Computational Biology and Chemistry 68 (2017) 143–152

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

A Bioinformatics approach to designing a Zika virus vaccine☆

Sumanta Dey[a,b], Ashesh Nandy[a,*], Subhash C. Basak[c], Papiya Nandy[a], Sukhen Das[a,b]

[a] Centre for Interdisciplinary Research and Education, 404 B Jodhpur Park, Kolkata 700068, India
[b] Department of Physics, Jadavpur University, Jadavpur, Kolkata 700032, India
[c] University of Minnesota Duluth-Natural Resources Research Institute and Department of Chemistry and Biochemistry, University of Minnesota Duluth,5013 Miller Trunk Highway, Duluth, MN 55811, USA

ARTICLE INFO

Article history:
Received 15 July 2016
Received in revised form 3 March 2017
Accepted 5 March 2017
Available online 10 March 2017

Keywords:
Zika virus
Zika vaccine
2D graphical representation
Alignment-free models
Epitope regions

ABSTRACT

The Zika virus infections have reached epidemic proportions in the Latin American countries causing severe birth defects and neurological disorders. While several organizations have begun research into design of prophylactic vaccines and therapeutic drugs, computer assisted methods with adequate data resources can be expected to assist in these measures to reduce lead times through bioinformatics approaches. Using 60 sequences of the Zika virus envelope protein available in the GenBank database, our analysis with numerical characterization techniques and several web based bioinformatics servers identified four peptide stretches on the Zika virus envelope protein that are well conserved and surface exposed and are predicted to have reasonable epitope binding efficiency. These peptides can be expected to form the basis for a nascent peptide vaccine which, enhanced by incorporation of suitable adjuvants, can elicit immune response against the Zika virus infections.

1. Introduction

Zika virus infections have reached epidemic proportions in the New World, especially affecting pregnant women and leading to high levels of microcephaly in newborns (Victoria et al., 2016). The incidences of such cases and other neurological disorders such as Guillaine-Barre syndrome coinciding with the spread of the Zika virus infections, the lack of preventive or therapeutic medications against the virus and the prospect of further expansion of the virus, have prompted the World Health Organization (WHO) to declare on 1st February 2016 these disorders as Public Health Emergency of International Concern (WHO, 2016a). Several countries have set-up machinery to combat the effects of the Zika virus (ZIKV) through public measures and heightened public awareness (Elachola et al., 2016; Roa, 2016; Rasmussen et al., 2016) while the National Institute of Allergy and Infectious Diseases (NIAID) under US National Institute of Health (NIH) is responding to the Zika virus crisis through vaccine, treatment, and clinical trials research (Anon., 2016a) to alleviate the suffering.

The Zika virus had been isolated almost 70 years ago from a monkey in Uganda's Zika forest but remained out of frontline research interest although it had been detected in several African and Asian countries in the subsequent period. It is a vector-borne disease spread through the bites of the *Aedes aegypti* and *Aedes albopictus* mosquitoes, whose ranges have been increasing in recent times because of global warming (IPCC 4th, 2007; Shuman, 2010). The virus first attracted limited attention when a Zika epidemic erupted in Yap island in Micronesia in 2007 (Duffy et al., 2009) and again in French Polynesia in 2013 (Heymann et al., 2016). However, a sudden increase in cases of microcephaly in newborns detected in Latin American countries late last year catapulted the virus to world attention (WHO, 2016a). Lack of any drugs or vaccine against the new disease led to inadequate containment of the disease. Development of new drugs from laboratory bench to market takes years of effort and billions of dollars; development of vaccines in the traditional way costs only slightly less in time and effort. A more rapid response to the Zika virus epidemic would appear to be highly desirable (Basak and Nandy, 2016).

Traditional vaccines consist of attenuated, inactivated or subunit cultures, all of them cultivated from the virus. While the former types have chances of regression or cause genetic

Abbreviations: MDPI, Multidisciplinary Digital Publishing Institute; DOAJ, directory of open access journals; HLA, Human Leukocyte Antigen; IEDB, Immune Epitope Database; ABCpred, artificial neural network based B-cell epitope prediction; NIAID, National Institute of Allergy and Infectious Diseases; NIH, National Institute of Health; NCBI, National Centre for Biotechnology Information.

☆ This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

* Corresponding author.
E-mail address: anandy43@yahoo.com (A. Nandy).

http://dx.doi.org/10.1016/j.compbiolchem.2017.03.002
1476-9271/

damage, subunits consist of individual chemically purified components which can evoke immune response but can be expensive and also have problems like side effects, possible leakage of infectious agents, etc. (Sobolev et al., 2005). A new strategy of developing synthetic peptide vaccines is taking shape (Moisa and Kolesanova, 2012) where the virus genome is scanned for the particular protein antigens that elicit immune response and these are selected for synthesis into a peptide vaccine. A more focused approach is to precisely locate the epitope regions within this antigen and present those for immune response (Purcell et al., 2007). Wide scale computer based approaches are essential to this purpose (Basak and Nandy, 2016). We have applied such methods to identify target regions in surface proteins of influenza virus and rotavirus (Ghosh et al., 2010, 2012; Sarkar et al., 2015). Initially we scan a library of sequences of the designated protein to determine segments that are unchanged or least changed among the various strains, then couple these with average solvent accessibility profiles of the sequences to select those segments that are most conserved and have highest solvent accessibility profile. We then determine the T-cell linear epitopes with acceptable binding affinity to human MHC class 1 and class II and finally select those peptides that pose no autoimmune threat, whereas for conformational epitopes we search for B-cell epitopes and try to find out whether our segments are also part of the conformational epitopes; in the process, the 3D crystal structure of the surface protein is utilized to ensure that the selected peptide regions are not covered against solvent accessibility due to neighboring proteins that together constitute the quaternary structure.

The Zika virus is a positive-sense non-segmented RNA genome about 10700 bases long that codes for three structural proteins – the capsid, the pre-membrane/membrane and envelope genes – and seven non-structural genes referred to as NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5. The Zika virus belongs to the *flavivirus* family which includes Japanese encephalitis, West Nile, Yellow fever, and dengue viruses. Vaccines are available against most of these latter viruses and the World Health Organization has organized resources to develop an inactivated vaccine against the Zika virus (WHO, 2016b). In the meantime, the National Institute of Health in the USA has started work on a Zika vaccine based on the West Nile virus vaccine (Anon., 2016d) and Bharat Biotech in India has claimed to have developed a Zika vaccine, Zikavac, that is ready for pre-clinical trials (Bagla, 2016).

Whereas peptide vaccines can be designed to be more focused, none of the above attempts to devise a Zika vaccine have claimed a peptide vaccine approach. We launched our project to devise a peptide vaccine against the Zika virus based on studies of the sequences of the Zika envelope (ZIKV-E) protein. The search for peptide targets in the Zika virus, however, suffered on two grounds: insufficient sequence data and lack of crystal structure at the time of our initial study, areas of concern we had touched upon in our recent comment and review (Nandy and Basak, 2016a,b). The recent publication of the structure of mature Zika virus (Sirohi et al., 2016) solved one issue, we improvised on the other and applied our methodology to arrive at a set of four possible peptide regions that could be utilized in designing a vaccine against the Zika virus. These are indicators only to the eventual peptide vaccine design; more refinement, selection of appropriate adjuvants, delivery systems and lab trials are essential steps that are yet to be undertaken.

2. Results and discussions

Since a major part of our downloaded sequences were partial cds (listed in Table S1), we first needed to understand which part of the full protein they represent. An alignment exercise through MEGA5.22 software showed that all these protein segments were almost perfectly aligned (221 matches out of 251 amino acids), so we were assured that they all represented the same part of the protein; nucleotide matches were understandably less, 557/753. To determine which part of the whole protein this fragment represented, we utilized an alignment free method by doing 2D graphical representation of the nucleotide sequences of the envelope segment from the DQ859059 Uganda MR766 1947 polyprotein gene, and one partial cds sequence, KF383016 Senegal 2001 (Fig. 1). (BLAST or MEGA approaches could have helped but alignments includes gaps, etc. which we avoid with our alignment-free model.) From the visual clues we could determine an approximate location and referring to the sequence of the Uganda gene we determined that the partial sequence start point coincided with nucleotide number 393 of the Ugandan gene and ending at around nucleotide number 1140. A check with amino acid (aa) sequences showed that the partial cds protein sequences of the Senegal sample had an extra peptide segment, DIGHETD (at aa no. 25 of fragment; corresponding position in Ugandan amino acid sequence is 155; see Table 1 for schematic), otherwise the match was very good; the extra 21 bases and other mutational changes cause the slight differences between the two plots in the corresponding areas. As an additional check we did the same exercise with KU926310 Brazil 2016 envelope gene (figure not shown) and found that the gene fragments matched with the corresponding segment of the Brazil 2016 gene which had the 21 nucleotides that were absent in the Ugandan sequence.

The ASA (average solvent accessibility) profiles were predicted from the SABLE server (Anon., 2016f); we smooth out the numbers by taking a running average over the numbers for 12 amino acids at a time. Getting the solvent accessibility profile of the protein

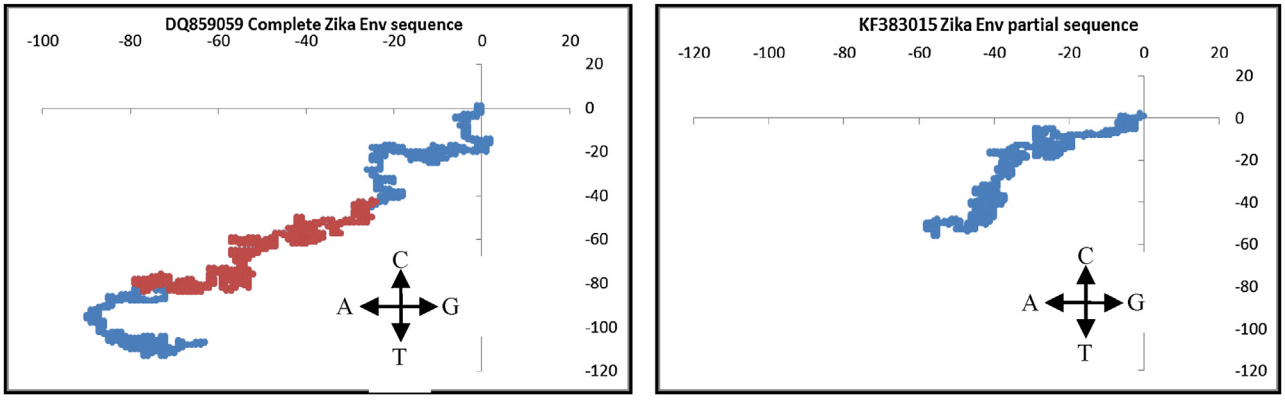

Fig. 1. The 2D graphical representation of a complete Zika envelope gene and a partial cds. The part of the full gene (from base no. 391 to 1143) that most closely matches the partial cds is coloured red.