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].


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 (∑H) for these GP1,2 sequences were: ∑H(ZEBOV)=27.7303 and ∑H(SEBOV)=37.7636, respectively. These two ∑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: ∑H(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 peptides were located in the complete ZEBOV and SEBOV GP1,2 NCBI datasets. Each of these invariant peptides is of length L >= 12 amino acids

peptide1 = TKRWGFRSVPYP
peptide2 = YEAGEWAENCYLEIKPDPGSECLP
peptide3 = GAFFLYDLASTVYIRG
peptide4 = CGLRQLANETTQLQLFLRATELRT
peptide5 = ILGPDCCEPHDWTKNITDKI
peptide6 = DNNWWTGWRQWPAGG

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
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. 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.
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.

<table>
<thead>
<tr>
<th>Invariant Peptide</th>
<th>Amino Acid Positions</th>
<th>Median Score</th>
<th>Mean Score</th>
<th>Standard Deviation</th>
<th>Minimum Score</th>
<th>Maximum Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>peptide 1</td>
<td>83-94</td>
<td>0.4575</td>
<td>0.5220</td>
<td>0.3644</td>
<td>-0.0580</td>
<td>1.0510</td>
</tr>
<tr>
<td>peptide 2</td>
<td>99-123</td>
<td>0.8440</td>
<td>0.6217</td>
<td>0.5561</td>
<td>-0.3740</td>
<td>1.3980</td>
</tr>
<tr>
<td>peptide 3</td>
<td>157-173</td>
<td>-0.8150</td>
<td>-0.9653</td>
<td>0.4250</td>
<td>-1.7700</td>
<td>-0.4040</td>
</tr>
<tr>
<td>peptide 4</td>
<td>556-581</td>
<td>-0.3385</td>
<td>-0.3672</td>
<td>0.6510</td>
<td>-1.4410</td>
<td>0.6060</td>
</tr>
<tr>
<td>peptide 5</td>
<td>603-623</td>
<td>0.3000</td>
<td>0.3065</td>
<td>0.3597</td>
<td>-0.4390</td>
<td>0.8030</td>
</tr>
<tr>
<td>peptide 6</td>
<td>642-657</td>
<td>0.3430</td>
<td>0.3054</td>
<td>0.2380</td>
<td>-0.2300</td>
<td>0.7560</td>
</tr>
</tbody>
</table>

Author Contributions
Joel K Weltman is the only author of this research.

Reference