OPEN ACCESS

International Journal of Applied Sciences and Biotechnology

A Rapid Publishing Journal

ISSN: 2091-2609

Indexing and Abstracting

CrossRef, Google Scholar, Global Impact Factor, Genamics, Index Copernicus, Directory of Open Access Journals, WorldCat, Electronic Journals Library (EZB), Universitätsbibliothek Leipzig, Hamburg University, UTS (University of Technology, Sydney): Library, International Society of Universal Research in Sciences (EyeSource), Journal Seeker, WZB, Socolar, BioRes, Indian Science, Jadoun Science, Jour-Informatics, Journal Directory, JournalTOCs, Academic Journals Database, Journal Quality Evaluation Report, PDOAJ, Science Central, Journal Impact Factor, NewJour, Open Science Directory, Directory of Research Journals Indexing, Open Access Library, International Impact Factor Services, SciSeek, Cabell's Directories, Scientific Indexing Services, CiteFactor, UniSA Library, InfoBase Index, Infomine, Getinfo, Open Academic Journals Index, HINARI, etc.

CODEN (Chemical Abstract Services, USA): IJASKD

Vol-4, Issue-1 (March, 2016)

Available online at:
http://www.ijasbt.org
&
http://www.nepjol.info/index.php/IJASBT/index

SEM-Biotech Publishing

Impact factor*: **1.422**
Scientific Journal Impact factor#: **3.419**
Index Copernicus Value: **6.02**
IBI Factor 2015**: **4.19**

*Impact factor is issued by Universal Impact Factor. Kindly note that this is not the IF of Journal Citation Report (JCR).
#Impact factor is issued by SJIF INNO SPACE; **Impact factor is issued by INFOBASE INDEX.

M.J. Islam et al. (2016) Int J ApplSci Biotechnol, Vol 4(1): 92-103
DOI: 10.3126/ijasbt.v4i1.14557

ISSN (Online): 2091-2609
DOI Prefix: 10.3126/ijasbt

International Journal of Applied Sciences and Biotechnology (IJASBT)

Research Article

IN SILICO APPROACH OF STRUCTURE PREDICTION AND FUNCTIONAL CHARACTERIZATION OF ZAIRE EBOLA (EBOV) AND IDENTIFICATION OF BINDING SITE FOR DRUG DEVELOPMENT

Md. Jahirul Islam*, Kaniz Fatema and Pipasha Biswas

Department of Biochemistry and Biotechnology, University of Science and Technology Chittagong (USTC) Chittagong-4202, Foy's Lake, Bangladesh

Corresponding author's email: jahirjoel87@gmail.com

Abstract

Zaire ebolavirus (EBOV) is one of the dangerous and a negative-stranded ssRNA virus. EBOV is a zoonotic pathogen that causes severe hemorrhagic fever in humans. Nowadays epidemic outbreak caused by EBOV is incurable with present technologies; thus figure out as a major health risk which needs enhanced surveillance. The study was conducted with seven proteins of Zaire ebola (EBOV) and gene sequences are available in NCBI database. The homology modeling was done by SWISS-MODEL, Phyre2 and HHpred. The obtained model was verified with structure validation programs such as PROCHECK, Verify3D and ERRAT. PROCHECK analysis of seven proteins showed that 85-96.6% of the residues are in the most favored region, the verify 3D value of 80-100% indicates that constructed model is good and ERRAT value of 87.442-100% indicates that overall good quality factor. In this study, we also reported phylogenic relationship, physico-chemical characteristics, secondary structure,3-D structure. Moreover, active sites were identified by CASTp suggests that these proteins can be utilized as a potential drug target. Furthermore, the initial findings were reinforced by the results from I-Mutant and mCSM as these tools predicted significant and functional instability of the mutated vp35 protein.

Keywords: Zaire ebola, EBOV, Computational Tools, Active site, Mutation Point.

Introduction

Ebola virus is a negative-sense, single-stranded RNA virus that causes a severe hemorrhagic fever in humans with high case fatality rates ranging from 47 to 91% (Fedmann *et.al*.,2011). The genus *Ebolavirus* is inside in the family *Filoviridae* and order *Mononegavirales*. Viruses within this genus called ebolaviruses (Kuhn JH *et.al*.,2010) and consists of five species- Bundibugyo ebolavirus (BDBV), Zaire ebolavirus (EBOV), Reston ebolavirus (RESTV), Sudan ebolavirus (SUDV) and Tai forest ebolavirus (TAFV) (WHO, 2014). Disease from *ebolavirus* is marked by fever, shock, and coagulation defects with 50–90% mortality occurring 7–12 days after infection (Bwaka *et. al.*1999).This species was first identified during an outreak on August, 1976 in Yambuku (Suzuki *et al.*, 1997).

From the genetic studies, Zaire ebolavirus (EBOV) is linear and is about 18,959nt as well as consist of seven genes(Bukreyev *et al.*1993). These genes are arranged as **3'-NP-VP35-VP40-GP-VP30-VP24-L** (Volchkov *et al*., 1999).Various proteins have various functions in the life cycle of Ebola. GP is a type-I transmembrane protein cleaved by furin proteases in GP1 and GP2 subunits (Neumann *et al*., 2007; Neumann *et a*l., 2002; Wool-Lewis *et al.*1999; Volchkov *et al.*,1998). VP24 is a peripheral viral membrane protein in viral binding that also plays an important role in the suppression of host interferon activity (Ziying Han *et al*.., 2003; Basler *et al*., 2009). The VP30 is essential maintaining the balance between transcription and replication process in ebolavirus replication cycle (Martinez *et al*., 2008). NP is a nucleoprotein that forms a ribonucleoprotein complex when binds to the viral RNA. The L protein (Large structural protein) helps in the synthesis of mRNA from negative-stranded ssRNA and hence is an RNA-dependent RNA polymerase (Watanabe *et al*., 2006).VP35 has been shown to prevent phosphorylation and dimerization of IRF-3 (Pires *et al.*, 2014), to block induction of IFN α/β expression (Basler CF *et al.*, 2003; Basler CF *et al.*, 2000), to inhibit activation of protein kinase R (PKR) (Feng *et a*l., 2007, Schumann *et al*., 2009), and to serve as a suppressor of RNA silencing (Haasnoot J, *et al*., 2007). VP35 possesses double-stranded RNA (dsRNA) activity. Two VP35 mutant points K309A and R312A were found to be greatly impaired in their dsRNA-binding activity (Cardenas *et al*., 2006). There are three basic residues, R305, K309, and R312, as numbered in the

Zaire species of ebolavirus, are critical for binding dsRNA and blocking IFN expression (Christopher *et al*., 2010).

Materials and Methods

Data and Materials

The protein sequences can be retrieved from NCBI (http://www.ncbi.nlm.gov/) 'protein' database. The search listed seven proteins as – Nuclueoprotein (GI:788304299), L (Large structural Protein) which is RNA depended RNA polymerase (GI:788304307), Glycoprotein (GI:788304302), VP35 (GI:788304300), VP40 (GI:788304301), VP30 (GI:788304305), VP24 (GI:788304306). The FASTA sequence was retrieved and used for the further bioinformatics analysis.

Multiple sequence alignment

These sequences were analyzed on ClustalW (http://www.ebi.ac.uk/clustalW/) for the multiple sequences alignment.

Construction of phylogenetic trees

The phylogenetic trees were first constructed using the neighbor joining method (Saitou N *et al*., 1987) from the MEGA6 package (Tamura *et al*., 2011). Sequences were also analyzed using MEGA6 and a ClustalW algorithm was used to align multiple sequences in parallel. Confidence on each node was assessed by 2000 bootstrap replications (Felsenstein *et al*., 1985). Also the maximum likelihood method from MEGA6 package was used to construct a phylogenetic tree and 2000 replicates were used for bootstrap statistical test.

Physico-chemical characterization

Different properties including number of amino acids, molecular wight, theoretical isoelectric point(pI), amino acid composition (%), number of positively (Arg+ Lys) and negatively charged (Asp+ Glu) residues, extinction co-efficient, instability index, aliphatic index and Gran Average of Hydropathicity (GRAVY) were calculated using ExPASy's PortParam tool (Gasteiger *et al*., 2005) (http://expasy.org/tools/protparam.html).

Homology study

Homology study was performed by using NCBI protein blast package contain blast-p, psi-blast, delta-blast algorithms, BLOSUM 62 matrix, Existence 11 Extension-1, with non-redundant protein sequence (nr).

Protein 3D structure prediction

Protein structure homology modeling has become a routine technique to generate 3D models for proteins when experimental structures are not available (Biasini *et al*., 2014). The 3D structures of seven proteins of Zaire ebola virus were predicted by different tools, using SWISS-MODEL (Bordoli *et al*., 2009) (http://www.swissmodel.expasy.org/), Phyre2 (Kelly LA et al., 2009) (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index), HHpred (Agarwal *et al*., 2008) (http://toolkit.tuebingen.mpg.de/hhpred) . The input data was in FASTA format.

3D Model Validation

The program PROCHEK (Laskowaski *et al.,* 1996) by Ramachandran plot analysis is used to define the stereo chemical quality of the generated model and it was also validated by ERRAT (Colovos *et al.,*1993) and Verify 3D programs (Eisenberg *et.al*.,1997). Finally, proteins were visualized by using Chimera 1.8.1 (Pettersen *et al*.,2004).

Active site prediction

To identify the ligand binding capacity with the determinded model, CASTp server (Dundas *et al.*2006) was used. The predicted active site in generated model will help in further work to study on docking site.

Protein-Protein Interaction networking

Protein interacts with other proteins to execute accurate functions. To identify virus-host protein-protein interactions. by VirHostNet 2.0 (Guirimand *et al*., 2015) (http://virhostnet.prabi.fr/#) . VirHostNet 2.0 is based on Cytoscape web library and provides most complete and accurate resource of virus-virus and virus-host protein-protein interactions networks.

Prediction of Change in Stability upon Mutation

I-Mutant 2.0 (Capriotti *et al*., 2005) and mCSM (Pires *et al*., 2014) were used to predict the change in stability due to mutation. This tool can automatically predict the change in structural stability analyzing the structure or the sequence of the protein. I-Mutant 2.0 and mCSM can used as classifier for predicting the sign of protein stability upon mutation and a regression estimator which predicts the change in Gibbs free energy. The resulting DDG value is the difference between the Gibbs free energy of mutated protein and wild type protein in kcal/mol.

Results

Pair-wise Distance

The pairwise distance method of phylogenetic analysis revealed on a measure of genetic distance between the sequences being classified. This exploration showed the divergence and percent identity of each sequence pair in the current alignment. Sequence comparison between the seven proteins of EBOV (Fig-1). The overall average is 9.7568 and there were a total of 225 positions in the final dataset.

Evolutionary Relationship

Phylogenetic tree shows evolutionary relationship among the seven proteins. The phylogenic tree (Fig-2) classified the proteins into five groups. Group-1 contains of two proteins (Nucleoprotein and Glycoprotein), Group-2 contains one protein (VP24), Group-3 contains of two proteins (VP30 and Large structure protein), Group-4 and Group-5 contains one protein (VP35 and VP40).

	1	2	3	4	5	6	7
1. Nucleoprotein(NP)		5.1332	3.0578	2.6721	4.1542	1.9252	2.4792
2. VP35	15.0714		1.4469	1.7978	1.5296	3.6230	2.4086
3. VP40(matrix_protein)	10.8421	6.7586		2.0669	2.2390	2.7060	1.6107
4. VP24	9.7143	8.0000	8.3750		5.5904	1.6054	1.7130
5. VP30	13.0625	7.0357	9.2273	16.3077		3.7245	0.5160
6. GP(glycoprotein_precursor)	8.3750	11.5000	10.2500	7.3333	12.2353		3.8989
7. L(polymerase)	9.7143	9.7143	7.3333	7.6538	3.3269	13.0625	

Fig-1: Estimates of Evolutionary Divergence between sequences

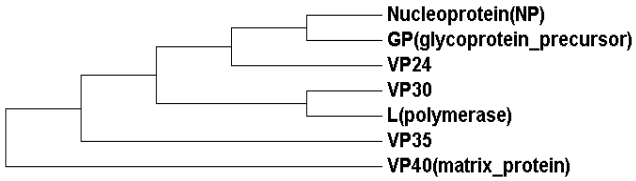

Fig-2: Phylogenetic relationship among the seven proteins of Zaire ebola

Physio-chemical Characterization

Isoelectric point (pI) is a pH in which net charge of protein is zero. pI of Nucleoprotein(NP), VP35 and Glycoprotein(GP) were observed to lie in the acidic range, while the rest of the proteins occur in alkaline range. From the study of instability index, it was found that Nucleoprotein (NP), VP35, VP30, Large Structural protein (L) were unstable, while the rest of the proteins were in stable. As instability index value less than 40 indicates stability of a protein (Table-2). Additionally, Aliphatic index (AI) refers to the relative volume of a protein occupied by its aliphatic side chains. The higher the Aliphatic index of proteins, the more thermally stable the proteins. Aliphatic index of VP24 (105.94), VP40 (95.43) and Large structural protein(90.10) classifies them as most thermostable, closely followed by VP35, VP30, Glycoprotein, Nucleoprotein. Grand average of Hydropathicity index (GRAVY) indicates the interaction of the proteins in water. GRAVY values of all the proteins were within a wide range of -0.036 to -0.687 (hydrophilic).

The amino acid composition of each protein sequence was calculated by using ExPASY's ProtParam tool (supplementary file, Table-3). High percentage of Leusine (above 10.1), Serine (above 6.7) were found in VP24,VP30 and VP40 compared to other amino acid. In all Zaire ebola protein (except Large structural protein) alanine content was found significant. Again VP35, VP30 and Large structural protein (L) have a arginine content above 3.4 . Moreover, a good percentage of Thrionine content was found in Glycoprotein (10.5%), VP40 (8.9%) and VP35 (7.9%).

Homology study

The homology model study (Table-4) shows Nucleoprotein (NP), VP40, VP35, VP24, Glyco-protein query covered 93-100% with Bundibugyo virus, Tai Forest virus, Reston virus. On the other hand, VP30 query covered 90-100% with Bundibugyo virus, Tai Forest virus, Lloviu cuevavirus. Moreover, Large structural protein (L) query covered 60-99% with Lloviu cuevavirus, Pocine para influenza virus, Human para influenza.

Model verification and validation

PROCHECK sever had been used for building Ramachandran plot that measure the accuracy of protein model. The results of PROCHECK (Ramachandran plot: % core) are depicted in Fig-3 and Table-1 and then verified by using Verify3D (% of the residues had an averaged 3D-1D score>0.2), ERRAT (Overall quality factor) were narrated in Table-1.

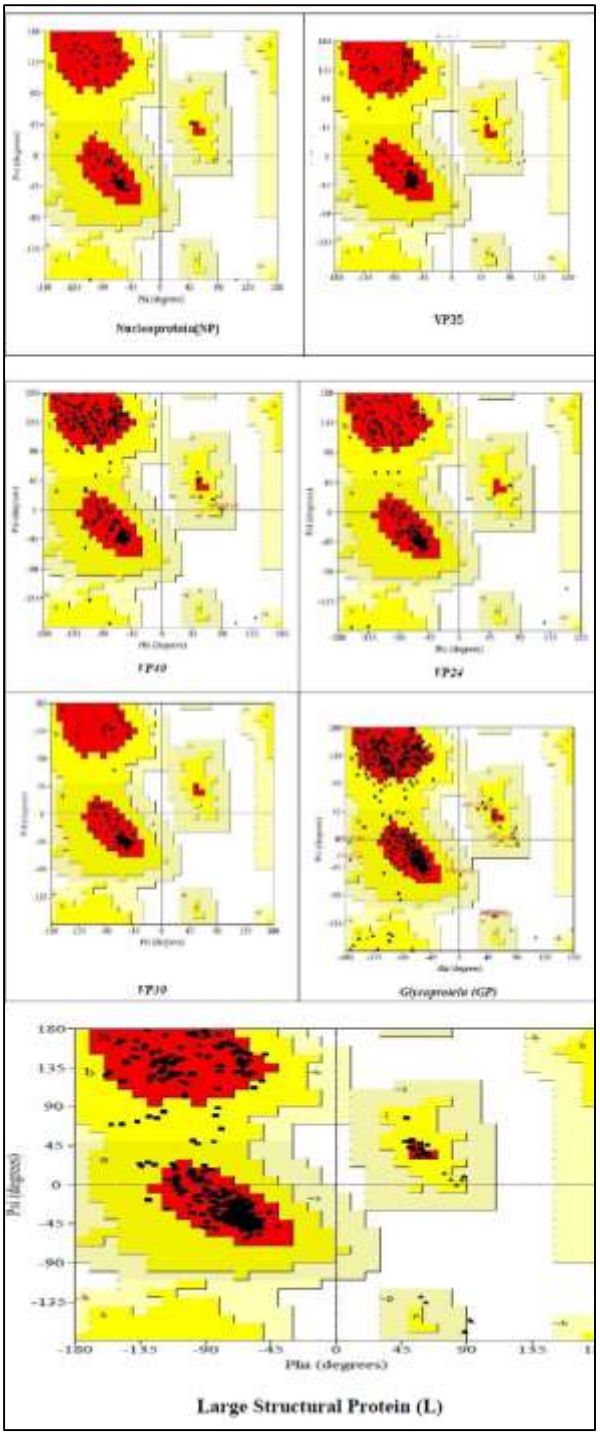

Fig. 3: Ramachandran plot of seven proteins of Zaire ebola (EBOV) obtained through the modeling tool.

Table 1: Model validation result of seven genes of Zaire ebola by different tools.

Gene Name	PROCHECK	Verify3D	ERRAT
NucleoProtein (NP)	95.2	100	100
VP35	95.2	100	98.291
VP40	90.7	92.47	87.442
VP24	93.1	98.99	89.529
VP30	96.6	87.30	100
Glycoprotein (GP)	85	93.95	95.093
Large structural protein (L)	95.1	80	97.808

Table 2: Physico-chemical parameters of seven proteins of Zaire ebola.

Gene Name	No. of Amino Acid	Molecular Weight	Theoretical pI	'-' charged residues (Asp+Glu)	'+' charged residues (Arg+Lys)	Extinction Coefficients	Instability Index	Aliphatic Index	(GRAVY)
NucleoProtein (NP)	739	83200.5	4.93	118	70	53540	48.97 unstable	73.52	-0.687
VP35	340	37448.5	6.01	38	34	25940	48.62 unstable	78.91	-0.402
VP40	326	35140.7	8.76	26	29	20065	39.04 stable	95.43	-0.060
VP24	251	28215.8	9.57	19	26	31970	36.51 stable	105.94	-0.036
VP30	288	32520.8	8.40	37	40	28460	52.08 unstable	78.36	-0.607
Glycoprotein (GP)	676	74434.5	6.30	70	64	101590	37.45 stable	76.35	-0.378
Large structural Protein (L)	2212	252495.1	8.64	209	231	287285	40.94 unstable	90.10	-0.235

Table 3: Composition of amino acid of seven proteins of Zaire ebola (%):

Protein	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Nucleo-Protein (NP)	7.0	4.6	4.3	8.1	0.5	7.2	7.8	5.7	3.9	3.7	9.1	4.9	2.7	3.5	5.7	6.4	5.7	0.5	2.8	5.8
VP35	7.6	5.3	4.1	5.3	2.4	6.8	5.9	5.9	2.1	6.8	7.4	4.7	2.6	2.6	7.1	4.7	7.9	0.9	1.8	5.6
VP40	7.1	3.4	4.3	5.2	0.6	3.4	2.8	6.4	2.1	8.0	10.1	5.5	2.5	3.1	11.3	6.7	8.9	0.6	1.8	6.1
VP24	6.4	4.0	6.8	3.6	0.4	4.8	4.0	5.2	2.8	7.6	14.7	6.4	3.2	4.4	3.6	8.4	6.8	2.0	1.2	4.0
VP30	7.3	8.7	2.1	5.6	2.8	6.2	7.3	4.5	3.1	3.5	11.1	5.2	1.0	2.8	5.2	9.4	6.6	1.4	1.4	4.9
Glyco-protein (GP)	7.2	4.9	5.6	5.2	1.8	4.0	5.2	7.8	2.5	6.1	7.7	4.6	0.6	4.3	5.3	7.0	10.5	2.1	2.4	5.3
Large structural protein (L)	5.3	5.3	5.0	4.7	1.9	4.8	4.8	4.6	3.6	6.6	11.3	5.1	1.6	5.2	4.6	8.3	7.1	1.3	3.8	5.1

Table 4: Homology parameters for the proteins studies

Protein	Accession Number	Homology	Query cover	Identity	E value
Nucleoprotein(NP)	GI:788304299	Bundibugyo virus	100%	75%	0.0
		Tai Forest virus	100%	75%	0.0
		Reston virus	100%	68%	0.0
VP35	GI:788304300	Tai Forest virus	93%	80%	4e-178
		Bundibugyo virus	96%	78%	9e-172
		Reston virus	93%	69%	1e-158
		Sudan virus	94%	69%	3e-153
VP40	GI:788304298	Bundibugyo virus	100%	83%	0.0
		Tai Forest virus	100%	82%	0.0
		Sudan virus	100%	75%	3e-174
		Reston virus	100%	74%	2e-172
		Lloviu cuevavirus	84%	53%	1e-95
VP24	GI:788304306	Tai Forest virus	100%	88%	2e-161
		Bundibugyo virus	100%	86%	4e-160
		Reston virus	100%	82%	3e-153
VP30	GI:788304305	Bundibugyo virus	100%	80%	5e-162
		Tai Forest virus	99%	78%	3e-158
		Lloviu cuevavirus	90%	53%	1e-78
Glycoprotein (GP)	GI:788304302	Bundibugyo virus	100%	65%	0.0
		Tai Forest virus	100%	65%	0.0
		Reston virus	95%	58%	0.0
Large structural protein (L)	GI:788304307	Lloviu cuevavirus	99%	55%	0.0
		Porcine parainflunenza virus	60%	26%	2e-116
		Human parainflunenza	60%	26%	8e-115

Table. 5: Active site prediction of selected proteins of Zaire ebola.

Protein	Area(Å^2)	Volume(Å^3)
Nucleoprotein(NP)	82.5	64
VP35	110	92.5
VP40	709.4	1365.6
VP24	163.1	254
VP30	362.7	1169.9
Glycoprotein(GP)	642.5	796.5
Large structural protein(L)	1351.2	2216.7

Protein 3D structure

The 3D structure of protein is very crucial for comprehending the protein functions, their sub-cellular localization as well as protein-protein interactions. Based on homology modeling, SWISS-MODEL, Phyre2 and HHpred results showed 3D models for each of the given sequences and ranked them according to the scores of Ramachandran plot and validation program Verify3D, ERRAT and were visualized by chimera (Fig. 4).

Active site Analysis

The active sites of seven proteins were predicted (Fig-5). Further, in this study, we have also reported the best active site area of the experimental enzymes as well as the number of amino acids involved in it; showed the number of pockets, with their area and volume (Table-5).Among all

the proteins, the highest volume of the pockets of Large structural protein(L) is 2216.7Å3 and area is 1351.2Å2. There are some variations in the area of pockets for VP40, it is 709.4Å2 and volume 1365.6 Å3. In case of Glycoprotein (GP), area and volume of the pockets is 642.5 Å2 and 796.5 Å3 respectively. For VP30, VP24, VP35 and Nucleoprotein(NP); the area of the pockets are 362.7 Å2 , 163.1 Å2 , 110 Å2, 82.5 Å2 and the volume of the pockets are 1169.9 Å3 , 254 Å3 , 92.5 Å3, 64 Å3 respectively

Network generation

The protein-protein interacting was occurred VP35,VP40,VP24,NP, L proteins of EBOV with 154 proteins of Human by VirHostNet 2.0 tools and interaction network was depicted in Fig-6.

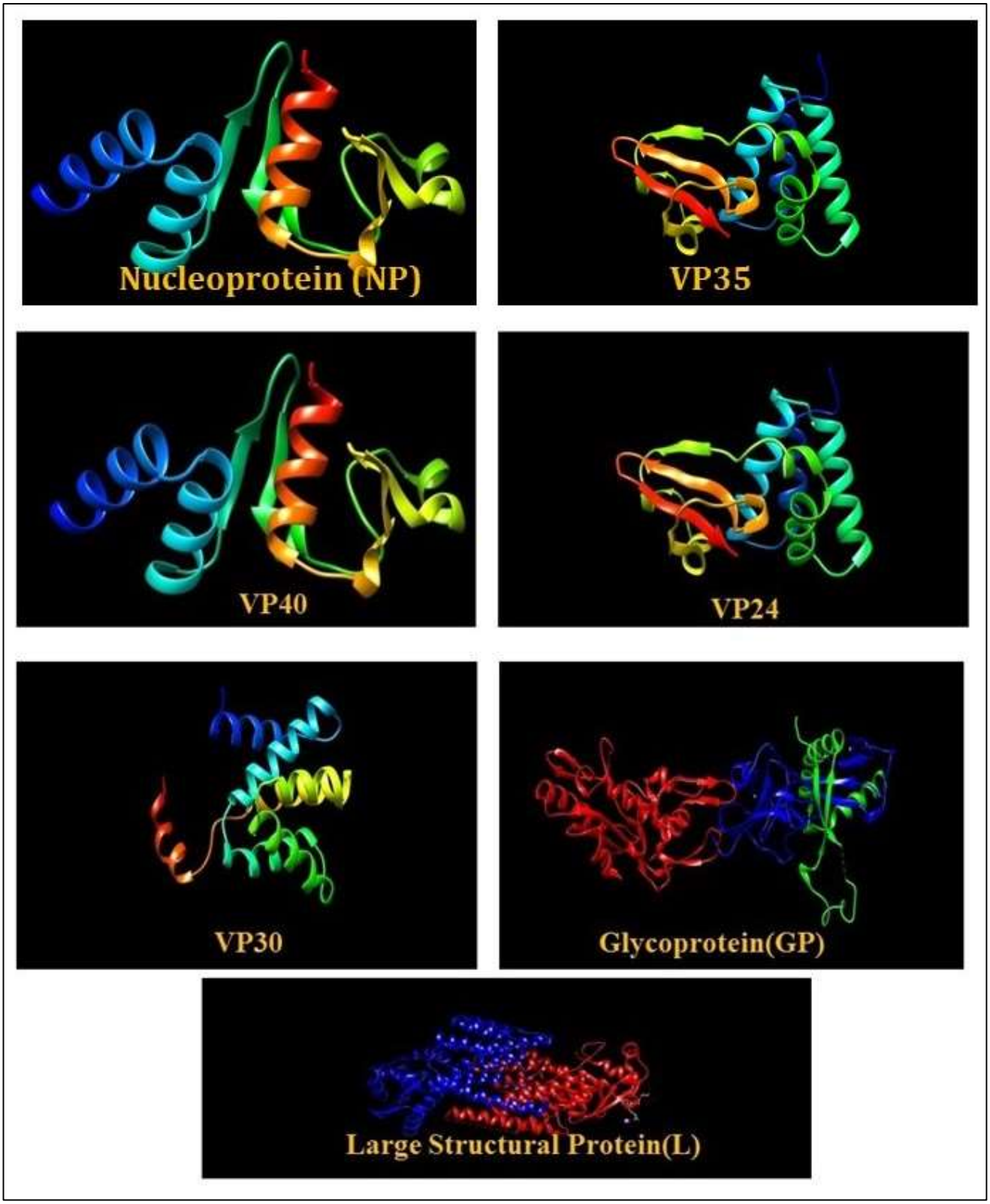

Fig-4: 3D structure of seven proteins of Zaire ebola**.**

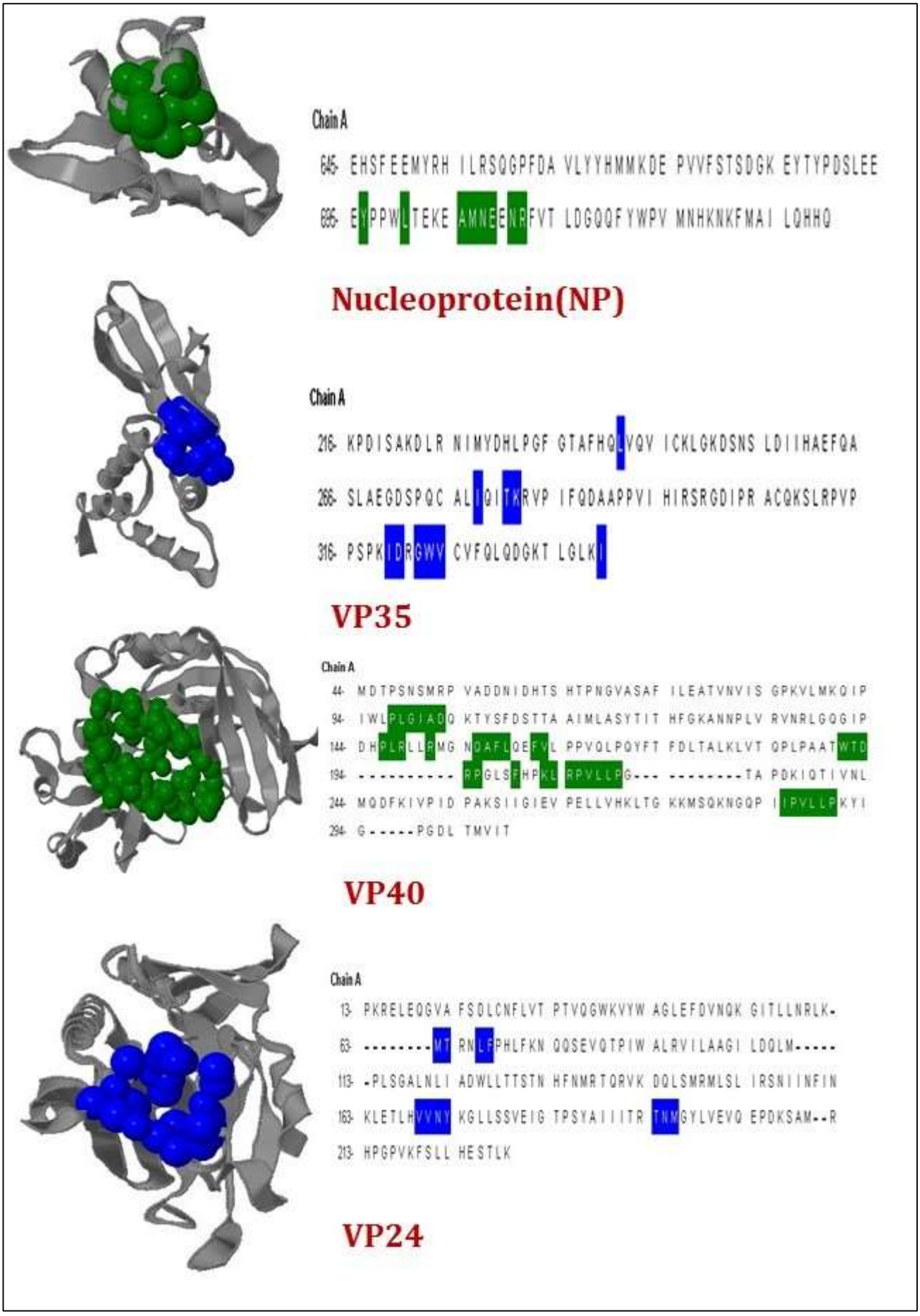

Fig. 5: Active site prediction of selected proteins of Zaire ebola

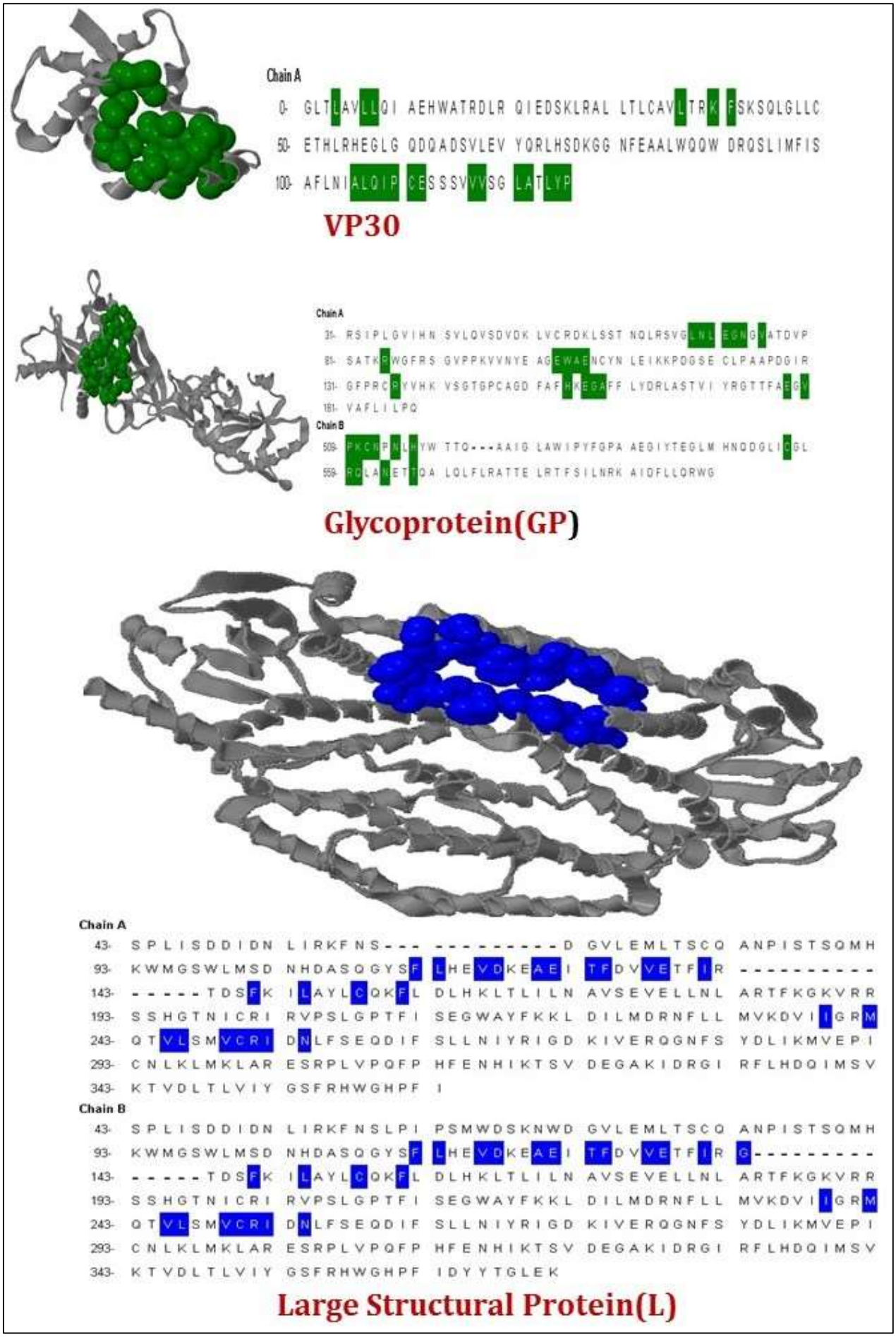

Fig. 5(contd.): Active site prediction of selected proteins of Zaire ebola

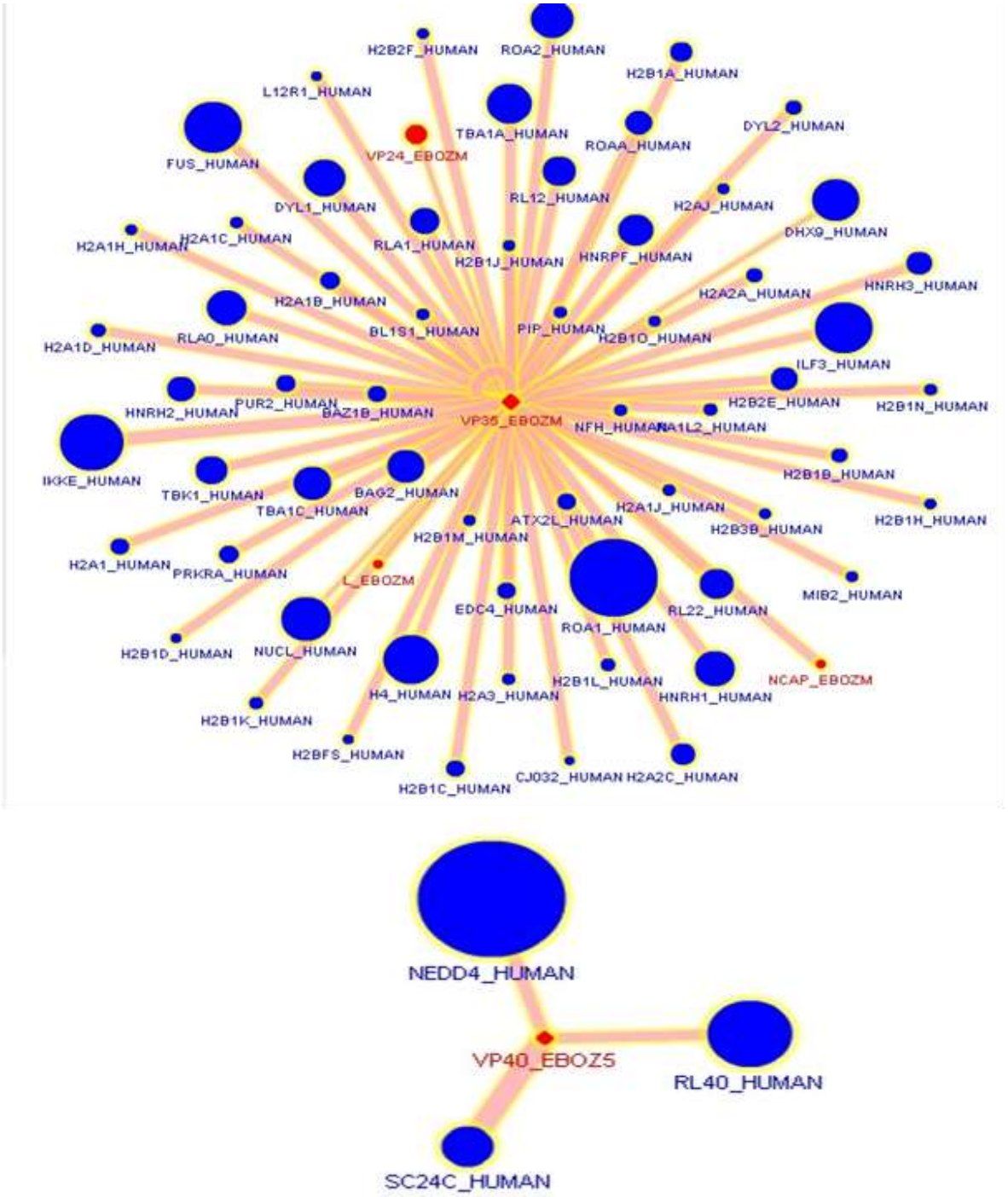

Fig. 6: Protein-protein interaction of Zaire Ebola virus (EBOV).

Prediction of Protein structure Stability

To assess the applicability of I-Mutant 2.0 and mCSM signatures in predicting the structure and the impact of mutations in protein stability. mCSM tool took pdb format file of wild type protein resulting RSA(%) and visualized mutant structure. 80% or 70% accurate prediction can be achieved by using protein structure or sequence, respectively, by these tools (Fig-7). Models with following mutations R305A, K309A and R312A of VP35 were submitted to the server for DDG stability prediction and RSA calculation. All the mutations decrease protein stability. The effects of the amino acid substitutions on the domain structure of protein were received in detail form I-Mutant 2.0 and mCSM server. The mutation R305A results (Table-6) in a alanine residue in place of Arginine at 305 position located in the VP35. The highest reliability index score 8 in R305A and this score followed by R312A and K309A. Mutation R305A accounted for the lowest DDG value (-0.56Kcal/mol) followed by R312A (-0.38kcal/mol) respectively.

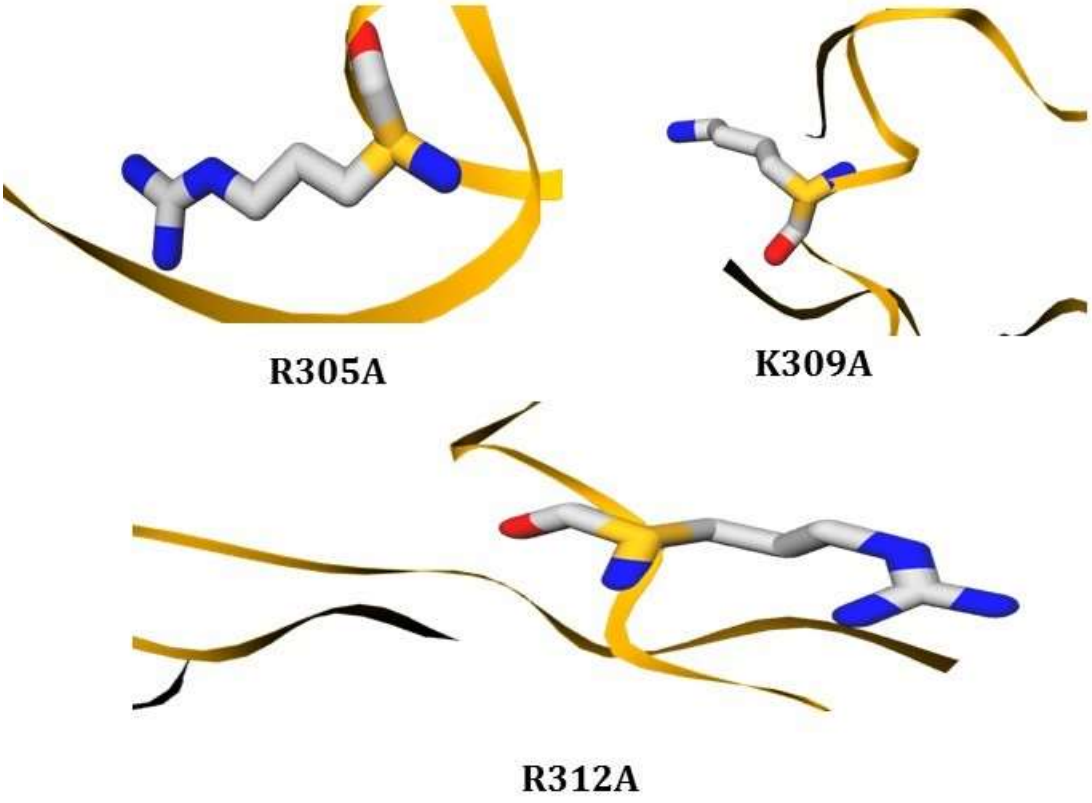

Fig. 7: Close-up of the mutation R305A, K309A and R312A of VP35 generated by mCSM.

Table. 6: I-mutant 2.0 and mCSM predictions for selected VP35 protein**.**

Mutation	Sign of DDG	DDG value prediction Kcal/mol	RI	RSA (%)
R305A	Decrease	-0.56	8	101.4
K309A	Decrease	-0.11	4	89.2
R312A	Decrease	-0.38	7	52.3

Discussion

In recent period, severe outbreak of Ebola virus covered not only in a naive region but also transmitted through the world (Feldmann *et al*.,2011).The aim of this study was to develop homology model with active site prediction and mutagenesis experiment using computational based biological approach (Pipasha *et al*.,2014). The future work is docking the active site for drug developing and Designing a potential siRNA against VP35. This study was executed with seven proteins of Zaire ebola (EBOV) and gene sequences are available in the viral gene bank database from NCBI. The pairwise distance indicates that the maximum genetic distance of 15.0714 occurred between Nucloprotein (NP) and VP35. Multiple sequences alignment and Phylogenic tree constructed using Clustal W and MEGA6 algorithm. The physical and chemical characterization of selecting protein done by Expasy's protparm Tool, where Vp35, Nucleoprotein and Glyocoprotein showed less than 7 isoelectric point. Whereas, Instability index value showed only Glycoprotein as a stable protein. Aliphatic index of VP24 (105.94), VP40 (95.43) and Large structural protein (90.10) classifies them as most thermostable, closely followed by VP35, VP30, Glycoprotein, Nucleoprotein. Grand average of Hydropathicity index (GRAVY) values of all the proteins were within a wide range of -0.036 to -0.687 that proved protein interaction with water. Alanine content was showed positive significant for collected Zaire ebola virus protein except Large structural protein.

VP35, VP30 and Large structural protein(L) had arginine content above 3.4,whereas good percentage of thrionine content was found in Glycoprotein (10.5%), VP40 (8.9%) and VP35 (7.9%).The homology model study (Table-4) shows Nucleoprotein (NP), VP40, VP35, VP24, Glycoprotein query covered 93-100% ,VP30 query covered 90-100% and Large structural protein (L) query covered 60-

99% with Lloviu cuevavirus, Pocine para influenza virus, Human para influenza.

Ramachandran plotting done by PROCHECK sever that measure the accuracy of protein model. The results of PROCHECK (Ramachandran plot: % core) are depicted in Fig-3 and Table-1 and then verified by using Verify-3D (% of the residues had an averaged 3D-1D score>0.2), ERRAT (Overall quality factor) and visualized by chimera (Fig-4).

Homology modeling prediction occurred by SWISS-MODEL, Phyre2 and HHpred results showed 3D models for each of the given sequences.

The active sites of seven proteins were analysed (Fig-5) to identify the best active site area with experimental enzymes, the number of amino acids, the number of pockets, with their area and volume (Table-5).

Among all the proteins, the highest volume of the pockets of Large structural protein(L) is 2216.7Å3 and area is 1351.2Å2. There are some variations in the area of pockets for VP40, it is 709.4Å2 and volume 1365.6 Å3. In case of Glycoprotein(GP), area and volume of the pockets is 642.5 Å2 and 796.5 Å3 respectively. For VP30, VP24, VP35 and Nucleoprotein(NP); the area of the pockets are 362.7 Å2 , 163.1 Å2 , 110 Å2, 82.5 Å2 and the volume of the pockets are 1169.9 Å3 , 254 Å3 , 92.5 Å3, 64 Å3 respectively. VirHostNet 2.0 tools and interaction network analyzed protein-protein interaction VP35,VP40,VP24,NP, L protein of EBOV with 154 proteins of Human (Fig-6).

To assess the applicability of I-Mutant 2.0 and mCSM signatures ,mCSM tool received pdb format file of wild type protein resulting RSA(%) and visualized mutant structure with 80% or 70% accurate prediction, can be achieved by using protein structure or sequence, respectively, by these tools (Fig-7) to predict the 3D structure and protein stability due to mutation. Models with following mutations R305A, K309A and R312A of VP35 were submitted to the server for DDG stability prediction and RSA calculation which showed all the mutations decrease protein stability.

Finally, after mutation amino acid substitutions on the domain structure of protein were received in detail form I-Mutant 2.0 and mCSM server where the mutation R305A results (Table-6) in Alanine residue in place of Arginine at 305 position located in the VP35. The highest reliability index score 8 in R305A and this score followed by R312A and K309A. Mutation R305A accounted for the lowest DDG value (-0.56Kcal/mol) followed by R312A (-0.38kcal/mol) respectively.

Conclusion

In this study, the 3D structure of seven proteins of Zaire ebola(EBOV) were predicted and validated by various bioinformatics tools and software. Analysis of evolutionary relationship reveals that all of the proteins shared a common ancestor. These findings help us to realize the characters of these proteins. It is obvious that VP35 is highly unstable and contain high percentage of Alanine. On the other hand, VP30, VP24, VP40 are predominated by Serine, Leucine, glycine, proline. These amino acids have important role in protein structure**.** In future, broad screening inhibitor against VP35 will help for effective drug designing.

Conflict of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

Agarwal V, Remmert M, Biegert A, Soding J (2008) PDB alert: automatic, recurrent remote homology tracking and protein structure prediction. *BMC Structural Biology* **8**:51. DOI: 10.1186/1472-6807-8-51

Basler CF and Amarasinghe GK (2009) Evasion of interferon responses by Ebola and Marburg viruses. *Journal of Interferon & Cytokine Research* **29**(9): 511-520. DOI: 10.1089/jir.2009.0076

Basler CF et al. (2000) The Ebola virus VP35 protein functions as a type I IFN antagonist. *Proc. Natl. Acad. Sci. USA* **97**: 12289–12294. DOI: 10.1073/pnas.220398297

Basler CF et al. (2003) The Ebola virus VP35 protein inhibits activation of interferon regulatory factor 3. *J. Virol*. **77**: 7945–7956. DOI: 10.1128/JVI.77.14.7945-7956.2003

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Gallo T, Bertoni C.M, Bordoli L and Schwede T (2014) SWISS-MODEL: modeling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Research 42: 252-258. DOI: 10.1093/nar/gku340

Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J and Schwede T (2009) Protein structure homology modeling using SWISS-MODEL Workspace. *Nature Protocols* **4**: 1. DOI: 10.1038/nprot.2008.197

Bukreyev A et al. (1993) The VP35 and VP40 proteins of filoviruses. Homology between Marburg and Ebola viruses; 322: 41. DOI: 10.1016/0014-5793(93)81107-b

Bukreyev AA, Volchkov VE, Blinov VM and Netesov SV (1993) The VP35 and VP40 proteins of filoviruses. *FEBS letters* **322**(1): 41-46.

Bwaka MA et al. (1999) Ebola hemorrhagic fever in Kikwit, Democratic Republic of the Congo: Clinical observations in 103 patients. *J. Infect. Dis*. **179**(Suppl 1): S1–S7. DOI: 10.1086/514308

Capriotti E, Fariselli P and Casadio R (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Research* 33(2): W306–W310. DOI: 10.1093/nar/gki375

Cárdenas WB et al. (2006) Ebola virus VP35 protein binds double-stranded RNA and inhibits alpha/beta interferon production induced by RIG-I signaling. *J. Virol*. **80**: 5168–5178. DOI: 10.1128/JVI.02199-05

Christopher R. Kimberlin, Zachary A. Bornholdt, Sheng Li, Virgil L. Woods Jr, Ian J. MacRae, Erica ollmann sapphire (2010) Ebolavirus VP35 uses a bimodal strategy to bind dsRNA for innate immune suppression.. *PNAS* **107**(1): 314-319. DOI: 10.1073/pnas.0910547107

Colovos C and Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sciences.* **2**: 1511–1519. DOI: 10.1002/pro.5560020916

Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpoz Y and Liang J (2006) CASTp: Computed atlas of surface topography of proteins with structure and topographical mapping of functionally annotated residues. *Nucleic Acids Res* **34**(web server issue):w116-118. DOI: 10.1093/nar/gkl282

Eisenberg D, Lüthy R and Bowie JU (1997) VERIFY3D, assessment of protein models with three-dimensional profiles. Methods Enzymol. 277: 396–404. DOI: 10.1016/S0076-6879(97)77022-8

Feldmann, H and Geisbert TW (2011 Ebola haemorrhagic fever. Lancet 377: 849–862. DOI: 10.1016/S0140-6736(10)60667-8

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783-791. DOI: 10.2307/2408678

Feng Z, Cerveny M, Yan Z, He B (2007) The VP35 protein of Ebola virus inhibits the antiviral effect mediated by double-stranded RNA-dependent protein kinase PKR. J. Virol. 81:182–192. DOI: 10.1128/JVI.01006-06

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD and Bairoch A (2005) Protein Identification and Analysis Tools on the ExPASy Server: The Proteomics Protocols Handbook. Humana Press: 571-607. DOI: 10.1385/1-59259-890-0:571

Guirimand T, Delmotte S and Navratil V (2015) VirHostNet 2.0: surfing on the web of virus/host molecular interactions data. Nucleic acids research 43(D1): D583-D587. DOI: 10.1093/nar/gku1121

Haasnoot J et al. (2007) The Ebola virus VP35 protein is a suppressor of RNA silencing. PLoS Pathog 3: e86. DOI: 10.1371/journal.ppat.0030086

10.1093/oxfordjournals.molbev.a025820

Kelley LA and Sternberg MJ (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. 4: 363–71. DOI: 10.1038/nprot.2009.2

Kuhn JH, Becker S, Ebihara H, Geisbert TW, Johnson KM, Kawaoka Y, Lipkin WI, Negredo AI, Netesov SV, Nichol ST, Palacious G, Peters CJ, Tenorio A, Volchkov VE, and Jahrling PB (2010). Proposal for a revised taxonomy of the family Filoviridae: Classification, names of taxa and viruses, and virus abbreviations. Archives of virology 155(12): 2083-2103. DOI: 10.1007/s00705-010-0814-x

Laskowaski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA (1999) PROCHEK-NMR programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR 8: 477-86 .

Martinez MJ, Biedenkopf N, Volchkova V, Hartlied B, Alazard-Dany N, Reynard O, Becker S and Volchkov V (2008) Role of Ebola virus VP30 in transcription reinitiation. J. Virol. 82: 12569-12573. DOI: 10.1128/JVI.01395-08

Neumann G, Feldmann H, Watanabe S, Lukashevich I and Kawaoka Y (2002) Reverse genetics demonstrates that proteolytic processing of theEbola virus glycoprotein is not essential for replication in cell culture. J. Virol. 76:406–410. DOI: 10.1128/JVI.76.1.406-410.2002

Neumann G, Geisbert TW, Ebihara H, Geisbert JB, Daddario-DiCaprio KM, Feldmann H and Kawaoka Y (2007) Proteolytic processing of the Ebola virus glycoprotein is not critical for Ebola virus replication in nonhuman primates. J. Virol. 81: 2995–2998. DOI: 10.1128/JVI.02486-06

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. 25(13): 1605-1612. DOI: 10.1002/jcc.20084

Pipasha Biswas (2014) In silico approach to develop structure and functional analysis of response receiver regulator protein of the strain pseudomonas fulva 12X. IOSR journal of pharmacy and biological science 9(1) : 79-86. DOI: 10.9790/3008-09147986

Pires DE, Ascher DB and Blundell TL (2014) mCSM: predicting the effects of mutations in proteins using graph-based signatures. Bioinformatics 30(3): 335-342. DOI: 10.1093/bioinformatics/btt691

Saitou N and Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4(4): 406-425.

Schümann M, Gantke T and Mühlberger E (2009) Ebola virus VP35 antagonizes PKR activity through its C-terminal interferon inhibitory domain. J. Virol. 83: 8993–8997. DOI: 10.1128/JVI.00523-09

Suzuki Y and Gojobori T (1997) The origin and evolution of Ebola and Marburg viruses. Mol. Biol. Evol. 14(8): 800-806. DOI:

Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. (10): 2731-2739. DOI: 10.1093/molbev/msr121

Volchkov VE, Feldmann H, Volchkova VA and Klenk HD (1998) Processing of the Ebola virus glycoprotein by the proprotein convertase furin. Proc. Natl. Acad. Sci. U.S.A. 95: 5762–5767. DOI: 10.1073/pnas.95.10.5762

Volchkov VE, Vochkova VA, Chepurnov AA, Blinov VM, DolinkO, Netesov SV and Feldmannn H (1999) Characterization of the L-gene and 5` trailer region of Ebola virus. J. Gen. Virol. 80: 355-362. DOI: 10.1099/0022-1317-80-2-355

Watanabe S et al. (2006) Functional mapping of the nucleoprotein of Ebola virus. 80: 3743. (PMID:16571791).

Wool-Lewis RJ and Bates P (1999) Endoproteolytic processing of the Ebola virus envelope glycoprotein: cleavage is not required for function. J. Virol. 73: 1419 –1426.

Ziying Han et al. (2003) Biochemical and Functional Characterization of the Ebola Virus VP24 Protein: Implications for a Role in Virus Assembly and Budding 77: 1793. (PMCID:PMC140957).