Journal of Medical Microbiology & Diagnosis

Weltman, J Med Microb Diagn 2017, 6:4
DOI: 10.4172/2161-0703.1000265

Research Article Open Access

A Bioinformatic Glimpse of Human-Origin Zika Virus Polyprotein

Joel K Weltman*

Faculty of Medicine, Alpert Medical School, Brown University Providence, USA

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

Rec Date: October 31, 2017; **Acc Date:** November 27, 2017; **Pub Date:** November 30, 2017

Abstract

An amino acid position with maximum Shannon information entropy and maximum cumulative mutual information is identified in Zika virus polyprotein. This amino acid position is used to sort the subset of Zika virus polyprotein mutations found exclusively in viruses isolated from human hosts but not from vector Aedes mosquitos. The identified mutational amino acid position is a component of a 20-mer peptide in the NS1 protein that has been reported with putative epitopic activity by Freire et al. It is suggested that the observed dual maxima bioinformatic parameters reported here for an exclusively human mutational site support the proposed function of that site in immunological activity.

Keywords: Zika virus ZIKV; Polyprotein; nonstructural protein NS1; Shannon information entropy H; Cumulative mutual information cMI; Secondary structure ss3; Epitopes

Introduction

Zika virus (ZIKV) is a single-stranded, positive sense RNA virus [1]. The Zika virus polyprotein consists of precursors of three structural proteins (capsid, precursor membrane (prM) and envelope (E) protein) and seven major nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) [2]. ZIKV causes microcephaly and other birth defects in a significant fraction of gestationally infected infants and causes Guillain–Barré syndrome in a very small percentage of infected adults [3,4]. The major transmitters of ZIKV to humans are Aedes mosquitos, especially species *Aedes aegypti* [5].

It was recently reported that the set of mutations in polyproteins obtained from ZIKV isolated from humans can be partitioned into two subsets [6]: Exclusive (x) subset and Common (c) subset. Mutations in the x subset occur exclusively in the human host. In contrast, mutations in the c subset occur in ZIKV isolated both from human hosts and from the Aedes mosquito vectors. The present report is focused on identification and proposed immunological significance of the bioinformatically predominant member of the exclusively human (x) subset. It is proposed that the mutations that occurred only in the x subset but not in the c subset may reflect biological processes occurring only in humans but not in mosquitos. Some of these processes may be metabolic, some may be conformational, and some may be immunologic. In the work reported here, an amino acid position with prominent bioinformatic properties is identified within the Zika virus polyprotein and a biologic function of that amino acid position is proposed.

Materials and Methods

Complete sets of full-length (3423 amino acid positions) ZIKV polyprotein sequences isolated either from humans (n=389) or from Aedes mosquitos (n=50) were downloaded from the NCBI Virus Variation Resource (https: //www.ncbi.nlm.nih.gov/genomes/VirusVariation/Database/nph-select.cgi?taxid=64320) on 04 Oct 2017. The following seven species were represented in the Aedes sequences: *Aedes aegypti* (n=32), *Aedes africanus* (n=9), *Aedes dalzieli* (n=2), *Aedes taylori* (n=2), *Aedes luteocephalus* (n=2), *Aedes albopictus* (n=2) and *Aedes opok* (n=1).

Polyprotein domains were assigned by alignment with the default MR 766 reference sequence [7]. Polyprotein sequence management was facilitated with Jalview 2.9.0b2 [8]. Computations were performed on the computer facilities of the Brown University Center for Computation & Visualization (CCV) using Python 2.7.3, Numpy 1.10.4, Scipy 0.15.1 and Matplotlib 1.4.3.

Information entropy (H) was computed by the equation of Shannon [9] and is expressed in bits. H was determined for all amino acid positions in the set of polyproteins isolated from humans, and independently for all amino acid positions in the set of polyproteins isolated from mosquitos. Amino acid positions where H>0.0 in the polyprotein were classified and sorted into Exclusive (x) and Common (c) subsets as previously described [6], depending upon whether (1) a positive H value occurred only at amino acid positions in ZIKV polyproteins obtained exclusively from human hosts or whether (2) the positive H value occurred at amino acid positions in ZIKV polyproteins common both to human hosts and to Aedes species of vector mosquitos.

Mutual information (MI), also in bits, was computed according to Cover and Thomas [10]. Cumulative mutual information (cMI) was computed with exclusion of autocorrelation. Z-tests were performed using 1000 pseudo-random trials and are reported with two-tail probabilities. Secondary structure (ss3) of the non-structural NS1 protein was computed online with RaptorX [11].

Sample Python code for computing the major components within H and cMI, and for sorting human and Aedes polyprotein sequences are in the Supplementary Information file.

Results and Discussion

There were 454 amino acid positions at which H>0.0 in the polyproteins of human origin and 561 of such positions in the

polyproteins of Aedes mosquito origin. Of these entropic amino acid positions, 160 were common to both the sequences of human and of Aedes mosquito origin. These 160 Common positions are not considered further in this study. The set of 294 amino acid positions at which H>0.0 exclusively in the ZIKV polyproteins of human origin were used for further study.

A plot of the information entropy (Hx) as a function of cumultative mutual information (cMIxx) for the Exclusive (x) human set of ZIKV polyprotein amino acid positions at which Hx>0.0 is shown in Figure 1. The xx notation indicates that both members of the pair used to compute mutual information are members of the Exclusive subset. Of the two hundred and ninety-four (294) ZIKV polyprotein amino acid positions at which H>0.0, one position (amino acid position 1118) possessed both the maximum Hx (0.8079 bits) and the maximum cMIxx (1.8648 bits) values of the entire dataset. Because of its maximum values, amino acid position 1118 was used as a reference point for sorting of the complete Hx set of mutations into subsets. To facilitate this sorting process, a straight line was constructed between the double maximum (1.8648, 0.8079) of position 1118 and the double minimum of the origin (0.0000, 0.0000). This straight line enabled sorting the complete set of amino acid positions into three disjoint subsets: on-the-line, above-the-line, and below-the-line. The on-the-line subset was defined as possessing an observed Hx/cMIxx slope within 90-110% of the value predicted by the observed slope of the line (slope=0.4332). The on-the-line subset fulfilling this criterion consisted of 34 positions. A second subset, with observed (Hx, cMIxx) values above the predicted line, consisted of 88 positions. The third subset consisted of 172 positions, all with (Hx, cMIxx) values below the predicted line. The number of positions that were on-the-line was significantly less than both the number of positions above the line ($Z=4.7932$, $p=1.6412 \times 10^{-06}$) and the number below the line ($Z=9.4377$, $p=3.8110 \times 10^{-21}$). In contrast, the number of positions below the line is significantly greater than the number above the line ($Z=5.2784$, $p=1.3031 \times 10^{-07}$). Thus, the sorting process yielded 3 subsets of amino acid positions, each with a statistically significant different count. The characteristics of the amino acid position upon which this sorting is based are described below.

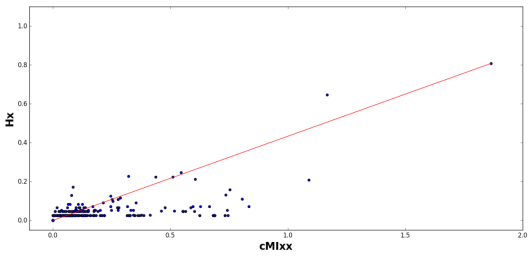

Figure 1: Shannon Information Entropy (Hx) as a Function of Cumulative Mutual Information (cMIxx) in amino acids of polyproteins isolated from the set of Zika viruses isolated exclusively from infected humans. The single x and double xx notation indicates membership of positions (x) and pairs of position (xx) in the exclusive human subset. autocorrelation was not included in the computation of mutual information.

The sorting of mutational positions described above is based upon the mutational position of both maximum information entropy and maximum mutual information. This mutational position was amino acid 1118 of the ZIKV polyprotein. Amino acid 1118 is a component of the nonstructural NS1 protein domain of the ZIKV polyprotein. Position 1118 was occupied by an amino acid in 387 of the 389-total number (99.49%) of sequences in the dataset prepared from humans. These 387 amino acids at position 1118 were (counts in parentheses): wild type=ARG (315), mutant1=TRP (61) and mutant2=GLN (11). Each of the two mutant counts differ significantly from zero: mutant1 $Z=7.9173$, $p=2.4278 \times 10^{-15}$ and mutant2 $Z=3.2354$, $p=0.0012$ with a combined (mutant1, mutant2) probability $p=2.9488 \times 10^{-18}$. It should be noted that the Hopp-Woods [12] hydrophilicity coefficients of these amino acids are: HW(ARG)=3.0, HW(TRP)=-3.4 and HW(GLN)=0.2. The values of the HW coefficients for the observed amino acids thus span the entire hydrophilicity spectrum. The expected effects of these amino acids on the secondary structure of the NS1 protein are shown in Figure 2. Each of the three amino acids at position 1118 is a component of an extended strand. However, there is an increased helical tendency in the wild type ARG1118 NS1 protein, both in the neighboring NH2-region and in several, more distal regions between position 1118 and the NH2-terminus of the protein.

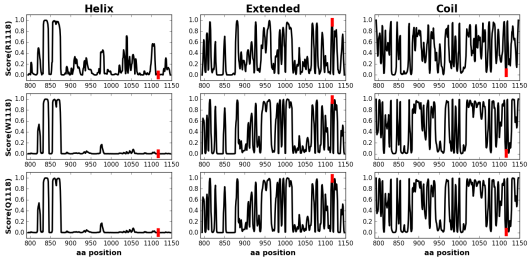

Figure 2: Secondary structure of ZIKV polyprotein non-structural NS1 Domains with amino acid 1118 mutations. top row=Arg1118 wild type; middle row=Trp1118 mutant; bottom row=Gln1118 mutant. Left column=Helix; middle column=Extended strand; right column=random Coil, loop. The red, vertical lines represent polyprotein amino acid position 1118.

Amino acid position 1118 is a member of the Exclusive human subset of mutation positions, i.e., no mutations were observed at position 1118 in the dataset of polyproteins of Aedes origin. The mutation rate at position 1118 in the sequences exclusively of human origin was 72/387=18.60%. Applying that mutation rate to the set of sequences of Aedes origin yields a predicted $50 \times 18.60\%=9.3$ mutations. Instead of the predicted 9.3 mutations, zero mutations were observed. The difference between zero observed mutations and 9.3 predicted mutations is statistically significant ($Z=3.0569$, $p=2.2362 \times 10^{-3}$). It is noted that a larger ZIKV dataset of mosquito origin may be needed to detect a possible shift of position 1118 from the Exclusive to the Common subset. Meanwhile, it may currently be concluded that position 1118 is indeed a member of the Exclusive human subset of mutations and that membership in the Exclusive subset of ZIKV mutations reflects biological processes that occur in human hosts but not in Aedes mosquito vectors. One class of such biological processes is the immunologic response.

ZIKV is known to induce antibody-mediated and cell-mediated immunological responses [13,14]. Most significant to the bioinformatic parameters reported here, polyprotein amino acid 1118 is a component

of a putative 20-mer conformational epitope (CE6) that has been reported [15] for polyprotein amino acids 1105-1124. These 20 amino acids are within the nonstructural protein NS1 domain of the polyprotein; these 20 amino acids occupy numerical positions 311-330 of the mature NS1. This 20-amino acid sequence was computationally identified [15] by structural, conformational and epitopic mapping of Zika virus polyproteins by means of the combined use of Ellipro [16], Epitopia [17] and Discotope [18]. The protein structure and geometric properties are used by Ellipro to computationally predict immunogenic regions of the protein [16]. Protein structure and amino acid sequence are used by Epitopia [17] to computationally predict B-cell antigenicity of the protein. The occurrence of discontinuous B-cell epitopes is computationally predicted by Discotope [18] on the basis of three-dimensional structure and surface accessibility of the protein. Thus, none of these epitope-prediction methods directly depend upon the propensity of a set of amino acids at a given position in a set of sequences to mutate, as does the Shannon information entropy reported here.

Polyprotein position 1118 is position 324 of the mature NS1 protein. CE6 is depicted below as peptide1, along with the variants described in this report as peptide2 and peptide3:

R1118, R324 = WCCREC TMPPLSFRAKDGCW [1]

W1118, W324 =WCCREC TMPPLSFWAKDGCW [2]

Q1118, Q324 = WCCREC TMPPLSFQAKDGCW [3]

The mutation site in polyprotein amino acid 1118 (numerically NS1 amino acid 324) is depicted in red. The data and analysis presented here support and expand the data and analysis presented by Freire et al. [15] and therefore suggest that the observed Shannon entropy may be associated with immunological activity. Unfortunately, not all immunological activity is favorable to the infected host. For example, Zika virus has been shown to inhibit and evade the immune response by interaction with several regulatory physiological processes at the molecular level [19] and to cross-react with antibodies against other flaviviruses, thereby worsening infection through antibody-dependent-enhancement [20,21].

Because of its serious and common effects on infants infected in utero and the serious, albeit rare CNS diseases it causes, Zika virus remains a significant public-health problem [7,22]. As of this writing, there is neither a preventive vaccine nor a treatment for infection by Zika virus. The three peptides reported here, with a highly mutating position at polyprotein amino acid 1118 may help provide a basis for the needed anti-Zika vaccine. Initial analysis of the immunological characteristics of these peptides in an experimental system should be relatively simple, rapid and cost-effective.

Conclusion

It is recognized that the sorting assignment of amino acid position 1118 to the Exclusive human subset may change with time, especially because of the relatively small size of the current set of Zika polyproteins of Aedes mosquito origin (n=50). As stated above, it is also recognized that the bioinformatic maxima of position 1118 reported here may be associated with non-immunological biological processes. Those processes may be manifest as networks of interacting genes detectable by bioinformatic techniques similar to those used here for ZIKV and reported previously for influenza A virus [23]. Indeed, the network analysis previously used for influenza A can be applied to the Exclusive and Common subsets of Zika polyprotein mutational amino acid positions. A network analysis can increase insight into the biological organization and the driving forces behind those mutations. However, in the context of infantile microcephaly and the other complications associated with Zika infection, the results reported here, in agreement with the findings of Freire et al. [15], suggest that the immunogenicity, toxicity and protective effectiveness of these three peptides should expeditiously be tested experimentally for potential clinical usefulness.

Acknowledgement

This research was conducted using computational resources and services of the Center for Computation and Visualization (CCV), Brown University.

References

1. Saiz J-C, Vázquez-Calvo A, Martín-Acebes MA (2016) Zika Virus: the Latest Newcomer. Front Microbiol 7: 496.
2. Cox BD, Stanton RA, Schinazi RF (2016) Predicting Zika virus structural biology: Challenges and opportunities for intervention. Antiviral Chem Chemother 24: 118–126.
3. Cao B, Diamond MS, Mysorekar IU (2017) Maternal-fetal transmission of Zika virus: Routes and signals for infection. J Interferon Cytokine Res 37: 287-294.
4. Ellington SR, Devine O, Bertolli J, Martinez Quiñones A, Shapiro-Mendoza CK, et al (2016) Estimating the number of pregnant women infected with Zika virus and expected infants with microcephaly following the Zika virus outbreak in Puerto Rico. JAMA Pediatr 170: 940-945.
5. Ayllón T, Campos R, Brasil P, Morone F, Câmara D, et al. (2017) Early Evidence for Zika virus circulation among Aedes aegypti mosquitoes, Rio de Janeiro, Brazil. Emerg Infect Dis 23: 1411-1412.
6. Weltman JK (2017) Exclusive and common Subsets of Zika virus polyprotein mutants. J Med Microb Diagn 6: 256.
7. Petersen LR, Jamieson DJ, Powers AM Honein MA (2016) Zika Virus. N Engl J Med 374: 1552-1563.
8. 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.
9. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27: 379-423, 623-656.
10. Cover TM, Thomas JA (2006) Entropy, relative entropy and mutual information. In: Elements of information theory (2nd edn), Chapter 2. Wiley, USA. pp. 13-56.
11. Källberg M, Wang H, Wang S, Peng J, Wang Z, et al (2012) Template-based protein structure modeling using the RaptorX web server. Nat Protoc 7: 1511-1522.
12. Hopp TP, Woods KR (1981) Prediction of protein antigenic determinants from amino acid sequences. Proc Natl Acad Sci USA 78: 3824-3828.
13. Robbiani DF, Bozzacco L, Keeffe JR, Khouri R, Olsen PC, et al (2017) Recurrent potent human neutralizing antibodies to Zika virus in Brazil and Mexico. Cell 169: 597-609.
14. Andrade DV, Harris E (2017) Recent advances in understanding the adaptive immune response to Zika virus and the effect of previous flavivirus exposure. Virus Res pii: S0168-1702: 30462-30468.
15. Freire MCLC, Pol-Fachin L, Coêlho DF, Viana IFT, Magalhães T, et al (2017) Mapping putative Bcell Zika virus NS1 epitopes provides molecular basis for Anti-NS1 antibody discrimination between Zika and Dengue viruses. ACS Omega 2: 3913−3920.
16. Ponomarenko JV, Bui H, Li W, Fusseder N, Bourne PE, et al. (2008) ElliPro: A new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics 9: 514.

17. Rubinstein ND, Mayrose I, Martz E, Pupko T (2009) Epitopia: A web-server for predicting B-cell epitopes. BMC Bioinformatics 10: 287.
18. Kringelum JV, Lundegaard C, Lund O, Nielsen M (2012) Reliable B cell epitope predictions: Impacts of method development and improved benchmarking. PLoS Comput Biol 8: e1002829.
19. Asif A, Manzoor S, Fatima-Tuz-Zahra, Salim M, Ashraf M, et al (2017) Virus: Immune evasion mechanisms, currently available therapeutic regimens and vaccines. Viral Immunol 30: 1-9.
20. Willis E, Hensley SE (2017) Characterization of Zika virus binding and enhancement potential of a large panel of flavivirus murine monoclonal antibodies. Virology 508: 1-6.
21. Xu X, Vaughan K, Weiskopf D, Grifoni A, Diamond MS, et al. (2016) Identifying candidate targets of immune responses in Zika virus based on homology to epitopes in other flavivirus species, (1st edn). PLOS Currents Outbreaks.
22. Oduyebo T, Polen KD, Walke HT, Reagan-Steiner S, Lathrop E, et al (2017) Update: Interim guidance for health care providers caring for pregnant women with possible Zika virus exposure-United States (Including U.S. Territories), July 2017. MMWR Morb Mortal Wkly Rep 66: 781-793.
23. Thompson WA, Weltman JK (2012) Intergenic subset organization within a set of geographically-defined viral sequences from the 2009 H1N1 influenza A pandemic. Amer J Mol Biol, 2: 32-41.