stability of protein. Aliphatic index of protein is the relative volume of a protein occupied by aliphatic side chains (Alanine, Valine, Isoleucine and Leucine) of vp40 protein. It showed a positive factor for the increase of thermo stability of globular proteins. GRAVY value is the sum of hydropathy value of all amino acids, divided by the number of residues in the sequence. Increasing positive score indicates greater hydrophobicity and vice versa. The GRAVY value with Mw 35182.8 of VP 40 is -0.052, indicating better interaction with water hence is hydrophilic protein. There are 26 negative charge residues and 29 positive charge residues and due to large number of positive residues the attribute of protein becomes acidic.

Prediction of Phosphorylation and N-Glycosylation site.

N-glycosylation occurs in asparagine amino acid residues present in asn/Xaa-Ser/Thr stretches where Xaa can be any amino acid except proline. This consensus tripeptide is referred as N-glycosylation sites. VP40 protein contains one serine, one threonine and two tyrosine phosphorylation sites and two N-Glycosylation potential sites (Figure1.). The potential crossing,the threshold of 0.5 represents a predicted phosphorylated site and N-Glycosylation site. Protein phosphorylation at serine, thereonine or tyrosine residue affects multitude of cellular signaling process. N-Glycosylated proteins would increase the performance of network such as protein interaction, protein folding.

Functional annotation.

The functions of VP40 protein at molecular, biological and cellular level under study were predicted by PFP from Kihara Bioinformatics Laboratory (Purdue University, USA) which uses references from Gene Ontology (Table1.). Function of protein predicted on the basis of topology similarity to structures with known function, geometric and residue similarity of predicted functional sites to regions of known structures and sequence homology to functionally annotated sequences.

Identifications of Active, Antigenic and RNA binding sites.

Active site, Antigenic site and RNA binding site prediction helps in description of cellular activities, drug target and whole cell engineering and also play an important role in function and structural organization of a cell. Determination of the active site residues (Figure 2) for the binding of ligand for inhibiting the activity of the function of target protein. In this study we get total nine best active sits positioned in between 44 and 326 amino acid residues. The largest active site has an area of 704.4 Å and volume of 869.3 Å. Antigenic determinant program predicts antigenic segment within a protein sequence that are to be expected antigenic against antibody response. There were total 13 antigenic sites determined using method of Kolaskar and Tongaonkar (Table2.). KYG (Kim. O.T.P, et. al., 2006) was used to determine the RNA interface residues on a protein surface (Wass. M. N, et. al., 2009).The susceptibility of residue occurrence in the interface of protein and RNA molecules observed in protein-RNA complex structures.The RNA binding interface residue prediction reveals that residues Lysine (at position 225, 221, 275, 274, 279 and 274) Asparagine (position at 200 and 280), Serine (position at 222), Threonine (position at 277), Proline (position at 50,165, 196), Leucine (51) and Asparagine (296) are RNA binding site (Figure-3).

Prediction of catalytic domains and motifs analysis

Motif scan tool scan all known motifs that present in a sequence. There are number of motifs present in one domain and it decides the function of protein. Motif scan predicted 11 different catalytic domains with their position in sequence and each motif has different function (Table3.).Motifs distribution in EBOV species and Marburg matrix protein shows in Figure5b. There are seven motifs sequence obtained and out of seven only two motifs sequence similar in VP40 (motifs 1 **(AAIMLASYTITHFGKTSNPLVRINRLGPGIPDHPLRLLRIGNQAFLQEFV)& 2 (DHPSHTPNSVASAFILEAMVNVISGPKVLMKQIPIWLPLGVSDQKTYSFD)** of Marburg as compared to other species of EBOV matrix protein (all seven motifs present).In case of all 17 negative sense RNA virus matrix proteins revealed that there is no motifs distribution in Zaire EBOV and Rabies VP40.Moreover, motif 3 (**LVHRLTGKKITTKNGQPIIPILLPKYIGMDPISQGDLTMVITQDCDTCHS**) is common in all negative sense RNA virus matrix proteins except Zaire, Rabies and turkey VP40 (Fig6 b&c.). Motifs distributions in Ebola virus proteins (including all five species) along with Marburg virus proteins indicates the presence of total seven best possible motifs and motif 3 (**VRRTAGLNEKLVFYNNTPLTLLTPWRKVLTTGSVFNANQVCNAVNLIPLD**) is common in all proteins of Zaire, Bundbugyo and Sudan EBOV and motif 3(**VRRTAGLNEKLVFYNNTPLTLLTPWRKVLTTGSVFNANQVCNAVNLIPLD**) is common in all proteins of Reston EBOV except GP and VP30.

Same as Reston, motif 3 is common in Tai EBOV proteins except GP and L. In case of Marburg motif3 is uncommon in VP35, VP40 and GP (Fig7 b & c.). These conserved motifs could be the important elements determining the molecular function among different matrix protein species.

Identification of nuclear localization signal or sequence and sub cellular localization

Activity of protein depends on localization of protein because posttranslational modification occurs on the basis of localization of protein in the cell. Interaction of protein differs due to posttranslational modification. Nuclear and sub-cellular localization analysis revealed that VP40 proteins of EBOV were distributed in mitochondria, cytoplasm, and peroxisome as well as in nucleus.

Table-1: Function prediction of VP40 protein using PFP tool.

Molecular Function	Biological Function	Cellular Component
RNA binding	Viral release from host cell	Virion membrane
Nucleic acid binding	Virus-host interaction	Host cell plasma membrane
Binding	Interaction with host	Host cell endomembrane system
	Viral reproduction	Host cell membrane
	Symbiosis, encompassing mutualism through parasitism	Ribonucleoprotein complex
	Interspecies interaction between organisms	Other organism membrane

Table-2: Antigenic sites prediction: There were 11 antigenic sites determined at the different position of amino acid residues.

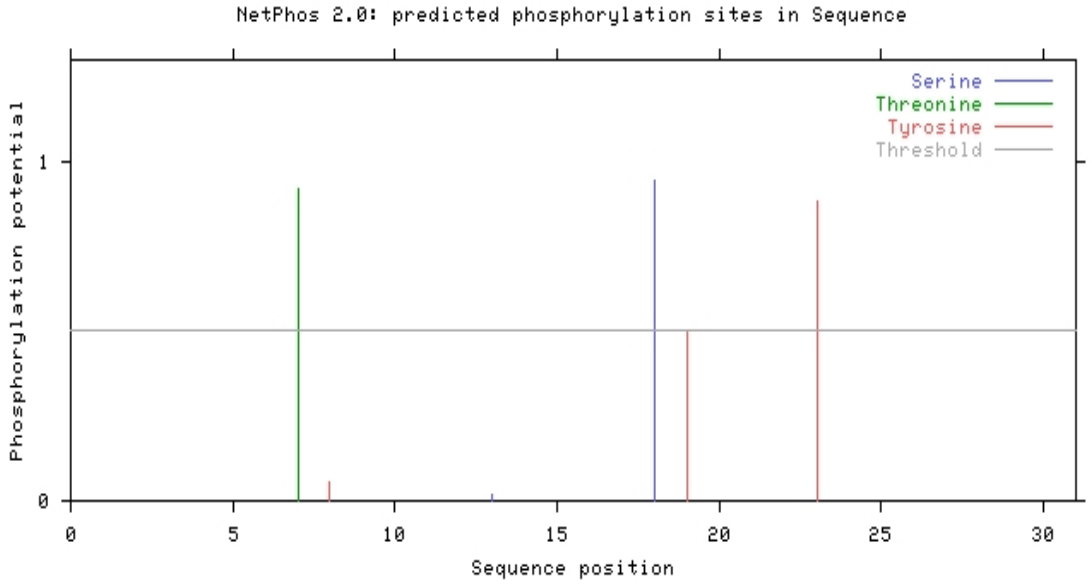
S.No	Start Position	Sequence	End Position
1	14	MEAIYPVRS	22
2	60	DHASHTPGSVSSAFILEAMVNVISGPKVLMKQIPIWLPLGVAD	102
3	113	AAIMLASYTI	122
4	129	TNPLVRVNR	137
5	142	IPDHPLRLLRI	152
6	154	NQAFLQEFVLPPVQLPQYFTFDLTALKLITQPLPA	188
7	203	LRPGISFHPKLRPILLP	219
8	244	LQDFKIVPI	252
9	259	MGIEVPETLVHK	270
10	282	QPIIPVLLPKYIGLDPVAP	300
11	302	DLTMVITQDCDTCHSPASLP	321

Secondary structure prediction

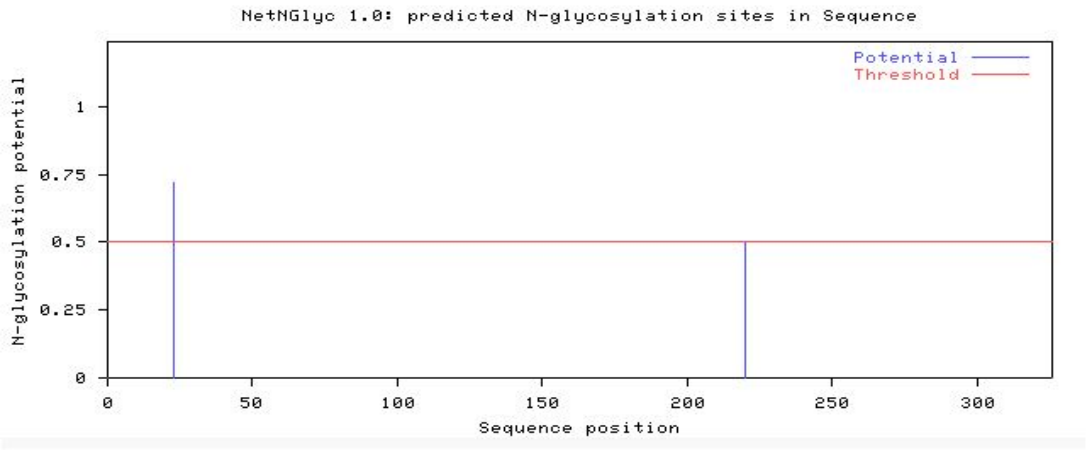
SOPMA is a secondary structure prediction program (self optimized prediction method) that uses multiple alignments methods aimed at predicting protein structure and function. The results of this analysis are depicted that it contains of 17.18 % alpha helix, 25.77% extended strand, 6.75% and 50.35% β-turn and random coil (Figure-4.).

Table-3: Catalytic arrangement in deduced amino acid sequence of VP40 protein using motif scan.

Catalytic Domain	Position	Function
Amidation	222-225, 272-275	Receptor recognition and signal transduction
Asn_Glycosyltion	23-26, 220-223	Post translational modification
CAMP_Phospho_site	274-277	Glycogen, sugar and lipid metabolism in cell.
Ck2_Phospho_site	190-193, 232-235, 243-246	Phosphorylates the acidic protein.
Myristyl	29-34, 67-72, 198-203	Plays vital role in membrane targeting and signal transduction.
PKC_Phospho_site	222-224, 272-274, 277-279	Receptor desensitization and regulating transcription.
Multiheme_cytc	306-320	It helps in transcription regulation
PFTA(Protein prenyltransferases)	1-11	Formation of stable complex
DUF1602	39-53	Unknown
Dak2(dihydroacetone kinase)	66-180	Glycerone kinase activity, cellular response to toxic substance
Vp40	1-295	It plays an important role in replication of EBOV



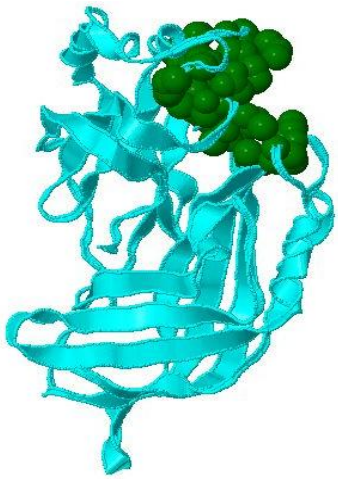
(a)



(b)

Figure-1: Prediction of (a) Phosphorylation sites and (b) N-Glycosylation potential sites in VP40 EBOV

(a)



(b)

Figure-2: Indicates information about active sites using CasTp. (a) Green color showed the active site area from 44 to 326 with the beta-sheet and alpha sheet in between them. (b.) The 3D structure of best active site, which shown in green color.

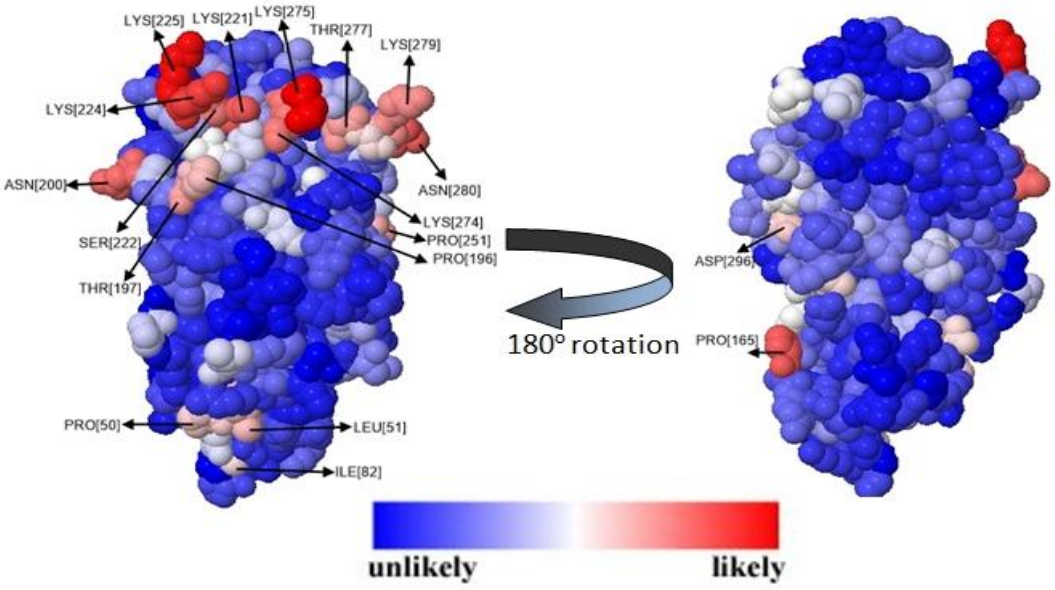


Figure-3: RNA interface residue prediction: The protein is color coded residue-wise as shown above. Red color showed highly likely to be an interface, light blue color showed unlikely to be an interface and deep blue color showed buried residue (not considered as an interface residue). Above structure rotate by 180 ° for showing the other surface area of structure.

```
        10        20        30        40        50        60        70
         |         |         |         |         |         |         |
MRRVILPTAPPEYMEAIYPVRSNSTIARGGNSNTGFLTPESVNGDTPSNPLRPIADDTIDHASHTPGSVS
hceeeccccchhhhhheeeccttceeeetccccceeecccccccccccccccccccccccccccttccc
SAFILEAMVNVISGPKVLMKQIPIWLPLGVADQKTYSFDSTTAAIMLASYTITHFGKATNPLVRVNRLGP
hhhhhhhhhhhccccteeeeccceeeeeeccttceeeeecchhhheeeeeeeeecccccccceeeeeccccc
GIPDHPLRLLRIGNQAFLQEFVLPPVQLPQYFTFDLTALKLITQPLPAATWTDDTPTGSNGALRPGISFH
cccccceeeeeetchhhhhheccccccccceeeeehhhhheecccccccccccccccccccccttcccc
PKLRPILLPNKSGKKGNSADLTSPEKIQAIMTSLQDFKIVPIDPTKNIMGIEVPETLVHKLTGKKVTSKN
ttcceeeecccccccccceecccchhhhhhhhhhttteeeecccccceeeecchhhhhhttceeecctt
GQPIIPVLLPKYIGLDPVAPGDLTMVITQDCDTCHSPASLPAVIEK
ccccceeeccceeecccccttceeeeeecccccccccccchhhhhh
```

Figure- 4: Secondary structure of EBOV VP40 protein predicted using SOPMA: The secondary structure like helical structure (h) indicates in blue color, extended strand (t) showed with green color, β-sheet (e) depicted in red color and random coils (c) are in yellow color

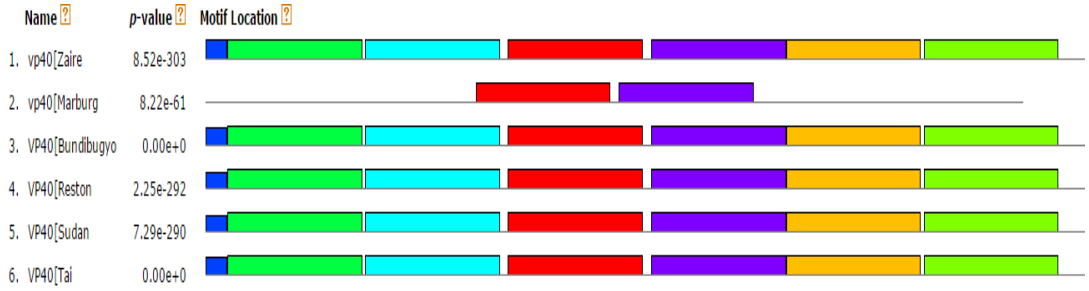
```
                    10        20        30        40        50        60        70        80
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|...
VP40[Bundibugyo     MRRAILPTAPPEYIEAVYPMRTVSTSINSTASGPNFPAPDVMMSDTPSNSLRPIADDNIDHPSHTPTSVSSAFILEAMVNVIS
VP40[Tai            MRRIILPTAPPEYMEAVYPMRTMNSGADNTASGPNYTTTGVMTNDTPSNSLRPVADDNIDHPSHTPNSVASAFILEAMVNVIS
vp40[Zaire          MRRVILPTAPPEYMEAIYPVRSNSTIARGGNSNTGFLTPESVNGDTPSNPLRPIADDTIDHASHTPGSVSSAFILEAMVNVIS
VP40[Reston         MRRGVLPTAPPAYNDIAYPMSILPTRPSVIVNETKSDVLAVPGADVPSNSMRPVADDNIDHSSHTPSGVASAFILEATVNVIS
VP40[Sudan          MRRVTVPTAPPAYADIGYPMSMLPIKSSRAVSGIQQKQEVLPGMDTPSNSMRPVADDNIDHTSHTPNGVASAFILEATVNVIS
vp40[Marburg        ---MASSSNYNTYMQYLNSPPYADHGANQLIPADQLSN---------QQGITPNYVGDLNLDDQFKGNVCHAFTLEAIIDISA
Clustal Consensus         .:    * :   .                             .: : *   . ::  .:   .*. ** *** ::: :

                             110       120       130       140       150       160       170       180
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|...
VP40[Bundibugyo     ADQKTYSFDSTTAAIMLASYTITHFGKTSNPLVRINRLGPGIPDHPLRLLRIGNQAFLQEFVLPPVQLPQYFTFDLTALKLIT
VP40[Tai            SDQKTYSFDSTTAAIMLASYTITHFGKTSNPLVRINRLGPGIPDHPLRLLRIGNQAFLQEFVLPPVQLPQYFTFDLTALKLIT
vp40[Zaire          ADQKTYSFDSTTAAIMLASYTITHFGKATNPLVRVNRLGPGIPDHPLRLLRIGNQAFLQEFVLPPVQLPQYFTFDLTALKLIT
VP40[Reston         ADQKIYSFDSTTAAIMLASYTVTHFGKISNPLVRVNRLGPGIPDHPLRLLRLGNQAFLQEFVLPPVQLPQYFTFDLTALKLIT
VP40[Sudan          ADQKTYSFDSTTAAIMLASYTITHFGKANNPLVRVNRLGQGIPDHPLRLLRMGNQAFLQEFVLPPVQLPQYFTFDLTALKLVT
vp40[Marburg        MSNFEYPLAHTVAALLTGSYTITQFTHNGQKFVRVNRLGTGIPAHPLRMLREGNQAFIQNMVIPRNFSTNQFTYNLTNLVLSV
Clustal Consensus    .:  *.:  *.**:: .***:*:* :  : :**:**** *** ****:** *****:*::*:*    .: **::** * * .

                             210       220       230       240       250       260       270       280
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|...
VP40[Bundibugyo     GILRPGISFHPKLRPILLPGKTGKRGSSSDLTSPDKIQAIMNFLQDLKLVPIDPAKNIMGIEVPELLVHRLTGKKITTKNGQP
VP40[Tai            GTLRPGISFHPKLRPILLPGRAGKKGSNSDLTSPDKIQAIMNFLQDLKIVPIDPTKNIMGIEVPELLVHRLTGKKTTTKNGQP
vp40[Zaire          GALRPGISFHPKLRPILLPNKSGKKGNSADLTSPEKIQAIMTSLQDFKIVPIDPTKNIMGIEVPETLVHKLTGKKVTSKNGQP
VP40[Reston         NALRPGLSLHPKLRPILLPGKTGKKGHASDLTSPDKIQTIMNAIPDLKIVPIDPTKNIVGIEVPELLVQRLTGKKPQPKNGQP
VP40[Sudan          GALRPGLSFHPKLRPVLLPGKTGKKGHVSDLTAPDKIQTIVNLMQDFKIVPIDPAKSIIGIEVPELLVHKLTGKKMSQKNGQP
vp40[Marburg        NTMHPAVSIHPNLPPIVLPTVKKQAYRQHKNPNNGPLLAISGILHQLRVEKVPEKTSLFRISLPADMFSVKEGMMKKRGENSP
Clustal Consensus   . ::*.:*:**:* *::**    :     . .    : :*   : ::::  :   ..:. *.:*  :.    *      :..*

                             310       320       330
                    ....|....|....|....|....|....|.
VP40[Bundibugyo     GDLTMVITQDCDTCHSPASLPPVSEK-----
VP40[Tai            GDLTMVITQDCDSCHSPASLPPVNEK-----
vp40[Zaire          GDLTMVITQDCDTCHSPASLPAVIEK-----
VP40[Reston         GDLTMVITQDCDSCHSPASHPYHMDKQNSYQ
VP40[Sudan          GDLTMVITPDYDDCHSPASCSYLSEK-----
vp40[Marburg        RQVVLAYANPTLSAV----------------
Clustal Consensus    ::.:. :     .
```

(a.)

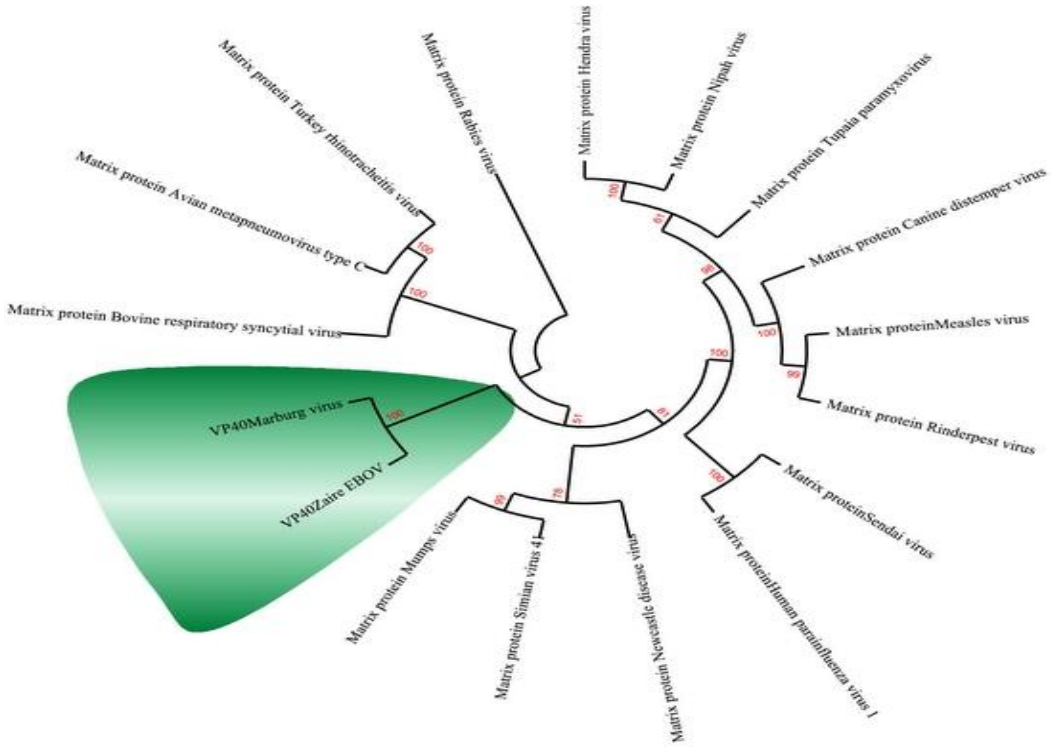


(b.)

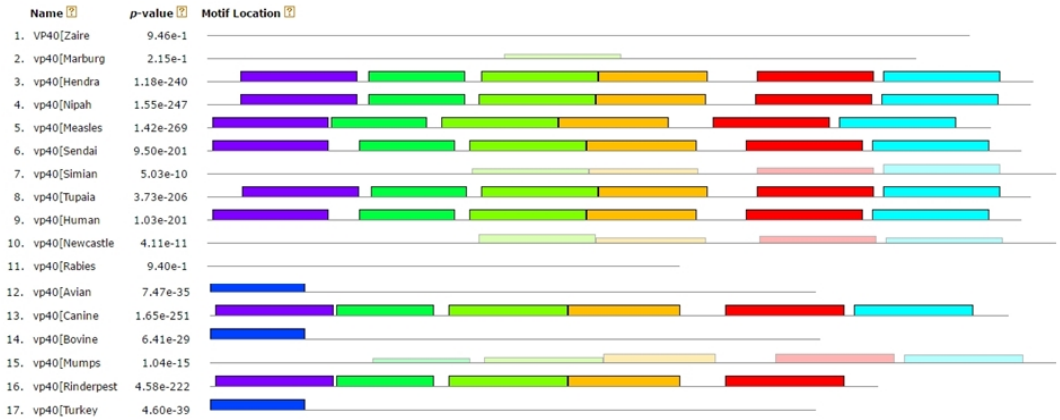
Motifs	Conserve motifs sequence
Motif1	LCYPLMDINEDLNRLLWRSRCKIVRIQAVLQPSVPQEFRIYDDVIINDDQ
Motif2	FMVHVGNFRRKKNEAYSADYCKMKIEKMGLVFALGGIGGTSLHIRSTGKM
Motif3	VRRTAGLNEKLVFYNNTPLTLLTPWRKVLTTGSVFNANQVCNAVNLIPLD
Motif4	EVYDFDQSSWYTKGSLAPILPTTYPDGRLIPQVRVIDPGLGDRKDECFMY
Motif5	TPQRFRVVYMSITRLSDNGYYTVPRRMLEFRSVNAVAFNLLVTLRID
Motif6	FLLGIVEDSDPLSPPRGRTFGSLPLGVGKSTAKPEELLKEV
Motif7	EADQAITQARIAPYAGLILIMTMNNPKGIFKKLGAGMQVIVELGPYVQAE

(c.)

Figure 5. Above information shows the result of MSA of EBOV and Marburg VP40 and conserve motifs analysis of matrix proteins. (a.) MSA of matrix protein revealed considerable variations around domain area however '*' shows residues in the column are identical in all sequence, ':' indicates conserve substitution, '.' represents semi conserved substitution and dashes between sequence shows deletion (b.) VP40 of Marburg virus has only two conserved motifs sequence (i.e. motif 1& 4) however, other species of EBOV shows 7 conserved motifs mentioned in table. (c.) Seven conserved motifs are found in all five.



(a)



(b)

Motifs	Conserve motifs sequence
Motif1	FMV VGNFRRKKNEAYSADYCKMKIEKMGLVFALGGIGGTSL IRSTGKM
Motif2	FAQMGFQKNLCFSLMDIN WLNRLTWNNSCEISRVAAVLQ SV REFMIY
Motif3	VRRTAGLNEKLVFYNNT LTLLT WRKVLTTGSVFNANQVCNAVNLI LD
Motif4	EVYDFDQSSWYTKGSLA IL TTYP DGRLI QVRVID GLGDRKDECFMY
Motif5	T QRFRVVYMSITRLSDNGYYTV RRMLEFRSVNAVAFNLLVTLRID
Motif6	FLLGIVEDSD LS RGRTFGSL LGVGKSTAK EELLKEV
Motif7	MESYIIDTYQGV YTAAVQVDLIEKDSN ATLTVWF LFQ

(C)

Figure 6. (a) Phylogenetic connection of VP40 matrix protein and their evolutionary relationship with 16 other viral matrix proteins of negative-sense RNA virus. This analysis revealed that VP40 matrix protein of EBOV is orthologous to Marburg VP40 matrix protein. The others 16 negative –sense RNA virus matrix protein was collected from NCBI database Accession Nos.: Hendra virus (AAB39504), Nipah virus (NP_112025), Measles virus (ACC86103), Newcastle virus (AAY78544), Siendai virus (AAB06280), Simian virus41 (Yp_138507), Tupaiaparamyxovirus (NP_054694), Human parainfluenza virus type I (AAA46861), Rabies virus (ADK90865), Avian metapneumo virus type C (AGB67520), Canine distemper virus (AAA87372), Bovine respiratory syncytial virus (AAC96305), Rinder pest virus (CAA53779), Mumps virus (BAA13025), Turkey rhinotracheitis virus (CAA41496), Marburg virus (CAA78116). (b) Figure represents conserve motifs analysis. (c) Seven best conserve motifs sequence found in all 17 negative strand RNA filamentous virus matrix protein.

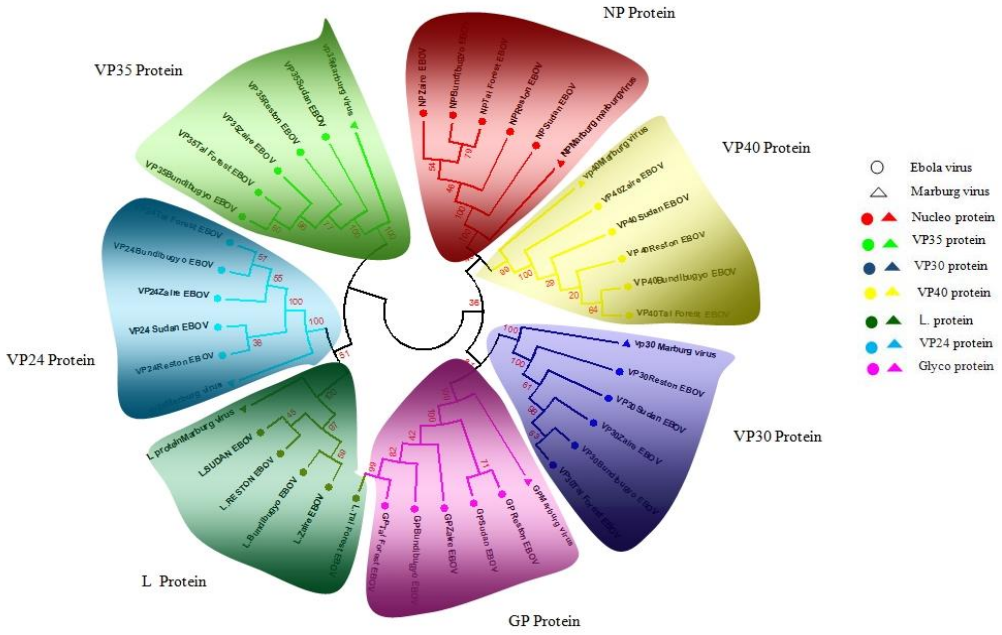
Multiple Sequence Alignment.

The multiple sequence alignment of all the five species of EBOV matrix protein VP40 including Marburg virus revealed that the lengths of the peptide and the VP40 domains are quite similar in all species of EBOV VP40 than Marburg matrix protein; nevertheless, substantial variation was found in the mature peptide region of all matrix protein (Figure-6a).

Phylogenetic Analysis

The Unrooted Phylogenetic tree of EBOV VP40 matrix protein along with other 16 negative sense RNA virus matrix proteins which showed that EBOV VP40 is orthologous to Marburg virus VP40 (Figure6a.). It also revealed that there are three major clusters and same VP40 matrix proteins are not grouped along with matrix protein of Hendra virus, Nipah virus, Measles virus, Newcastle virus, Siendai virus, Simian virus41, Tupaiaparamyxo virus, Human parainfluenza virus type I, Rabies virus, Avian metapneumo virus type C, Canine distemper virus, Bovine respiratory syncytial virus, Rinder pest virus, Mumps virus, and Turkey rhinotracheitis virus.

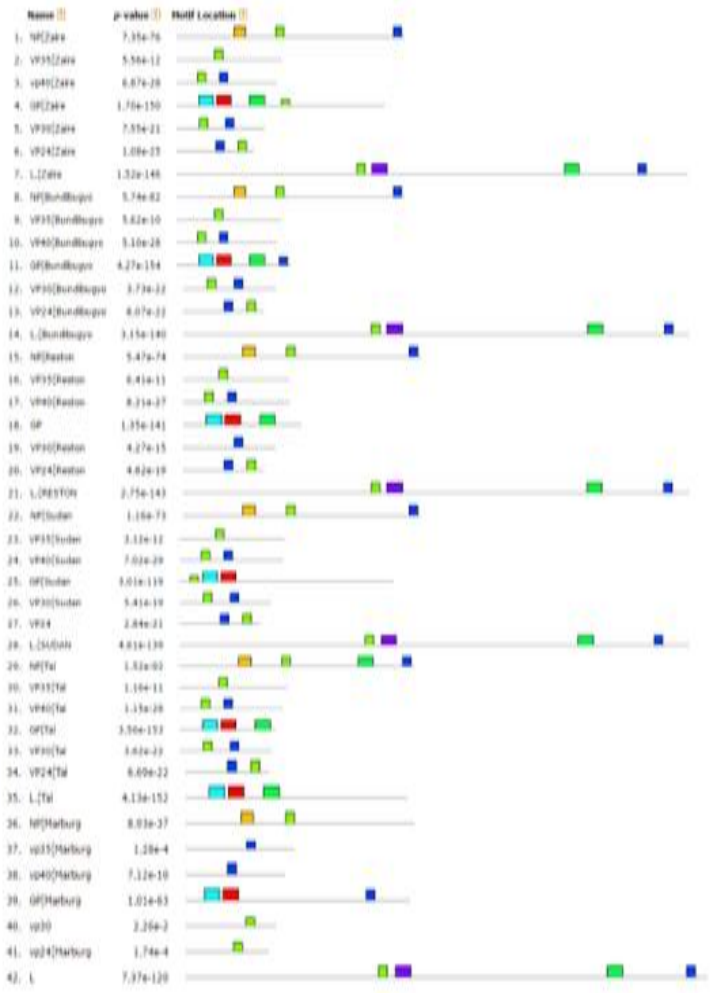
The statistical importance of evolutionary tree was determined by high bootstrap value. For the better understanding of evolutionary significance of EBOV protein, which showed probable homologous proteins with analogous functions from common ancestors, a broad Phylogenetic tree was reconstructed from the 49 EBOV proteins including all structural proteins of five species of EBOV along with 7 Marburg virus proteins (Figure7a.)



(a.)

Motifs	Conserve motifs sequence
Motif1	AAIMLASYTITHFGKTSNPLVRINRLGPGIPDHPLRLLRIGNQAFLQEFV
Motif2	DHPSHTPNSVASAFILEAMVNVISGPKVLMKQIPIWLPLGVSDQKTYSFD
Motif3	LVGRLTGKKITTKNGQPIIPILLPKYIGMDPISQGDLTMVITQDCDTCHS
Motif4	VQLPQYFTFDLTALKLITQPLPAATWTDDTPTGSNGALRPGISFHPKLRP
Motif5	ILLPGRAGKKGSNSDLTSPDKIQAIMNFLQDLKIVPIDPTKNIMGIEVPE
Motif6	APPEYMEAVYPMRTMNSGADNTASGPNYTTTGVMTNDTPSNSLRPVADDN
Motif7	MRRVILPT

(b)



(c).

Figure 7. (a.) Phylogenetic relationship of Ebola virus proteins (including all five species) along with Marburg virus proteins, which codes seven different structural proteins. The sequence of all seven different structural protein were retrieved from NCBI Entrez protein database and their accession number are as follows:- Zaire EBOV [NP (AAG40164), VP35 (AAQ55046), VP40 (AAG40166), GP (AAM76034), VP30 (AHX24670), VP24 (AAM76037), L (AER59721)], Bundibugyo EBOV [NP (AC128620), VP35 (AC128621), VP40 (AC128622), GP (AC128623) ,VP30 (AC128625) ,VP24 (AC128626), L (AC128627), Reston EBOV [NP (BAB69003), VP35 (BAB69004), VP40 (AAN04450), GP (BAB69007), VP30 (BAB69008), VP24 (BAB69009), L (Q8JPX5)], Sudan EBOV [NP (ACR33187), VP35 (ABY75322), VP40 (Q5XX06), GP (Q7T9D9), VP30 (AGL50929), VP24 (AAU43889), L (Q5XX01)], Taiforest EBOV [NP (AC128629), VP35 (AC128630), VP40 (AC128631), GP (YP_003815428), VP30 (AC128634), VP24 (AC128635), L (Q66810)], Marburg EBOV [NP(AAR85460), VP35 (CAA78115), VP40 (CAA78116), GP (ADM72998), VP30 (CAA78118), VP24 (CAA78119), L (ABA87130)] (b.) Best 7 conserve motifs presents in all EBOV along with Marburg virus. (c.) Representation of motifs analysis in Ebola virus proteins (including all five species) along with Marburg virus proteins.

CONCLUSION

In this study we conclude that EBOV VP40 protein is hydrophilic in nature and its attribute is acidic. Presence of phosphorylation and N-Glycosylation sites indicates that it has role in interferon signaling, protein interaction and protein folding. Occurrence of several RNA binding sites is significant, which have role in various processes like cellular function, transport and localization. They may play essential role in post-transcriptional control of RNA such as polyadenylation, mRNA localization and translation. Evidently identification of the existence of active site and antigenic determinant in EBOV VP 40 protein will thus eventually help in effective drug designing in future. A total of 10 different catalytic domains were identified in this proteinwhich decide its function. The Phylogenetic analysis illustrates that VP40 EBOV out of 16 other negative sense RNA virus matrix proteins is orthologous to Marburg virus VP40.

ACKNOWLEDGEMENT

We are grateful for providing Internet facility from Centre for Genetic Disorders at Banaras Hindu University.

REFERENCES

Blom N, Gammeltoft S. and Brunak S. (1999). Sequence and structure based prediction of eukaryotic protein phosphorylation sites. J MolBiol: Vol. 294, 5, 1351–1362.

Bailey T.L, Williams N,Misleh C. and Li W.W. (2006). MEME: Discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Research: Vol. 34, W369–W373.

Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y. and Liang J. (2006). CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res: Vol. 34, 116–8.

Fabozzi G, Nabel C.S, Dolan M.A. and Sullivan N.J. (2011). Ebola Virus Protein Suppress the effects of small interfering RNA by direct interaction with the mammalian RNA interference pathway. Journal of virology: Vol. 85, 6: 2512–2523.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins M.R, Appel R.D. and Bairoch A. (2005). Protein identification and analysis tools on the ExPASy server.The proteomics protocols handbook.Humana Press, Totowa, NJ.

Gupta R, Jung E. and Brunak S. (2004). Prediction of N-glycosylation sites in human proteins. Pac SympBiocomput: Vol. 7, 310–322.

Geourjon C. and Deléage G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. ComputApplBiosci: Vol. 11, 6, 681-684.

Hartman A.L, Towner J.S. and Nichol S.T. (2010).Ebola and Marburg hemorrhagic fever.Clin Lab Med: Vol. 30, 1, 161-177.

Horton P, Park K, Obayashi T. and Nakai K. (2007). WoLF PSORT: protein localization predictor. Nucleic Acids Res: Vol. 35, W585–W587.

Jia F, Wu B, Li H, Huang J. and Zheng C. (2013). Genome-wide identification and characterization of F-box family in maize. Mol Genet Genomics: Vol. 288, 559–577.

Kim O. T. P, Yura K. and Go N. (2006). Amino acid residue doublet propensity in the protein-RNA interface and its application to RNA interface prediction. Nuc. Acids Res: Vol. 34, 22, 6450-6460.

Larkin M.A, Blackshields G, Brown N.P, Chenna R, McGettigan P.A, McWilliam H, Valentin F, Wallace I.M, Wilm A, Lopez R, Thompson J.D, Gibson T.J. and Higgins D.G. (2007). Clustal W and Clustal X version 2.0.Bioinformatics: Vol. 23, 21, 2947-2948.

Leroy E.M, Kumulungui B, Pourrut X, Rouquet P, Hassanin A, Yaba P, Délicat A, Paweska J.T, Gonzalez J.P. and Swanepoel R. (2005). Fruit bats as reservoirs of Ebola virus. Nature: Vol. 438, 575-576.

Nanbo A, Watanabe S, Halfmann P. and Kawaoka Y. (2013). The spatio-temporal distribution dynamics of Ebola virus proteins and RNA in infected cells. Scientific Reports: Vol. 3, 1206.

Noda T, Sagara H, Suzuki E, Takada A, Kida H. and Kawaoka Y. (2002). Ebola Virus VP40 Drives the Formation of Virus-Like Filamentous Particles Along with GP. Journal of virology: Vol. 76, 10, 4855-4865.

Nguyen B.A.N, Pogoutse A, Provart N. and Moses A.M. (2009). NLStrada-mus: a simple Hidden Markov Model for nuclear localization signal prediction. BMC Bioinf: Vol. 10, 202.

Okumura A, Pitha P.M. and Harty R.N. (2008). ISG15 inhibits Ebola VP40 VLP budding in an L-domain-dependent manner by blocking Nedd4 ligase activity. PNAS: Vol. 105, 10, 3974 –3979

Silva L.P, Vanzile M, Bavari S, Aman J. M. J. and Schriemer D.C. (2012). Assembly of Ebola Virus Matrix Protein VP40 Is Regulated by Latch-Like Properties of N and C Terminal Tails. PLoS One: Vol. 7, 7, 39978.

Sigrist C. J, Cerutti L, de Castro E, Langendijk-Genevaux P.S, Bulliard V, Bairoch A. and Hulo N. (2010). PROSITE, a protein domain database for functional characterization and annotation. Nucleic Acids Res: Vol. 38, D161-6.

Tamura K, Peterson D, Stecher G, Nei M. and Kumar S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. MolBiolEvol: Vol. 28, 10, 2731–2739.

Wauquier N, Becquart P, Padilla C, Baize S. and Leroy E.M. (2010). Human Fatal Zaire Ebola Virus Infection Is Associated with an Aberrant Innate Immunity and with Massive Lymphocyte Apoptosis. PLoSNegl Trop Dis: Vol.4, 10, 837.

Wass M. N. and Sternberg M. J. (2009).Prediction of ligand binding sites using homologueous structures and conservation at CASP8. Proteins: Vol 77, 9, 147-151.

Xavier F. Gomis-Rüth, Dessen A, Timmins J, Bracher A, Kolesnikowa L, Becker S, Klenk H.D. and Weissenhorn W. (2003). The Matrix Protein VP40 from Ebola Virus Octamerizes into Pore-like Structures with Specific RNA Binding Properties. Structure: Vol. 11, 4, 423-433.