

Ligand screening approach to find potential inhibitors of GP1 from Ebola Virus

Garshasb Rigi^{*1}, Raham Armand¹, Ibrahim Torktaaz²

¹Department of Biology, Faculty of Science, Behbahan Khatam Alanbia University of Technology, Behbahan, Iran

²R&D center, Ilya Biotech Co. Ltd., Tehran, Iran

*Corresponding author: email address: garshasbbiotech@gmail.com (G.rigi)

ABSTRACT

Ebola is one of the members of filoviruses family. It causes severe hemorrhagic with the human mortality rate of 50-90%. In the initial steps of infection, it enters the host cell by Glyco protein1 (GP1). GP1 is a big subunit with the molecular weight of 130 kDa. Its N-terminal domain is responsible for attaching the host cell. In the present study, a database containing 100.000 drug like chemicals which was obtained from Zinc has been screened. The top successive hits were then analyzed regarding Lipinski rules, oral toxicity value and recorded biological properties. Finally, 3 new ligands were introduced as new theoretical inhibitors of Ebola virus entry. Hit #1 (Heptacyclo[18.7.0.02,10.03,8.011,19.012,17.021,26]heptacosal(20),2(10),3,5,7,11(19),12,14,16,21,23,25-didecaene-9,18,27-trione) indicated the binding affinity of -10.7 kcal/mol with previous biological reports indicating no toxicity on human cell line. Hit#3 (1,1',3,3'-tetraethyl-5'-(4-nitrophenyl)spiro[1,3-diazinane-5,6'-5H-furo[2,3-d]pyrimidine]-2,2',4,4',6 pentone) with the binding affinity of -9.7 kcal/mol theoretically passed pharmacological filters and the toxicity class of 4 theoretically revealed that this hit also can be a potential anti GP1 agent. Moreover, hit#4(10-(Hydroxymethyl)-7,18-dimethoxy-6,17-dimethyl-21-methyl-11.21 diazapentacyclo[11.7.1.02,11.04,9.015,20]henicosa-4(9),6,15(20),17-tetraene-5,8,12,16,19-pentone) also indicated the binding affinity of -9.7 kcal/mol and was matched with Lipinski rules and the toxicity class 4, suggesting that it can be safe in low dose usage. Based on the pharmacological properties, hit #1, 3 and 4 are presented as the new hypothetical drugs against GP1. Moreover, hit#1 has previously been recorded as nontoxic activity on human cell line and makes it an appropriate candidate for further *in vitro* and *in vivo* studies.

Keywords: Ebola virus; Glyco protein1, Virtual screening; Lipinski; Drug

INTRODUCTION

Ebola virus is one of two family members' filoviruses which causes severe hemorrhagic fever with a human fatality of 50-90%. There is not any vaccine or cure yet existing in combination with frequent appearance of this virus, and its high spread among wildlife; ease of importing viruses into the cell has made it an important public health concern. Ebola virus (EBOV) creates Ebola disease (EVD), a disease with a high mortality rate that is died more than 11000 people from February to December 1, 2014 as reported by Centers for Disease Control and Prevention 2014. Gomes et al reportage was the first published report of the risk assessment

[1]. Recently, 1500 genomes of Ebola virus have been studied from several patients: the results and data of the respected author teams are attainable free of the charge on scientific databases [2]. Import and virus entry into cells is mediated by a glycoprotein (GP), the virion surface glycoprotein [3]. GP is known as class I fusion protein [4, 5] which ligates to the cell surface and fuses with an inner membrane of the cell. Like many other class I fusion proteins such as influenza virus hemagglutinin, intact GP trimer of heterodimers consist of a subunit ligating to the receptor (GP₁) and a fusion subunit (GP₂) [6]. Two subunits are produced through cleavage of a pre-protein as GP via the Golgi compartment and stay via a

disulfide bond and non-covalent interactions. GP₁ is a big subunit (definite molecular mass of 130 kDa) that contains a cap including 4 N-linked carbohydrate addition sites and a large domain that is O-glycosylated which is a mucin-like domain. The mucin-like domain is inessential for infection of cells in culture medium; pseudovirions and virus-like particles with their mucin-like domain (GP) is deleted by genetic engineering techniques and are absolutely qualified for entry and infection [7-9]. The two carbohydrate substance motifs of GP₁ are located on a top domain, which includes a receptor binding region (RBR) [7, 8, 10-12]. In turn, the top of the GP₁ is located on the head of a base that is widely interacting with fusion subunit (GP₂), which makes a fusion ring [13]. Endosomal protease enzymes named cathepsin B (Cat B) and cathepsin L (cat L) are needed for Ebola virus to import to the cytoplasm and thus to infect host cells [8, 9, 14]. Cat B and Cat L proteases cut GP within misfolded 13-14 rings of GP₁, producing an average of 20 kDa and then original and special 19 kDa core protein [7, 9, 15]. It seems that the interaction produces an average 50-kDa protein that is similar to the GP. According to these great cutting events, GP₂ is completely conserved and holds disulfide connected to the GP₁. Production of 19 kDa GP deletes the mucin-like motif, glycans hat, and the outer string [16] of the top domain of GP₁. The Receptor binding region (RBR) and backbone of GP₁ and also GP₂ are all kept native. In the previous study, a key protein of Ebola virus which has the important action in both virus assembly and budding functions was targeted. Prohibition of VP₄₀ assembly excludes octamer structure foundation and disturbs a vital stage in the virus life cycle. A single VP₄₀ construction in the biological situation was also targeted, and a drug model which prevents oligomerization of VP₄₀ subunits was found; so the potential inhibitors for Ebola virus was predicted [17]. But in this study, according to the GP₁ based mechanism of entry of Ebola virus into the host cell, we tried to find an appropriate ligand with acceptable pharmacological quality with the ability to bind GP₁ and inhibit the virus entry to the host cell.

MATERIALS AND METHODS

Structures

Currently, there is not any crystallography structure available for GP₁ from Ebola Virus. So its nature is predicted based on homology modeling. To do this, the sequence of Zaire Ebola virus (accession number: 965569707) for identification of homolog crystal templates was used. The modeling process followed by I-Tasser webserver (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>); it was carried out using Chain I, Crystal Structure of Sudan Ebolavirus Glycoprotein (PDB ID: 3VE0) as the main template [18-20]. The predicted structure was checked by Prosa Z score for quality control [21]. The final structure was submitted to FindSite web server (www.cssb.biology.gatech.edu/finalsite) to find its potential binding sites [22, 23]. After identification of pockets, Zinc database was docked against the pockets and top 1000 ligands was extracted for further screening process [24].

Selecting high-affinity ligands

In order to simulate biological conditions, the final predicted model was put into water box in the presence of neutralizing ions. The system was then minimized regarding energy level by GROMACS 4.5.6 simulation package [25]. The most important residues which are involved in receptor binding are first 150 N-Terminal domain residues. Subsequently, a radius of 20 Å along with coordinates of X: 73, Y: 97 and Z: 127 was applied to cover the entire area. Virtual screening was also performed by PyRx software [26, 27].

Analysis of pharmacological properties

Top successive hits were analyzed regarding pharmacological properties by FAF Drugs 3 server [28]; the unspecific bindings were also measured by PROTOX web server [29].

RESULTS

In the present study, the 3D structure of GP₁ was initially predicted and after structure verifications, a database containing 100,000 ligands was screened and 4 hits were found which could potentially inhibit the viral entry by blocking the N-terminal domain of GP₁. In order to reach high binding affinity poses, the virtual screening radius was put in the N-terminal domain of GP₁. Top successive hits were filtered

regarding Lipinski rule of 5 and finally 4 hits were selected for the further study. The structures of selected hits are depicted in figure 1. Ligand 1 (IUPAC name: Heptacyclo[18.7.0.0.2,10.0.3,8.0.11,19.0.12,17.0.21,26]heptacosal(20),2(10),3,5,7,11(19),12,14,16,21,23,25-didecaene-9,18,27-trione)

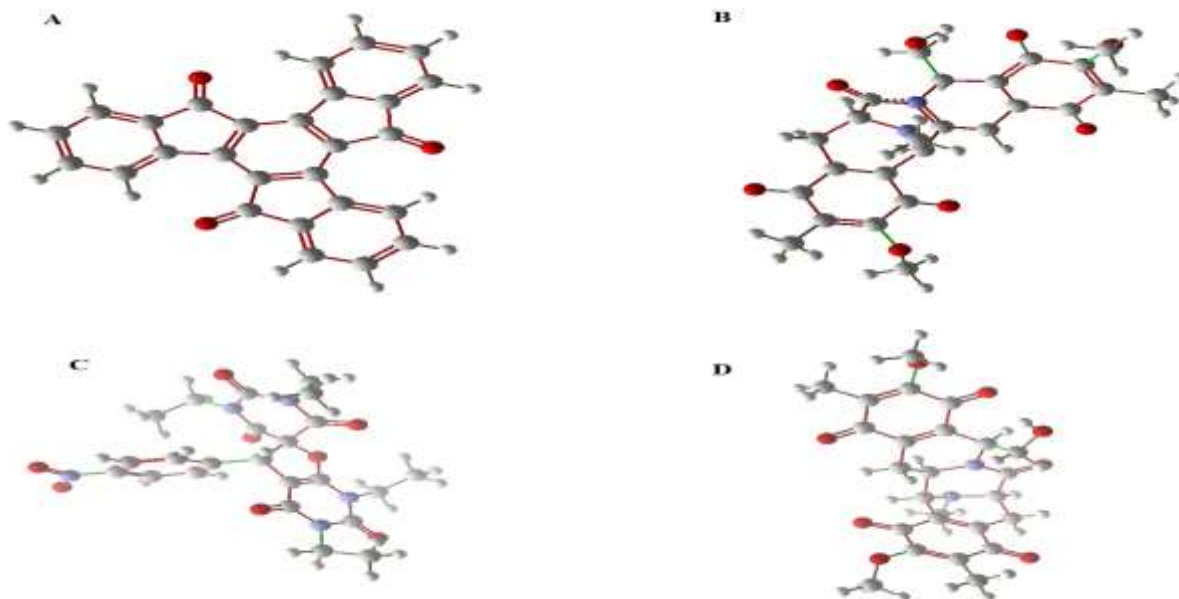


Figure1. The structure of 4 successive hits which binds to the GP1 from Ebola virus, A: hit#1, B: hit#2, C: hit#3 and D: hit#4

Interestingly, this hit matches with the Lipinski rule of 5. The molecular formula ($C_{27}H_{12}O_3$) indicates that its components are carbon and oxygen. Based on the ligand map of hit#1 which is depicted in figure 2a, it has several interactions with amino acids: Trp 13, Arg 14, Pro 16, Asn 49, Lys 63, His 65, Leu 66 and Thr 68. The protox result of this hit predicted LD50 of 385mg/kg with the toxicity class 4 and no toxicity target has been identified. Also hit#2 (IUPAC name: 5,8,16,19-Tetrahydroxy-1(hydroxymethyl)-7,18-dimethoxy-6,17-dimethyl-21-methyl-11 λ 5.21diazapentacyclo[11.7.1.0.2,11.0.4,9.0.15,20]henicosal-2(11),6,8,17-tetraen-12-one) could reach a binding affinity of -10.2 and has no record in the PubChem but matches with the Lipinski rule of 5. Its ligand map (figure 2b) indicates that it is in contact with amino acids: Trp 13, Arg 14, Pro 15, Pro16, Asn 49, Asn 50, Lys 63, His 65 and Leu 66. The protox results indicated that although this hit matches Lipinski rules, but it has a predicted LD50 of 6mg/kg with the toxicity class of 2 and its pharmacophore has interference with

with the molecular weight of 384.38238 g/mol, reaching a binding affinity of -10.7 kcal/mol. Its bio assay is reported by PubChem as inactive in NCI human tumor cell line growth inhibition bioassay. (Data for the NCI-H23 Non-Small Cell Lung cell line).

Amine Oxidase A and Prostaglandin G/H Synthase 1. So this hit cannot be suggested as an anti-Ebola agent. Hit#3 (IUPAC name: 1,1',3,3'-tetraethyl-5'-(4-nitrophenyl)spiro[1,3-diazinane-5,6'-5H-furo[2,3-d]pyrimidine]-2,2',4,4',6-pentone) does not have any biological assay record in PubChem but the protox results predicts that it has the toxicity class of 4 with the LD50 of 500 mg/kg. Moreover, no toxicity target has been identified. The ligand map (figure 2c) indicates that it is in contact with amino acids: Arg 14, Pro 15, Pro 16, Asn 18, Asn 49, Lys 63, Glu 64, His 65, Leu 66, Ala 67 and Thr 68 which indicates that this hit covers amino acid position 63-68 sequentially. This ligand can also be introduced as the new inhibitor of GP1 from Ebola Virus. This hit could reach the binding affinity of -9.7 kcal/mol. Hit#4 (IUPAC name: 10-(Hydroxymethyl)-7,18-dimethoxy-6,17-dimethyl-21-methyl-11.21-diazapentacyclo[11.7.1.0.2,11.0.4,9.0.15,20]henicosal-4(9),6,15(20),17-tetraene-5,8,12,16,19-pentone) also indicates that the binding affinity of -9.7

does not have any recorded biological properties in PubChem. Its LD50 of 445 mg/kg with the toxicity class 4 indicates that it can be safe in low dosages. The ligand map of hit#4 depicted in figure 2d indicates that it has several contacts with residues: Pro 15, Pro 16, Pro 17,

Asn 18, Leu 24, Ile 32, Asn 49, Asn 50, Lys 63, His 65 and Leu 66. Based on the pharmacological properties, this ligand can also be introduced as a new hypothetical inhibitor of GP1 from Ebola virus.

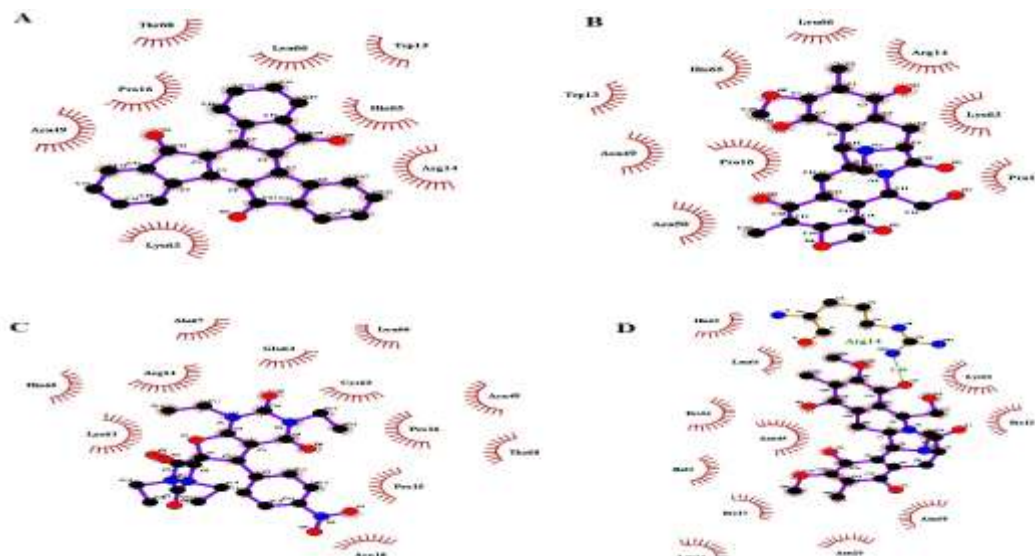


Figure 2. The ligand map of 4 successive hits, A: hit#1 is in contact with Trp 13, Arg 14, Pro 16, Asn 49, Lys 63, His 65, Leu 66 and Thr 68, B: hit#2 is in contact with Trp 13, Arg 14, Pro 15, Pro16, Asn 49, Asn 50, Lys 63, His 65 and Leu 66, C: hit#3 is in contact with Arg 14, Pro 15, Pro 16, Asn 18, Asn 49, Lys 63, Glu 64, His 65, Leu 66, Ala 67 and Thr 68 which indicates that this hit covers amino acid position 63-68 sequentially and D: hit#4 is contacted with residues: Pro 15, Pro 16, Pro 17, Asn 18, Leu 24, Ile 32, Asn 49, Asn 50, Lys 63, His 65 and Leu 66

DISCUSSION

Ebola virus (EBOV) login needs surface glycoprotein (GP) for anchoring and fusion of the viral membrane and the host cell. GP₁ which is accountable for binding to the receptor is on target cells.

Nathan et al reported that a glycoprotein construct, which mimics the cathepsin enzyme of the EboGP glycan caps and mucin-like domains, is able to counteract tetherin. Incorporating these outcomes proposes an important role for the EboGP glycan cap in tetherin antagonism [30]. Jeffrey et al. found the composite of crystal structure and Ebola virus GP in its trimeric, pre-ligation (GP₁ + GP₂) which is confined to a counteracting antibody [31]. The findings of the other study provided the the assumption that priming of EBOV GP, originally the core of the 19 kDa, stoutened GP fusion dependent to structural varieties thereupon. Supplementary

results supported that low pH and additional endosomal element operate as extensive ligation or fusion [32]. In the present study, the 3D structure of GP1 was first predicted and after structure verifications, a database containing 100.000 ligands was screened and 4 hits were found which could potentially inhibit the viral entry by blocking the N-terminal domain of GP1. In order to reach high binding affinity poses, the virtual screening radius was put in the N-terminal domain of GP1. Top successive hits were filtered regarding Lipinski rule of 5 and finally 4 hits were selected for the further study.

CONCLUSION

GP1 from Ebola virus is responsible for host cell binding. In this study, 3 new hypothetical ligands are introduced which could bind the N-terminal domain of GP1. Based on the pharmacological properties, hit #1, 3 and 4 are

presented as the new hypothetical drugs against GP1. Moreover, hit#1 has previously been recorded as nontoxic activity on human cell line which makes it an appropriate candidate for further *in vitro* and *in vivo* studies.

ACKNOWLEDGMENTS

The authors would like to thank the Behbahan Khatam Alanbia University of Technology for providing the funding for this project (Project no. 94-02-25).

"The authors declare no conflict of interest."

REFERENCES

- Gomes MFC, Pastore y Piontti A, Rossi L, Chao D, Longini I, Halloran ME, et al. Assessing the International Spreading Risk Associated with the 2014 West African Ebola Outbreak. *PLoS Currents*. 2014;6:ecurrents.outbreaks.cd818f63d40e24aef769dda7df9e0da5.
- Gire SK, Goba A, Andersen KG, Sealfon RS, Park DJ, Kanneh L, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science (New York, NY)*. 2014;345(6202):1369-72.
- Lee JE, Saphire EO. Neutralizing ebolavirus: structural insights into the envelope glycoprotein and antibodies targeted against it. *Current opinion in structural biology*. 2009;19(4):408-17.
- Harrison SC. Viral membrane fusion. *Nature structural & molecular biology*. 2008;15(7):690-8.
- White JM, Delos SE, Brecher M, Schornberg K. Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. *Critical reviews in biochemistry and molecular biology*. 2008;43(3):189-219.
- Lee JE, Fusco ML, Hessel 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.
- Dube D, Brecher MB, Delos SE, Rose SC, Park EW, Schornberg KL, et al. The primed ebolavirus glycoprotein (19-kilodalton GP1,2): sequence and residues critical for host cell binding. *Journal of virology*. 2009;83(7):2883-91.
- Kaletsky RL, Simmons G, Bates P. Proteolysis of the Ebola virus glycoproteins enhances virus binding and infectivity. *Journal of virology*. 2007;81(24):13378-84.
- Schornberg KL, Shoemaker CJ, Dube D, Abshire MY, Delos SE, Bouton AH, et al. Alpha5beta1-integrin controls ebolavirus entry by regulating endosomal cathepsins. *Proceedings of the National Academy of Sciences of the United States of America*. 2009;106(19):8003-8.
- Brindley MA, Hughes L, Ruiz A, McCray PB, Jr., Sanchez A, Sanders DA, et al. Ebola virus glycoprotein 1: identification of residues important for binding and postbinding events. *Journal of virology*. 2007;81(14):7702-9.
- Manicassamy B, Wang J, Jiang H, Rong L. Comprehensive analysis of ebola virus GP1 in viral entry. *Journal of virology*. 2005;79(8):4793-805.
- Mpanju OM, Towner JS, Dover JE, Nichol ST, Wilson CA. Identification of two amino acid residues on Ebola virus glycoprotein 1 critical for cell entry. *Virus research*. 2006;121(2):205-14.
- Ito H, Watanabe S, Sanchez A, Whitt MA, Kawaoka Y. Mutational analysis of the putative fusion domain of Ebola virus glycoprotein. *Journal of virology*. 1999;73(10):8907-12.
- Chandran K, Sullivan NJ, Felbor U, Whelan SP, Cunningham JM. Endosomal proteolysis of the Ebola virus glycoprotein is necessary for infection. *Science (New York, NY)*. 2005;308(5728):1643-5.
- Hood CL, Abraham J, Boyington JC, Leung K, Kwong PD, Nabel GJ. Biochemical and Structural Characterization of Cathepsin L-Processed Ebola Virus Glycoprotein: Implications for Viral Entry and Immunogenicity. *Journal of virology*. 2010;84(6):2972-82.
- Delos SE, La B, Gilmartin A, White JM. Studies of the "Chain Reversal Regions" of the Avian Sarcoma/Leukosis Virus (ASLV) and Ebolavirus Fusion Proteins: Analogous Residues Are Important, and a His Residue Unique to EnvA Affects the pH Dependence of ASLV Entry. *Journal of virology*. 2010;84(11):5687-94.
- Abazari D, Moghtadaei M, Behvarmanesh A, Ghannadi B, Aghaei M, Behruznia M, et al. Molecular docking based screening of predicted potential inhibitors for VP40 from Ebola virus. *Bioinformation*. 2015;11(5):243-7.

18. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols*. 2010;5(4):725-38.
19. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nature methods*. 2015;12(1):7-8.
20. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC bioinformatics*. 2008;9:40.
21. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*. 2007;35(Web Server issue):W407-W10.
22. Brylinski M, Skolnick J. A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proceedings of the National Academy of Sciences*. 2008;105(1):129-34.
23. Skolnick J, Brylinski M. FINDSITE: a combined evolution/structure-based approach to protein function prediction. *Briefings in Bioinformatics*. 2009;10(4):378-91.
24. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: A Free Tool to Discover Chemistry for Biology. *Journal of Chemical Information and Modeling*. 2012;52(7):1757-68.
25. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*. 2013;29(7):845-54.
26. Dallakyan S, Olson AJ. Small-molecule library screening by docking with PyRx. *Methods in molecular biology (Clifton, NJ)*. 2015;1263:243-50.
27. Manicassamy B, Wang J, Jiang H, Rong L. Comprehensive Analysis of Ebola Virus GP1 in Viral Entry. *Journal of virology*. 2005;79(8):4793-805.
28. Lagorce D, Sperandio O, Baell JB, Miteva MA, Villoutreix BO. FAF-Drugs3: a web server for compound property calculation and chemical library design. *Nucleic acids research*. 2015;43(W1):W200-7.
29. Drwal MN, Banerjee P, Dunkel M, Wettig MR, Preissner R. ProTox: a web server for the in silico prediction of rodent oral toxicity. *Nucleic acids research*. 2014;42(Web Server issue):W53-8.
30. Vande Burgt NH, Kaletsky RL, Bates P. Requirements within the Ebola Viral Glycoprotein for Tetherin Antagonism. *Viruses*. 2015;7(10):5587-602.
31. Lee JE, Fusco ML, Hessel 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.
32. Brecher M, Schornberg KL, Delos SE, Fusco ML, Saphire EO, White JM. Cathepsin cleavage potentiates the Ebola virus glycoprotein to undergo a subsequent fusion-relevant conformational change. *Journal of virology*. 2012;86(1):364-72.