Journal of Health Study and Medicine
2016, nr 4
ss. 5-35

The deadly octet: the structure and function of the Ebola virus proteins

Malwina Radwańska[1,2], Zygmunt Derewenda[1]

[1] Department of Molecular Physiology and Biological Physics, University of Virginia School of Medicine, Charlottesville, VA 22908-0736, USA
[2] Department of Chemistry, University of Technology, Wroclaw

Abstract
The Ebola hemorrhagic fever epidemic of 2014 is a powerful reminder of the threats posed by viral diseases, as their geographical range and reservoirs are altered by global warming and other factors. Viruses contain very small genomes, encoding only few proteins in the simplest viruses (eight in the case of the Ebola virus), but the multifunctionality of these proteins ensures that the virus can accomplish a multitude of biological functions, including cell entry, membrane fusion, replication, suppression of immune response, packaging of new virions and exit from the cell. In this short review we summarize the recent progress in the understanding of how the eight proteins encoded by the Ebola RNA accomplish these specific functions at the structural level.

Key words
ebola virus, hemorrhagic fever, protein structure, macromolecular crystallography, structural biology

Introduction

The Ebola virus (EBOV) causes a horrifying hemorrhagic fever in humans and other primates with fatality rate up to 90% [1]. It was first described in 1976 as a distant cousin of the Marburg virus (MARV) [2], although it has been hypothesized that an outbreak occurred in Athens in 430 BC [3]. In spite of the fact that EBOV is very dangerous, there was only a relatively modest effort over the past decades to elucidate its biology and to design therapeutics or vaccines. This situation was rationalized by the fact that EBOV caused only sporadic and relatively limited epidemics during the last few decades, and all were endemic to some of the poorest countries in the world, in Central and Eastern Africa. Consequently, there was no financial incentive for pharmaceutical companies to develop treatments and vaccines in the absence of a significant market [4]. The perception of EBOV as a limited and endemic threat changed dramatically in early 2014 with the West African outbreak of the disease, notably when a small number of infected individuals traveled to western countries, including the USA, spreading panic [5]. These few cases were prominently featured in the media, overshadowing the real epidemic and the fact that a record 29,000 cases were reported at the end of the outbreak in Africa, with more than 11,000 deaths [6]. The hitherto unanticipated threat to the Western countries stimulated a surge in research and funding opportunities and rapid progress was achieved in the characterization of the virus and its components, in the design of therapeutic antibodies [7] and — most recently — introduction of an effective vaccine[8]. The search for drugs able to deal with the virus after infection is still ongoing.

EBOV has a cylindrical or tubular-shaped, non-segmented morphology, characteristic of all members of *Filoviridae* family [9]. There are five known strains Zaire, Sudan, Taï Forest, Bundibugyo and Reston [10]. EBOV is a single-strand RNA (ssRNA) virus, containing a non-coding (negative) strand, ~19 kB in size. The virus is covered by a lipid envelope with protruding spikes on the surface, as visualized by electron microscopy. The ssRNA is packaged inside the envelope within a helical nucleocapsid, made up of proteins and ssRNA [11].

The life cycle of the EBOV is reasonably well understood. Fruit bats are thought to serve as the primary reservoir for the virus across all tropical areas of the world [12]. They are not harmed by the virus. EBOV also infects other animal species, although the details are still scarce. Eating

bush meat — including bats — is thought to be the principal mechanism for infection of humans [13], with subsequent epidemics resulting from contact with bodily fluids of the infected or dead individuals. Once inside the human body, EBOV targets specific cells, including those in the liver and immune system, as well as endothelial cells. Relatively non-specific receptors on those cells recognize and bind either carbohydrates which decorate the viral surface, or PtdSer (phosphatidylserine) which is an element of the viral envelope [14]. The adherence to cell surface triggers internalization of virus particles through the process of macropinocytosis, as well as clathrin-dependent endocytosis [15]. Once the virus reaches endosomes, it undergoes complex proteolysis of the surface glycoprotein, its membrane fuses with that of the attacked cell and the contents enter the cytoplasm. The replication process involves the synthesis of the positive strand of RNA which serves as a template for new RNA genomes [16]. During transcription, the negative RNA strand is transcribed into seven monocistronic mRNAs, corresponding to the seven genes encoded by the RNA, each capped and polyadenylated. The viral proteins are synthesized by hijacking the ribosome machinery of the infected cells, and accomplish their diverse biological functions owing to exceptional multifunctionality. They are responsible for invasion, membrane fusion, suppression and evasion of the immune response, transcription, replication and packaging of new virions, and exiting the cell. This process is so effective, that a healthy man infected with a single virus can die as a result of internal hemorrhage within 6-16 days of developing symptoms, typically no more than 21 days after infection. At death, the viral titers in blood (viremia) may exceed 10^9 copies per ml [17].

This short review summarizes the most important aspects of the recent progress achieved towards the understanding of the structure-function relationships among the proteins encoded by the EBOV viral RNA. Space constraints preclude a fully comprehensive presentation, and our review is by necessity limited to only a fraction of the available material; we apologize for omissions. We start with an overview of the biological functions served by the protein products of the seven EBOV genes.

The biological functions of the EBOV proteins.

The order of the seven genes in the EBOV ssRNA is as follows: NP – VP35 – VP40 – GP– VP30 – VP24 – L (with the exceptions of NP, GP and L, the

genes are denoted as **V**iral **P**rotein (VP) and followed by a number indicating molecular weight) (Fig. 1). Once the virions are packaged, multiple copies of seven of the proteins are included, each found in a distinct location. These proteins are known as 'structural proteins' as they are part of the virus particle. The eighth protein, an alternate product of the GP gene, is secreted by infected cells and not found within the virus; it is the only non-structural protein.

The **glycoprotein** (GP) is the only exposed protein on the viral surface, visible in electron micrographs as 7-10 nm long spikes spaced at approximately 10 nm, and protruding from the lipid bilayer [18]. It mediates attachment to specific cell receptors, and enables the virus to enter the cell through endocytosis into macropinosomes. Interestingly, this is the only gene that has two alternate transcripts; the second transcript leads to the expression of the **soluble GP (sGP)**, the secreted, and non-structural protein [19]. The abundance of this protein, which is similar to GP, confuses the immune responses of the host and acts as a decoy protecting the virus [20].

VP40 is a matrix protein, linking the inside layer of the membrane with the nucleocapsid, and is one of the most versatile protein in the EBOV proteome. Aside from its structural function, it plays a role in the regulation of transcription, morphogenesis of the virus, as well as packaging and budding of mature virions. It is also implicated in suppression of RNA silencing, although the mechanism is not clear [21].

The remaining 'structural proteins', i.e. nucleoprotein **(NP)**, **VP35**, **VP40**, **VP30** and **VP24**, along with the viral **L (large) polymerase**, are contained within the nucleocapsid which also contains helically packaged ssRNA. The regular structure of the nucleocapsid has been elucidated by 3D sub-tomographic analysis [11]. There are nearly equimolar amounts of the VP24, VP30, VP35 and NP proteins in the nucleocapsid. There is evidence that VP24 as well as VP35 interact with NP, and all three proteins are necessary for the formation of the nucleocapsid [11]. The key protein for RNA packaging is the nucleoprotein (NP) [22]. **VP30** protein is a transcription activator, which plays an important role in the virus replication process [23]. In contrast, both **VP24** and **VP35** are intimately involved in blocking the host's interferon-mediated immune response, through several diverse and complementary mechanisms. VP35 inhibits induction of the α and β interferons by blocking phosphorylation and activation of the

interferon regulatory factor 3 (IRF3), a transcription factor which regulates interferon synthesis [24]. Moreover, VP35 also interferes with the activity of the RIG-I like receptors (retinoic acid inducible gene I receptors), which under normal conditions detect viral RNA during infection and replication, and initiate interferon response [25]. Another immune response pathway blocked by VP35 is mediated by the EIF2AK2/PKR kinase (eukaryotic translation initiation factor 2 alpha kinase 2/protein kinase R) [26]. Finally, VP 35 is also a cofactor of the L polymerase, and thus plays a role in transcription and replication [27]. The complementary and independent action of VP24 is to block the already existing interferon response, by binding to karyopherin α, a nuclear transporter protein, which — when activated by interferon — carries thetyrosine-phosphorylated transcription factor STAT1 (PY-STAT1) to the nucleus, and thus stops the immune response downstream of interferon [28].

Last, but not least, the EBOV ssRNA encodes an RNA dependent **RNA polymerase**, or the **L protein** [29]. The enzyme is multifunctional, with RNA-directed RNA polymerase, mRNA guanylyl transferase, mRNA (guanine-N(7)-)-methyltransferase and poly(A) synthetase activities. It functions either as transcriptase, generating the mRNAs required for protein production, or as replicase, copying the template positive strands into negative strands of RNA. The transcriptase synthesizes subgenomic RNAs, assuring their capping and polyadenylation. The replicase mode is switched on at a particular intracellular concentration of NP. In this mode, the L polymerase replicates the whole viral genome without recognizing the transcriptional signals.

Structure and mechanism of the EBOV proteins

How are the eight EBOV proteins able to accomplish all their deadly functions? Many of the answers can be deduced from the molecular structures of the proteins, and the structures of their complexes with viral and host partner proteins, and with RNA. We will now review the current state of knowledge of the structural biology of EBOV proteins, mostly derived from X-ray crystallographic studies. Much of the information is very recent, emerging after the 2014 outbreak.

The EBOV glycoprotein (GP). GP is a 676-residue class I membrane glycoprotein responsible for viral entry into the cell and fusion with the host

membrane (Fig. 2A). It is not a product of direct transcript of the GP gene; instead, the corresponding mRNA results when the L polymerase stutters on a specific sequence, adding one more adenine and extending the open reading frame past the first 295 residues (there is also a third product, the small soluble GP, which, for lack of space, we do not describe further) [30].

GP is arguably the most complex of all the EBOV proteins. It is expressed in infected cells as a precursor polypeptide (GP0) (Fig. 2B), which is then cleaved by the human furin endoprotease into two disulfide-linked chains: the receptor binding subunit, GP1, residues 1-501 (~130 kDa, when fully-glycosylated) and the fusion subunit GP2, residues 502-676 (~20 kDa, non-glycosylated) (Fig 2A). In the maturing EBOV particles the GP1/2 forms trimers inserted into the membrane; these are the spikes visible in electron micrographson on the viral surface (Fig 2 E-F). The first insight into the atomic structure of this ensemble came from the crystal structure of GP1/2 bound to a Fab fragment of a human antibody, elucidated by X-ray crystallography to a resolution of 3.4 Å [18] (the GP1 in this study was modified to remove the mucin-like domain, MLD, which is heavily glycosylated and interferes with crystallization, as well as the trans membrane helical anchor). Further studies using cryo-electron tomography provided information about the localization of the MLD in the full-length protein [31]. More recently, the structure of the intact sGP dimer in complex with two Fab variable antibody domains were characterized by single-particle cryo-EM methods at 5.5Å resolution (Fig. 2D) [32]. Moreover, these studies revealed the precise domain architecture of the surface glycoprotein spikes and rationalized some of the functions. The biological GP1/2 trimer assumes a chalice-like shape with GP2 serving as the base and GP1 as the cup (Fig 2F). GP2 contains the fusion loop (IFL) which is critical for viral entry into the cytoplasm (see below), but remains hidden on the surface of a free virus; and the C-terminal helical fragment which serves as an anchor in the viral membrane. The GP1 is made up of four distinct domains (Fig. 2A): the globular base, the receptor-binding domain (RBD), the glycan cap and the MLD which is partly disordered and heavily modified by O- and N-linked glycosylation. In contrast, the glycan cap is more structured than MLD, and contains four N-linked glycosylation sites. MLD domains are located at the apex of the GP spikes, and are associated with multiple functions including enhance-

ment of viral attachment to host cell surfaces and protection of GP protein regions from antibody recognition.

After the uptake of the virus into macropinosomes and subsequent transport into endosomes, GP1 undergoes another proteolytic cleavage by the cathepsins B and L proteases (catB/catL). The cleavage site removes the glycan cap as well as the MLD, generating a fusion-competent trimer of heterodimers: the ~19 kDa N-terminal fragment of GP1 and intact GP2 (GPCL). The structure of the GPCL has been recently visualized by X-ray crystallography at 3.3 Å resolution (Fig. 2F) [33]. Proteolysis has no significant impact on the structure of GPCL compared to the pre-cleavage ensemble, but importantly it uncovers a unique binding site in GP1 RBD which allows for recognition of the host endosomal receptor, the Niemann-Pick C1 (NPC1) [34, 35]. Very recently, a single-particle cryo-EM study visualized the molecular basis of the interaction of the GPCL with NPC1 at 6.6 Å resolution [36]. The same study showed that the affinity of GP_{CL} for NPC1 is greatly enhanced at pH 6.0compared to pH 7.5, which explains the binding in the acidic environment of the endosomes [36]. Further atomic details of the interaction were visualized by the crystal structure of GP_{CL} bound to the isolated C-domain of NPC1 at 2.3 Å resolution (Fig. 2C) [37].Following the binding of GPCL to NPC1, the GP2 subunit undergoes a conformational change, leading to exposure of the internal fusion loop (IFL). The loop has an architecture of an anti-parallel, β-stranded scaffold with a centrally located ensemble of hydrophobic residues (Leu529, Trp531, Ile532, Pro533, Tyr534, and Phe535). Studies using nuclear magnetic resonance (NMR) revealed that at low pH (i.e. in endosomes) the fusion loop adopts a conformation in which the above residues become exposed and form a hydrophobic 'fist' that facilitates penetration into the membrane [38, 39]. This event, in turn, initiates fusion of the host and viral membranes, which is catalyzed by the energetically and low-pH favored conformational rearrangement of two α-helical structural elements of GP2. These helices from the three GP2 units in the trimer come together to form 6-helix bundle stabilized at low pH, overcoming the kinetic energy barrier of fusion of two membranes; with fusion completed, the contents of the viral particle enter the cell cytosol [38, 39].

The EBOV Matrix Protein (VP40). This is the most abundant protein in EBOV particles, playing a complex role in assembly and budding of

the virus-like particles (VLPs) from the cell membrane [40]. Like other viral matrix proteins, VP40 provides a link between the lipid membrane and the nucleocapsid core, having the capability to associate with both. This protein does not undergo any proteolytic conversion, but instead is characterized by unique conformational malleability, a prerequisite for formation of diverse oligomeric structures that underlie the protein's multifunctionality [41]. The VP40 is synthesized as a 326 residue long monomer, composed of two loosely associated globular domains (N- and C-terminal domains, i.e. NTD and CTD, respectively) joined by a short flexible linker (Fig. 3A). Both are involved in trafficking to and interactions with membranes [42]. At the very N-terminus, VP40 contains a disordered stretch of ~40 amino acids that contain a PPXY motif and a PTAP motif [43]. Initially, it has been thought that VP40 exists as a monomer in solution, as suggested by the first determined crystal structure [44]. More recent studies generated compelling evidence that the protein is dimeric, and this association is mediated by the NTD (Fig. 3B) [41, 45]. The dimeric structure is required for trafficking of VP40 to the cell membrane. There are two structural features on CTD which are important for association with membrane and subsequent conformational changes that drive VP40 oligomerization. One is the hydrophobic loop composed of Ile293, Leu295 and Val298 residues, which was shown not only to bind to, but actually penetrate the plasma membrane, assisting in the localization of VP40 to the membrane, release of VLPs and exit of viral particle [46]. The other element includes a conserved basic patch (Lys274 and Lys275), located within a disordered loop at the opposite site of the CTD, also shown to mediate VLPs release [41]. In addition, CTD also promotes trafficking of VP40, owing to its interactions with the Sec24C protein, a member of the coat protein complex II (COPII) associated with vesicular transport system [47].

Upon interaction with the membrane VP40 undergoes conformational change resulting in spatial displacement of NTD from CTD, made possible owing to the flexibility of the linker [48]. This leads to the oligomerization of VP40 and formation of linear hexamers. The crystal structure of VP40 hexamer revealed the oligomerization core made up of NTDs derived from four central protomers, flanked by remaining two protomers (Fig. 3C) [41]. At this point, conformational changes trigger displacement of four CTDs from central core in opposing directions (two are

upwards and two are downwards). The VP40 hexamers may further associate through their CTDs, forming long, interrupted filaments [41].

To add to the complexity of the VP40 oligomerization pattern, it was also shown that the protein may octamerize into a ring with RNA-binding properties, notably through the NTD domains (Fig. 3D). This was initially visualized in a crystal structure of an octameric ring of isolated NTD domains with bound RNA trinucleotide [49]. Interestingly, RNA does not appear to be required for this assembly, as evidenced by the latest structure of RNA-free ring of isolated NTDs. The formation of the ring is made possible through a conformational change that involves the displacement of N-terminal 69 residues from dimeric interface, which in turn exposes the RNA-binding site. In this context, RNA plays a critical role in separation of the CTD from NTD that triggers octamerization, but not in the ring assembly itself. The VP40 ring has a role in regulation of viral transcription inside infected cells and this function is dependent on its RNA-binding capability [41].

Finally, VP40 is also known to interact with selected cellular proteins. Two of them, Nedd4 (ubiquitin ligase) and Tsg101 (a regulator of vesicular trafficking), have been found to bind the sequence elements in the N-terminal motif, i.e. PPXY and PT/SAP, respectively [50]. These interactions indicate a complex way in which the budding and exit of EBOV from the cell are regulated. There is currently no structural information about the underlying mechanisms.

The EBOV nucleoprotein (NP). NP is a large protein (739 residues) which is critical for ssRNA packaging. Together, NP, VP24 and VP35 are sufficient for the formation of nucleocapsids that are morphologically identical to those formed in infected cells [22]. On the other hand, a complex of ssRNA, NP, VP30 VP35 and the L polymerase constitutes the polymerase complex, sufficient for replication and transcription of the viral RNA.

NP contains two distinct globular domains (Fig. 4A) [22, 51] and is subject to post-translational modifications, i.e. O-glycosylation and sialylation [22]. Significant information has emerged recently about the molecular structures of the two globular moieties within the NP. The isolated RNA-packaging N-terminal domain has been crystallized and the structure was solved at 1.8 Å resolution [52]. The domain shows a dumbbell like structure with the N- and C-terminal lobes separated by a positive-

ly-charged groove that is presumed to constitute the RNA binding site, where the protein binds the phosphate backbone of the helical RNA [52]. However, to date there is no detailed structure known of any complex of EBOV NP with RNA. The independently determined structure of a similar NTD construct in complex with a peptide derived from VP35 (residues 20-48), showed how the latter occludes a large portion of the RNA-binding surface, and prevents newly synthesized NP in the cell from interactions with non-cognate RNA (Fig. 4B) [53].This explains the regulatory function of the VP35 protein at the molecular level and hints at the possible way the proteins are associated in the replication complex.

The smaller, C-terminal fragment of the EBOV NP is a major antigenic determinant, raising the possibility that it could be effective in virus detection and diagnostics [54]. The isolated C-terminal domain from several strains has been crystallized and the structures were solved at high-resolution (Fig. 4C) [55, 56]. These studies revealed a new fold, hitherto unseen in any protein, although the structure itself did not hint at specific functional properties.

The two globular domains of NP are joined by a disordered and highly flexible linker that endows the protein with plasticity, and serves as an attachment point for other protein. We will discuss this in the section focusing on VP30. The interactions with both VP35 and VP30 point to NP as providing a scaffold for the assembly of the replication complex.

Finally, there is the question of possible interactions of NP with cellular proteins. A recent study of the cellular interactome of EBOV NP strongly hints at a chaperone HSP70, as a protein that modulates stability and NP may assist it in its physiological functions [57]. The structural basis of this interaction are not known.

The EBOV VP30. This protein is made up of 288 amino acids, and — like VP40 — has the ability to form oligomers. It contains several motifs important for its physiological function, including a Zn-finger domain (the site for binding Zn^{2+} ions, essential for the interaction with RNA [58]), a hexamerization domain and a unique, globular C-terminal domain (CTD) which forms a tight dimer in isolation (Fig. 5A). The domain, encompassing residues 114-265, was crystallized and the structure was determined at 2.0 Å resolution [59]. The domain contains a core of six α-helices; there is an additional, seventh helix which crosses over and 'embraces'

the second domain in the dimer, ensuring a large interface between monomers (Fig. 5B). Recent studies revealed how this protein is anchored to the nucleocapsid through an interaction with NP, and specifically with a Pro-rich peptide which resides within the linker between the two NP domains [60]. The structures of the Zaire and Sudan VP30 with a bound peptide (residues 602-614) from the NP protein were recently published, and show how the NP peptide is bound within a long, shallow cavity on the face of the monomer distal to the homodimeric interface. The integrity of this interaction was shown to be critical to the regulation of the viral RNA synthesis.

At this moment there is no other structural information about VP30 and in particular there is no structural data on the interaction with viral RNA, except for experimental evidence that a basic sequence residing between residues 26 and 40 is involved in addition to the Zn-finger motif [58]. The interaction between VP30 and RNA is necessary for the activation of transcription [23, 61].

The VP24. This is the smallest product of EBOV genome, comprised of 251 residues, folded into a single globular domain (Fig. 6A). Significant information has recently emerged about the structure of VP24 and the structural basis of its interaction with karyopherin-α (KPNA), an interesting protein containing so-called armadillo repeats. Crystal structures of Sudan and Reston VP24 protein revealed a compact, single α/β domain architecture, which is characterized by a unique pyramidal fold (Fig. 6B). There are three distinct faces of this triangular molecular pyramid. The top of the pyramid is made up of several α-helices and a small, three-stranded antiparallel β-sheet, encircled by the N-terminus, while the interior contains a five-stranded antiparallel β-sheet, resting on several more α-helices that make up the bottom. One of the faces with a distinct hydrophobic characteristics mediates an intimate interaction with the KPNA, as visualized by a crystal structure of the complex of VP24 with the C-terminal domain of KPNA (containing the armadillo repeats 8-10) at 3.15 Å resolution (Fig. 6C) [62]. The consequence of this interaction is that a unique, non-classical nuclear localization signal (ncNLS) binding site, necessary to anchor PY-STAT1 to KPNA, is blocked. Thus, KPNA is rendered incapable of transporting the latter transcription factor to the nucleus, and this stops the signaling pathway that is activated by interferon.

VP24 is also known as the secondary matrix protein, as some of its functions tend to overlap with that of VP40, i.e. it is essential for virion assembly and budding, and has the ability to associate strongly with lipid bilayers. Also, the interaction of VP24 and NP, not yet characterized, is considered to be important factor in the assembly and function of the viral ribonucleoprotein complex (RNP) in addition to its likely association with membranes. The structural basis for these phenomena is not understood.

The VP35. Like VP24, VP35 is the key player in the suppression of the interferon-mediated immune defenses, through a number of mechanisms (see above). VP35 is a 340-residue long protein with three distinct elements: a nucleoprotein targeting N-terminal peptide, a coiled-coil motif responsible for the oligomerization of the protein, and the C-terminal globular domain, responsible for the interferon inhibition, through interaction with dsRNA (double-stranded RNA) (Fig. 7A). Although no structure of the coiled-coil domain from the EBOV VP35 is available, two crystal structures of this domain from the Marburg virus protein have been recently deposited in the Protein Data Bank (www.rcsb.org/pdb) (5TOH, 5TOI), although no publication has yet appeared. The structures show a trimeric parallel architecture, as expected from the earlier studies, and explain how the protein oligomerizes. The crystal structure of the C-terminal domain (residues 221-34) has been characterized for the Zaire VP35 [63]. The domain is organized into two modules, or sub-domains, i.e. a four-helix bundle at the N-terminus and a five-stranded □-sheet at the C-terminus. This is a unique, hitherto unseen fold for an RNA binding protein. Importantly, this structure revealed a cluster of basic residues with Arg312, known to be critical for the dsRNA binding. The crystal structure of the complex of two of these domains with an 8-base-pair dsRNA, determined at 2.0 Å resolution, shows exactly how the protein coats the dsRNA, to prevent its recognition by the host proteins (Fig. 7B) [64]. Specifically, one domain binds to the dsRNA backbone through the basic patch, which the other coats the blunt--end of RNA with a hydrophobic patch on its surface (Fig. 7C). The mutual disposition of the two domains is stabilized by a direct interaction between them, involving several conserved polar residues (Fig. 7D).

VP35 also plays a structural role in the assembly of the nucleocapsid through its N-terminal peptide, which binds with high affinity to the N-

-terminal domain of NP [53]. This interaction regulates NP assembly and viral genome binding, as in the absence of VP35 NP undergoes self-assembly.

The L polymerase. At the time of the writing of this review, the structure of the L protein, RNA-dependent RNA polymerase from the Ebola virus is still unknown.

Conclusion

The new wealth of structural information regarding the Ebola virus proteins paves the way for the future development of small molecule drugs that could provide effective antiviral remedies, complementing the vaccines and therapeutic antibodies. The so-called structure-based drug design depends on the availability of high resolution structures of target proteins in the search of small molecules that bind to specific pockets on the surface, inhibiting the function. Many such efforts are underway, targeting a number of viral proteins, including GP [65, 66], VP35 [67] and VP40 [68].

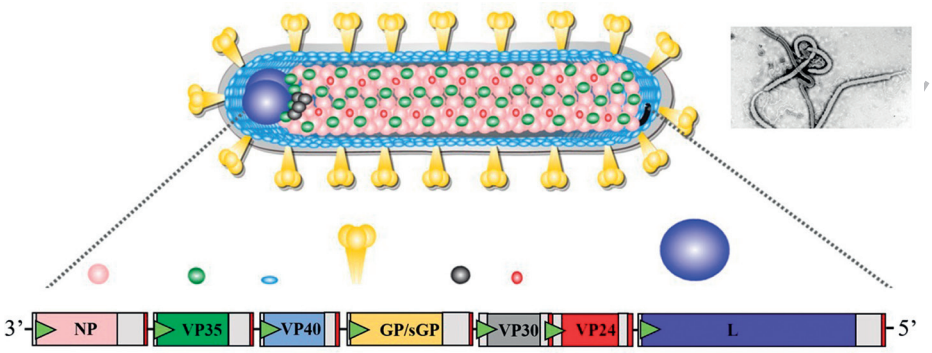

Figure 1. Ebola virus structure and genome organization.

Schematic representation of the Ebola virus and its ssRNA genome with the map of the genes (based on reference [69]). An electron micrograph of the virus is in the upper right corner (reproduced from http://www.ebolavirusnet.com/ebola-virus.html with permission). The colors of the open reading frames correspond to the colors used to depict the viral proteins in the digram. Untranslated regions of the given genes are represented as grey boxes. Transcription start signals are shown as green triangles, while red bars indicate transcription stop signals.

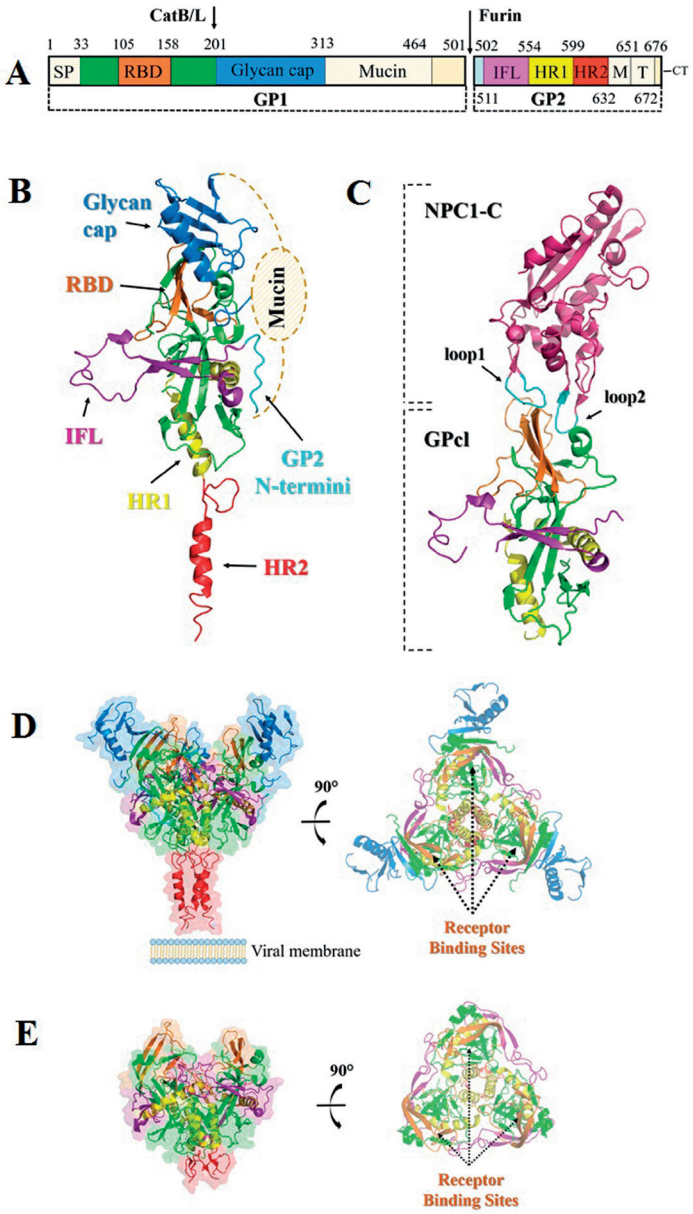

Figure 2. The structure of the EBOV glycoprotein GP

(A) Domain structure of EBOV GP (adapted from reference [31]).Wheat and white regions correspond to disordered or conctruct-deleted regions respectively. Arrows indicate cleavage sites for furin and cathepsin B/L. Disulfide bonds and glycosylation sites are

omitted for clarity. SP – signal peptide, RBD – receptor binding domain, Mucin – Mucin-like domain (further denoted as MLD); IFL – internal fusion loop, HR1 – first heptad repeat, HR2 – second heptad repeat, M – membrane – proximal external region, T – transmembrane domain, CT – cytoplasmic tail.

(B) Cartoon representation of Zaire EBOV $GP_{\Delta MLD\Delta TM}$GP0 monomer (PDB entry 5JQ3). The structural elements are colored accordingly to the domain scheme in A. Yellow oval corresponds to the mucin domain which was removed in the construct used for crystallization purposes.

(C) The structure of NPC1-C bound to Zaire EBOV GPcl (PDB entry 5F1B). The NPC1-C (colored in raspberry) binds to the RBD of GPcl at a perpendicular angle through its two protruding loops colored in cyan. Structural elements of GPcl are colored as in (B).

(D) Cryo-EM structure of Zaire EBOV sGP dimer in complex with c13C6 and BDBV91 Fabs (PDB entry 5KEM). sGP dimer (sGPA, sGPB) is colored as in (B) with C13C6 Fab in grey and BDBV91 Fab in brown. There are two copies of each Fab bound to the sGP dimer.

(E) Cartoon representation of the biological trimer of Zaire EBOV GP1/2 heterodimers (PDB entry 5JQ3) viewed perpendicular to the threefold axis (left) with a semitransparent surface, and with semitransparent rendering along (right) the threefold axis, towards the viral membrane. The structural elements are colored as in (B). Arrows indicate receptor binding sites.

(F) Receptor binding-competent conformation of Zaire EBOV GPcl (PDB entry 5HJ3) modeled in the same way as (E).

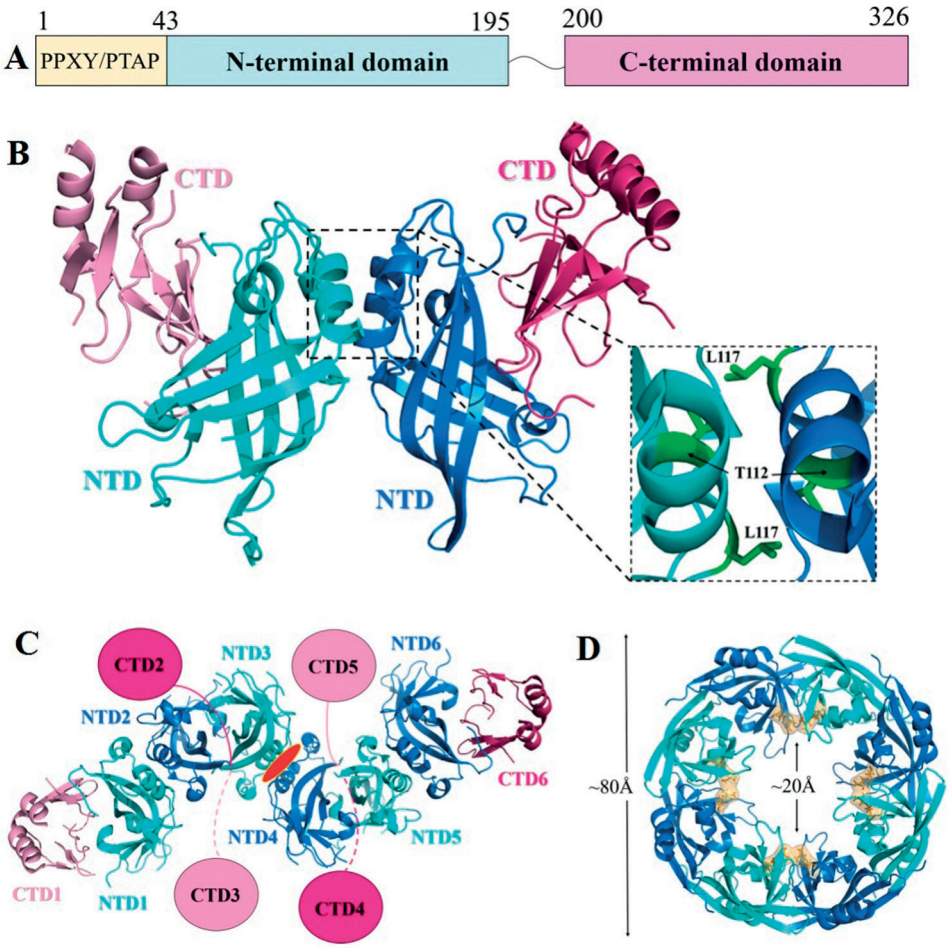

Figure 3. The structure of VP40.

(A) Schematic representation of the EBOV VP40 domain structure showing also the N--terminal stretch of 43residues containing PPXY/ PTAP motifs (yellow) as well as the NTD (cyan) and CTD (pink). Waved line indicates the disordered flexible linker connecting NTD and CTD.

(B) Crystal structure of dimeric Zaire EBOV VP40ΔN showing NTD-to-NTD interface (PDB entry 4LDB). NTDs from each protomer are colored cyan or blue and the CTDs colored light pink or raspberry. A dashed box shows close-up view of the dimeric interface, centered on Thr112 and Leu117 residues (green). The construct used for crystallization lacks the first 43 N-terminal residues.

(C) Crystal structure of Zaire EBOV VP40ΔN hexamer displayed with the NTDs alter-

nating in colors (cyan or blue) and the CTDs alternate in pink or raspberry (PDB entry 4LDD) and oligomerization interface centered on Trp95 residue (red oval). The NTDs and CTDs are numbered corresponding to the protomer from which they derive. The displaced CTDs are showed as circles (to scale).
(D) Crystal structure of RNA-free Zaire EBOV VP40 octameric ring (PDB entry 4LDM) with NTD protomers colored in alternating cyan and blue. Residue Arg134 is depicted by orange sticks with semitransparent surface.

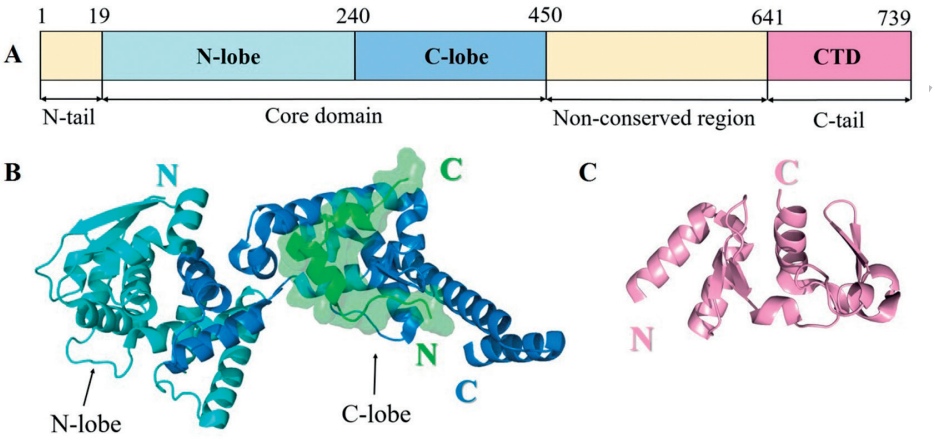

Figure 4. The structure of NP.

(A) Domain architecture of EBOV NP showing the location of the N-terminal core domain with N-lobe and C-lobe colored in cyan and blue, respectively and the CTD colored in pink. The disordered N-terminal peptide and non-conserved regions are colored in yellow.

(B) Crystal structure of Zaire EBOV NP N-terminal core domain (ΔNP_{NTD}; residues 39-384) in complex with peptide derived from EBOV VP35 (PDB entry 4YPI). The NP N-lobe (residues 37–146, cyan) contains a flexible hinge region (residues 147–239) that attaches to the 240–285 residues of the C-lobe (blue). The peptide derived from EBOV VP35 (residues 20-48, green) interacts entirely with the binding surface of the NP C-lobe.

(C) Cartoon representation of the structure of Zaire EBOV NP C-terminal domain (PDB entry 4QAZ). The NP CTD (residues 645-739, pink) exhibits a novel protein fold with an architecture distantly resembling that of some members of β-grasp superfamily.

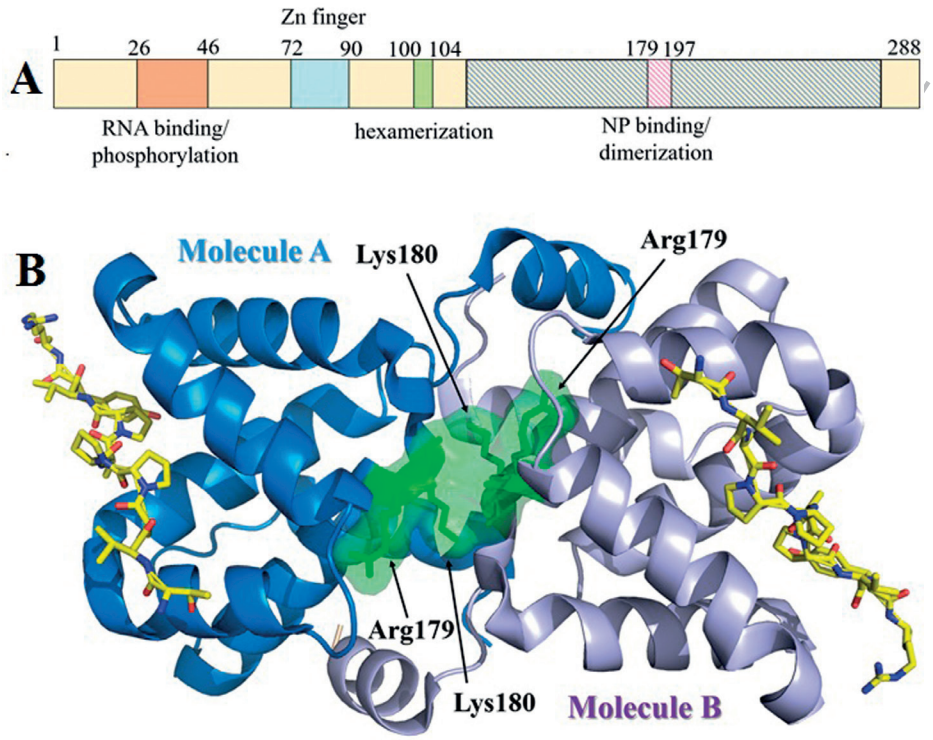

Figure 5. The structure of VP30.

(A) A schematic diagram of the EBOV VP30 gene showing the location of functional regions such as RNA binding/phosphorylation domain (orange box), zinc finger domain (cyan box), the hexamerization motif (green box), as well as region involved in binding to NP and VP30 dimerization (pink box). Hash-marked region correspond to the constructs used for crystallization purposes.

(B) Crystal structure of the Zaire EBOV VP30 C-terminal domain bound to NP (PDB entry 5T3T). In the crystal structure, VP30 CTD (114-265 residues) forms a dimer of two globular domains (Molecule A, blue; Molecule B, light blue) which is shown in cartoon representation. NP (602-614 residues, yellow sticks) binds to the narrow cavity on the VP30 globular domain located onwards the dimeric interface. Two key residues (Arg179, Lys180) of the putative druggable pocket, essential for transcription activation and nucleocapsid association are shown as green sticks with semitransparent surface.

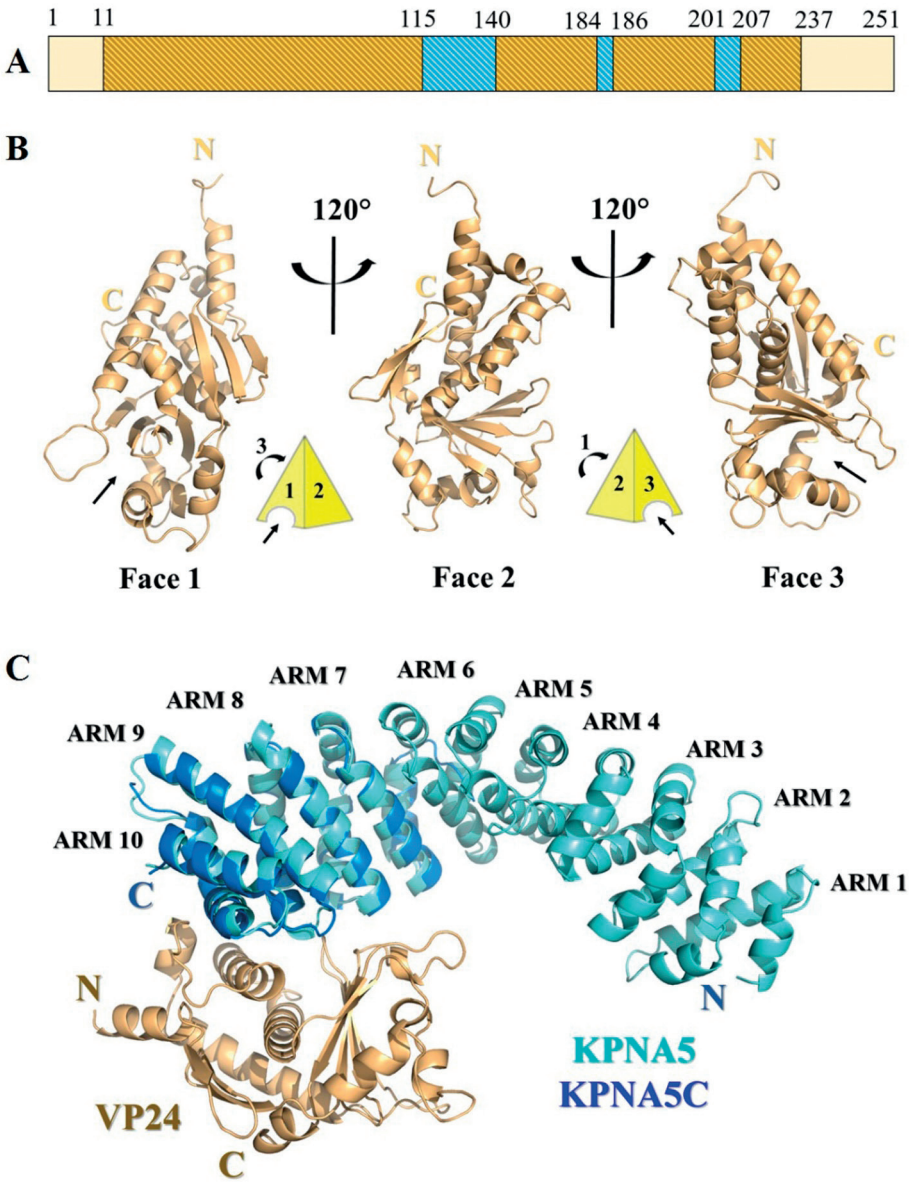

Figure 6. The structure of VP24.

(A) A schematic representation of EBOV VP24 gene. Blue boxes indicate three amino acid clusters that interact with karyopherin α 5 C-terminus (KPNA5C). Hash-marked region correspond to the construct used for crystallization purposes.

(B) Cartoon representation of the crystal structure of Zaire EBOV VP24 (PDB entry 4M0Q). EBOV VP24 (gold) adopts a novel three-sided pyramidal fold with Faces 1, 2 and 3 as illustrated. In the structure, there are two highly conserved pockets located adjacently on the protein surface. The first, hydrophobic pocket lies on the Face 1 and the second, more hydrophilic pocket is located on the Face 3.
(C) Alignment of the crystal structure of Zaire EBOV VP24 in complex with the C-terminal domain of KPNA5 (repeats 8-10) (PDB entry 4U2X) on the structure of full-length KPNA5 (PDB entry 1BK5). VP24 (gold) interacts with KPNA5 8,9,10 armadillo repeats (ARMs; colored blue) which is a specific binding site for KPNA transporters. This interaction has no effect on the overall structure of the full-length KPNA5 (cyan).

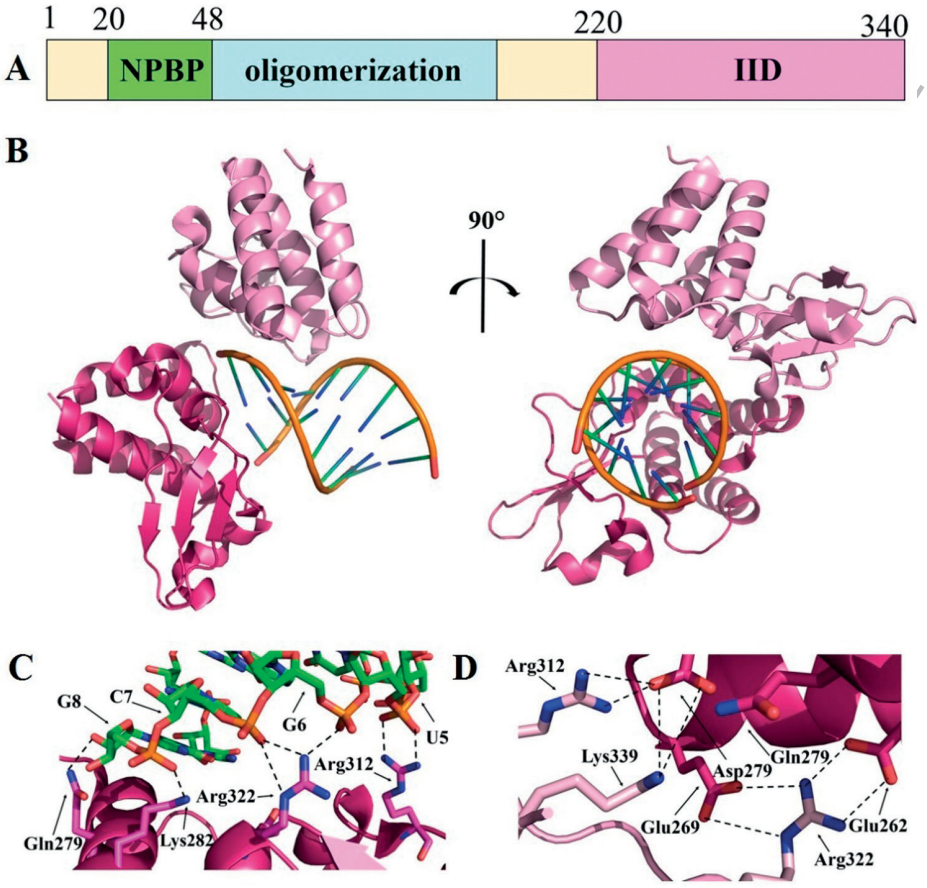

Figure 7. The structure of VP35

(A) A diagram of the EBOV VP35 gene showing the oligomerization domain (cyan box) involved in replication and transcription, and the interferon-inhibitory domain (IID, pink box) which bind dsRNA. There is also a highly conserved region located N-terminally to the oligomerization domain that is known to interact with EBOV NP (NPBP, green box). (B) Crystal structure of Zaire EBOV VP35 IID bound to the 8 bp dsRNA(PDB entry 3L25). The structure shows how dsRNA recognized by the VP35 IID domain, shown in ribbon representation. The end-capping (initial binding event) VP35 IID monomer is shown in raspberry, whereas the second monomer that binds to the dsRNA backbone (secondary binding event) is shown in light pink. The phosphate backbone of the dsRNA is represented by an orange ribbon and the bases are shown as sticks. There are four VP35 IID molecules bound to dsRNA in the crystallographic asymmetric unit, but only two D molecules are shown for clarity.

(C) The end-capping VP35 IID monomer (raspberry) interacts with the dsRNA in a sequence independent manner. In this case sidechains of Arg312 and Arg322 residues form hydrogen bonds with the dsRNA backbone.
(D) Close-up view on the binding interface between two VP35 IID monomers. The end-capping monomer (raspberry) interacts with ds-RNA backbone binding monomer (light pink) in a head-to-tail orientation, where residues Arg312 and Arg322 form hydrogen bonds with the Asp271 and Glu262 respectively.

References

1. Baseler L, Chertow DS, Johnson KM, Feldmann H, Morens DM. The Pathogenesis of Ebola Virus Disease. Annu Rev Pathol 2016.

2. Rougeron V, Feldmann H, Grard G, Becker S, Leroy EM. Ebola and Marburg haemorrhagic fever. J Clin Virol 2015;64:111-9.

3. Kazanjian P. Ebola in Antiquity? Clin Infect Dis 2015;61(6):963-8.

4. Sykes C, Reisman M. Ebola: Working Toward Treatments and Vaccines. P T 2015;40(8):521-5.

5. Towers S, Afzal S, Bernal G, Bliss N, Brown S, Espinoza B, et al. Mass Media and the Contagion of Fear: The Case of Ebola in America. PLoS One 2015;10(6):e0129179.

6. 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(1):17-22.

7. Qiu XG, 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(7520):47-+.

8. Henao-Restrepo AM, Camacho A, Longini IM, Watson CH, Edmunds WJ, Egger M, et al. Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease: final results from the Guinea ring vaccination, open-label, cluster-randomised trial (Ebola Ca Suffit!). Lancet 2016.

9. Olejnik J, Ryabchikova E, Corley RB, Muhlberger E. Intracellular events and cell fate in filovirus infection. Viruses 2011;3(8):1501-31.

10. Kuhn JH, Bao Y, Bavari S, Becker S, Bradfute S, Brauburger K, et al. Virus nomenclature below the species level: a standardized nomenclature for filovirus strains and variants rescued from cDNA. Arch Virol 2013.

11. Booth TF, Rabb MJ, Beniac DR. How do filovirus filaments bend without breaking? Trends Microbiol 2013.

12. Leendertz SA, Gogarten JF, Dux A, Calvignac-Spencer S, Leendertz FH. Assessing the Evidence Supporting Fruit Bats as the Primary Reservoirs for Ebola Viruses. Ecohealth 2016;13(1):18-25.

13. Mann E, Streng S, Bergeron J, Kircher A. A Review of the Role of Food and the Food System in the Transmission and Spread of Ebolavirus. PLoS neglected tropical diseases 2015;9(12):e0004160.

14. Moller-Tank S, Maury W. Ebola virus entry: a curious and complex series of events. PLoS Pathog 2015;11(4):e1004731.

15. Aleksandrowicz P, Marzi A, Biedenkopf N, Beimforde N, Becker S, Hoenen T, et al. Ebola virus enters host cells by macropinocytosis and clathrin-mediated endocytosis. J Infect Dis 2011;204 Suppl 3:S957-67.

16. Muhlberger E. Filovirus replication and transcription. Future Virol 2007;2(2):205-15.

17. Faye O, Andronico A, Faye O, Salje H, Boelle PY, Magassouba N, et al. Use of Viremia to Evaluate the Baseline Case Fatality Ratio of Ebola Virus Disease and Inform Treatment Studies: A Retrospective Cohort Study. PLoS Med 2015;12(12):e1001908.

18. Lee JE, Fusco ML, Hessell AJ, Oswald WB, Burton DR, Saphire EO. Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. Nature 2008;454(7201):177-82.

19. Falzarano D, Krokhin O, Wahl-Jensen V, Seebach J, Wolf K, Schnittler HJ, et al. Structure-function analysis of the soluble glycoprotein, sGP, of Ebola virus. Chembiochem 2006;7(10):1605-11.

20. de La Vega MA, Wong G, Kobinger GP, Qiu X. The multiple roles of sGP in Ebola pathogenesis. Viral Immunol 2015;28(1):3-9.

21. Madara JJ, Han Z, Ruthel G, Freedman BD, Harty RN. The multifunctional Ebola virus VP40 matrix protein is a promising therapeutic target. Future Virol 2015;10(5):537-46.

22. Watanabe S, Noda T, Kawaoka Y. Functional mapping of the nucleoprotein of Ebola virus. J Virol 2006;80(8):3743-51.

23. Biedenkopf N, Schlereth J, Grunweller A, Becker S, Hartmann RK. RNA Binding of Ebola Virus VP30 Is Essential for Activating Viral Transcription. J Virol 2016;90(16):7481-96.

24. Kuhl A, Pohlmann S. How Ebola virus counters the interferon system. Zoonoses and public health 2012;59 Suppl 2:116-31.

25. Hausmann S, Marq JB, Tapparel C, Kolakofsky D, Garcin D. RIG-I and dsRNA-induced IFNbeta activation. PLoS One 2008;3(12):e3965.

26. Audet J, Kobinger GP. Immune evasion in ebolavirus infections. Viral Immunol 2015;28(1):10-8.

27. Prins KC, Binning JM, Shabman RS, Leung DW, Amarasinghe GK, Basler CF. Basic residues within the ebolavirus VP35 protein are required for its viral polymerase cofactor function. J Virol 2010;84(20):10581-91.

28. Mateo M, Reid SP, Leung LW, Basler CF, Volchkov VE. Ebolavirus VP24 binding to karyopherins is required for inhibition of interferon signaling. J Virol 2010;84(2):1169-75.

29. Volchkov VE, Volchkova VA, Chepurnov AA, Blinov VM, Dolnik O, Netesov SV, et al. Characterization of the L gene and 5' trailer region of Ebola virus. J Gen Virol 1999;80 (Pt 2):355-62.

30. Mehedi M, Falzarano D, Seebach J, Hu X, Carpenter MS, Schnittler HJ, et al. A new Ebola virus nonstructural glycoprotein expressed through RNA editing. J Virol 2011;85(11):5406-14.

31. Tran EE, Simmons JA, Bartesaghi A, Shoemaker CJ, Nelson E, White JM, et al. Spatial localization of the Ebola virus glycoprotein mucin-like domain determined by cryo-electron tomography. J Virol 2014;88(18):10958-62.

32. Pallesen J, Murin CD, de Val N, Cottrell CA, Hastie KM, Turner HL, et al. Structures of Ebola virus GP and sGP in complex with therapeutic antibodies. Nat Microbiol 2016;1(9):16128.

33. Bornholdt ZA, Ndungo E, Fusco ML, Bale S, Flyak AI, Crowe JE, Jr., et al. Host-Primed Ebola Virus GP Exposes a Hydrophobic NPC1 Receptor-Binding Pocket, Revealing a Target for Broadly Neutralizing Antibodies. MBio 2016;7(1):e02154-15.

34. Carette JE, Raaben M, Wong AC, Herbert AS, Obernosterer G, Mulherkar N, et al. Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. Nature 2011;477(7364):340-3.

35. Miller EH, Obernosterer G, Raaben M, Herbert AS, Deffieu MS, Krishnan A, et al. Ebola virus entry requires the host-programmed recognition of an intracellular receptor. EMBO J 2012;31(8):1947-60.

36. Gong X, Qian H, Zhou X, Wu J, Wan T, Cao P, et al. Structural Insights into the Niemann-Pick C1 (NPC1)-Mediated Cholesterol Transfer and Ebola Infection. Cell 2016;165(6):1467-78.

37. 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(1-2):258-68.

38. Gregory SM, Harada E, Liang B, Delos SE, White JM, Tamm LK. Structure and function of the complete internal fusion loop from Ebolavirus glycoprotein 2. Proc Natl Acad Sci U S A 2011;108(27):11211-6.

39. Gregory SM, Larsson P, Nelson EA, Kasson PM, White JM, Tamm LK. Ebolavirus entry requires a compact hydrophobic fist at the tip of the fusion loop. J Virol 2014;88(12):6636-49.

40. Jasenosky LD, Neumann G, Lukashevich I, Kawaoka Y. Ebola virus VP40-induced particle formation and association with the lipid bilayer. J Virol 2001;75(11):5205-14.

41. Bornholdt ZA, Noda T, Abelson DM, Halfmann P, Wood MR, Kawaoka Y, et al. Structural Rearrangement of Ebola Virus VP40 Begets Multiple Functions in the Virus Life Cycle. Cell 2013;154(4):763-74.

42. Liu Y, Cocka L, Okumura A, Zhang YA, Sunyer JO, Harty RN. Conserved motifs within Ebola and Marburg virus VP40 proteins are important for stability, localization, and subsequent budding of virus-like particles. J Virol 2010;84(5):2294-303.

43. Timmins J, Scianimanico S, Schoehn G, Weissenhorn W. Vesicular release of ebola virus matrix protein VP40. Virology 2001;283(1):1-6.

44. Dessen A, Volchkov V, Dolnik O, Klenk HD, Weissenhorn W. Crystal structure of the matrix protein VP40 from Ebola virus. Embo J 2000;19(16):4228-36.

45. Clifton MC, Bruhn JF, Atkins K, Webb TL, Baydo RO, Raymond A, et al. High-resolution Crystal Structure of Dimeric VP40 From Sudan ebolavirus. J Infect Dis 2015;212 Suppl 2:S167-71.

46. Soni SP, Adu-Gyamfi E, Yong SS, Jee CS, Stahelin RV. The Ebola virus matrix protein deeply penetrates the plasma membrane: an important step in viral egress. Biophys J 2013;104(9):1940-9.

47. Yamayoshi S, Noda T, Ebihara H, Goto H, Morikawa Y, Lukashevich IS, et al. Ebola virus matrix protein VP40 uses the COPII transport system for its intracellular transport. Cell Host Microbe 2008;3(3):168-77.

48. Adu-Gyamfi E, Digman MA, Gratton E, Stahelin RV. Investigation of Ebola VP40 assembly and oligomerization in live cells using number and brightness analysis. Biophys J 2012;102(11):2517-25.

49. Gomis-Ruth FX, Dessen A, Timmins J, Bracher A, Kolesnikowa L, Becker S, et al. The matrix protein VP40 from Ebola virus octamerizes into pore-like structures with specific RNA binding properties. Structure 2003;11(4):423-33.

50. Timmins J, Schoehn G, Ricard-Blum S, Scianimanico S, Vernet T, Ruigrok RW, et al. Ebola virus matrix protein VP40 interaction with human cellular factors Tsg101 and Nedd4. J Mol Biol 2003;326(2):493-502.

51. Noda T, Hagiwara K, Sagara H, Kawaoka Y. Characterization of the Ebola virus nucleoprotein-RNA complex. J Gen Virol 2010;91(Pt 6):1478-83.

52. Dong S, Yang P, Li G, Liu B, Wang W, Liu X, et al. Insight into the Ebola virus nucleocapsid assembly mechanism: crystal structure of Ebola virus nucleoprotein core domain at 1.8 A resolution. Protein Cell 2015;6(5):351-62.

53. Leung DW, Borek D, Luthra P, Binning JM, Anantpadma M, Liu G, et al. An Intrinsically Disordered Peptide from Ebola Virus VP35 Controls Viral RNA Synthesis by Modulating Nucleoprotein-RNA Interactions. Cell Rep 2015;11(3):376-89.

54. Sherwood LJ, Hayhurst A. Ebolavirus nucleoprotein C-termini potently attract single domain antibodies enabling monoclonal affinity reagent sandwich assay (MARSA) formulation. PLoS One 2013;8(4):e61232.

55. Baker LE, Ellena JF, Handing KB, Derewenda U, Utepbergenov D, Engel DA, et al. Molecular architecture of the nucleoprotein C-terminal do-

main from the Ebola and Marburg viruses. Acta Crystallogr D Struct Biol 2016;72(Pt 1):49-58.

56. Dziubanska PJ, Derewenda U, Ellena JF, Engel DA, Derewenda ZS. The structure of the C-terminal domain of the Zaire ebolavirus nucleoprotein. Acta Cryst D 2014;70(Pt 9):2420-9.

57. Garcia-Dorival I, Wu W, Armstrong SD, Barr JN, Carroll MW, Hewson R, et al. Elucidation of the Cellular Interactome of Ebola Virus Nucleoprotein and Identification of Therapeutic Targets. J Proteome Res 2016;15(12):4290-303.

58. John SP, Wang T, Steffen S, Longhi S, Schmaljohn CS, Jonsson CB. Ebola virus VP30 is an RNA binding protein. J Virol 2007;81(17):8967-76.

59. Hartlieb B, Muziol T, Weissenhorn W, Becker S. Crystal structure of the C-terminal domain of Ebola virus VP30 reveals a role in transcription and nucleocapsid association. Proc Natl Acad Sci U S A 2007;104(2):624-9.

60. Kirchdoerfer RN, Moyer CL, Abelson DM, Saphire EO. The Ebola Virus VP30-NP Interaction Is a Regulator of Viral RNA Synthesis. PLoS Pathog 2016;12(10):e1005937.

61. Schlereth J, Grunweller A, Biedenkopf N, Becker S, Hartmann RK. RNA binding specificity of Ebola virus transcription factor VP30. RNA Biol 2016;13(9):783-98.

62. Xu W, Edwards MR, Borek DM, Feagins AR, Mittal A, Alinger JB, et al. Ebola Virus VP24 Targets a Unique NLS Binding Site on Karyopherin Alpha 5 to Selectively Compete with Nuclear Import of Phosphorylated STAT1. Cell Host Microbe 2014;16(2):187-200.

63. Leung DW, Ginder ND, Fulton DB, Nix J, Basler CF, Honzatko RB, et al. Structure of the Ebola VP35 interferon inhibitory domain. Proc Natl Acad Sci U S A 2009;106(2):411-6.

64. Leung DW, Prins KC, Borek DM, Farahbakhsh M, Tufariello JM, Ramanan P, et al. Structural basis for dsRNA recognition and interferon antagonism by Ebola VP35. Nat Struct Mol Biol 2010;17(2):165-72.

65. Basu A, Li B, Mills DM, Panchal RG, Cardinale SC, Butler MM, et al. Identification of a small-molecule entry inhibitor for filoviruses. J Virol 2011;85(7):3106-19.

66. Johansen LM, DeWald LE, Shoemaker CJ, Hoffstrom BG, Lear-Rooney CM, Stossel A, et al. A screen of approved drugs and molecular probes identifies therapeutics with anti-Ebola virus activity. Science translational medicine 2015;7(290):290ra89.

Address for correspondence
prof. Zygmunt Derewenda
Department of Molecular Physiology and Biological Physics,
University of Virginia, 1340 Jefferson Park Avenue, Charlottesville, Virginia, 22903, United States
email: zsd4n@eservices.virginia.edu