Journal of Medical Microbiology & Diagnosis

Weltman, J Med Microb Diagn 2016, 5:3
http://dx.doi.org/10.4172/2161-0703.1000235

Research Article Open Access

Computer-Assisted Vaccine Design by Analysis of Zika Virus E Proteins Obtained either from Humans or from Aedes Mosquitos

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 No: 401-2457588; E-mail: joel_weltman@brown.edu

Rec Date: Jun 09 2016; **Acc Date:** Jul 02 2016; **Pub Date:** Jul 12 2016

Abstract

Information entropy (H) and predicted B cell epitope score (Bepipred) were determined for the envelope E protein of Zika viruses (ZIKV) isolated from infected humans and Aedes mosquitos with the aim of identifying E protein regions that may be useful as immunological targets of anti-ZIKV vaccines. Total H of mosquito origin E proteins was 4.2380 greater than that of E proteins of human origin, suggestive of constraints on ZIKV mutation in the human host. Seven invariant peptides (H=0.0) of length 10 amino acids, or greater, were identified. These peptide sequences where H=0.0 were screened for predicted epitopes. The seven invariant peptides were comprised of 93 amino acid residues, 31 of which demonstrated predicted B-cell epitopic activity. The predicted epitopic residues were distributed predominantly to 5 of the 7 invariant peptides. It is proposed that these 5 invariant (H=0) peptides in the E proteins of both human and Aedes mosquito ZIKV represent domains with constrained mutational/evolutionary potential and that epitopes predicted to reside in such invariant domains thus may be stable immunological targets for development of an anti-ZIKV vaccine.

Keywords: *Homo sapiens*; human; Aedes mosquito; Zika virus; Envelope E protein; Information entropy H; B cell epitopes Bepipred; Vaccines

Introduction

An anti-ZIKV (Zika virus) vaccine would be especially important because of the association of ZIKV infection with microcephaly [1,2 and http://www.cdc.gov/media/releases/2016/s0413-zika-microcephaly.html].The envelope E protein is a potential target for the development of vaccines against the Zika virus (ZIKV) [3]. An extensive report of predicted ZIKV envelope E protein epitopes has recently been reported by Mahfuz et al. [4].

A bioinformatic analysis of the ZIKV envelope E protein that considered Shannon entropy [H; 5] in addition to B cell epitope prediction [Bepipred score; 6] was recently published [7]. In that report it was proposed that mutational activity within predicted epitopes may be evidence of mutational escape [8] of the ZIKV from the human immune response. The current manuscript again utilizes the combined use of H and B cell epitope metrics, but extends that usage to ZIKV isolated from Aedes mosquitos as well as to ZIKV from *Homo sapiens*. Furthermore, in contrast to the previous study, this research is focused on detection of potential epitopic sites within mutationally stable peptides, where H=0.0 because such mutationally stable epitopic sites would be efficient and cost-effective targets for protective vaccines. If a viral peptide region is observed to be mutationally stable in species as diverse as humans and mosquitos, such mutational stability is suggestive of structural and/or other biological constraints upon mutation. Computational studies of effects of protein secondary structure on mutational stability and immunogenicity of ZIKV E protein are currently in progress.

Materials and Methods

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. The Mann-Whitney non-parametric U test was performed with Scipy.stats; a one-tail p-value is reported.

The complete set of 280 ZIKV envelope E protein sequences was downloaded via the NCBI Zika Virus Resource (http://www.ncbi.nlm.nih.gov/genome/viruses/variation/Zika/) on 15 May 2016. Parsing and sorting of the sequence data were performed with Python and monitored with Jalview 2.9.0b2 [9]. In the sequence download, the minimum sequence length was 66 amino acids and the maximum sequence length was 3423 amino acids. The complete set of downloaded sequences was parsed into two subsets, one containing 31 sequences, each of length 251 amino acids, and a second subset consisting of 65 sequences, each of length 3423 amino acids. Amino acid subsequence 421-671 of each sequence in the second subset was isolated and combined with the sequences of the first subset, yielding a total of 96 full length (L=251) envelope E protein sequences. This set of full length E protein sequences contained 53 E protein sequences that were isolated from infected *Homo sapiens*, i.e. that were of human origin and 39 E protein sequences that were isolated from infected Aedes mosquitos, i.e. that were of mosquito origin. These 92 E protein sequences, of either human or mosquito origin, were used in this study in fasta format [10].

Information entropy (H) was computed by the equation of Shannon [5]. Predicted linear epitope scores were obtained with Bepipred [6] using the IEDB Analysis Resource (www.iedb.org/). The reported Bepipred threshold score was 0.350.

Results

The distributions of H in ZIKV E protein of human origin, in ZIKV E protein of Aedes mosquito origin and in a dataset of combined E protein sequences of human and Aedes origin are shown in Figure 1. There were 20 amino acid positions with H>0.0 in the E proteins of human origin, 38 positions with H>0.0 in Aedes E proteins and 49 positions with H>0.0 in the combined *Homo sapiens*: Aedes species dataset. The maximum H was 0.2695 bits in E proteins (position 200) from humans, 1.0000 bits in E proteins (position 26) from Aedes species and 0.9832 bits in the combined human:Aedes dataset (position 187). The median, mean and standard deviations of H at the positive (H>0.0) positions were 0.1350, 0.1418 and 0.0293 bits for the human dataset, 0.1720, 0.3162 and 0.2542 bits for the Aedes dataset and 0.0865, 0.2198 and 0.2522 bits for the combined human:Aedes dataset. The non-zero H values in the E proteins of Aedes origin were significantly greater than the non-zero H values in the E proteins of human origin (MannWhitney U=25.0, p=4.2851e-10). Seven invariant peptides, each at least 10 amino acids in length, were identified in both the E proteins of human origin and the E proteins of Aedes origin. These seven invariant peptides were comprised of a total of 93 amino acids, where H=0.0 at each of the amino acid positions. The Bepipred score was positive for predicted epitope activity at 31 of these invariant 93 amino acids (Figure 2). As shown in Figure 2, the epitopic amino acids were clustered into 5 of the 7 invariant peptides. These invariant peptides are further defined in Table 1.

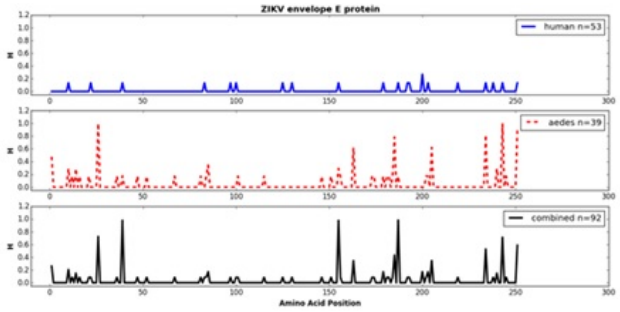

Figure 1: Distribution of Information entropy (H) the ZIKV Envelope E Protein. (top) human origin ZIKVE protein (middle) Aedes species origin E protein (bottom) combined human and Aedes E protein; ordinate: H (bits); abscissa: amino acid position (1-251).

Discussion

ZIKV infection is important because of the association of the virus infection with microcephaly [1,2]. Because of this prenatal risk to infants, there is currently an urgent need for an effective anti-ZIKV vaccine [3,4].

It was recently proposed that epitope prediction, together with Shannon entropy observed in envelope E protein produced by ZIKV from humans [7] provide information potentially useful for vaccine design; it was suggested that these entropic epitopic positions may represent displays of mutational escape of the Zika virus [8] in the human hosts. In the current report, joint application of epitope prediction and observed Shannon entropy was extended to the envelope E protein of ZIKV of Aedes mosquito origin. Despite the significantly greater H values of ZIKV envelope E proteins of Aedes-species origin (Figure 1), 7 invariant peptides (H=0.0) with length of at least 10 amino acids were identified in proteins both of human and of mosquito origin. Epitopic regions were predicted in five of these invariant peptides (Figure 2 and Table 1). Two of these epitopic, invariant peptides (53-66, 102-114) contained one proline residue and two contained two proline residues (206-218, 220-233). Proline is consistent with epitope activity [11].

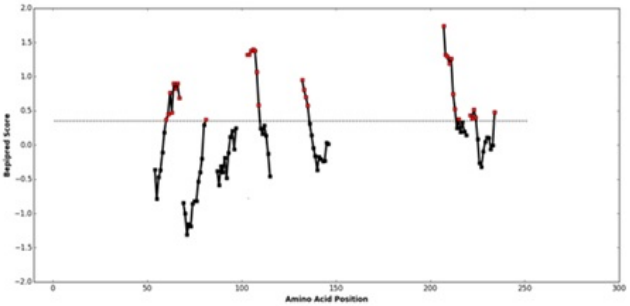

Figure 2: Predicted B-cell Linear Epitope Activity in Invariant (H=0.0) ZIKV Envelope E Protein Peptide Sequences at Least 10 Amino Acids in Length. Ordinate: Bepipred score; abscissa: amino acid position (1-251). The cutoff Bepipred score (0.350) is shown as a dotted horizontal line. Amino acid positions with significant Bepipred scores are in red; amino acid positions with non-epitope Bepipred scores are in black.

Amino Acid Positions	Peptide Length	Amino Acid Sequence
53-66	14	F G S L G L **D C E P R T G L**
68-80	13	F S D L Y Y L T M N N K **H**
86-96	11	E W F H D I P L P W H
102-114	13	**G T P H W N N** K E A L V E
131-145	15	**Q E G A** V H T A L A G A L E A
206-218	13	**D G P C K V P** A **Q** M A V D
220-233	14	**Q T L T** P V G R L I T A N **P**

Table 1: Predicted Epitopic Activity in Invariant (H=0.0) Amino Acid Sequences in ZIKV Envelope E Proteins from *Homo sapiens* and Aedes Species. Amino acid positions with predicted epitope scores above the threshold Bepipred value are shown in bold.

The absence of mutation in peptides with immunological potential may reflect structural and other biological constraints upon ZIKV. The significantly lower H values in ZIKV of human origin (Figure 1) suggest constrained mutation in humans and/or relatively unconstrained mutation in Aedes mosquitos. It is proposed here that use of information entropy together with B cell epitope prediction can facilitate development of an anti-ZIKV vaccine, one with a lowered susceptibility to viral mutational escape.

The analysis presented here and the analysis in the previous report [7] have antithetical points of view regarding the inclusion of entropic

(H>0.0) amino acid positions in the peptide antigens. Certainly, anti-ZIKV vaccines in which H=0.0 at all positions would be more cost-effective. In the laboratory application of the analyses presented here and in the previous report, immunological and other biological effects of omission and inclusion of mutating amino acid positions, i.e, positions at which H>0.0, should be determined.

References

1. Rasmussen SA, Jamieson DJ, Honein MA, Petersen LR (2016) Zika Virus and Birth Defects-Reviewing the Evidence for Causality. N Engl J Med 374:1981-1987.
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. Faye O, Freire CC, Iamarino A, Faye O, de Oliveira JVC, et al. (2014) Molecular evolution of Zika virus during its emergence in the 20(th) century. PLoSNegl Trop Dis 8:e2636.
4. Mahfuz M, Shawan AK, Mahmud HA, Hasan M, Parvin A, et al. (2014) In Silico Modeling and Immunoinformatics Probing Disclose the Epitope Based Peptide Vaccine Against Zika Virus Envelope Glycoprotein. Indian J Pharm Biol Res 2:44-57.
5. Shannon CE (1948) A Mathematical Theory of Communication. Bell Syst Tech J 27: 379-423.
6. Larsen JEP, Lund O, Nielsen M (2006) Improved method for predicting linear B-cell epitopes. Immunome Res 24: 2.
7. Weltman JK (2016) An Immuno-Bioinformatic Analysis of Zika virus (ZIKV) Envelope E Protein. J Med MicrobDiagn 5: 228.
8. Vossen MT, Westerhout EM, Söderberg-Nauclér C, Wiertz EJ (2002) Viral immune evasion: a masterpiece of evolution. Immunogenetics 54:527-542.
9. 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.
10. Pearson WR, Lipman, DJ (1988) Improved tools for biological sequence comparison. Proc Natl Acad Sci 85: 2444-24488.
11. 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.