Research Article   **Open Access**

An Immuno-Bioinformatic Analysis of Zika virus (ZIKV) Envelope E Protein

Joel K Weltman*

Alpert Medical School, Brown University, USA

Abstract

By means of the combined use of B cell epitope prediction (Bepipred) and measurement of information entropy

(H) in envelope (E) protein of Zika virus (ZIKV) isolated from infected humans, five amino acid sequences were identified as containing probable epitopes. These five predicted epitopic sequences contained ten amino acid positions where H>0.0. It is proposed that some of the observed entropic positions may reflect mutational escape of the ZIKV from the immune response of the infected host and that such information, applied together with conventional epitope prediction, can guide and facilitate design of anti-ZIKV vaccines.

Keywords: Zika virus (ZIKV); Envelope protein E; Information entropy; B cell epitopes; Vaccines

Introduction

The envelope E protein is a potential target for the development of vaccines against the Zika virus (ZIKV) [1]. An anti-ZIKV vaccine would be especially important because of the probable association of ZIKV infection with microcephaly [2]. Design of an effective anti-ZIKV vaccine may be facilitated by our understanding of the mutational patterns and evolutionary trajectory of the virus. An extensive report of predicted ZIKV envelope E protein epitopes has recently been reported by Mahfuz et al. [3]. Reported here is a bioinformatic analysis of the ZIKV envelope E protein that considers Shannon entropy [4] in addition to B cell epitope prediction [5].

Materials and Methods

Complete sets of full-length ZIKV polyprotein sequences (L=3423, n=10) and envelope E protein sequences (L=251, n=30), without error (X) at any amino acid positions, were downloaded via the NCBI Zika Virus Resource: (http://www.ncbi.nlm.nih.gov/genome/viruses/variation/Zika/) on 02 Feb 2016. These two download sets formed the full-length (L=251), error-free dataset of ZIKV envelope E protein sequences (n=40) used for this study. All of the sequences in the dataset had been isolated from humans. Computations and graphing were performed with Anaconda 2.4.0 (64-bit), Python 2.7.10, Numpy 1.10.1, Scipy 0.16.0, matplotlib 1.4.3. Consensus sequences were determined with Jalview (2.9.0b2) [6]. Information entropy (H) was computed by the equation of Shannon [4]. Z-tests were performed using 1000 pseudo-random trials and are reported with two-tail probabilities. Predicted linear epitope scores were obtained with Bepipred [5]. The reported Bepipred threshold score (0.185) was subtracted from each raw Bepipred score and only differences greater than zero were considered.

Results

Of the 251 positions of the entire E protein, there were 31 amino acid positions at which H>0.0. Total H, summed over these 31 positions was ΣH=12.4904 bits. The mean, median and standard deviation of H at these 31 positions were 0.4029, 0.2864 and 0.2753, respectively. These H values were associated with a total of 134 mutants summed over the entire dataset. The distribution of the H values throughout the E protein is shown in Figure 1. The maximum H values in the dataset occurred at position 26 (H=0.9982, 19 mutants, Z=4.3935, p=1.1156e-05), position

243 (H=0.9928, 18 mutants, Z=4.3277, p=1.5066e-05), position 251 (H=0.9097, 13 mutants, Z=3.6134, p=0.0003) and position 234 (H=0.8112, 10 mutants, Z=3.0992, p=0.0019). The distribution of Bepipred scores in the E protein is shown in Figure 1. Fifteen amino acid positions were located at which both the Bepipred score and the H value were greater than zero. Details for these 15 positions are given in Table 1.

Discussion

Analysis of the ZIKV envelope protein on the bioinformatic level is especially important at this time because of the significant
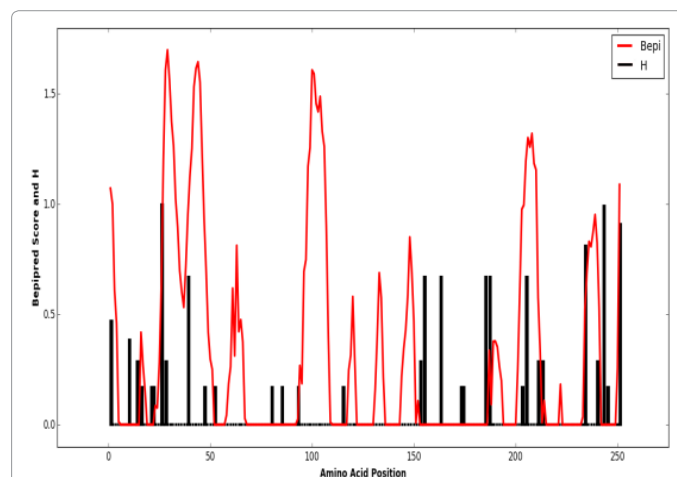

Figure1: Distribution of Information Entropy (H) and Predicted B cell Epitopes in the ZIKV Envelope (E) Protein. Ordinate: H (bits) and Bepipred score; abscissa: amino acid position (1-251). Bepipred scores were corrected by subtracting the threshold score (0.185) and using the resulting values that were greater than zero.

*Corresponding author: Joel K Weltman, Clinical Professor Emeritus of Medicine, Alpert Medical School, Brown University, Providence, RI 02912, USA, Tel No:401-2457588; E-mail: joel_weltman@brown.edu

n	Amino Acid Position	Consensus and Mutant Amino Acid Counts	Information Entropy (H)	Predicted B Cell Epitope Activity (Bepipred Score)	Mutation
1	1	36 GLN, 4 SER	0.4690	1.071	Q->S
2	16	39 SER, 1 THR	0.1687	0.418	S->T
3	26	21 THR, 19 ILE	0.9982	0.640	T->I
4	28	38 HIS, 2 TYR	0.2864	1.608	H->Y
5	39	33 VAL, 7 ILE	0.6690	0.950	V->I
6	47	39 GLU, 1 THR	0.1687	0.932	E->T
7	93	39 LEU, 1 PHE	0.1687	0.025	L->F
8	187	33 VAL, 7 ILE	0.6690	0.336	V->I
9	203	39 ALA, 1 VAL	0.1687	0.977	A->V
10	205	33 THR, 7 ARG	0.6690	1.197	T->R
11	211	38 VAL, 2 ILE	0.2864	0.577	V->I
12	213	38 ALA, 2 VAL	0.2864	0.010	A->V
13	234	30 VAL, 10 ILE	0.8113	0.525	V->I
14	240	38 GLU, 2 LYS	0.2864	0.825	E->K
15	251	27 PRO, 13 THR	0.9097	1.088	P->T

Table 1: Amino Acid Positions in the Envelope Protein within Predicted B Cell Epitopes and with Information Entropy Greater Than Zero. H is in bits. The Bepipred scores were corrected by subtracting the threshold value (0.185) from each score and using only positive results. The total number of sequences equals 40. Mutations were defined and counted as deviants from the consensus amino acid at each position.

public health implications and because of the current need of an effective anti-ZIKV vaccine [1,3]. The present report utilizes epitope prediction in conjunction with observed Shannon entropy in ZIKV E protein produced by virus obtained from humans. It is proposed here that certain of these entropic positions may represent displays of mutational escape of the ZIKV [7] in the human hosts from whom the Zika viruses were isolated. The homogeneity of the mutant populations at the entropic amino acid positions (Table 1) may reflect structural and functional constraints on ZIKV mutational processes. Predicted B cell epitope activity (Bepipred score) was prominent in five amino acid sequences which also contained amino acid positions at which H>0.0, reflecting viral mutational activity in the infected human hosts (Figure 1). These five ZIKV envelope amino acid (aa) sequences are:

(aa 25-51) D**TG**H**ETDENRAKVEVTPNSPRAEATLG

(aa 92-108) P**L**PWHAGADTGTPHWNN

(aa 185-193) TK**V**PAETLH

(aa 201-214) QY**A**G**T**DGPCK**V**PAQ

(aa 233-241) P**V**ITEST**E**N

Amino acid positions in these five sequences where H>0.0 are indicated by a bold font. There were ten such entropic, mutating positions. Consideration of actual, observed entropic events in viruses isolated from infected humans, in combination with epitope prediction, may therefore facilitate design of anti-ZIKV vaccines.

Moreover, two of these sequences (92-108, 233-241) are initiated by a proline residue, a favourable feature in epitopes [8]. It is proposed here that use of information entropy together with B cell epitope prediction can facilitate development of an anti-ZIKV vaccine.

References

1. Faye O, Freire CC, Iamarino A, Faye O, de Oliveira JV, et al. (2014) Molecular evolution of Zika virus during its emergence in the 20(th) century. PLoS Negl Trop Dis 8:e2636.

2. Omer SB, Beigei RH (2016) Pregnancy in the Time of Zika; Addressing Barriers for Developing Vaccines and Other Measures for Pregnant Women. JAMA 315:1227-1228.

3. Mahfuz M, Shawan AK, Mahmud HA, Hasan M, Parvin A, et al. (2014) In Silico Modeling and Immunoinformatics Probing Disclose the Epitope Based PeptideVaccine Against Zika Virus Envelope Glycoprotein. Indian JPharm Biol Res 2:44-57.

4. Shannon Claude E (1948) A Mathematical Theory of Communication. Bell System Technical Journal 27: 379-423.

5. Larsen JEP, Lund, ONielsen M (2006) Improved method for predicting linear B-cell epitopes. Immunome Res 2: 2.

6. 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.

7. Vossen MT, Westerhout EM, Söderberg-Nauclér C, Wiertz EJ (2002) Viral immune evasion: a masterpiece of evolution. Immunogenetics 54: 527-542.

8. Tchernychev BT, Cabilly S, Wilchek M (1997) The epitopes for natural polyreactive antibodies are rich in proline. Proc Natl Acad Sci USA 94:6335-6339.