

Rational Design of Peptide Vaccines against the Zika Virus through sequence Descriptors: Techniques and Problems

Ashesh Nandy^{1*} and Subhash C Basak²

¹Centre for Interdisciplinary Research and Education, Jodhpur Park, Kolkata, INDIA

²University of Minnesota Duluth-Natural Resources Research Institute and Department of Chemistry and Biochemistry, University of Minnesota Duluth, Duluth, USA

Received: February 22, 2016; Accepted: March 02, 2016; Published: March 15, 2016

*Corresponding author: Centre for Interdisciplinary Research and Education, 404B Jodhpur Park, Kolkata 700068, INDIA, E-mail: anandy43@yahoo.com

Short Communication

The sudden emergence of an epidemic of Zika virus infections in South America has raised concerns of its virulence and transmission potential, especially in view of mass gatherings at carnivals, Olympic Games, Hajj and others within this year [1]; the as yet unproven link between the Zika epidemic and heightened cases of microcephaly among newborns for which the World Health Organization (WHO) have declared on February 1, 2016 the suspected association of the virus to cases of microcephaly and Guillain-Barre disease to be Public Health Emergency of International Concern [2] has added to the anxieties. The main vector of the Zika virus is the *Aedes aegypti* mosquito prevalent in tropical countries but whose range is increasing with the effects of global warming, recognized as causing shifts in distribution of vector-borne diseases [3]; spread of the virus through mass gatherings to other countries becomes an additional worrying possibility. The issues here are of serious concern since the *Aedes aegypti* mosquito is also a primary carrier of the dengue virus and Brazil has not had much success in controlling that [4]. Coming soon after the scare of Ebola virus, and with the background of the 2009 H1N1 pandemic, the MERS and SARS epidemics, the near-pandemic of H5N1 bird flu and fatalities associated with the H7N9, the Zika virus epidemic has focused questions on containment and mitigation of viral activity and further zoonotic issues that may arise.

Such recurrent viral incidences and emergence necessitate a combinatorial analysis and perspective on confronting and containing the menace. Given the nature and diversity of infective viruses, it would seem difficult for such a task to be contemplated by a priori experimental means, whereas theories behind such phenomena are yet to be devised. In such a circumstance, phenomenological approaches as offered by numerical characterizations of bio-molecular sequences arising from their graphical representations hold some promise of approaching the issue. Numerical characterizations of DNA, RNA and protein sequences provide descriptors that are characteristic of each sequence, albeit depending upon the particular methodology

used [5]. Developing upon initial work on DNA sequences, these descriptors have been used to characterize genes to genomes [5] and proteins [6] to proteomes [7], and have seen applications in many areas. While Nandy [8-10], Nandy and Nandy [11] and Larionov et al [12] have used 2D graphical systems to determine DNA systematics, Liao et al [13,14] used the novel techniques for alignment-free phylogeny to determine sequence ancestry, Wiesner and Wiesnerova [15] found the new methodology giving better insights into germ-plasm identifiers, Gonzalez-Diaz and his group presented several papers using the concept for alignment-free prediction of polygalacturonases [16], alternative "in silico" technique for chemical research in toxicology [17] and predicting antimicrobial drugs and targets [18], Nandy et al [19] were able to model influenza hemagglutinin and neuraminidase interdependence which provided predictability to new possible viral assortments [20]. The technique of numerical characterization was extended to proteins initially by Randic et al [21] and led to several approaches being proposed [6, 22, 23], analyzing phylogenetic relationships between protein families [24], hydropathy profiles of amino acids [25] and others. Thus, numerical characterization of bio-molecular sequences can be considered to have wide-ranging applicability leading to acceptable results, and occasionally new insights.

An important application was done in anti-viral vaccine design by Nandy and his group [26, 27] who used protein sequence descriptors to determine surface exposed conserved segments of viral coat envelope proteins for eventual generation of immune response. These segments were analyzed further for epitope potential and auto-immune protection and the segments that passed these tests were predicted to be usable as peptide vaccines, individually or in a cocktail of several peptides for a higher level of protection. Such peptide candidate vaccines have been predicted for influenza [26] and rotaviruses [27] and more viruses such as human papillomavirus have been investigated. These techniques provide a methodology for the rational design of peptide vaccines, but the veracity of these predictions need experimental tests before they can be considered as viable vaccine candidates. However, since synthetic peptides are

relatively easy and inexpensive to manufacture, if proved viable such procedures hold possibilities of rapid deployment in case of epidemic outbreaks.

The Zika virus is a case in point, but there is a problem. The bioinformatic analysis of the kind mentioned here require voluminous data, which unfortunately are not available in this instance, primarily because the scientific community had been caught unawares, although some analyses have been done with limited data [28,29]. However, paucity of data is a blight in another instance where the viral causes are well documented: human papillomavirus. Although this virus exists in over 170 variants, and several of them cause cancer, there are none to very few sequences of some of the important variants available in the NIH GenBank database. Although available vaccines against the primary types of this dsDNA (double-stranded DNA) virus have proved very effective, they leave out 20-30% of the cases [30] and continue to be very expensive, often beyond the capacity of affordability of a vast swathe of human population. It is possible that adequate data are available but are being held back for unexplained reasons, but lack of such data hinders intensive analyses and prolongs the sufferings of a large section of humanity.

In the case of the Zika virus where data have been wanting in the first place, the questions of surveillance and preparedness come to the fore. Given the very large number of viruses that can or may attack humans, the possibility of data domination in the near future seems remote. Added to that is the recent experience of increasing incidences of zoonotic diseases gaining ground and imperiling whatever little resource we can expect to mobilize.

With the huge lead time and costs involved in developing new medications for therapeutic regimes against a disease, using the human body's built in mechanisms to generate immune responses would seem the better alternative where this is applicable, feasible or advisable. Using sequence descriptors to this end and developing fast response synthetic peptide vaccine would appear a possible alternative. However, current techniques require a very large number of sequences to be analyzed to get at the usable peptide segments, which may be difficult in the case of new viral infections where data may be few and far between. It would seem imperative to develop more robust analytical techniques which may be able to identify the appropriate peptide segments with much lesser amount of data. Some research is being undertaken to realize this goal, but until such time that the new methods are proven, it would be a great service to humanity if the molecular databanks are entrusted with data that may be lying elsewhere.

References

1. Elachola H, Gozzer E, Zhuo J, Memish ZA. A crucial time for public health preparedness: Zika virus and the 2016 Olympics, Umrh, and Hajj. *Lancet*. 2016;387(10019):630-632. doi: 10.1016/S0140-6736(16)00274-9.
2. WHO. "WHO Director-General summarizes the outcome of the Emergency Committee regarding clusters of microcephaly and Guillain-Barré syndrome". World Health Organization. 2016.
3. Climate Change 2007 Impacts, Adaption and Vulnerability. In: Editor name, Martin L. Parry. 8.2.8 Vector-borne, rodent-borne and other infectious diseases. 2007;403-405.
4. Horton R. Offline: Brazil—the unexpected opportunity that Zika presents. *Lancet*. 2016;387(10019):633. DOI: doi.org/10.1016/S0140-6736(16)00268-3
5. Nandy A, Harle M, Basak SC. Mathematical descriptors of DNA sequences: development and applications, *ARKIVOC*. 2006;9:211-238.
6. Randic M, Zupan J, Balaban AT, Topic DV, Plavsic D. Graphical Representation of Proteins. *Chem. Rev*. 2011;111(2):790-862. DOI: 10.1021/cr800198j
7. Basak SC, Gute BD. Mathematical descriptors of proteomics maps: Background and applications. *Curr. Opin Drug Discov Dev*. 2008;11(3):320-326.
8. Nandy A. Two dimensional graphical representation of DNA sequences and intron-exon discrimination in intron-rich sequences. *Comput Appl Biosci*. 1996;12(1):55-62.
9. Nandy A. Graphical analysis of DNA sequence structure: III. Indications of evolutionary distinctions and characteristics of Introns and Exons. *Current Sc*. 1996;70(7):661-668.
10. Nandy A. Empirical Relationship between Intra-Purine and Intra-Pyrimidine Differences in Conserved Gene Sequences. *PLoS ONE*. 2009;4(8):e6829. doi:10.1371/journal.pone.0006829
11. Nandy A, Nandy P. Graphical analysis of DNA sequence structure: II. Relative abundances of nucleotides in DNAs, gene evolution and duplication, *Current Sc*. 1995;68(1):75-85.
12. Larionov S, Loskutov A, Ryadchenko E. Chromosome evolution with naked eye: Palindromic context of the life origin. *CHAOS*. 2008;18(1):013105. doi: 10.1063/1.2826631.
13. Liao B, Tan M, Ding K. Application of 2-D graphical representation of DNA sequence. *Chemical Physics Letters*. 2005;414(4-6):296-300
14. Liao Bo, Liu Y, Li R, Zhu W. Coronavirus phylogeny based on triplets of nucleic acids bases. *Chem. Phys. Lett*. 2006;421(4-6):313-318.
15. Wiesner I, Wiesnerova D. 2D random walk representation of Begonia × tuberhybrida multiallelic loci used for germplasm identification. *Biologia Plantarum*. 2010;54(2):353-356. DOI: 10.1007/s10535-010-0062-7
16. Aguero-Chapin G, Varona-Santos J, De la Riva GA, Antunes A, Gonzalez-Villa T, Uriarte E, et al. Alignment-free prediction of polygalacturonases with pseudofolding topological indices: experimental isolation from coffea arabica and prediction of a new sequence. *J. Proteome Res*. 2009;8(4):2122-2128. doi: 10.1021/pr800867y.
17. Gonzalez-Diaz H, Prado-Prado F, Ubeira FM. Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach. *Curr Top Med Chem*. 2008;8(18):1676-1690.
18. Cruz-Monteagudo M, González-Díaz H, Borges F, Dominguez ER, Cordeiro MN. 3D-MEDNEs: an Alternative "in silico" technique for chemical research in toxicology. 2. Quantitative Proteome-Toxicity Relationships (QPTR) based on Mass Spectrum Spiral Entropy. *Chem Res Toxicol*. 2008;21(3):619-632. doi: 10.1021/tx700296t
19. Nandy A, Sarkar T, Basak SC, Nandy P, Das S. Characteristics of influenza HA-NA Interdependence determined through a graphical technique. *Curr Comput Aided Drug Design*. 2014;10(4):285-302.
20. Nandy A, Basak SC. An emerging immunogenomics and computational approach for peptide vaccinology: Rational design of peptide vaccines. *Curr Comput Aided Drug Des*. 2014;10(4):283-284.

21. Randić M. 2-D graphical representation of proteins based on virtual genetic code. SAR and QSAR in Environmental Research. 2004;15(3):147-157.
22. Nandy A, Ghosh A, Nandy P. Numerical Characterization of Protein Sequences and Application to Voltage-Gated Sodium Channel Alpha Subunit Phylogeny, *In Silico Biology*. 2009;9(3):77-87.
23. Czerniecka A, Bielińska-Waz D, Waz P, Clark T. 20D-dynamic representation of protein sequences. *Genomics* 2016;107(1):16-23. doi: 10.1016/j.ygeno.2015.12.003.
24. Bai F, Wang TM. On Graphical and Numerical Representation of Protein Sequences. *Journal of Biomolecular Structure & Dynamics*. 2006;23(5):539-545.
25. Xie X, Zheng L, Yu Y, Liang L, Guo M, Song J, Yuan Z. Protein sequence analysis based on hydrophathy profile of amino acid. *Journal of Zhejiang University SCIENCE B*. 2012;13(2):152-158. doi: 10.1631/jzus.B1100052
26. Ghosh A, Nandy A, Nandy P. Computational analysis and determination of a highly conserved surface exposed segment in H5N1 avian flu and H1N1 swine flu neuraminidase. *BMC Struct Biol*. 2010;10:6. doi:10.1186/1472-6807-10-6.
27. Ghosh A, Chattopadhyay S, Sarkar MC, Nandy P, Nandy A. In Silico Study of Rotavirus VP7 Surface Accessible Conserved Regions for Antiviral Drug/Vaccine Design. *PLoS ONE*. 2012. DOI: 10.1371/journal.pone.0040749
28. Faye O, Freire CCM, Iamarino A, Faye O, de Oliveira JVC, Mawlouth Diallo, et al. Molecular Evolution of Zika Virus during Its Emergence in the 20th Century. *PLoS Negl Trop Dis*. 2014;8(1):e2636. doi:10.1371/journal.pntd.0002636
29. Shawaan MMAK AK, Mahmud HA, Hasan MM, Parvin A, Rahman MN, Rahman SMB. In Silico Modeling and Immunoinformatics Probing to Disclose the Epitope Based Peptide Vaccine Against Zika Virus Envelope Glycoprotein. *Indian J. Pharm. Biol. Res*. 2014;2(4):44-57.
30. Ghittoni R, Accardi R, Chiocca S, Tommasino M. Role of human papillomaviruses in carcinogenesis. (website) *Ecanermedicallscience*. 2015;9:526. doi: 10.3332/ecancer.2015.526. eCollection 2015.