CrossMark

Analysis of Ebola virus polymerase domains to find strain-specific differences and to gain insight on their pathogenicity

Seema Patel[1] · Snigdha Patel[2]

Abstract Ebola virus, a member of the family *Filoviridae* has caused immense morbidity and mortality in recent times, especially in West Africa. The infection characterized by chills, fever, diarrhea, and myalgia can progress to hemorrhage and death. Hence, it is a high priority area to better understand its biology in order to expedite vaccine development pipelines. In this regard, this study analyzes the domains in RNA polymerase of fifteen publicly-available Ebola isolates belonging to three strains (Zaire, Sudan and Reston). The protein FASTA sequences of the isolates belonging Zaire, Sudan and Reston strains were extracted from UniProt database and submitted to the interactive web tool SMART for the polymerase domain profiles. Subsequent in silico investigation furnished interesting results that sure can contribute to the understanding of Ebola pathogenesis. The key findings and patterns have been presented, and based on them hypotheses have been formulated for further empirical validation.

Keywords Ebola · Polyprotein · RNA polymerase · Protein domains · Pathogenesis · Immune manipulation

✉ Seema Patel
seemabiotech83@gmail.com

1 Bioinformatics and Medical Informatics Research Center, San Diego State University, 5500 Campanile Dr, San Diego, CA 92182, USA

2 Strand Life Sciences Pvt. Ltd., Bangalore, Karnataka 560024, India

Introduction

A wide range of pathogens belonging to virus, bacteria, protozoa, and fungi Kingdom afflict mankind. Viruses are particularly challenging to control for their rapid antigen variations and immune evasion [1]. Ebola virus belonging to the family *Filoviridae* has been the cause of high mortality in recent times [2, 3]. Another *Filoviridae* member Marburg virus has also been associated with lethal human pathogenesis [4, 5]. A study reports 50–90 % fatality following infection with the hemorrhagic strains of these viruses [6]. As per World Health Organization (WHO) report out of 5335 cases recorded till September 2014, 50 % (2622 cases) led to death [4]. To contain the transmission and treat the vulnerable community in Africa, Centers for Disease Control and Prevention (CDC) has taken steps by deploying health care staffs [7]. For the enormous public health risk, Ebola virus has been the focus of intense research in recent times. It was first discovered in West Africa in 1976 and now it is endemic to various countries in this region, including Sudan, Zaire, Uganda, Guinea, Liberia, Sierra Leone, and Congo. In fact, the name Ebola traces its origin to Ebola river in Congo [8–12]. The virus spreads by contact with infected person's body fluid such as blood, saliva, urine, semen etc. [13]. Symptoms of the infection include chills, fever, diarrhea, malaise, and myalgia, which can progress to hemorrhage and death [8]. It is a zoonotic disease with bats as major vectors [14]. Fruit bats from Pteropodidae family have been validated as reservoirs (with the detection of virus-specific antibody in the bat serum) [15, 16] and insectivorous free-tailed bats from Molossidae family have been suspected as vectors for this virus [17]. Also, evidences suggest transmission of the virus from chimpanzee and monkeys, even pigs [8]. As characteristic of most

viruses, Ebola virus has diversified into several strains. The most-studied strains include Zaire, Sudan, Côte d'Ivoire (from Tai Forest reserve), Bundibugyo, and Reston, all of which have originated in Africa, except the last one, which evolved in Philippines [8, 18–20]. Chronologically, the outbreak-associated strains are Zaire (strain Mayinga-76), Sudan (strain Maleo-79), Tai Forest (strain Cote d'Ivoire-94), Zaire (strain Gabon-94), Zaire (strain Kikwit-95), Sudan (strain Uganda-00), and Zaire (diverse lineage, 2014) [21, 22]. Among the existing Ebola strains, Zaire is the most aggressive one, and is linked to most outbreaks [23].

Ebola virus is an enveloped, non-segmented, single stranded, negative-sense RNA virus with genome length spanning 19,000 bases [8]. The RNA is coated in nucleocapsid, which in turn is covered in a glycoprotein-embedded membrane. The polyprotein comprises of seven parts, such as leader sequence, nucleoprotein, virion proteins (VP35 and VP40), glycoprotein, virion proteins (VP30, VP24), and RNA-dependent RNA polymerase [8]. These components of polyprotein are mostly conserved in terms of their amino acid length, such as nucleoprotein (739aa), VP35 (340aa), VP40 (326aa), glycoprotein (676–677aa), VP30 (288aa), VP24 (251aa), and polymerase (2212aa). The polyprotein has a 32aa-long conserved coiled coil region (LVSVTQHLAHLRAEIRELTNDYNQQRQSRTQT) at 2082–2113aa region [24].

VP24 and VP35 act as transcription activators [25]. The former perturbs interferon signaling and latter is an interferon antagonist, thus together they are capable of blocking production of interferons via STAT1 inhibition [25, 26]. VP40 is the matrix protein, which mediates virus-like particle budding [27]. Glycoprotein is the virulence factor that can be liberated or anchored to membrane [6]. These conjugated proteins are secreted into host extracellular space, in diverse truncated isoforms [28]. Full-length glycoproteins measure 150–170-kDa, and they are inserted into the viral membrane, through transcriptional editing [29]. These trimeric proteins with *O*-linked oligomannose glycans adhere to host cells and mediate fusion with host membrane [6]. Attachment to the endothelial cells via Niemann-Pick C1 receptors (C-type lectin membrane proteins) is followed by replication of the virus [30]. Antigen-presenting cells (APCs) like macrophages and dendritic cells are targeted by the virus, which creates a barrage of cytokines such as interferon (IFN-α), interleukin (IL-2, IL-10), and tumor necrosis factor-α (TNF-α) [31]. Also, excess of T lymphocyte (T helper and T cytotoxic cells) and Natural killer cells (NKC) apoptosis has been reported [23]. In advanced form of the infection, complement cascade is activated, which clots blood, causes endothelial leakage, multi-organ failure, hypotension, and leads to respiratory collapse [32]. Thus, antigenic subversion,

characterized by immune suppression and inflammation is described as a potent pathogenesis mechanism of this virus [8, 28], more or less, akin to other deadly viral pathogens like dengue, SARS (Severe acute respiratory syndrome) etc. [32].

Excessive fluid loss, leading to hyponatremia, hypokalemia, hypocalcemia, hypomagnesemia, hypoalbuminemia, and hypoxemia (abnormally low oxygen level in blood) are characteristic of Ebola fever, which if untreated can cause, shock and hemorrhage [33]. So, 'fluid replacement therapy' for replenishing the depleted electrolytes is a major support in averting the adverse effects [34]. In serious cases, vasoactive agents, hemodialysis and mechanical ventilation are recommended to prevent respiratory and circulatory collapse [35]. There are no Ebola-specific therapeutics yet [36]; however, several promising candidates are under intense trial. Monoclonal antibodies (MAbs) have been validated to target glycoproteins on the virus membrane. Though the MAb-glycoproein interaction is still enigmatic, it has been revealed that MAbs bind to epitopes in glycoproteins base, glycan cap, or mucin-like domain [37]. In this regard, a combination of MAbs, termed as ZMapp has shown considerable therapeutic promise [38, 39]. It can mitigate viremia and related abnormalities up to 5 days post-infection [38]. Favipiravir (T-705) (an ant-influenza drug) has shown efficacy towards this virus [40]. Ribavirin is another drug effective against many RNA viruses such as hepatitis C (HCV), Lassa virus, and respiratory syncytial viruses (RSV) [41]. Studies have found synergistic effect of above two drugs in management of hemorrhagic Ebola fever [42]. A synthetic adenosine analogue BCX4430 is capable of inhibiting viral RNA polymerase function, as demonstrated in animal models [43]. Also, small interfering RNAs (sRNAs) are being tested to target the virus [12]. In this regard, phosphorodiamidate morpholino oligomers (PMO), a type of synthetic antisense molecules blocking mRNA coding for VP24 proteins has shown promise [44]. Convalescent plasma (plasma from Ebola survivors) is under evaluation for a possible therapeutic [45]. To fine-tune the emerging drugs and to develop novel therapeutics, a keen knowledge of protein domain configuration of Ebola virus is paramount.

Materials and methods

Sequence retrieval from UniProt database

This investigation used polymerase protein FASTA sequences of Ebola virus available in publicly-available database UniProt (http://www.uniprot.org/uniprot/) [46]. Care was taken to pull out sequences belonging to different strains of Ebola i.e. Zaire, Sudan and Reston.

Usage of SMART platform for protein domain list

For the protein domain information of the polymerase sequences, public platform SMART (Simple Modular Architecture Research Tool) [24] was used. Using HMMer (for alignment) and BLAST (for bit score), SMART identifies and annotates domains, assigning them to families and illustrating their topologies [24].

Custom scripts development for domain distribution

Subsequently, the domain profiles in the polymerase sequences and their distribution patterns were analyzed using scripts developed in Bash language. The scripts were constructed using the commands like awk, sort, grep, comm and while loop. The scripts included ebola_protein_domains.sh, ebola_data_manipulations.sh and ebola_protein_common.sh. The script ebola_protein_domains.sh sorts the polymerase domains of each isolates alphabetically, counts the total number of domain for each isolate and then conducts comparison of domain profile between each pair of isolates. The pair-wise comparison was meant to find domain unique to an isolate. The script ebola_data_manipulations.sh uses the output of ebola_protein_domains.sh as input and finds the domains common to each pair of isolates. The script ebola_protein_common.sh uses each isolate polymersase domain list and searches pathogenically-critical domains like YARHG, WH1, RICTOR_M, Pro-kuma_activ, IENR1, DDHD, DALR_2, WSN, VWC, Telomerase_RBD, RasGAP, PA2c, MIT, YqgFc, TLC, STI1, RUN, RL11, RAP, R3H, LamG, HALZ, B41, HOLI, PLCYc, Hr1, H4, GGDEF, LPD_N, LON etc.

On executing the scripts, the generated output files were ebola_data, ebola_data_analysis and ebola_domain_consensus. Relevant and interesting findings were extracted from these result files. Domains common to all, shared among some and unique to some polymerase sequences; strain-specific signature and anomaly; relevance of the domains to pathogenesis were analyzed. Based on the data, clusters were formed and tabulated. Also, hypotheses were formulated and insights were discussed, which is likely to be of relevance in better management of Ebola infection.

Results

Ebola polymerases domain distribution

The 15 Ebola strains were A0A0A7LUV3, A0A068J465, A0A0B5EB22, A0A0D5W8U2, A0A0E3TN89, A0A0F7I

MH5, A0A0G2Y8I7, A0A0G2YD12, A0A068J9B1, Q5XX01, Q6V1Q2, Q8JPX5, Q91DD4, Q05318, and X5H5B6. The SMART-predicted number of domains in the polymerase ranged from 54 to 70 (some of them are overlapping), of which minimum was found in Q5XX01 (a Sudan Ebola virus) and maximum in Q91DD4 (a Reston Ebolavirus). All the Zaire strains contained domains in the range of 61–69. In total, the number of unique domains observed in Ebola virus is 158 (though some of they were overlapped due to limitations of homology-based predictions). This information has been presented in Table 1. Out of them only a few i.e. 9 (only 5.7 %) domains are present in all the isolates. These universally-occurring domains are WH2, TBC, SNc, SMI1_KNR4, RICTOR_V, PX, Pfam:FtsJ, MBD, and IGR. These domains have well-conserved positions such as Pfam:FtsJ (215–340aa), PX (402–524aa), SNc (483–608baa), TBC (512–883aa), WH2 (726–738aa), SMI1_KNR4 (898–986aa), IGR (980–1033aa), MBD (1016–1069aa), RICTOR_V (1077–1120aa), Pfam:FtsJ (1813–2007aa). Pfam:FtsJ is present more than once (i.e. at 215–340aa and 1813–2007aa). Figure 1 illustrates these essential domains.

The domains present in any of the 14 isolates include YARHG, WH1, RICTOR_M, Pro-kuma_activ, MYSc, IENR1, HTH_ASNC, FABD, DDHD, and DALR_2. YARHG is harbored at 391–447aa, though Reston isolates Q8JPX5 and Q91DD4 have this domain at 382–447aa and Sudan isolate Q5XX01 lacks it. WH1 lies at 995–1103aa and absent from Sudan isolate Q5XX01. RICTOR_M is present at 839–1119aa and absent from Sudan isolate Q5XX01. Pro-kuma_activ lies between 813 and 918aa, It is absent in Sudan isolate Q5XX01. Domain MYSc spans 619–1056aa, though it lies between 515 and 1057aa in the Sudan isolate and a Reston isolate Q8JPX5. It is missing in another Reston isolate Q91DD4. The Zaire isolate Q05318 (strain Mayinga-76) has this domain at 1606–1974aa. IENR1 occupies position 1316–1366aa, though Reston isolate Q8JPX5 and Q91DD4 have it at 1285–1332aa. It's absent in Sudan isolate Q5XX01. Total of 24 domains (YqgFc, TLC, STI1, RUN, RL11, RAP, R3H, PI3Ka, PhBP, MGS, Lipid_DES, LIM, LamG, HhH1, HALZ, Grip, Glyco_10, Elp3, DEP, Cyclin_C, Citrate_ly_lig, CAT, Brr6_like_C_C, B41) have only presence in any 12 isolates. B41 at 1570–1806aa is missing in 3 isolates (1 Sudan and 2 Reston isolates). There is positional shift of this domain in Zaire isolates Q6V1Q2 (Kikwit-95) and Q05318 (Mayinga-76), which occurs at 1589–1806aa. Y1_Tnp at 1219–1326aa is lacking in 4 isolates (1 Zaire, 1 Sudan and 2 Reston isolates). HOX at 1872–1924aa is lacking in 4 isolates (1 Zaire, 1 Sudan and 2 Reston isolates). HOLI at 1603–1817aa is lacking in 4

Table 1 Polymerase domain count of the 15 Ebola isolates belonging to Zaire, Sudan and Reston strains

No.	UniProt accession no.	Strain	Breakout (year)	No. of domains
1	A0A0A7LUV3	Zaire	Liberia-14	62
2	A0A068J465	Zaire	Liberia-14	63
3	A0A0B5EB22	Zaire	Liberia-14	61
4	A0A0D5W8U2	Zaire	Liberia-14	62
5	A0A0E3TN89	Zaire	Liberia-14	63
6	A0A0F7IMH5	Zaire	Liberia-14	61
7	A0A0G2Y8I7	Zaire	Liberia-14	62
8	A0A0G2YD12	Zaire	Liberia-14	64
9	A0A068J9B1	Zaire	Liberia-14	63
10	Q5XX01	Sudan	Uganda-00	54
11	Q6V1Q2	Zaire	Kikwit-95	69
12	Q8JPX5	Reston	Reston-89	68
13	Q91DD4	Reston	Philippines-96	70
14	Q05318	Zaire	Mayinga-76	67
15	X5H5B6	Zaire	Guinea-14	64

There polymerase protein domain ranged from 54 to 70. Liberia, Uganda, Kikit, Philippines, Myinga and Guinea are places of the virus outbreak. The hyphenated numbers are the year of breakout. Liberia-14 was a severe outbreak, most of the isolates are from that outbreak. The Liberia-14 isolates are from Zaire strain

Fig. 1 The core domains in the polymerase protein of all Ebola virus strains

isolates (1 Zaire, 1 Sudan and 2 Reston isolates). PLCYc at 200–304aa is lacking in 5 isolates (2 Zaire, 1 Sudan and 2 Reston isolates). Hr1 at 2047–2106aa is lacking in 6 isolates (3 Zaire, 1 Sudan and 2 Reston isolates). H4 at1887–1948aa is lacking in 6 isolates (4 Zaire and 2 Reston isolates). In the Sudan isolate, this domain has appositional shift i.e. at 2089–2113aa. GGDEF at 1479–1639aa is lacking in 6 isolates (3 Zaire, 1 Sudan and 2 Reston isolates). LPD_N at 1604–2118 is lacking in 7 isolates (4 Zaire, 1 Sudan and 2 Reston isolates). Other domain profile information has been furnished in Table 2.

AARP2CN domain is present only in Sudan isolate Q5XX01. A2M_recep domain is present in Reston strain Q8JPX5 and Q91DD4. Q91DD4 contained a DDHD domain that Q8JPX5 lacked. ALBUMIN at 994–1167aa and VWC_out at 1106–1162aa are domains, confined to only in Sudan isolate Q5XX01. LON 235–398 is present only in Reston isolates Q8JPX5 and Q91DD4. Zalpha at 2086–2151aa is present only in Reston isolate Q91DD4.

The domains present only in any two isolates are ZM, uDENN, Tubulin, TRCF, Sec63, RTC4, RhoGEF, PRP, PLAc, Pfam:SQHop_cyclase_C, PepX_N, MADS, LON, KH, JHBP, IlGF, IGc1, HTH_ARSR, FYVE,

Fmp27_GFWDK, Flavin_Reduct, FISNA, DSL, DM10, Cullin_Nedd8, CULLIN, CTLH, CRA, CO_deh_flav_C, CarD_TRCF, calpain_III, BEN, BAG, Alpha_kinase, AgrD, ACTIN, and A2M_recep. The domains present in only one isolate are zf-AD, Zalpha, VWC_out, TyrKc, Spc7, RICTOR_phospho, RICTOR_N, Ribosomal_L2_C, RIBOc, POL3Bc, PKD, Pfam:Mononeg_mRNAcap, Pfam:Methyltrans_Mon, NADH-G_4Fe-4S_3, MAGE, L51_S25_CI-B8, KR, ITAM, HTH_DTXR, HPT, HELICc2, G_gamma, FN1, Flo11, FA58C, eRF1_1, EMP24_GP25L, DSPc, DIRP, Cyt-b5, CUE, CENPB, CASc, B_lectin, BH4, BAR, B561, AT_hook, ALBUMIN, AARP2CN. The 2 LON domains were found only in the Reston isolates; 1 Zalpha was in Q91DD4 (Reston); VWC_out and ALBUMIN were detected in Q5XX01 (Sudan) only. Some DUFs (domains of unknown functions) (5 types) are found, which included DUF1041, DUF1237, DUF1866, DUF4208, and DUF862, present in Q91DD4 (Reston), Q6V1Q2 (Zaire), Q91DD4 (Reston), Q5XX01 (Sudan), Q8JPX5 (Reston), respectively. Domains unique to a strain in pairwise comparisons were considered accessory or dispensable domains. Pair-wise common domains varied in number, the range being 8–63. B41 was

Table 2 Domain distribution in the analyzed Ebola isolates

No. of isolates	Domains	No. of domains	Missing in isolates
15	WH2, TBC, SNc, SMI1_KNR4, RICTOR_V, PX, Pfam:FtsJ, MBD, IGR	9	–
14	YARHG	11	1 Sudan
	WH1		1 Sudan
	RICTOR_M		1 Sudan
	Pro-kuma_activ		1 Sudan
	MYSc		1 Reston (Q91DD4)
	IENR1		1 Sudan
	Zpr1, HTH_ASNC, FABD, DDHD, DALR_2		–
13	WSN	6	1 Zaire and 1 reston isolate
	VWC		2 Reston isolates
	Telomerase_RBD, RasGAP, PA2c, MIT		–
12	YqgFc, TLC, STI1, RUN, RL11, RAP, R3H, PI3Ka, PhBP, MGS, Lipid_DES, LIM, LamG, HhH1, HALZ, Grip, Glyco_10, Elp3, DEP, Cyclin_C, Citrate_ly_lig, CAT, Brr6_like_C_C	24	–
	B41		1 Sudan and 2 Reston isolates
11	Y1_Tnp	5	1 Zaire, 1 Sudan and 2 Reston isolates
	LIGANc		–
	IBN_N		–
	HOX		1 Zaire, 1 Sudan and 2 Reston isolates
	HOLI		1 Zaire, 1 Sudan and 2 Reston isolates
10	PLCYc	2	2 Zaire, 1 Sudan and 2 Reston isolates
	CPDc		–
9	Hr1	3	3 Zaire, 1 Sudan and 2 Reston isolates
	H4		4 Zaire and 2 Reston isolates
	GGDEF		3 Zaire, 1 Sudan and 2 Reston isolates
8	LPD_N	1	4 Zaire, 1 Sudan and 2 Reston isolates

This table shows the domains in the polymerase proteins common to a set of isolates. Nine domains (WH2, TBC, SNc, SMI1_KNR4, RICTOR_V, PX, Pfam:FtsJ, MBD, IGR) are present in all 15 isolates; whereas variable number of domains are present in other isolate sets. The domain names and their biological significance have been outlined in the "Discussion" section

absent in Sudan strain Q5XX01 as well as the Reston strains Q8JPX5 and Q91DD4. COG6 (conserved oligomeric complex) domain is unique to Zaire strain A0A0E3TN89. Some Zaire strains such as A0A0F7IMH5 (Libria-14) and Q05318 (Mayinga-76) have the chitin-binding domain ChtBD3.

Despite belonging to the same Zaire strain, the members varied considerably in their polymerase domain profiles. Some of these auxiliary domains included ChtBD3, LPD_N, PKD, Hr1, H4, GGDEF, C4, BAG, B_lectin. Only

the Isolate Q05318 (Mayinga-76 strain) and Isolate Q6V1Q2 (Kikwit-95 strain) contained the BAG domain. Isolate Q05318 (Mayinga-76 strain) also has B_lectin that is missing in other Zaire isolates.

SMART annotations of some crucial domains have been presented within the parentheses [24]. The core domains are WH2 (WASP-Homology 2 is an actin-binding motif), TBC (GTPase activator proteins), SNc (Staphylococcal nuclease homologues), SMI1_KNR4 (yeast cell wall assembly regulator SMI1 and the cell

proliferation protein KNR4), RICTOR_V (Rictor is a scaffolding protein important for maintaining mTORC2 integrity), PX (phox domain is involved in cell signaling, vesicular trafficking, protein sorting and lipid modification, among others), Pfam:FtsJ (a methyltransferase with viral RNA capping role), MBD (methyl-CpG binding domain), and IGR (Conserved motif in fungal protein).

Consensus domains

The domains present in all Zaire strain isolates, but missing in some Reston and all Sudan strain isolate isolates are YARHG (an extracellular domain in kinases named after conserved YARHG motif), WH1 (WASP homology region 1), RICTOR_M (scaffolding protein domain), Pro-kuma_activ (pro-kumamolisin, activation domain), MYSc (Myosin large ATPases), IENR1 (intron encoded nuclease repeat motif), Zpr1 (ZPR1-type zinc finger domains), HTH_ASNC (helix_turn_helix ASNC type), FABD (F-actin binding domain), DALR_2 (domain of cysteinyl-tRNA-synthetases), DDHD (Four conserved residues forming metal binding site), WSN (Worm-specific (usually) N-terminal domain), VWC (von Willebrand factor type C domain), Telomerase_RBD (RNA binding domain), RasGAP (GTPase-activator protein for Ras-like GTPases), PA2c (Phospholipase A2), MIT (microtubule interacting and trafficking molecule domain), YqgFc (ribonuclease with RNase H fold), TLC (TRAM, LAG1 and CLN8 homology domains), STI1 (heat shock chaperonin-binding motif), RUN (domain involved in Ras-like GTPase signaling), RL11 (Ribosomal protein L11/L12), RAP (RNA-binding domain common in Apicomplexans), R3H (putative single-stranded nucleic acids-binding domain), PI3Ka (phosphoinositide 3-kinase family, accessory domain), PhBP (insect pheromone/odorant binding protein domains), MGS (domain of methylglyoxal synthetase), Lipid_DES (sphingolipid Delta4-desaturase), LIM (Zinc-binding domain present in Lin-11, Isl-1, Mec-3), LamG (Laminin G domain), HhH1 (helix-hairpin-helix DNA-binding motif class 1), HALZ (homeobox associated leucin zipper), Grip (golgin-97, RanBP2alpha, Imh1p and p230/golgin-245), Glyco_10 (glycosyl hydrolase family 10), Elp3 (elongator protein 3, MiaB family, Radical SAM), DEP (domain in Dishevelled, Egl-10, and Pleckstrin), Cyclin_C (proteins controlling the progression of cell cycle by activating cyclin-dependent kinase (Cdk) enzymes), Citrate_ly_lig (domain of citrate lyase ligase), CAT (chloramphenicol acetyltransferase), Brr6_like_C_C (Cysteine-rich C terminus of fungal protein), and B41 (Plasma membrane-binding domain). Further information about these domains can be obtained at http://smart.embl-heidelberg.de/browse.shtml [24].

Discriminating domains

The following domains are lacking in all Sudan and Reston isolates; also missing in a Zaire isolate. The domains include Y1_Tnp (transposase IS200 like), LIGANc (DNA ligase), IBN_N (importin-beta N-terminal domain), HOX (Homeodomain is a DNA binding protein), HOLI (ligand binding domain of hormone receptors) etc. The domains absent from all Sudan and Reston isolates, also from 2 Zaire isolates are PLCYc (phospholipase C domain Y), CPDc (catalytic domain of ctd-like phosphatases), PKD (repeats in polycystic kidney disease 1 (PKD1) and other proteins).

Some other discriminating domains are Hr1 (Rho effector or protein kinase C-related kinase homology region 1 homologues), H4 (histone H4), GGDEF (Diguanylate cyclase, present in a variety of bacteria), and LPD_N (Lipoprotein N-terminal Domain).

Domain present in only few isolates include HisKA (dimerisation and phosphoacceptor domain of histidine kinases), C4 (C-terminal tandem repeated domain in type 4 procollagens), AARP2CN (domain in asparagine and aspartate rich protein 2), BAG (present in regulator of Hsp70 proteins), A2M_recep (receptor-binding domain (RBD) of alpha-2-macroglobulin proteins), B_lectin (Bulb-type mannose-specific lectin), GAF (cGMP-specific phosphodiesterases) etc.

Discussion

This virus is highly contagious and it has shown the potential to spread as an epidemic. Our understanding of this virus is still nascent and vaccine development is yet to succeed. In this scenario, precaution is the best strategy, which can be achieved by educating the vulnerable group, in the virus-endemic reason. Also, limiting interaction with wildlife vectors like primates and bats is required to obviate any outbreaks. Meanwhile, research understandings should be continued to unravel pathogenesis mechanisms and factors. This study has contributed in this objective, key inferences of which have been discussed here.

Domain architectures are decisive in catalytic functions of proteins, including their pathogenicity roles [47]. Results of this study indicate that despite the similar component structures in Ebola virus, the domain distribution vary immensely and might be the cause of variable virulence vigor of different strains. Some critical findings have been analyzed and interpreted here.

DDHD, a domain with four conserved amino acid residues forming metal binding site is a conserved domain. It is lacking in a Reston strain, suggesting the loss as one of

the likely reason for the loss of pathogenesis in this strain. Studies in other pathogens have shown that this domain has conserved aspartate and histidine residues, modification of which leads to loss of phospholipase activity and membrane trafficking [48]. B41, a plasma membrane-binding domain appears to be another critical domain for pathogenesis. It is lacking in Sudan isolate Q5XX01 and both Reston isolates Q8JPX5 and Q91DD4. It clearly indicates role of this domain in attachment to host membrane, absence of which in the Sudan and Reston isolate might be rendering them less aggressive than Zaire strains. So, it can be hypothesized that B41 domain located approximately at 1570–1806aa, the quintessential weapon of Zaire strain can be targeted to deter the viruses from anchoring to host endothelial cells. Zaire isolate Q05318 (a Mayinga-76) and isolate Q6V1Q2 (a Kikwit-95 strain) contained this domain at 1589-1806aa, indicating their closer relatedness. There is sparse published literature on this critical domain; however one publication justifies its immense relevance in immune functions, which has been cited here. A conserved neuronal protein GRP1-associated scaffolding protein (GASP) has a B41 domain (as part of a FERM domain), implicated in binding to membrane as well as cytoskeletal elements like actin [49].

Zaire isolate Q05318 (Mayinga-76 strain) and Q6V1Q2 (Kikwit-95 strain) contained the BAG domain (heat shock protein regulator), normally lacking in other Zaire isolates. This domain plays role as co-chaperone for Hsp70 chaperones for proper protein folding with quality control and degradation pathways [50]. Role of this domain in regulating the heat shock protein quality check pathways can be correlated to the pathogenesis of the isolates harboring it.

Four Zaire isolates show anomalous behavior such as A0A0A7LUV3 (Liberia-14), A0A0F7IMH5 (Liberia-14), Q6V1Q2 (Kikwit-95), Q05318 (Mayinga-76). The last two Zaire isolates have similar features (a BAG domain, shifted B41 domain), which suggest their phylogenetic proximity. Also, these two strains have been linked to large outbreaks. It leads to the hypotheses that the BAG domain might be their advantage. Isolate Q05318 (Mayinga-76) also has a mannose-specific lectin (B_lectin) and a chitin-binding domain (ChtBD3) which has been associated with host pathogenesis. ChtBD3 is present in isolate A0A0F7IMH5 (Libria-14) as well. ChtBD3 domain is present in serotype 3 of dengue virus, a deadly Flavivirus [51]. As Mayinga-76 strain was associated with the very first outbreak, it can be hypothesized that this lectin and chitin-binding domain in the ancestral strain led to human infection, which evolved over the time to lose it and diversify into other strains.

By comparison with Q8JPX5 (Reston isolate) polymerase domain sequence, DUF1041 at 2091–2177aa in Q91DD4 (Reston isolate) was predicted to be Zalpha domain. By comparison with Q8JPX5 (Reston isolate) domain sequence, DUF1866 at 1774–1886aa in Q91DD4 (Reston isolate) is likely to be either Cyt-b5 or CASc domain. DUF862 at 2030–2145aa position in Q8JPX5 (Reston isolate) lies just above HisKA domain. In another Reston isolate Q91DD4, this location is occupied by Telomerase_RBD (2064–2168aa). So, it was gathered that DUF862 is Telomerase_RBD, which has undergone heavy polymorphism. In Q6V1Q2 (Zaire isolate), DUF1237 (1543–1811aa) overlaps with B41 domain, which indicates the domain might be just a modified form of B41 domain. The domains occurring before this DUF are exactly same (IENR1, DEP, LamG, Lipid_DES, YqgFc) in another Zaire isolate A0A0G2Y8I7. Based on position and location comparison with other isolates, DUF4208 at 242–328aa in Sudan isolate Q5XX01 could be PLCYc, RUN, or Cyclin_C domains.

Many of the critical domains are missing from Sudan and Reston strain, which suggests their comparatively weaker pathogenic potential compared to Zaire strain of the virus. Sudan isolate has the least number of domains (54), and it lacks in otherwise well-conserved domains like VWC, YARHG, WH1, RICTOR_M, Pro-kuma_activ, IENR1, B41, among others. Reston isolates lack in otherwise well-conserved domain like B41, DDHD, Y1_Tnp, HOX, HOLI, PLCYc, Hr1, H4, GGDEF and LPD_N. VWC, a von Willebrand factor C domain is known to be involved with many developmental and pathological conditions via platelet activation [52]. However, role of this domain in infectious diseases is deficient, obliterating many critical links in pathogenesis. Table 3 contains the pertinent data.

There is considerable domain variation in this virus, even within isolates of same strain. Some regions of the polymerase protein are conserved, some are variable. By comparison of the two Reston isolates, it was seen that, up to CPSase_L_D3 domain at 1312–1361aa (i.e. 55–57 domains), the polymerase is conserved in both. Domain HisKA is present in Sudan isolate at 316–375aa and in Reston isolates at 2076–2138aa, while lacking in the Zaire strains. HisKA is a crucial sensor kinase in pathogens like bacteria [53], yet its absence in the aggressive Zaire strain seems enigmatic, which ought to be investigated.

Domain MIT, involved in microtubule manipulation is present in all Zaire isolates (at 2128–2199aa) and Sudan isolate (at 2125–2196aa isolate), while missing from Reston isolates. It might be another likely reason that Reston isolates cannot infect human.

Based on the findings, some investigation-worthy hypotheses have been made. The virus protein domain profiles and their functions revealed that the pathogenesis mechanism is not much different from other lethal viruses such as dengue. In this regard, drug repurposing to control Ebola virus seems pragmatic [54]. Limitation of this work

Table 3 Ebola strain-specific domains, based on the studied isolates

Strain	Outbreak country and year	Lacking domain	Peculiar domain
Zaire	Liberia-14	HisKA	ChtBD3 (in some)
	Kikwit-95		B41 (in all Zaire, but in a different position in *Q6V1Q2* (*Kikwit-95*), and *Q05318* (*Mayinga-76*))
	Mayinga-76		BAG (in *Q6V1Q2* (*Kikwit-95*), *Q05318* (*Mayinga-76*) only)
	Guinea-14		B_lectin (in *Mayinga-76*)
Sudan	Uganda-00	VWC, YARHG, WH1, RICTOR_M, Pro-kuma_activ, IENR1, B41	–
Reston	Reston-89	DDHD, B41, Y1_Tnp, HOX, HOLI, PLCYc, Hr1, H4, GGDEF and LPD_N. VWC	–
	Philippines-96		

The missing domains and the peculiar domains in the strains have been presented here. This information can be relevant in comprehending strain-specific pathogenesis mechanism. From this Table, it can be observed that Zaire strains have several domains unique to it. These domains might be responsible for aggressive nature of this strain

is that most of the analyzed isolates are from Zaire strain, and only few are from Sudan and Reston strains. Also, presence or absence of only a few domains have been discussed here, though based on the results, literature search can yield other relevant clues. Also, the work carried out here can be replicated with more Ebola virus polymerase and other protein sequences to garner further insights on pathogenicity determinants and strain-specific features.

This study furnished critical information regarding the polymerase protein domain diversity within the Ebola virus and related it to their variable virulence characteristics. The comparative analysis illuminated on many proteomic features of the lethal virus. It is clear that domain organization dictates virulence profile of different strains. Analyzing more isolates will eliminate inadvertent bias in interpretations. Presently, Ebola might be restricted to certain parts of the world, but the case fatality rate is highest among all pathogens at 90 %. In this regard, the work presented here is crucial in expanding our understanding of this Filovirus.

References

1. Lucas M, Karrer U, Lucas A, Klenerman P. Viral escape mechanisms–escapology taught by viruses. Int J Exp Pathol. 2001;82:269–86.
2. Martina BEE, Osterhaus ADME. "Filoviruses": a real pandemic threat? EMBO Mol Med. 2009;1:10–8.
3. Roca A, Afolabi MO, Saidu Y, Kampmann B. Ebola: a holistic approach is required to achieve effective management and control. J Allergy Clin Immunol. 2015;135:856–67.
4. Goeijenbier M, van Kampen JJA, Reusken CBEM, Koopmans MPG, van Gorp ECM. Ebola virus disease: a review on epidemiology, symptoms, treatment and pathogenesis. Neth J Med. 2014;72:442–8.
5. Thi EP, Mire CE, Ursic-Bedoya R, Geisbert JB, Lee ACH, Agans KN, et al. Marburg virus infection in nonhuman primates: Therapeutic treatment by lipid-encapsulated siRNA. Sci Transl Med. 2014;6:250ra116.
6. Lee JE, Saphire EO. Ebolavirus glycoprotein structure and mechanism of entry. Future Virol. 2009;4:621–35.
7. Frieden TR, Damon IK. Ebola in West Africa—CDC's role in epidemic detection, control, and prevention. Emerg Infect Dis. 2015;21:1897–905.
8. Feldmann H, Geisbert TW. Ebola haemorrhagic fever. Lancet. 2011;377:849–62.
9. Dixon MG, Schafer IJ. Ebola viral disease outbreak—West Africa, 2014. MMWR Morb Mortal Wkly Rep. 2014;63:548–51.
10. Victory KR, Coronado F, Ifono SO, Soropogui T, Dahl BA. Ebola transmission linked to a single traditional funeral ceremony—Kissidougou, Guinea, December 2014–January 2015. MMWR Morb Mortal Wkly Rep. 2015;64:386–8.
11. Gatherer D. The 2014 Ebola virus disease outbreak in West Africa. J Gen Virol. 2014;95:1619–24.
12. Kilgore PE, Grabenstein JD, Salim AM, Rybak M. Treatment of ebola virus disease. Pharmacotherapy. 2015;35:43–53.
13. Fischer RJ, Judson S, Miazgowicz K, Bushmaker T, Munster VJ. Ebola virus persistence in semen ex vivo. Emerg Infect Dis. 2016;22:289–91.
14. Calisher CH, Childs JE, Field HE, Holmes KV, Schountz T. Bats: important reservoir hosts of emerging viruses. Clin Microbiol Rev. 2006;19:531–45.
15. Pigott DM, Golding N, Mylne A, Huang Z, Henry AJ, Weiss DJ, et al. Mapping the zoonotic niche of Ebola virus disease in Africa. Elife. 2014;3:e04395.
16. Gonzalez JP, Pourrut X, Leroy E. Ebolavirus and other filoviruses. Curr Top Microbiol Immunol. 2007;315:363–87.
17. Marí Saéz A, Weiss S, Nowak K, Lapeyre V, Zimmermann F, Düx A, et al. Investigating the zoonotic origin of the West African Ebola epidemic. EMBO Mol Med. 2015;7:17–23.
18. Baker LE, Ellena JF, Handing KB, Derewenda U, Utepbergenov D, Engel DA, et al. Molecular architecture of the nucleoprotein C-terminal domain from the Ebola and Marburg viruses. Acta Crystallogr Sect D Struct Biol. 2016;72:49–58.
19. Wamala JF, Lukwago L, Malimbo M, Nguku P, Yoti Z, Musenero M, et al. Ebola hemorrhagic fever associated with novel virus strain, Uganda, 2007–2008. Emerg Infect Dis. 2010;16:1087–92.

20. Bausch DG, Schwarz L. Outbreak of ebola virus disease in Guinea: where ecology meets economy. PLoS Negl Trop Dis. 2014;8:e3056.

21. Dudas G, Rambaut A. Phylogenetic analysis of Guinea 2014 EBOV Ebolavirus outbreak. PLoS Curr. 2014;2014:6.

22. Na W, Park N, Yeom M, Song D. Ebola outbreak in Western Africa 2014: what is going on with Ebola virus? Clin Exp Vaccine Res. 2015;4:17–22.

23. Wauquier N, Becquart P, Padilla C, Baize S, Leroy EM. Human fatal zaire Ebola virus infection is associated with an aberrant innate immunity and with massive lymphocyte apoptosis. PLoS Negl Trop Dis. 2010;4:e837. doi:10.1371/journal.pntd.0000837

24. Ponting CP, Schultz J, Milpetz F, Bork P. SMART: identification and annotation of domains from signalling and extracellular protein sequences. Nucleic Acids Res. 1999;27:229–32.

25. Ilinykh PA, Lubaki NM, Widen SG, Renn LA, Theisen TC, Rabin RL, et al. Different temporal effects of Ebola virus VP35 and VP24 proteins on global gene expression in human dendritic cells. J Virol. 2015;89:7567–83.

26. Mühlberger E. Filovirus replication and transcription. Future Virol. 2007;2:205–15.

27. Lu J, Qu Y, Liu Y, Jambusaria R, Han Z, Ruthel G, et al. Host IQGAP1 and Ebola virus VP40 interactions facilitate virus-like particle egress. J Virol. 2013;87:7777–80.

28. Mohan GS, Li W, Ye L, Compans RW, Yang C. Antigenic subversion: a novel mechanism of host immune evasion by Ebola virus. PLoS Pathog. 2012;8:e1003065.

29. Sullivan N, Yang Z-Y, Nabel GJ. Ebola virus pathogenesis: implications for vaccines and therapies. J Virol. 2003;77:9733–7.

30. Wang H, Shi Y, Song J, Qi J, Lu G, Yan J, et al. Ebola viral glycoprotein bound to its endosomal receptor Niemann-Pick C1. Cell. 2016;164:258–68.

31. McElroy AK, Spiropoulou CF. Biomarkers for understanding Ebola virus disease. Biomark Med. 2014;8:1053–6.

32. Stoermer KA, Morrison TE. Complement and viral pathogenesis. Virology. 2011;411:362–73.

33. Uyeki TM, Mehta AK, Davey RT, Liddell AM, Wolf T, Vetter P, et al. Clinical management of Ebola virus disease in the United States and Europe. N Engl J Med. 2016;374:636–46.

34. West TE, von von Saint André-von Arnim A. Clinical presentation and management of severe Ebola virus disease. Ann Am Thorac Soc. 2014;11:1341–50.

35. Funk DJ, Kumar A. Ebola virus disease: an update for anesthesiologists and intensivists. Can J Anaesth. 2015;62:80–91.

36. Jahrling PB, Hensley LE, Barrett K, Lane HC, Davey RT. State-of-the-art workshops on medical countermeasures potentially available for human use following accidental exposures to Ebola virus. J Infect Dis. 2015;212(Suppl):S84–90.

37. Davidson E, Bryan C, Fong RH, Barnes T, Pfaff JM, Mabila M, et al. Mechanism of binding to Ebola virus glycoprotein by the ZMapp, ZMAb, and MB-003 cocktail antibodies. J Virol. 2015;89:10982–92.

38. Qiu X, Wong G, Audet J, Bello A, Fernando L, Alimonti JB, et al. Reversion of advanced Ebola virus disease in nonhuman primates with ZMapp. Nature. 2014;514:47–53.

39. Choi WY, Hong K-J, Hong JE, Lee W-J. Progress of vaccine and drug development for Ebola preparedness. Clin Exp Vaccine Res. 2015;4:11–6.

40. Oestereich L, Lüdtke A, Wurr S, Rieger T, Muñoz-Fontela C, Günther S. Successful treatment of advanced Ebola virus infection with T-705 (favipiravir) in a small animal model. Antiviral Res. 2014;105:17–21.

41. Crotty S, Andino R. Implications of high RNA virus mutation rates: lethal mutagenesis and the antiviral drug ribavirin. Microbes Infect. 2002;4:1301–7.

42. Westover JB, Sefing EJ, Bailey KW, Van Wettere AJ, Jung K-H, Dagley A, et al. Low-dose ribavirin potentiates the antiviral activity of favipiravir against hemorrhagic fever viruses. Antiviral Res. 2016;126:62–8.

43. Warren TK, Wells J, Panchal RG, Stuthman KS, Garza NL, Van Tongeren SA, et al. Protection against filovirus diseases by a novel broad-spectrum nucleoside analogue BCX4430. Nature. 2014;508:402–5.

44. Warren TK, Whitehouse CA, Wells J, Welch L, Heald AE, Charleston JS, et al. A single phosphorodiamidate morpholino oligomer targeting VP24 protects rhesus monkeys against lethal Ebola virus infection. MBio. 2015;6:e02344-14.

45. van Griensven J, De Weiggheleire A, Delamou A, Smith PG, Edwards T, Vandekerckhove P, et al. The use of Ebola convalescent plasma to treat Ebola virus disease in resource-constrained settings: a perspective from the field. Clin Infect Dis. 2016;62:69–74.

46. The UniProt Consortium. The universal protein resource (UniProt). Nucleic Acids Res. 2008;36:D190–5.

47. Suen S, Lu HH-S, Yeang C-H. Evolution of domain architectures and catalytic functions of enzymes in metabolic systems. Genome Biol Evol. 2012;4:976–93.

48. Inoue H, Baba T, Sato S, Ohtsuki R, Takemori A, Watanabe T, et al. Roles of SAM and DDHD domains in mammalian intracellular phospholipase A1 KIAA0725p. Biochim Biophys Acta. 2012;1823:930–9.

49. MacNeil AJ, Pohajdak B, Australasian Society for Immunology Inc. Getting a GRASP on CASP: properties and role of the cytohesin-associated scaffolding protein in immunity. Immunol Cell Biol. 2009;87:72–80. doi:10.1038/icb.2008.71.

50. Bracher A, Verghese J. GrpE, Hsp110/Grp170, HspBP1/Sil1 and BAG domain proteins: nucleotide exchange factors for Hsp70 molecular chaperones. Subcell Biochem. 2015;78:1–33.

51. Messina JP, Brady OJ, Scott TW, Zou C, Pigott DM, Duda KA, et al. Global spread of dengue virus types: mapping the 70 year history. Trends Microbiol. 2014;22:138–46.

52. James PD, Goodeve AC. von Willebrand disease. Genet Med. 2011;13:365–76.

53. Willett JW, Kirby JR. Genetic and biochemical dissection of a HisKA domain identifies residues required exclusively for kinase and phosphatase activities. PLoS Genet. 2012;8:e1003084.

54. Zhao Z, Martin C, Fan R, Bourne PE, Xie L. Drug repurposing to target Ebola virus replication and virulence using structural systems pharmacology. BMC Bioinform. 2016;17:90.