Opinion **Open Access**

A Multifactorial Analysis of Ebola Virus Glycoprotein Receptor Binding Domain

Joel K Weltman[*]

Alpert Medical School, Brown University, USA

[*]**Corresponding author:** Joel K Weltman, Clinical Professor Emeritus of Medicine, Alpert Medical School, Brown University Providence, RI 02912 USA, Tel: 4012457588; E-mail: joel_weltman@brown.edu

Abstract

An analysis of the receptor-binding domain (RBD) of the ZEBOV-Makona glycoprotein is presented, based upon the following four factors: information entropy (H), protein conformation, thermal imprint (B factor) and predicted epitope activity (Bepipred Score). It was found that the position of maximum information entropy (Hmax) was located within a helical pentapeptide component of a 31-mer peptide in which H=0 at each amino acid except at the amino acid position where H=Hmax. It is proposed that identification of these RBD peptide components and characteristics can help facilitate efficient design of an anti-Ebola vaccine.

Keywords: Ebola; Glycoprotein; RBD; Information entropy; Conformation; B factor; Bepipred; Epitope; Vaccine

Introduction

The current epidemic of Ebola virus disease (EVD), caused by the Zaire Ebola virus ZEBOV-Makona variant, has killed more than 11,000 people since it began in 2014 (http://apps.who.int/ebola/current-situation/ebola-situation-report-20-january-2016). The receptor binding domain (RBD), extending between amino acids 54 and 201 of the viral glycoprotein (GP) sequence, mediates the binding of the virus to the target cell [1].

Presented here is an analysis of the biology of the ZEBOV-Makona GP RBD based upon the following four factors: information entropy, protein conformation, amino acid thermal flexibility and linear epitope activity. It is proposed that the simultaneous consideration of these factors leads to insights that can facilitate the design of effective anti-Ebola vaccines.

Materials and Methods

The entire dataset (n=965) of full-length (676 amino acids) ZEBOV-Makona variant GP protein sequences was downloaded via the NCBI Ebolavirus Resource (http://www.ncbi.nlm.nih.gov/genome/viruses/variation/ebola) on 21 Dec 2015; 94.09% of these sequences (n=908) were found to be free of error amino acids and were satisfactory for use. The RBD, ranging from GP amino acid 54 to GP amino acid 201 (length=148 amino acids), inclusively, was utilized for this study [1]. The amino acid numbering system of the full-length, intact GP was used for the RBD in this report.

Computations and graphing were performed with Anaconda 2.4.0 (64-bit), Python 2.7.10, Numpy 1.10.1, Scipy 0.16.0 and matplotlib 1.4.3. Consensus sequences were determined with Jalview (2.9.0b2) [2]. Information entropy (H) was computed with the equation of Shannon [3]. Three dimensional conformation of the GP consensus sequence was determined by homology modelling with RaptorX [4] and visualized with CCP4mg 2.10.4 [5]. Local thermal motion, expressed as

the Debye-Waller B factor [6] was obtained from the PDB protein conformation files produced by RaptorX. Predicted linear epitope scores were obtained with Bepipred [7]. Z-tests were performed using 1000 pseudo-random trials and reported with two-tail probabilities.

Results

H distribution in the GP RBD is shown in Figure 1. Total H in the RBD equals 0.7173 bits, distributed among nine amino acid positions (73, 74, 75, 82, 107, 111, 150, 163, 181). Maximum H (0.4412 bits), accounting for 61.50% of the total H, occurred at position 82. This maximum H value was associated with 825 GP sequences with amino acid valine at position 82 (RBD subset V82) and 83 GP sequences with amino acid alanine at position 82 (RBD subset A82); the difference between these observed V82 and A82 RBD subset sequence counts is highly significant (z=25.1332, p=2.1596e-139).
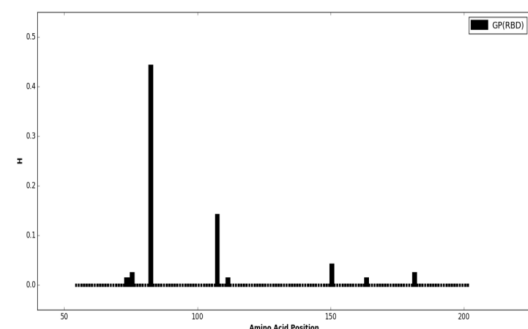
Figure 1: Distribution of Information Entropy (H) in the Zaire Ebola Virus-Makona Glycoprotein (GP) Receptor Binding Domain (RBD). H is in bits. The RBD extends from GP amino acid position 54 to GP amino acid position 201 [1].

The spatial conformation of the ZEBOV-Makona GP RBD is shown in Figure 2 for subset V82 RBD (Figure 2A) and subset A82 RBD (Figure 2B). In both cases, there is an alpha helix structure ranging from amino acid 79 to amino acid 83, with no other alpha helices observed within either of the subset RBDs.
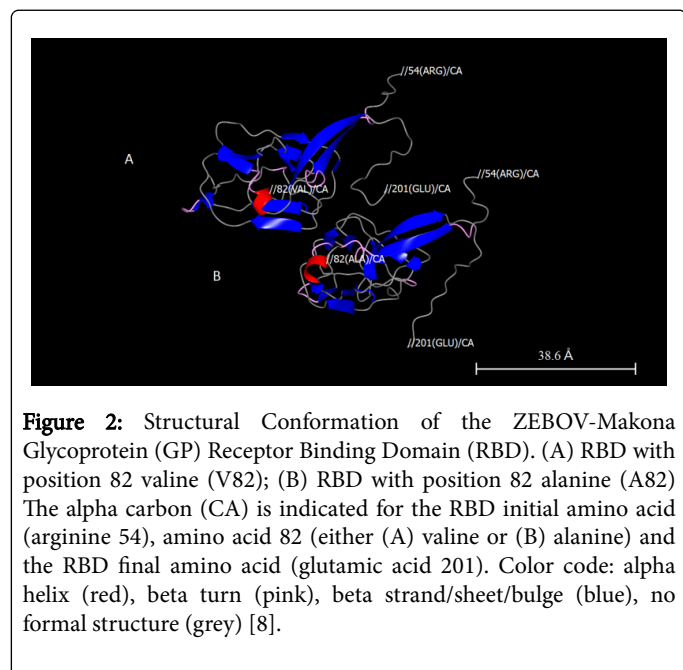
Figure 2: Structural Conformation of the ZEBOV-Makona Glycoprotein (GP) Receptor Binding Domain (RBD). (A) RBD with position 82 valine (V82); (B) RBD with position 82 alanine (A82) The alpha carbon (CA) is indicated for the RBD initial amino acid (arginine 54), amino acid 82 (either (A) valine or (B) alanine) and the RBD final amino acid (glutamic acid 201). Color code: alpha helix (red), beta turn (pink), beta strand/sheet/bulge (blue), no formal structure (grey) [8].

The distribution of Debye-Waller B factor values in the five amino acids comprising the RBD helix is given in Table 1. As shown in Table 1, the B factors at four (79, 80, 81 and 82) of the five amino acids of the V82 RBD alpha helix were statistically indistinguishable from the corresponding four B factor values in the A82 RBD helix. In contrast, at amino acid 83, the final amino acid of the helix, the B factor value was 3.5707-times greater in the V82 RBD than in the A82 RBD; as shown in Table 1; the difference in B factor values at amino acid 83 was statistically highly significant.
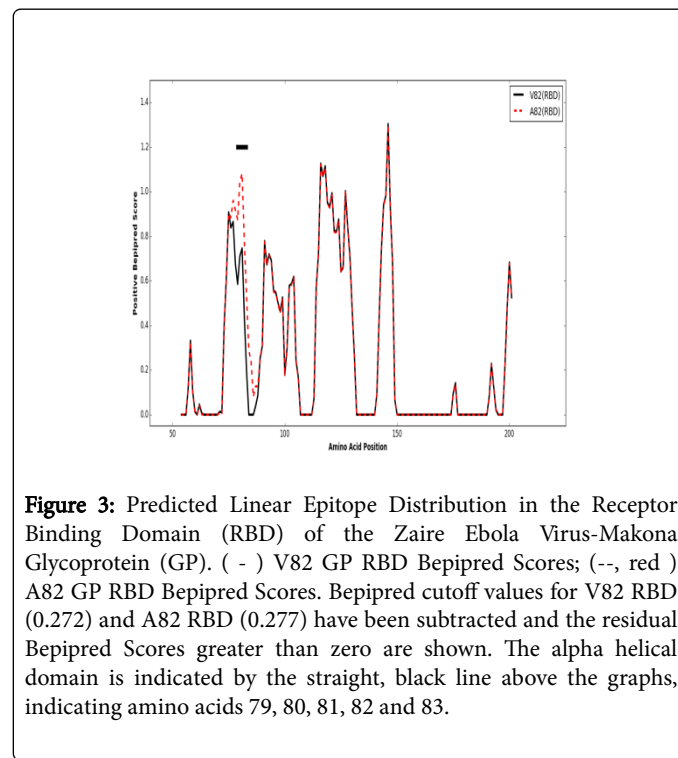
Predicted epitope distributions in the V82 and A82 RBDs are shown in Figure 3 as Bepipred scores. The V82 and A82 Bepipred scores were identical to each other throughout the RBD except at the 13 amino acids 76-88, where A82 Bepipred Score > V82 Bepipred Score at each amino acid position.

Amino Acid Position	V82 RBD Amino Acid	A82 RBD Amino Acid	B Factor	B Factor	Z	p
			V82 RBD	A82 RBD		
79	VAL	VAL	33.83	33.98	0.0184	0.9853
80	PRO	PRO	68.36	69.02	0.057	0.9546
81	SER	SER	28.34	29.51	0.1566	0.8756
82	VAL	ALA	20.58	18.9	0.264	0.7918
83	THR	THR	92.73	25.97	6.262	3.80E-10

Table 1: B Factor Distribution in the Alpha Helix of the GP RBD. B factor is measured in Å2 units. Z-tests were performed with 1000 pseudo-random trials; two-tail p-values are reported.

Discussion

Information entropy (H) within a peptide component of a virus may be regulated by conformational features of that peptide, by the immune response of the host against epitopes within that peptide, and by yet other factors [9]. A conformationally helical peptide would be expected to apply structural constraints [10] upon the occurrence of mutations within that peptide; however, the maximum value of H in the RBD occurred within an alpha helix, despite such structural constraints (Figures 1 and 2). A maximum value of H, despite structural constraints of the helix, may reflect a vigorous immunological response by the host, stimulating mutational escape by the virus [11]. Such an immunological explanation is supported by the observed B factors (Table 1) and Bepipred Scores (Figure 3). The B factor is a crystallographic parameter that represents the thermal imprint and the flexibility of an amino acid component of a protein. The greater the value of B, the greater is the thermal imprint and the flexibility of that amino acid. It has recently been shown that B factor values for epitopic amino acid residues are less than the corresponding values for non-epitopic residues [12]. The B factor value obtained for position 83 within the A82 RBD helix was significantly less than the corresponding amino acid B factor value for position 83 within the V82 RBD helix (Table 1). This result is consistent with the greater observed A82 Bepipred Scores in the helical and perihelical RBD regions (Figure 3). It should be noted that the B factor is not used in the computation of the Bepipred Score. Therefore, the B factor values in Table 1 and the Bepipred Scores are both consistent with a greater antigenicity of the A82 RBD helix compared to the antigenicity of the V82 RBD helix. These results suggest that the numerically greater V82 subset sequence count can be an example of immunological escape [11], i.e., the V82 RBD Ebola virus sequences have escaped the human hosts' immune system in the ongoing the 2014-2016 epidemic of EVD.

Figure 3: Predicted Linear Epitope Distribution in the Receptor Binding Domain (RBD) of the Zaire Ebola Virus-Makona Glycoprotein (GP). (-) V82 GP RBD Bepipred Scores; (--, red) A82 GP RBD Bepipred Scores. Bepipred cutoff values for V82 RBD (0.272) and A82 RBD (0.277) have been subtracted and the residual Bepipred Scores greater than zero are shown. The alpha helical domain is indicated by the straight, black line above the graphs, indicating amino acids 79, 80, 81, 82 and 83.

Conclusions

A pentapeptide containing a position of maximum H was identified in the Zaire Ebola virus GP RBD. The two observed versions of this pentapeptide were VPSVT and VPSAT. This pentapeptide is helical in conformation and is predicted to be epitopic. The pentapeptide containing the position of maximum H resides in a polypeptide domain of 31 amino acids in which H=0 at each of the 30 other amino acid positions in that domain. The two versions of this domain are:

ATDVPSVTKRWGFRSGVPPKVVNYEAGEWAE

ATDVPSATKRWGFRSGVPPKVVNYEAGEWAE

The helical pentapeptides in the above sequences are VPSVT and VPSAT. Data are presented that suggest that the A-version of the peptide is more antigenic than the V-version. It proposed that these 5-mer and 31-mer peptides may be useful as components of anti-Ebola vaccines. The safety and efficacy of a vaccine containing both versions of the most variable amino acid position in the peptide should be determined, especially considering the contradictory effects of valine and alanine on helix formation [13].

References

1. 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.

2. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics 25: 1189-1191.

3. Shannon CE (1948) A Mathematical Theory of Communication. Bell System Technical Journal 27: 379-423.

4. Källberg M, Wang H, Wang S, Peng J, Wang Z, et al. (2012) Template-based protein structure modeling using the RaptorX web server. Nature Protocols 7: 1511-1522.

5. McNicholas S, Potterton E, Wilson KS, Noble MEM (2011) Presenting your structures: the CCP4mg molecular-graphics software. Acta Cryst. D 67: 386-394.

6. Lipkin HJ (2004) Physics of Debye-Waller Factors.

7. Larsen JEP, Lund O, Nielsen M (2006) Improved method for predicting linear B-cell epitopes. Immunome Research 2: 2.

8. Kabsch W, Sanders C (1983) Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen Bonded and Geometrical Features. Biopolymers 22: 2577-2637.

9. Litwin S, Jores R (1992) Shannon information as a measure of amino acid diversity. In Theoretical and Experimental Insights into Immunology 66: 289-296.

10. Wang Z, Atchley WR (2007) Spectral Analysis of Sequence Variability in Basic-Helix-loop-helix (bHLH) Protein Domains. Evol Bioinform Online. 2006 2: 187-196.

11. Woo HJ, Reifman J (2012) A quantitative quasispecies theory-based model of virus escape mutation under immune selection. PNAS 109: 12980-12985.

12. Ren J, Liu Q, Ellis J, Jinyan Li J (2014) Tertiary structure-based prediction of conformational B-cell epitopes through B factors; Bioinformatics 30: i264-i273.

13. Gregoret LM, Sauer RT (1998) Tolerance of a protein helix to multiple alanine and valine substitutions. Fold Des 3: 119-126.