

## Identification of Invariant Peptide Domains within Ebola Virus Glycoprotein GP1, 2

Weltman Joel K

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

### Abstract

Six invariant oligopeptide sequences of length 12 amino acids, or greater, were identified in GP1,2 of ZEBOV Zaire and SEBOV Sudan Ebola viruses. Five of these invariant peptides had positive scores for predicted B-cell epitope scores. In contrast, the amino acid positions of the remaining invariant peptide all displayed negative predicted epitope scores. Invariant regions of Ebola GP1,2 oligopeptides may reflect structural, functional and immunological constraints on the virus, and thus, may be useful as immunological and pharmacological targets.

**Keywords:** Ebola virus; EBOV; Glycoprotein GP1,2; Entropy; Epitope; Bepipred; Antibodies; B-cell; Immunology

### Introduction

A bioinformatic analysis of Ebola virus (EBOV) glycoprotein (GP1,2) [1] is presented here based upon the combined distributions of information entropy [2] and predicted B-cell epitope score [3] in full length GP1,2 protein sequences. Most recognized cases of human infection with EBOV have been caused by ZEBOV (Zaire) and SEBOV (Sudan) [4,5]. Accordingly, this study focuses on GP amino acid sequences of those two EBOV strains.

### Materials and Methods

Computation of information entropy ( $H$ ) [2] and Z-tests were performed with the Enthought Canopy 1.4.1 distribution of 64-bit Python 2.7.6. Z-tests were performed using 1000 pseudo-random trials. The complete sets of ZEBOV (N=148) and SEBOV (N=18) GP1,2 protein sequences were downloaded from the NCBI Genbank [6] in FASTA format [7] on Oct 26, 2014. All of the sequences were full length, ie, 676 amino acids. Consensus sequences for the sequence sets were determined with Jalview [8].

Bepipred [3] analysis was performed on the Immune Epitope Database and Analysis Resource website at <http://www.iedb.org/>.

### Results

The  $H$  distributions in the ZEBOV and SEBOV GP1,2 sequence sets are shown in Figure 1(top) and Figure 1(middle), respectively. The summed total entropy values ( $\Sigma H$ ) for these GP1,2 sequences were:  $\Sigma H(\text{ZEBOV})=27.7303$  and  $\Sigma H(\text{SEBOV})=37.7636$ , respectively. These two  $\Sigma H$  values are indistinguishable ( $Z=1.2320$ ,  $p=0.2178$ ). The  $H$  distribution in the combined {consensus ZEBOV, consensus SEBOV} GP1,2 dataset is shown in Figure 1 (bottom). The summed  $H$  value obtained for the combined, consensus GP1,2 dataset was:  $\Sigma H(\text{ZEBOV, SEBOV})=304.0$ . This increased summed entropy value was significantly greater than that obtained for either the ZEBOV GP1,2 dataset ( $Z=15.1948$ ,  $p=3.8289e-52$ ) or the SEBOV GP1,2 dataset ( $Z=14.1756$ ,  $p=1.2978e-45$ ). These results show that the patterns of distribution of  $H$  in the two GP1,2 datasets are clearly different from each other.

There were 166 GP1,2 sequences in the combined datasets of ZEBOV and SEBOV GP1,2 sequences. The probability ( $p$ ) of a chance mutation at a position at which  $H = 0$  in the ZEBOV, SEBOV and combined consensus sequences is therefore  $p < 1/166 = 0.0060$ . GP1,2 oligopeptides were selected in which at least 12 contiguous amino acid

positions were identical, i.e.,  $H=0$  for the ZEBOV dataset, the SEBOV dataset and for the combined consensus dataset. The probability of such peptides having occurred by chance is thus  $p=(0.0060)^L$ , where  $L$  is the length of the peptide chain. The following invariant peptide sequences, where  $H=0$  at each position with the identical amino acids in both the ZEBOV and SEBOV GP1,2 sequences were located in the complete ZEBOV and SEBOV GP1,2 NCBI datasets. Each of these invariant peptides is of length  $L \geq 12$  amino acids

peptide1 = TKRWGFRSGVPP

peptide2 = YEAGWAENCYNLEIKKPDGSECLP

peptide3 = GAFFLYDRLASTVIYRG

peptide4 = CGLRQLANETTQALQLFLRATTELRT

peptide5 = ILGPDCCIEPHDWTKNITDKI

peptide6 = DNWWTGWRQWIPAGIG

For each of these six invariant peptides, length ( $L$ ), and resulting probability ( $p=(0.0060)^L$ ) of the chance occurrence of an invariant peptide of that length is given in Table 1.  $L$  varies from a minimum of 12 amino acids (peptide1) to a maximum of 26 amino acids (peptide 4). For each of these six invariant peptides the calculated  $p$  value was less than  $2.2840e-27$ . Bepipred scores for each of these six invariant peptides are shown in Table 2. Two of these peptides (peptide 1 and peptide 2) have median, mean and maximum Bepipred scores above the recommended cutoff value of 0.350.

### Discussion

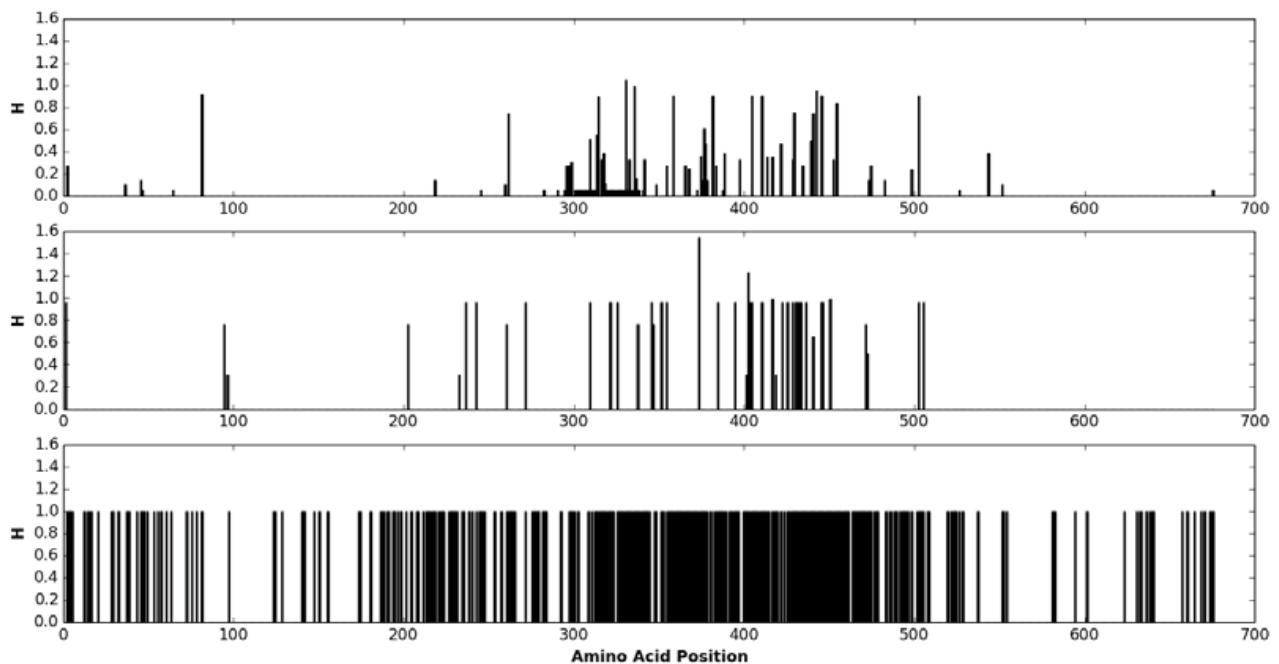
The glycosylated GP1, 2 protein of the EBOV virus is cleaved into a GP1 protein and a GP2 protein [9,10]. These proteins remain disulfide-linked after the proteolytic cleavage [11]. The GP1 protein

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

**Received** November 06, 2014; **Accepted** December 26, 2014; **Published** December 28, 2014

**Citation:** Weltman JK (2014) Identification of Invariant Peptide Domains within Ebola Virus Glycoprotein GP1, 2. J Med Microb Diagn 4: 176. doi:10.4172/21610703.1000176

**Copyright:** © 2014 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.



**Figure 1:** Information Entropy (H) Distribution in Ebola Virus (EBOV) Glycoprotein GP1,2, (top) ZEBOV GP1,2 protein H distribution; (middle) SEBOV GP1,2 protein H distribution; (bottom) H distribution of the {ZEBOV GP12, SEBOV GP12} set of two consensus sequences.

Invariant Peptide	Peptide Length (L)	p
peptide 1	12	< 2.2840e-27
peptide 2	25	< 3.1427e-56
peptide 3	17	< 1.8120e-38
peptide 4	26	< 1.8932e-58
peptide 5	21	< 2.3863e-47
peptide 6	16	< 3.0079e-36

**Table 1:** Length (L) and Probability (p) of Invariant Peptides of Length 12-mer or Greater in ZEBOV and SEBOV Ebola Virus GP1,2 Protein. L is the number of amino acid positions in the peptide. p is the probability of a chance mutation at a position at which H = 0 in the total, combined and consensus ZEBOV and SEBOV sequence datasets. Amino acid sequences of the six invariant peptides are given in the Results text.

is responsible for the binding of the EBOV virus to the membrane of the target cell. The GP2 protein is responsible for internalization of the virus into the target cell.

The immunologic activity predicted in this current study differs from that of a set of antigens originally reported by Wilson et al for EBOV GP1,2 [12]. In the present study, the bioinformatic entropy metric H was used to identify six peptidic regions in ZEBOV GP1,2 and SEBOV GP1,2 that were invariant over sufficient length potentially to function as linear epitopes. Six GP1,2 totally invariant peptides were identified (Table 1). The small probability values obtained for the random occurrence of these contiguous invariant amino acid positions suggests that biological constraints exist that inhibited the mutational process within these segments of the GP1, 2 proteins of both Zaire and Sudan EBOV viruses. Two of these invariant peptides (peptide1 and peptide2) also had positive Bepipred Scores for activity as linear epitopes in the B-cell system (Table 2). Both of these peptides reside within the receptor-binding domain (RBD) of the GP1 protein. One of these predicted epitopic oligopeptides (peptide 2) lies within a domain of GP1 proteins that has been reported to be recognized by human antibodies in recovered ZEBOV patients and/or survivors [13].

Proline has been reported to play a significant role in Ebola antigen structure and function [14]. Three of the peptides identified here (peptide 1, peptide 2 and peptide 5) each contain two proline residues. Proline has been reported to be a common component of epitope-containing segments of proteins, in association with specific effects of proline on protein secondary structure [15,16].

Significant progress has been made towards a nucleotide-based anti-EBOV vaccine [17,18]. The invariant peptides identified in this report may be useful in a vaccine development program. Two of these peptides (peptide 1 and peptide 2) reside within the receptor-binding domain (RBD) of the GP1 protein. Induction of antibodies to these peptides by a vaccine may thus serve as an indicator of a presumptive preventive response to the vaccine. Moreover, the peptide itself may serve as the inducer of a preventive response in a peptide-based vaccine.

By means of simultaneous application of H=0 and Bepipred score 0.350 as selection parameters, the following four invariant oligopeptide sequences were reported to have been present in all ZEBOV and SEBOV GP1,2 sequences in the NCBI GenBank: **FRSGVPP**, **YEAGEWAE**, **KKPDGSECLP** and **HDWTKN** [19]. These four potential B-cell epitopic oligopeptides are components of the longer, yet invariant peptides reported here that were identified by use of H entropy as the sole selection parameter, with subsequent measurement of Bepipred scores of the identified and selected invariant peptides. The additional invariant regions reported here may not be potential B-cell epitopes, but they may reflect other important biological functions such as T-cell antigenicity and structural interactions. If such GP1,2 oligopeptides remain invariant over time, they may indeed prove to be useful therapeutic and preventive targets, as well as means of increasing our understanding of EBOV viral function.

#### Acknowledgments

The Brown University Center for Computation and Visualization (CCV) provided resources and services for studies that enabled this research.

Invariant Peptide	Amino Acid Positions	Median Score	Mean Score	Standard Deviation	Minimum Score	Maximum Score
peptide 1	83-94	0.4575	0.5220	0.3644	-0.0580	1.0510
peptide 2	99-123	0.8440	0.6217	0.5561	-0.3740	1.3980
peptide 3	157-173	-0.8150	-0.9653	0.4250	-1.7700	-0.4040
peptide 4	556-581	-0.3385	-0.3672	0.6510	-1.4410	0.6060
peptide 5	603-623	0.3300	0.3065	0.3597	-0.4390	0.8030
peptide 6	642-657	0.3430	0.3054	0.2380	-0.2300	0.7560

**Table 2:** Predicted B-cell Epitope Scores of Invariant Peptides of Length 12-mer or Greater in ZEBOV and SEBOV Ebola Virus GP1,2 Protein. Amino acid sequences of the six invariant peptides are given in the Results text.

### Author Contributions

Joel K Weltman is the only author of this research.

### Reference

- Beniac DR, Melito PL, Devarenes SL, Hiebert SL, Rabb MJ, et al. (2012) The organisation of Ebola virus reveals a capacity for extensive, modular polyploidy. *PLoS One* 7: e29608.
- Shannon, Claude E (1948) "A Mathematical Theory of Communication". *Bell System Technical Journal* 27: 379-423.
- Larsen JE, Lund O, Nielsen M (2006) Improved method for predicting linear B-cell epitopes. *Immunome Res* 2: 2.
- Dudas G, Rambaut A (2014) Phylogenetic Analysis of Guinea 2014 EBOV Ebola virus Outbreak. *PLoS Curr* 6.
- Sanchez A, Pierre E, Rollin PE (2005) Complete genome sequence of an Ebola virus (Sudan species) responsible for a 2000 outbreak of human disease in Uganda. *Virus Res* 113: 16-25.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler, DL (2006) GenBank. *Nucleic Acids Res* 34 (Database issue): D16-D20. doi: 10.1093/nar/gkj157
- Lipman DJ, Pearson WR (1985) Rapid and sensitive protein similarity searches. *Science* 227: 1435-1441.
- Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25: 1189-1191.
- Dube D, Brecher MB, Delos SE, Rose SC, Park EW, et al. (2009) The primed ebolavirus glycoprotein (19-kilodalton GP1,2): sequence and residues critical for host cell binding. *J Virol* 83: 2883-2891.
- Lee JE, Fusco ML, Ann AJ (2008) Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. *Nature* 454: 177-182
- 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.
- Wilson JA, Hevey M, Bakken R, Guest S, Bray M, et al. (2000) Epitopes involved in antibody-mediated protection from Ebola virus. *Science* 287: 1664-1666.
- Becquart P, Muhlakoiv T, Nkoghe D, Leroy EM (2014) Identification of Continuous Human B-Cell Epitopes in the VP35, VP40, Nucleoprotein and Glycoprotein of Ebola Virus: e96360.
- Rutledge RD, Huffman BJ, Cliffl DE, Wright DW (2008) Design and synthesis of an antigenic mimic of the Ebola glycoprotein. *J Mater Res* 23: 3161-3168.
- Yu H, Chen JK, Feng S, Dalgarno DC, Brauer AW, et al. (1994) Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell* 76: 933-945.
- Vanhoof G, Goossens F, De Meester I, Hendriks D, Scharpé S (1995) Proline motifs in peptides and their biological processing. *FASEB J* 9: 736-744.
- Stanley DA, Honko AN, Asiedu C (2014) Chimpanzee adenovirus vaccine generates acute and durable protective immunity against ebolavirus challenge. *Nature Medicine*; advance online publication.
- Jones SM, Feldmann H, Ströher U, Geisbert JB, Fernando L, et al. (2005) Live attenuated recombinant vaccine protects nonhuman primates against Ebola and Marburg viruses. *Nat Med* 11: 786-790.
- Weltman JK (2014) Combined Use of Information Entropy and Bepipred Scores for Screening Ebola Virus Glycoprotein (GP) Sequences; in *Proceedings of the 1st Int. Electron. Conf. Entropy Appl*: 3-21