

## Mapping Zaire Ebola Virus Glycoprotein Organization onto Information Entropy

Joel K Weltman

Clinical Professor Emeritus of Medicine, Alpert Medical School, Brown University, USA

**Keywords:** Zaire ebola virus; ZEBOV; Glycoprotein GP1,2; Information entropy; Receptor binding domain (RBD); Fusion loop (FL); Escape mutants

### Editorial

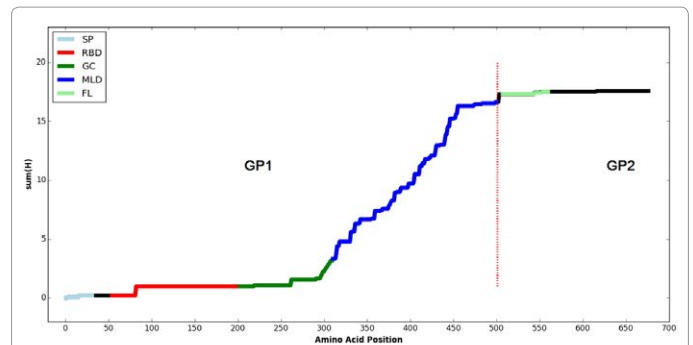
Ebola glycoprotein GP1,2 performs functions which are essential for the binding of virus to the target cell membrane and subsequent internalization and processing by the target cell [1]. This journal has recently published an analysis of the information entropy (H) of a dataset comprised of Zaire and Sudan Ebola virus (ZEBOV and SEBOV) glycoprotein GP1,2 sequences [2]. The current epidemic of Ebola virus disease (EVD) is caused exclusively by viruses of the ZEBOV strain (<http://www.who.int/mediacentre/factsheets/fs103/en/>). Accordingly, this Editorial focuses on the potential significance and usefulness of the H distribution [3] in analyzing ZEBOV GP1,2 proteins.

The complete set of ZEBOV GP1,2 amino acid sequences (N=139) was downloaded in FASTA format using the NCBI Ebolavirus Resource (<http://www.ncbi.nlm.nih.gov/genome/viruses/variation/ebola/>) on March 26, 2015. Determination of information entropy and graphing of the results were performed with 64-bit Enthought Python 2.7.6 as previously described. In this research, data are presented as summations of H in order to enhance the visualization of the organization of H distribution over the entire length of 676 amino acids that comprise the ZEBOV GP1,2 protein. In this representation, mutating regions are represented as regions in which  $\frac{\Delta \text{sum}(H)}{\Delta n} > 0.0$  where n is amino acid position. Non-mutating regions in the protein sequences are represented as regions in which  $\frac{\Delta \text{sum}(H)}{\Delta n} = 0.0$ .

The H distribution obtained for the ZEBOV GP1,2 set of sequences is shown in (Figure 1). The slope  $\frac{\Delta \text{sum}(H)}{\Delta n}$  of the distribution was equal to zero from positions 17 to 81 and from positions 83 to 218. There was a steep increase in slope beginning at position 296, ending at position 455. There was zero slope from position 504 to 543, with rare mutation from position 545 to the C-terminal phenylalanine.

The following regions of the GP1,2 protein have been reported to have special features and functions, some of which are essential for Ebola virus replication, infectivity, transmission and mutational escape: signal peptide [4], receptor binding domain [5,6], glycan cap [7,8], mucin-like domain [8] and fusion loop [6,9,10]. These regions have the following distribution (amino acid positions in parentheses): signal peptide (1-31), receptor binding domain (54-201), glycan cap (202-309), mucin-like domain (310-500) and fusion loop (507-560).

The distribution of sum(H) as a function of n for ZEBOV GP1,2 is shown in (Figure 1). The slope is close to zero until position 82, which is a mutating site. There is a zero slope, ie, with no mutation, from positions 83 to 218. A much greater  $\frac{\Delta \text{sum}(H)}{\Delta n}$  slope, indicating the accumulation of mutants, began at position 262 and persisted through position 503 of the GP1.2 sequence dataset. The  $\frac{\Delta \text{sum}(H)}{\Delta n}$  slope was again zero from position 504 until position 543, and remained close to zero until, and including the C-terminus at position 676.



Note: Ordinate=sum(H) in bits, abscissa=amino acid position. SP=signal peptide, RBD=receptor binding domain, GC=glycan cap, MLD=mucin-like domain, FL=fusion loop. Vertical red line denotes the cleavage site of GP1,2 precursor protein into GP1 and GP2 proteins by furin peptidase at amino acid 501 [10].

**Figure 1:** Summed Information Entropy (H) and Functional Organization of ZEBOV Glycoprotein GP1,2.

The mucin-like domain (MLD) is the region of ZEBOV GP1,2 that displayed the greatest accumulation of mutants. In contrast, the receptor binding domain (RBD) and the fusion loop (FL) both occupied ZEBOV GP1,2 regions that displayed little, if any, tendency to mutate.

The tendency towards zero H displayed by both the RBD and the FL regions of GP1,2 suggests that these regions may be under biological constraints that restrict the ability of mutations to occur or for mutated sequences to persist. It is proposed that these invariant positions may be useful as immunologic or therapeutic targets with an inherent stability that mitigates against mutational escape.

The amino acid sequence (83-218) that includes mainly the RBD, length=136 with H=0 at all positions, is:

TKRWGFRSGVPPKVVNYEAGAWAENCYNLEIKKPDGSE-  
CLPAAPDGIIRGFPRCRYVHKVSGTGPCAGDFAFHKEGAFFLY-  
DRLASTVIYRGTTFEAEGVVAFLILPQAKKDFSSHPLREPVNAT-  
EDPSSGYSTTI

The amino acid sequence (504-543) that includes mainly the FL, length=40 with H=0 at all positions, is:

IVNAQPKCNPNLHYWTTQDEGAAIGLAWIPYFGPAAEGIY  
IVNAQPKCNPNLHYWTTQDEGAAIGLAWIPYFGPAAEGIY

**\*Corresponding author:** Weltman JK, Clinical Professor Emeritus of Medicine, Alpert Medical School, Brown University, USA, Tel: 401/245-7588; E-mail: [joel\\_weltman@brown.edu](mailto:joel_weltman@brown.edu)

Received April 15, 2015; Accepted April 16, 2015; Published April 18, 2015

**Citation:** Weltman JK (2015) Mapping Zaire Ebola Virus Glycoprotein Organization onto Information Entropy. J Med Microb Diagn 4: e128. doi:10.4172/21610703.1000e128

**Copyright:** © 2015 Weltman JK. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

As of this writing, the 2013-2015 epidemic of Ebola virus disease (EVD) in West Africa, caused by ZEBOV, has produced 25,556 cases and 10,587 deaths (<http://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/case-counts.html>). It is proposed that these two invariant GP1,2 peptides detected by information entropy may be useful for development of anti-ZEBOV preventive vaccines and pharmaceutical agents.

#### References

1. Miller EH, Chandran K (2012) Filovirus entry into cells - new insights. *Curr Opin Virol* 2: 206-214.
2. Weltman JK (2015) Identification of Invariant Peptide Domains within Ebola Virus Glycoprotein GP 1,2. *J Med Microb Diagn* 4: 176
3. Shannon CE (1948) A Mathematical Theory of Communication. *Bell System Technical Journal* 27: 379-423.
4. Marzi A, Akhavan A, Simmons G, Gramberg T, Hofmann H, et al. (2006) The signal peptide of the ebolavirus glycoprotein influences interaction with the cellular lectins DC-SIGN and DC-SIGNR. *J Virol*. 80:6305-6317.
5. Dube D, Brecher MB, Delos SE, Rose SC, Park EW, et al. (2009) The primed ebolavirus glycoprotein (19-kilodalton GP2): sequence and residues critical for host cell binding. *J Virol* 83: 2883-2891.
6. Lee JE, Fusco ML, Hessel AJ, Oswald WB, Burton DR, et al. (2008) Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. *Nature* 454: 177-182.
7. Lee JE, Saphire EO (2009) Ebolavirus glycoprotein structure and mechanism of entry. *Future Virol* 4: 621-635.
8. Lennemann NJ, Rhein BA, Ndungo E, Chandran K, Qiu X, et al. (2014) Comprehensive functional analysis of N-linked glycans on Ebola virus GP1. *MBio* 5: e00862-00813.
9. Gregory SM, Harada E, Liang B, Delos SE, White JM, et al. (2011) Structure and function of the complete internal fusion loop from Ebolavirus glycoprotein 2. *Proc Natl Acad Sci U S A* 108: 11211-11216.
10. Volchkov VE, Feldmann H, Volchkova VA, Klenk HD (1998) Processing of the Ebola virus glycoprotein by the proprotein convertase furin. *Proc Natl Acad Sci U S A* 95: 5762-5767.