OPEN ACCESS

International Journal of Applied Sciences and Biotechnology

A Rapid Publishing Journal

ISSN 2091-2609

Available online at:

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

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-2(4) December, 2014

SEM-Biotech Publishing

Impact factor*: **1.422**
Scientific Journal Impact factor#: **3.419**
Index Copernicus Value: **6.02**

*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.

For any type of query and/or feedback don't hesitate to email us at: editor.ijasbt@gmail.com

Mundaganur D.S. et al. (2014) Int J Appl Sci Biotechnol, Vol 2(4): 544-552
DOI: 10.3126/ijasbt.v2i4.11588

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

International Journal of Applied Sciences and Biotechnology (IJASBT)

Research Article

BIOINFORMATICS STUDY ON ZAIRE *EBOLAVIRUS* (EBOV) PROTEIN FOR BETTER UNDERSTANDING THE VACCINE DEVELOPMENT

Mundaganur D.S[1], Mundaganur D.Y[2] and Ashokan Kannarath[3]

[1]Department of zoology, Willingdon College, Sangli, India
[2]Department of Zoology, Miraj Mahavidyalaya, Miraj, India
[3]Department of Zoology, P.V.P.College, K.Mahankal, Sangli, Maharashtra, India-416405

Corresponding author's email: akvsangli@gmail.com

Abstract

Nine, Ebola viruse EBOV (*Zaire ebolavirus*), proteins are extracted from the NCBI repository and their study was carried out. The physic-chemical properties and evolutionary link with other such viruses by homology modeling were carried out. All the proteins show rich in leucine domain an ideal requirement for fast attachment of the virus to the receptor molecule on the host cell surface. The prediction of trans-membrane sequence for the entire glycoprotein component reveals the ability of the virus to enter the host with ease. The lack of adequate homology model for the viral proteins indicates its novel origin and lack of well traceable evolutionary link. We studied the homology based model by using various available tools and find similar approach in all, hence finally concentrated only on one method. The model predicted shows well acceptable region on Ramchandran plot. This discrepancy is only due to the fact that we validate the model to Ramchandran plot and the model predicted were not under the well acceptable 'e' value range i.e. >1. Therefore we suggests that vaccine production against this deadly virus should be concentrated on the structure and functions of glycoprotein like low quality secreted glycoprotein (NP_066248), low quality spike glycoprotein(NP_066246)0, small secreted glycoprotein (NP_066247) and RNA dependent RNA polymerase (NP_066251).

Key words: Ebola virus (EBOV); *Zaire ebolavirus*; Protein; Amino acid composition; Homology Model; Ramchandran plot; vaccine.

Introduction

Ebola virus is a single stranded RNA virus having filamentous structure. Ebola first appeared in Nazara, Sudan and Yambuka, Democratic republic of Congo simultaneously in 1976. The Yambuka is a village near the Ebola river hence the name ebola. The genus ebola consists of five species- *Bundibugyo ebolavirus* (BDBV), *Zaire ebolavirus* (EBOV), *Reston ebolavirus* (RESTV), *Sudan ebolavirus* (SUDV) and *Tai forest ebolavirus* (TAFV). The EBOV an SUV is mainly associated with ebola outbreaks in Africa, while RSTV and TAFV is not associated with any ebola outbreak (WHO, 2014). The viral replication needs RNA dependent RNA polymerase and is performed by viral RNA dependent RNA polymerase only (Volchov *et al*, 1999). The ZEBOV consists of linearly arranged seven genes. These genes are arranged as 3'-NP-VP35-VP40-GP-VP30-VP24-L. Order (Volchov *et al*, 1999). The start and stop codon for transcription is 3′-UAAUU (Volchov *et al*, 1999). Between the NP and VP35 genes have the sequence (3′-GAU) and (3′-AGO) VP40 and GP genes and 142 bases separate the VP30 and VP24 genes (Volchkov *et al*, 1999). The 5` end of the leading strand shows stem and loop structure. Structural proteins of *Ebolavirus* and *Marburgvirus* contain large regions of homology (Volchkov *et al*, 1999). The fatality rate of EBOV is great, about 90% as compared to other species and death occurs within one day of the infection (Lung *et al*, 2010). It is observed that a co-factor protein called VP35 is needed for the EBOV RNA polymerase and viral replication (Lung *et al*, 2010). Virulence s determined by the attachment of the VP35 to the dsRNA .The end cap for VP35 is a hydrophobic molecule. Without the end cap the VP35 cannot attach to the dsRNA (Lung *et al*, 2010). This end cap helps in the expression of RIG-I-like receptor (RLR). This receptor is the ideal candidate for successful in control in the host immune system. Hence the VP345 protein is the main target for the EBOV treatment. The attack of the EBOV causes to secrete large quantity of chemokines, cytokines, and growth factors and also lymphocyte apoptosis (Warren *et al*, 2010). Current research also supports the lymphocyte apoptosis theory of disease progress (Waquier *et al*., 2010). In this paper we attempted to characterize the seven protein including two low quality proteins by bioinformatics tools, servers and software.

Materials and Methods

In the present study all the proteins available in the NCBI for the search "ebola proteins" was retrieved. The search listed nine proteins as-nucleoprotein (NP_066243), RNA

dependent-RNA polymerase(NP_066251), matrix protein (NP_066245), low quality secreted glycoprotein (NP_066248), low quality spike glycoprotein(NP_066246)0, small secreted glycoprotein (NP_066247), membrane associated protein(NP_066250), minor nucleoprotein (NP_066249), polymerase glycoprotein (NP_066244). The FASTA sequence was retrieved and used for the further bioinformatics analysis.

Protein statistics

Protein statistics was studied by using CLC sequence viewer version 6.8.1. The parameter included is sequence information like sequence type, length, weight, isoelctric point, aliphatic index, instability index, GRAVY, half-life, EC; annotations like CDS, protein and regions; amino acid percentages (Ashokan and Pillai, 2008).

Homology study

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

Trans-membrane sequence prediction

For the prediction of membrane protein SOSUI tools from the repository of EXPASY is used. It was developed for the discrimination of membrane proteins and soluble ones together with the prediction of trans-membrane helices, the accuracy of the classification of proteins was 99% and the corresponding value for the trans-membrane helix prediction was 97%. The system SOSUI is available through internet access: http://www.tuat.ac.jp/mitaku/sosui/ (Mitaku *et al*, 2002, Hirokawa *et al*, 1998). The SOSUI system is a web based tool and users can input their query sequence on the submission section. Results are typically returned in 1 min. Two predictions and two graphs are presented in the output page: (i) the type of protein; (ii) the region of trans-membrane helices when the protein is a membrane type; (iii) a graph of the hydropathy plot; (iv) helical wheel diagrams of all trans-membrane helices. Input sequence length is limited to the range of 200–5000 amino acids. Diagrams are displayed by a Java Applet pro- gram (Hirokawa *et al*, 1998).

3D Model predictions

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). SWISS-MODEL is fully automated; with user-friendly web interfaces generate reliable models without the need for complex software packages or downloading large databases. The SWISS-MODEL template library provides annotation of quaternary structure and essential ligands and co-factors to allow for building of complete structural models, including their oligomeric structure (Arnold *et al*, 2006). The improved SWISS-MODEL pipeline makes extensive use of model quality estimation for selection of the most suitable templates and provides estimates of the expected accuracy of the resulting models. The accuracy of the models generated by SWISS-MODEL is continuously evaluated by the CAMEO system. The new web site allows users to interactively search for templates, cluster them by sequence similarity, structurally compare alternative templates and select the ones to be used for model building. In cases where multiple alternative template structures are available for a protein of interest, a user-guided template selection step allows building models in different functional states. SWISS-MODEL is available at http://swissmodel.expasy.org (Bordoli *et al*, 2009).

3D model validation

For the validation of the generate d3D model we used VADAR. WASAR (Volume, Area, Dihedral Angle Reporter) is a compilation of more than 15 different algorithms and programs for analyzing and assessing peptide and protein structures from their PDB coordinate data. The results have been validated through extensive comparison to published data and careful visual inspection. The VADAR web server supports the submission of either PDB formatted files or PDB accession numbers. VADAR produces extensive tables and high quality graphs for quantitatively and qualitatively assessing protein structures determined by X-ray crystallography, NMR spectroscopy, 3D-threading or homology modeling (Willard, *et al*, 2003).

3D model viewer

We used Raswin for viewing and changing the pattern of 3D structure. This software has been created from several sources (Max, 1983, Arthur 1991). Much of the code is from RasMol 2.6, as created by Roger Sayle. The torsion angle code, new POVRAY3 code and other features are derived from the RasMol2.6x1 revisions by Arne Mueller. The Ramchandran printer plot code was derived from fisipl created by Frances C. Bernstein. (Kabsch and Sander, 1983)

Results

Physical and chemical parameters of the proteins

The physical and chemical parameters of the proteins were studied by using protparam tool form Expassy (Table1 and 2). All the protein extracted- nucleoprotein (NP_066243), RNA Dependent-RNA polymerase (NP_066251), matrix protein (NP_066245), low quality secreted glycoprotein (NP_066248), low quality spike glycoprotein (NP_066246)0, small secreted glycoprotein (NP_066247), membrane associated protein (NP_066250), minor nucleoprotein (NP_066249), polymerase glycoprotein (NP_066244)- shows invariably 140 amino acids, molecular weight about 16,000, iso-electric point ranges form 5.6-9.4, negatively charged residues range from 10-19, positively charged residues range from 11-25, EC ranges from 11.5 thousands to 30 thousand, half life is invariably 30 hours, instability index ranges from 13.9-39.6, aliphatic index ranges from 70.4 -101 and GRAVY ranges from 0.01 – 0.4

Table1: Various physical parameters of the protein extracted from EBOV

Sl. No	Protein	N. of AA	Mol. Wt	pI	(-)vely charged	(+)vely charged	Formula	No. of atoms	EC M^{-1} cm^{-1}	HL In hrs	II	AI	GRAVY
1	Nucleoprotein	140	15915	5.8	19	16	$C_{712}H_{1127}N_{191}O_{208}S_7$	2245	13075	30	55.2	94.71	-0.201
2	RNA dependent-RNA polymerase	139	15850	6.26	15	14	$C_{716}H_{1124}N_{182}O_{207}S_8$	2237	12295	30	33.8	98.13	-0.068
3	Matrix protein	140	15820	8.9	11	13	$C_{718}H_{1126}N_{194}O_{199}S_5$	2242	30480	30	32.72	101	-0.046
4	Low quality secreted glycoprotein	140	15952	9.4	19	25	$C_{682}H_{1102}N_{220}O_{213}S_5$	2222	4720	30	57	60.6	-01.022
5	Low quality spike glycoprotein	139	15536	9.23	13	18	$C_{696}H_{1097}N_{197}O_{197}S_5$	2192	21220	30	38.10	87.6	-0.202
6	Small secreted glycoprotein	140	15665	9.38	13	19	$C_{702}H_{1109}N_{199}O_{198}S_5$	2213	21220	30	39.6	86.93	-.0229
7	Membrane associated protein	140	15665	9.4	13	19	$C_{702}H_{1109}N_{199}O_{198}S_5$	2213	21220	30	39.6	86.9	-0.229
8	Minor nucleoprotein	139	14938	8.1	10	11	$C_{664}H_{1059}N_{181}O_{200}S_5$	2109	11460	30	13.9	88.4	-0.035

Table 2: Amino acid composition of various protein extracted from EBOV

	Amino aids																					
	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	Pyl	Sec
1	7.1	5.0	2.1	5.0	1.4	6.4	8.6	6.4	2.9	5.7	11.4	6.4	3.6	4.3	3.6	6.4	2.1	0.7	3.6	7.1	00	00
2	5.0	3.0	2.0	7.0	4.0	6.0	3.0	2.0	2.0	5.0	12	6.0	1.0	4.0	6.0	6.0	5.0	00	5.0	7.0	00	00
3	7.0	4.0	6.0	5.0	0.0	5.0	2.0	5.0	2.0	6.0	14	5.0	2.0	4.0	4.0	5.0	7.0	3.0	1.0	4.0	00	00
4	7.0	11	2.0	6.0	2.0	5.0	7.0	4.0	4.0	2.0	7.0	6.0	0.7	2.9	7.1	9.3	5.0	00	2.1	5.0	00	00
5	4.3	7.9	5.0	5.0	2.9	2.9	4.3	8.6	1.4	5.0	9.3	5.7	0.7	5.0	6.4	7.1	5.0	2.1	2.1	9.3	00	00
6	4.3	7.9	5.0	5.0	2.9	2.9	4.3	8.6	1.4	5.0	9.3	5.7	7.0	5.0	6.4	7.1	5	2.1	2.1	9.3	00	00
7	8.6	5.0	5.8	4.3	00	1.4	2.9	6.5	2.2	7.9	7.2	2.9	3.6	2.9	9.4	9.4	9.4	0.7	2.9	7.2	00	00
8	6.4	5.0	5.7	3.6	2.9	8.6	6.4	5.0	1.4	5.7	6.4	3.6	4.3	1.4	5.7	8.6	11.	0.7	1.4	5.7	00	00
9	6.0	5.0	5.2	3.7	3.0	8.5	6.3	5.0	1.3	5.6	6.4	3.2	4.1	1.2	5.6	8.3	11	0.7	1.3	5.4	00	00

Table 3: Homology parameters for the proteins studies

Protein	Accession number	Homology	%	E-value
Nucleoprotein	(NP_066243)	*Marburg marburgvirus*	33	7e-07
		Raven virus - Raven, Kenya, 1987	33	7e-07
		Lake Victoria marburgvirus - DRC1999	33	7e-07
RNA dependent-RNA polymerase	(NP_066251)	*Bundibugyo ebolavirus*	75	7e-67
		Marburg marburgvirus	45	7e-24
Matrix protein	(NP_066245)	*Bundibugyo ebolavirus*	87	7e-86
		Marburg marburgvirus	41	7e-24
		Lake Victoria marburgvirus - Angola2005	41	8e-24
Low quality secreted glycoprotein	(NP_066248)	Lake Victoria marburgvirus - Leiden	41	8e-09
		Marburg marburgvirus	56	7e-07
Low quality spike glycoprotein	(NP_066246)	*Lloviu cuevavirus*	58	7e-39
Small secreted glycoprotein	(NP_066247)	*Lloviu cuevavirus*	58	7e-39
Membrane associated protein	(NP_066250)	*Lloviu cuevavirus*	58	7e-39
Minor nucleoprotein	(NP_066249)	*Marburg marburgvirus*	41	9e- 10
Polymerase glycoprotein	(NP_066244)	*Reston ebolavirus* - Reston	48	8e-24

The chemical parameters (Table 2) shows phenyl alanine and serine is totally absent in all the proteins studied, tryptophan lacks in RNA dependent-RNA polymerase (NP_066251 and cystein lacks in Membrane associated protein. Lucien is the maximum in nucleoprotein in alal the nine proteins studied in EBOV.

The homology model study (Table 3) shows nucleoprotein is more related to *Marburg marburgvirus*, RNA dependent-RNA polymerase to *Bundibugyo ebolavirus* and *Marburg marburgvirus*, , matrix protein also to *Bundibugyo ebolavirus* and *Marburg marburgvirus* ,low quality secreted glycoprotein to *Marburg marburgvirus*, low quality spike glycol-protein to, small secreted glycoprotein and membrane associated protein to *Lloviu cuevavirus*,, minor nucleoprotein to *Marburg marburgvirus*, and polymerase glycoprotein to *Reston ebolavirus* - Reston.

Trans-membrane protein analysis shows (Table 4) it is absent in Nucleoprotein, Matrix protein, Low quality secreted glycoprotein, Minor nucleoprotein and Polymerase glycoprotein. All the other four proteins show trans-membrane sequence. Out of these four proteins RNA dependent-RNA polymerase shows (*LSDVPVATLPIDFIVPVLLKALS*) different sequence from Low quality spike glycoprotein, Small secreted glycoprotein and Membrane associated protein (*TSFFLWVIILFQRTFSIPLGVIH)*. But in all the proteins which exhibit trans-membrane shows sequence of length 23 amino-acids starting from 17th N-terminal and end in 39th C-terminal. TM in RNA dependent-RNA polymerase shows begins from 63rd N-terminal and end in 85th C-terminal.

The 3D model (Table 5 and Fig. 1) analysis shows that all the nine protein has no exact and satisfactory model with QMEAN >1.The protein Low quality secreted glycoprotein has no model not at all available in the repository. But the Validity study of the model shows al the models are well in the acceptable level as in the Ramchandran plot. (Fig.2).

Table 3: Homology parameters for the proteins studies

Protein	Accession number	Homology	%	E-value
Nucleoprotein	(NP_066243)	*Marburg marburgvirus*	33	7e-07
		Raven virus - Raven, Kenya, 1987	33	7e-07
		Lake Victoria marburgvirus - DRC1999	33	7e-07
RNA dependent-RNA polymerase	(NP_066251)	*Bundibugyo ebolavirus*	75	7e-67
		Marburg marburgvirus	45	7e-24
Matrix protein	(NP_066245)	*Bundibugyo ebolavirus*	87	7e-86
		Marburg marburgvirus	41	7e-24
		Lake Victoria marburgvirus - Angola2005	41	8e-24
Low quality secreted glycoprotein	(NP_066248)	Lake Victoria marburgvirus - Leiden	41	8e-09
		Marburg marburgvirus	56	7e-07
Low quality spike glycoprotein	(NP_066246)	*Lloviu cuevavirus*	58	7e-39
Small secreted glycoprotein	(NP_066247)	*Lloviu cuevavirus*	58	7e-39
Membrane associated protein	(NP_066250)	*Lloviu cuevavirus*	58	7e-39
Minor nucleoprotein	(NP_066249)	*Marburg marburgvirus*	41	9e- 10
Polymerase glycoprotein	(NP_066244)	*Reston ebolavirus* - Reston	48	8e-24

Table 4: Transmembrane sequence of the protein extracted from EBOV

Sl.No	Proteins	N terminal	transmembrane region	C terminal	Type	length
1	Nucleoprotein	No Trans membrane protein				
2	RNA dependent-RNA polymerase	63	LSDVPVATLPIDFIVPVLLKALS	85	Primary	23
3	Matrix protein	No Trans membrane protein				
4	Low quality secreted glycoprotein	No Trans membrane protein				
5	Low quality spike glycoprotein	17	TSFFLWVIILFQRTFSIPLGVIH	39	Primary	23
6	Small secreted glycoprotein	17	TSFFLWVIILFQRTFSIPLGVIH	39	Primary	23
7	Membrane associated protein	17	TSFFLWVIILFQRTFSIPLGVIH	39	Primary	23
8	Minor nucleoprotein	No Trans membrane protein				
9	Polymerase glycoprotein	No Trans membrane protein				

Table 5: 3D model Parameters deducted for various proteins assessed for EBOV

No	Protein	Template	Identity In %	Description	QMOE	QMEAN4	Oligo-state	Ligands
1	Nucleoprotein	4bk9.1.A	16.13	2-DEHYDRO-3-DEOXYPHOSPHOGLUCONATE ALDOLASE/4-HYDROXY-2-OXOGLUTARATE ALDOLASE	0.10	-0.70	MONOMER	None
2	RNA dependent-RNA polymerase	4dkw.1.A	25.93	Large terminase protein	0.08	-0.07	MONOMER	None
3	Matrix protein	4m0q.1.B	100	Membrane-associated protein VP24	0.98	-1.92	Homo-dimer	None
4	Low quality secreted glycoprotein	Nil	Nil	Nil	Nil	Nil	Nil	Nil
5	Low quality spike glycoprotein	3ve0.1.A	81.48	Envelope glycoprotein	0.75	-2.34	MONOMER	Non
6	Small secreted glycoprotein	3s88.1.B	81.65	Envelope glycoprotein	0.74	-2.00	MONOMER	Non
7	Membrane associated protein	3s88.1.B	81.65	Envelope glycoprotein	0.74	-2.00	MONOMER	Non
8	Minor nucleoprotein	4ld8.1.A	86.60	Matrix protein VP40	0.72	-0.88	Homo-dimer	None
9	Polymerase glycoprotein	3nmd.1.B	20.93	cGMP Dependent Protein Kinase	0.15	-0.96	MONOMER	None

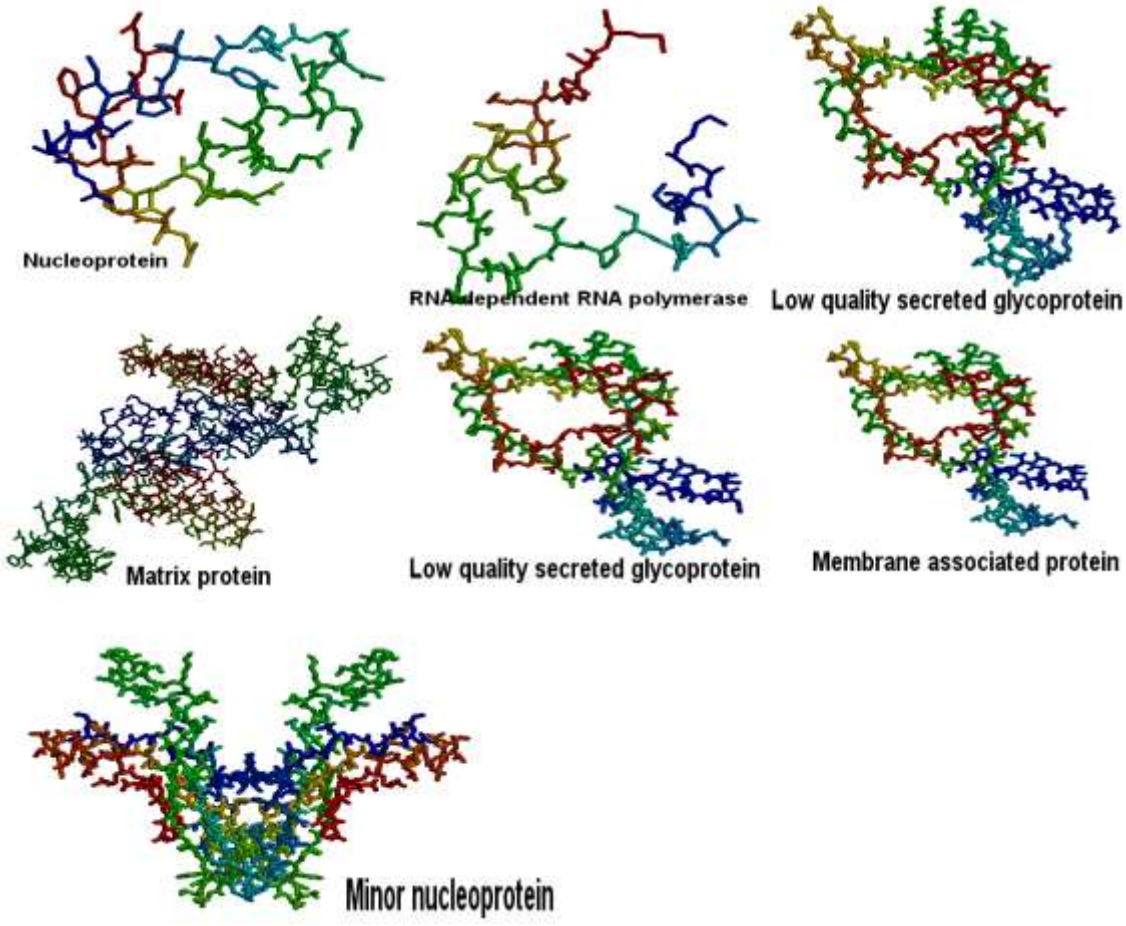

Fig. 1: 3D Model generated for various proteins of EBOV

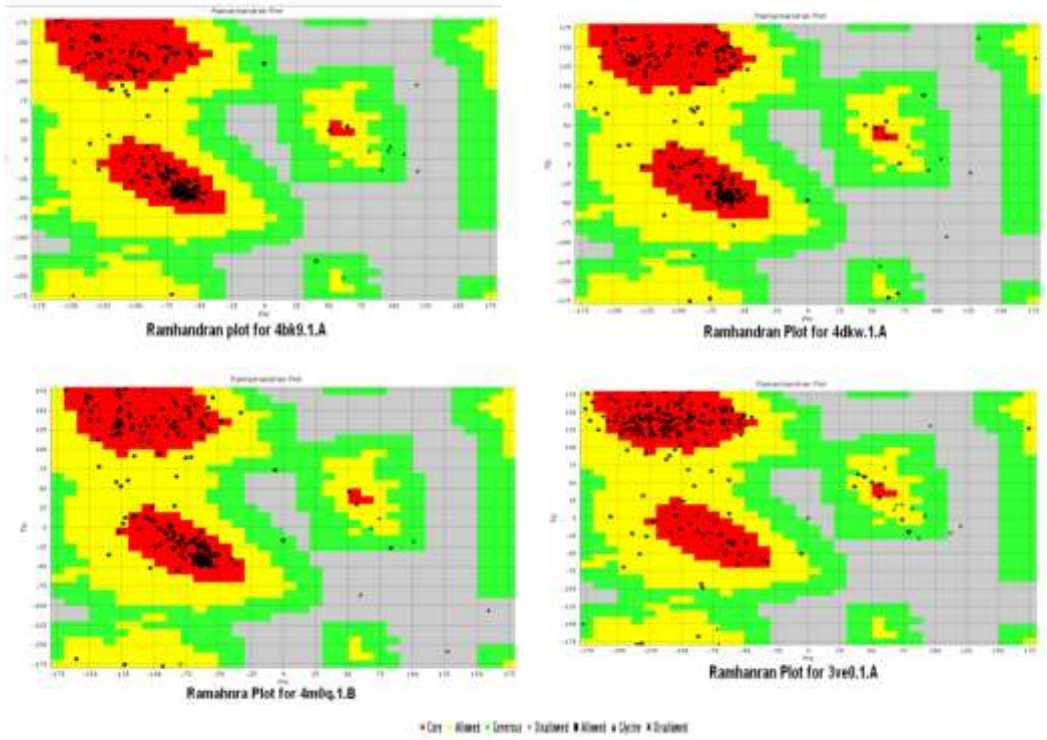

Fig. 2: Ramchandran Plot generated for selected EBOV proteins against the pdb structure

Discussion

Ebola hemorrhagic fever is a severe and often deadly illness that can occur in humans and primates. It made worldwide news because of its destructive potential. It is caused by a virus belonging to the family called Filoviridae. Scientists have identified five types of Ebola virus. Four have been reported to cause disease in humans: Ebola-Zaire virus, Ebola-Sudan virus, Ebola-Ivory Coast virus, and Ebola-Bundibugyo. The human disease has so far been limited to parts of Africa. The Reston type of Ebola virus has recently been found in the Philippines. The disease can be passed to humans from infected animals and animal materials. Ebola can also be spread between humans by close contact with infected body fluids or through infected needles in the hospital. During the incubation period, which can last about 1 week (rarely up to 2 weeks) after infection, symptoms include: Arthritis, Backache (low-back pain) ,Chills, Diarrhea , Fatigue, Fever, Headache, Malaise, Nausea, Sore throat, Vomiting. Late symptoms include:, Bleeding from eyes, ears, and nose, Bleeding from the mouth and rectum (gastrointestinal bleeding), Eye swelling (conjunctivitis), Genital swelling (labia and scrotum), Increased feeling of pain in the skin, Rash over the entire body that often contains blood (hemorrhagic), Roof of mouth looks red (Bausch, 2011, Peters *et al*, 2009) . The present investigation is concentrated on nine available protein sequences in the repository of NCBI. The physical parameter shows that all proteins are small with amino acids about 140 in number having molecular weight around 16,000. Along with these data and iso-electric point value will enable the researchers to separate the protein with little efforts. All the proteins are rich in leucine domain, it may be related with to increase the affinity to attach strongly to the host cell surface receptor as described by Luo *et al*, 1999. The homology study shows that EBOV is more related to the *Marburg marburgvirus*, except the proteins low quality spike glycoprotein (NP_066246)0, small secreted glycoprotein (NP_066247), membrane associated protein (NP_066250) these are more related with *Lloviu cuevavirus*, thus a bi-focal approach may be needed to treat the virus and to control the infection. All the glycoprotein shows trans-membrane sequences identical, hence a common vaccine or medicine may in the future can help to control the disease in effective manner. The lack of exact matches in the sense of 'e' value shows that the virus is new and has no evolutionary link with other such virus. This phylogenic relation is to be clarified in a better way. The more acceptable regions in the ramchndarn plot is only due to the fact that we assess the validity of the predicted model with ramchandran plot, hence concerned to the model the plot is valued and not with the structure of the viral protein.

References

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. J. Mol. Biol. **215**:403-410. DOI: 10.1016/S0022-2836(05)80360-2

Arnold K, Bordoli L, Kopp J, and Schwede T (2006) The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. *Bioinformatics* **22:**195-201. DOI: 10.1093/bioinformatics/bti770

Arthur ML (1991) Protein Architecture: A Practical Approach, IRL Press Publishers,.

Ashokan KV and Pilali MM (2008) *In silico* characterization of silk fibroin protein using computatational tools and servers. *Asian Journal of experimental sciences* **22**(3): 265-274.

Bausch DG (2011) Viral hemorrhagic fevers. In: Goldman L, Ausiello D (eds.) *Goldman's Cecil Medicine*. 24th ed. Philadelphia, PA: Saunders Elsevier; chap 389.

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

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K and Madden TL (2008) BLAST+: architecture and applications. *BMC Bioinformatics* **10**:421. DOI: 10.1186/1471-2105-10-421

Hirokawa T, Boon-Chieng S, and Mitaku S (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14**: 378-379. DOI: 10.1093/bioinformatics/14.4.378

Kabsch W and Sander C (1983) Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **22**:2577-2637. DOI: 10.1002/bip.360221211

Leung W, Shaffer CD, Cordonnier T, Wong J, Itano MS and *et al.* (2010). Evolution of a Distinct Genomic Domain in Drosophila: Comparative Analysis of the Dot Chromosome in Drosophila melanogaster and Drosophila virilis. Genetics 185(4): 1519-- 1534. DOI: 10.1534/genetics.110.116129

Luo Z, Matthews A.M and Weiss S.R (1999). Amino Acid Substitutions within the Leucine Zipper Domain of the Murine Coronavirus Spike Protein Cause Defects in Oligomerization and the Ability To Induce Cell-to-Cell Fusion. *J Virol*, **73**(10): 8152–8159.

Max N (1983) Computer Representation of Molecular Surfaces, IEEE Computer Graphics and Applications, pp. 21-29. DOI: 10.1109/MCG.1983.263183

Mitaku S, Hirokawa T and Tsuji T (2002) Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane-water interfaces. *Bioinformatics* **18:** 608-616. DOI: 10.1093/bioinformatics/18.4.608

Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R. and Schäffer AA (2008) Database indexing for production MegaBLAST searches. *Bioinformatics* **15**:1757-1764. DOI: 10.1093/bioinformatics/btn322

Peters CJ (2009) Marburg and Ebola virus hemorrhagic fevers. In: Mandell GL, Bennett JE, Dolin R (eds.) *Principles and Practice of Infectious Diseases*. 7th ed. Philadelphia, PA: Churchill Livingstone Elsevier; chap 164.

Sanhz JP, Bouillt MN, Colineau E, Blaise A, Amanowics M, Blurlet P, Fournier JM, Charvolin T and Larroque J (1993) *Physica B*, **186-188**: 675-677.

Schaffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, Koonin EV, Altschul SF. (2001) Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* 29(14):2994-3005. DOI: 10.1093/nar/29.14.2994

Volchkov VE, Vochkova VA, Chepurnov AA, Blinov VM, Dolink O, Netesov SV and Feldmannn H (1999) Characterization of the L-gene and 5` trailer region of Ebola virus. *J. Gen. Virol.* **80**: 355-362.

Waquier N, Becquart P, Padilla C, Baize S and Leroy EM (2010) Human Fatal Zaire Ebola Virus Infection Is Associated with an Aberrant Innate Immunity and with Massive Lymphocyte Apoptosis. *PLOS Neglected Tropical disease* **4**(10):1-10. DOI: 10.1371%2Fjournal.pntd.0000837

Warren L, Manos PD, Ahfeldt T, Loh YH, Li H, Lau F, Ebina W, and *et al.* (2010). Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* **7**: 618–630. DOI: 10.1016/j.stem.2010.08.012

WHO: www.who.int/mediacentre/factsheets/fs103/en/

Willard L, Ranjan A, Zhang H, Monzavi H, Robert FB, Brian DS and David SW (2003) VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res.* **31** (13): 3316-3319. DOI: 10.1093/nar/gkg565