MDPI sciforum

MOL2NET, International Conference Series on Multidisciplinary Sciences, 5th edition
USINEWS-03: US-IN-EU Worldwide Science Workshop Series, UMN, Duluth, USA, 2019

A novel approach to Peptide Vaccine Design for Ebola virus

Subhamoy Biswas[1], Tathagata Dey[2], Shreyans Chatterjee[3], Smarajit Manna[4, 5], Ashesh Nandy[5*], Sukhen Das[5,6], Papiya Nandy[5], Subhash C Basak[7]

[1]Electrical Engineering Department, Jadavpur University, Kolkata
[2]Computer Science Department, Government College of Engineering and Textile Technology, Serampore
[3]Microbiology Department, St. Xavier's College, Kolkata
[4]Jagadis Bose National Science Talent Search, Kolkata
[5]Centre for Interdisciplinary Research and Education, Kolkata
[6]Physics Department, Jadavpur University, Jadavpur, Kolkata
[7]Department of Chemistry and Biochemistry, University of Minnesota, Duluth, MN, USA
* Author to whom correspondence should be addressed; E-Mail: anandy43@yahoo.com;
Tel.: +919433579452.

Abstract:

The Ebola viral disease with high fatality ratio is making a comeback in the Democratic Republic of Congo (DRC), after its rampage in West Africa in 2014-16, that has spawned fears of leading to a pandemic. Vaccines such as the experimental rVSV-ZEBOV has provided protection in 70-80% of the cases, but such vaccines are in short supply and doubts exist of its availability and sustainability in pandemic cases.

Peptide vaccines promise to amend this lacuna as a chemical construct that can be scaled up to requirement in manufacturing set-up, are easy to produce in pure form and store as well as transport much more easily and economically than traditional vaccines. Although no peptide vaccines have been licensed yet for human use, the rapid growth of applications of *in silica* approaches to peptide vaccine design and application to a myriad of virus infections, and subsequent follow-up experimental work, have led to expectations to licensures in the near future. We have proposed a protocol to automate the search procedure for suitable peptide vaccines using mathematical and computational modelling approaches that ensure long life of such vaccines even in the face of rapid mutational changes in viral sequences. In this paper we outline the mathematical model we have used and the recent improvements in the techniques to ensure the best recommendations for peptide vaccine libraries, especially against the Ebola virus that threatens to spill over the Congo border and cause epidemic and pandemics in a globalized world.
.

1. Introduction

Current efforts at computational approaches to develop drugs and vaccines have taken on urgency in view of the fact that viral epidemics have been occurring with increasing frequency in recent years. In the 21st century, the epidemics of SARS (Severe Acute Respiratory Syndrome) in 2002-03 [1], H1N1 swine flu of 2009 [2], the Ebola virus crises in 2014-16 in West Africa and its latest manifestation in the Democratic Republic of Congo (DRC) [3], the Zika rampage of 2015-16 [4] and near-epidemic of the Nipah virus in South Asia [5,6] have caused high fatality, strained public health resources and led to significant economic damages. In all these cases the virus was either new or a mutated progeny of an earlier form and for which no drugs or vaccines existed. Traditional methods of developing new drugs and vaccines cost billions of dollars and take several years of lead time [7-9], which clearly is inadequate to contain viral epidemics that generally die out in 6 to 18 months.

Advances in computational technology, immune-informatics, genetic databases and related sciences in recent years have led to the development of the science of vaccinomics that relates to the understanding of how vaccines work and how focused and technological approach can be taken to accelerate the process of vaccine design [10,11,12}. Peptide vaccines where surface proteins of a virus particle, the virion, is checked for its potential to elicit the best immune response, is one product of the vaccinomics approach [13]. The availability of adequate viral sequence data and novel programming techniques that bypasses alignment procedures have enabled development of new and fast protocols for in silico approaches to peptide vaccine design. We have used these novel alignment-free techniques and protocols (Fig.1) to design *in silico* peptide vaccine libraries for Influenza [14,15], Rotavirus [16], Human Papilloma virus [17] and Zika virus [18], but the methods require more robust software approaches; we are in this paper very briefly reporting one such modification to our standard approach; another approach that separates viral sequences originating from different geographical areas using principal component analysis (PCA) and self-organizing maps (SOM), which holds promise of good separation between different segments also, has been recently communicated [19]. We use the Ebola virus (EBOV) as an example for application of the modified technique.

There are 6 known virus variants of the genus Ebola virus: Bundibugyo Ebola virus, Reston Ebola virus, Sudan Ebola virus, Tai Forest Ebola virus, Bombali Ebola virus and Zaire Ebola virus. Ebola viruses cause a type of hemorrhagic fever in human known as Ebola Virus disease (EVD) which has a high mortality rate. Of the six known observed variants, Zaire Ebola virus has the highest mortality rate and has caused the largest number of outbreaks: from over 11000 deaths in West Africa in the 2014-16 period to the recent outbreak in the Congo, where, as of 21st November 2019, a total of 3298 EVD cases were reported, including 3298 confirmed and 118 probable, and mortality of 2196 cases (overall fatality ratio 67%) [20].

Collect all information on the virus and its structure

Determine surface situated proteins of the virion, select one or more for vaccine design purposes

Get average solvent accessibility (ASA) information on selected protein sequence

Get protein variability index by Graphical Sliding Window Method

Determine protein sequence segments that have high ASA and low protein variability

Use 3D structural model of the protein to ensure the selected segments are indeed surface situated

Determine through web servers like IEDB if the selected segments have acceptable epitope potential

Run peptide segments from above step through a BLAST or similar software to ensure no autoimmune threats

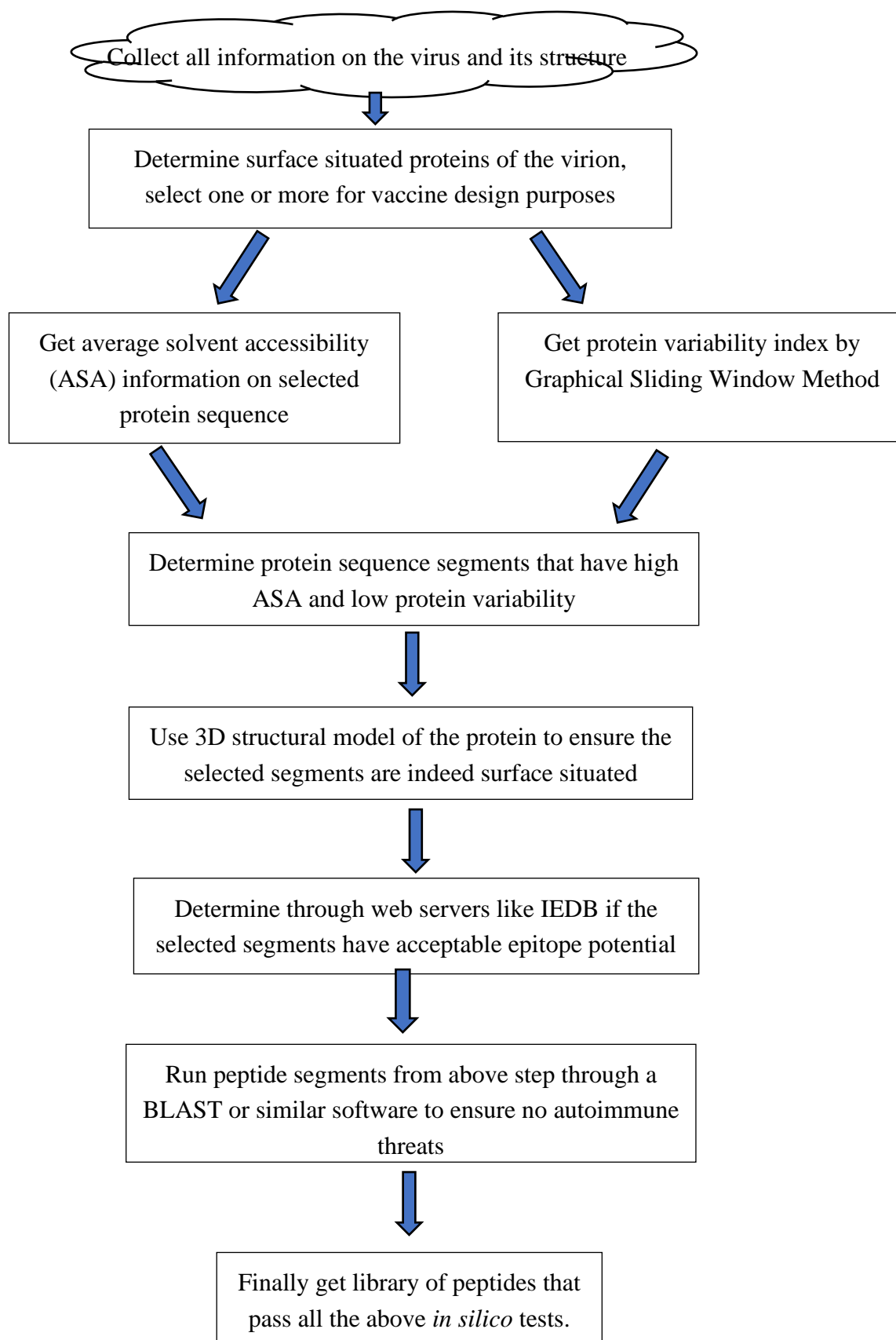Finally get library of peptides that pass all the above *in silico* tests.

Fig.1 Flowchart for *in silico* peptide vaccine design

Till now many potential vaccine candidates have been developed. But none have been approved for clinical use in human. But rVSV-ZEBOV has been extensively used in 2018-2019 on a compassionate use protocol [21] and reported 70-80% effectiveness against EBOV. Recently a new Ebola virus vaccine has been prequalified by the WHO for use in the DRC, but the results are awaited. Classical vaccines though used widely (but not very effective against a highly pathogenic virus like EBOV) have certain disadvantages; viz., 1) The supply is very less compared to their demand and 2) Classical vaccines are effective for only a particular strain of the virus and cannot be used for the same virus if it mutates to a new strain. On the other hand, newly developed peptide vaccines which are peptides serving to immunize an organism against a pathogen [22] is advantageous as it can meet the supply demand ratio, is cost effective and can be easily modified to suit mutated viral strains. In this paper we very briefly report on our efforts to design suitable peptide vaccines for the Zaire Ebola virus, and also take this opportunity to record our new modification to determining suitable prospects for peptide vaccines by a quantitative method.

2. Results and Discussions

The EBOV genome is a single-stranded, negative sense RNA 19000 nucleotides long [23]. The virion of EBOV has a virally encoded glycoprotein (GP) projecting as spikes (hence known as Spike Glycoprotein) from its lipid surface. EBOV can enter through two pathways. They are Niemann-Pick C1 (NPC1) and Hepatitis A Virus cellular receptor 1(HAVcr-1). The HAVcr-1 binds to the Spike glycoprotein present in the envelope of Ebola virus.

The available Zaire Ebola glycoprotein sequences that met our criteria defined in the Materials and Methods section were subjected to protein segment variability profile examination described below. The sequences were further tested for their ASA index and the two indices plotted against amino acid position number. This enabled a preview of the most likely conserved, surface exposed segments, most suitable for immune response. Applying the new Polygon representation algorithm, we got a more comprehensive quantitative assessment of the regions where protein variability is the least and matched well with high solvent accessibility.

These regions were then to be compared with the 3D structure of the EBOV to ensure that surface protein segments are not covered by neighbouring proteins and segments. Unfortunately, the only crystal structure of the Zaire EBOV available in the Protein Data Bank (PDB), 5HJ3.pdb, shows only a limited part of the protein structure and so our comparison task was rather constraint, but what was available showed that the segments covered by our investigations and falling within the protein segment covered by the crystal structure, were clearly surface exposed. Fig.2 shows one such segment predicted by the IEDB analysis described hereafter; the segment is marked in yellow in the cleft on the right hand side centre of the figure.

Having a list of peptide segments that matched our criteria, we next turned to the IEDB (Immune Epitope Analysis Resource) to determine if any of these had good epitope potential. Table 1 gives a list

of the first six high ranking epitopes determined by a MHC II analysis for B-cells with HLA profiles from Africa that covered 90% of the listed communities. Column 1 lists the start and end of the contiguous segment that ranked high in the IEDB results, column 2 shows the peptide, while column 3 lists the segments determined by our own analyses using the Polygon Representation. The matches are quite close. The available crystal structure did not allow representation of our peptides, but one segment predicted by IEDB analysis is shown in Fig.2 for representational purposes.

Table 1: High rank epitope prediction by IEDB and correspondence with selection of peptides through Polygon Representation

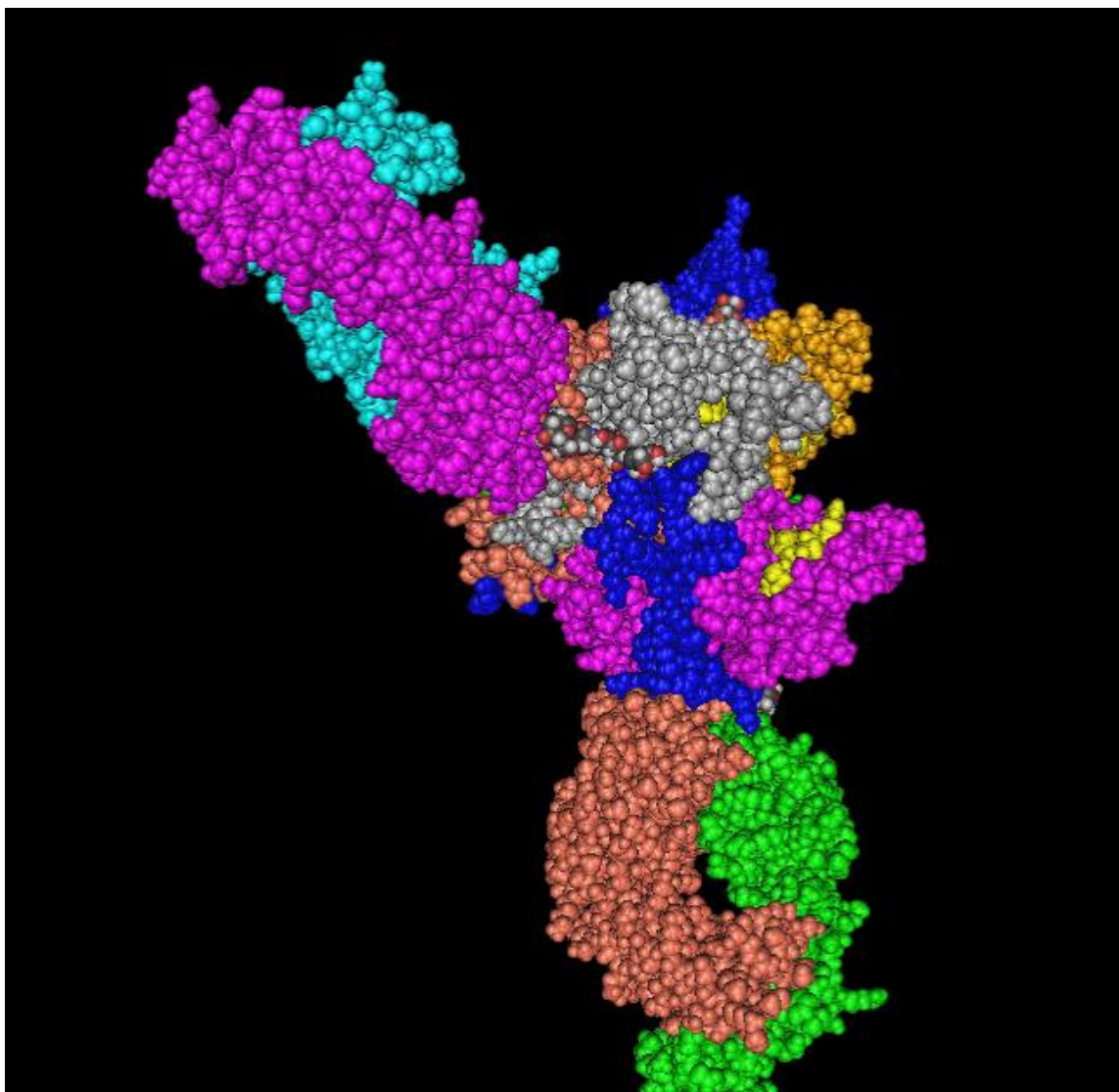| IEDB server epitope prediction		Polygon Representation
Region start-end	**Epitope of good rank**	**Region start-end**
36-87	GVIHNSTLQV…VPSATKR	50-66
97-112	VNYEAGEWAENCYNLE	109-128
116-124	PDGSEC-PA	
138-150	VHKVSGTGPCAGO	-
167-178	STVIYRGTTFAE	-
205-232	ATEDPSSG…FGTNETEY	187-209

Fig.2. One of the epitopes predicted by IEDB is seen in the cleft in the right hand side centre in yellow.

There are several more steps that need to be completed before the final list of possible vaccine candidates can be presented for wet-lab analysis. Our work so far has been to predict linear epitopes; we have to analyse for non-linear epitopes as well as for the T-cell interactions. We also need to use other web servers to support or negate the conclusions we need to draw from these analyses. Last but not least, we need to take the semi-final list of probable peptides and eliminate one or more of them that can lead to auto-immune diseases. Then finally we can draw up a list of best probable epitopes from amongst the list of peptides thrown up by our analyses so far. This work is in progress.

3. Materials and Methods

The protein sequence data of all Zaire Ebola viruses available in the NCBI GenBank database were downloaded and only those that were complete and reasonably annotated were filtered for the glycoprotein (GP) sequences for further analysis.

As per our protocol, the methodology required an in-house software to determine the conserved segments in the GP and a number of external servers to analyse various attributes. For the protein conserved segment search, we consider a protein sequence numerical characterization technique [24]. Since there are 20 different amino acids, so, in abstract space, a hypothetical 20D coordinate system can be considered where each amino acid is assigned one axis. As explained in detail in Ref [24], the sequence is read considering one amino acid at a time and the corresponding points are plotted by moving unit-step forward in respective directions. Likewise, the weighted averages along the axes are given as $\mu_1 = \Sigma(x_{1i})/N$, $\mu_2 = \Sigma(x_{2i})/N$, ..., $\mu_{20} = \Sigma(x_{20i})/N$. Now one can define a protein graph radius (p_R) as the magnitude of the resulting vector $(\mu_1, \mu_2, ..., \mu_{20})$ obtained, which is given as $p_R = \sqrt{(\mu_1{}^2 + \mu_2{}^2 + ... + \mu_{20}{}^2)}$. This value of p_R can be used to numerically characterize amino acid sequences with minor adjustments (see ref [16]) to yield a value that is characteristic of a given sequence, which implies that two sequences having the same p_R must be identical. This is the core of the conserved segment search method.

To determine conserved segments in the protein sequence, we select a window of 12 amino acid length, directed mainly by the MHC I and II pocket size, and determine the p_R value of the window stretch, then shift the window by one amino acid at a time and determine the new p_R value and so on till the end of the sequence. Doing this for all the sequences in our database, we can analyze each window's p_R values. In any window, some sequences will have the same sequence structure and so will have the same p_R value. We can scan all the sequences in each window to determine how many varieties of values are there: the lower the number of varieties, the more conserved is the segment.

The next step is to find the average solvent accessibility (ASA) of each residue in the sequences through an appropriate server; we use SABLE and iTASSER. The ASA profile is then compared to the protein segment variability profile to determine those segments where the protein variability is the least and hydrophilicity index is the highest that imply regions that are most conserved and exposed to solvent. Both profiles are smoothed out by a moving average of 12 aa to make following the graphs easier.

So long we had been doing this important step of identifying the conserved surface exposed regions by eye estimation. In this paper we report a 3D Polygon Representation method where this is automated. We represent on the three axes the parameters ASA, 1/PV and ASA-PV where PV is the protein variability index. At any amino acid position number on the sequence, we can get three parameters from which we can draw a triangle and get an area as a number to compare between the values at other amino acid position in the sequence (Fig.3). The higher the area, the closer we will be to the ideal of most conserved, most surface exposed segments of the virion protein sequence. The method was tested against our earlier work with rotaviruses. While the new method identified regions that included those we had identified by eye estimation, it showed a few more segments that could have been used, and also overlapped our earlier identified regions. This gave us confidence to utilize the program for this important Ebola virus peptide design work.
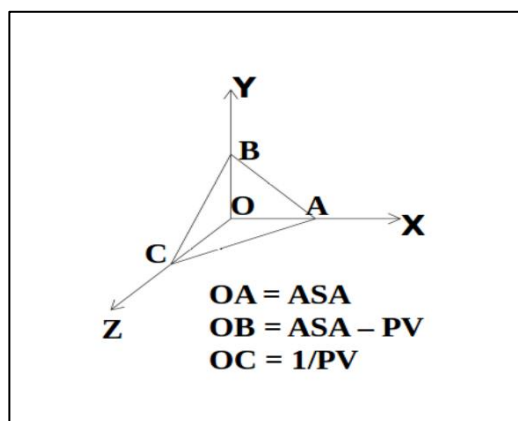
Fig. 3. Polygon Representation

The methodology for the rest of the protocol in Fig,1 is fairly straight forward and have been enumerated in the Results and Discussions section.

4. Conclusion

To summarise, we have outlined in this paper a novel methodology to determine quantitatively a set of peptide segments that are highly conserved and also highly exposed to the solvent so that they can act readily as Ebola virus vaccine candidates. The software algorithm we have outlined in this paper is quite novel and has proved useful in analyzing the Zaire Ebola viruses. This novel approach has been compared with our earlier work on rotaviruses and shown that while supporting what we predicted several years ago still works, but the new algorithm predicts a few more peptide segments that we had missed in the eye estimation work. This kind of analysis will provide a better foundation for future analysis for peptide vaccine designs.

In addition to the novel kind of analysis reported in this paper for the computer-assisted design of Ebola virus peptide vaccine libraries, another line of our recent computational biology research, based on alignment-free sequence descriptors, principal component analysis (PCA) and self-organizing maps (SOM) [19], clustered a diverse collection of viral sequences into homogeneous subsets predominantly present in distinct geographical regions of the world. Results of such research could be used in future analyses of heterogeneous collections of viral strains.

Conflicts of Interest

The authors declare no conflict of interest.

References

1 Smith, R.D. Responding to global infectious disease outbreaks: Lessons from SARS on the role of risk perception, communication and management. *Social Science & Medicine* **2006**, *63(12)*, 3113–3123.

2 Dawood, F.S.; Iuliano, A.D.; Reed, C.; Meltzer, M.I.; Shay, D.K.; Cheng, P.Y.; Bandaranayake, D.; Breiman, R.F.; Brooks, W.A.; Buchy, P.; Feikin, D.R.; Fowler, K.B.; Gordon, A.; Hien, N.T.; Horby, P.; Huang, Q.S.; Katz, M.A.; Krishnan, A.; Lal, R.; Montgomery, J.M.; Mølbak, K.; Pebody, R.; Presanis, A.M.; Razuri, H.; Steens, A.; Tinoco, Y.O.; Wallinga, J.; Yu, H.; Vong, S.; Bresee, J.; Widdowson, M.A. Estimated global mortality associated with the circulation: a modelling study. *Lancet Infect Dis* **2012**, *12(9)*, 687-695.

3 WHO 2018. Ebola virus disease. http://www.who.int/news-room/factsheets/detail/ebola-virus-disease. (Accessed 7th August, 2018).

4 Nandy, A.; Basak, S.C. The Epidemic that Shook the World—The Zika Virus Rampage. *Explor. Res. Hypothesis Med*. **2017**, *2*, 43–56. doi:10.14218/ERHM.2017.00018.

5 Luby, S.P.; Hossain, M.J.; Gurley, E.S.; Ahmed, B.N.; Banu, S.; Khan, S.U…..Rahman, M. Recurrent zoonotic transmission of Nipah virus into humans, Bangladesh, 2001–2007. *Emerg Infect Dis*, **2009**, 15, 1229-1235.

6 Roy, P.; Nandy, A. Nipah Virus – An Epidemic in the Making and Vaccine Strategy. *J Bacteriol Vaccin Res* **2018**, *1(1)*, 1004.

7 Pharmaceutical Research and Manufacturers of America. Biopharmaceutical Research Industry Profile. 2014; Available from: http://www.phrma.org/sites/default/files/pdf/2014_PhRMA_PROFILE.pdf. (Accessed December 1, 2014)

8 Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, Shacht AL. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov*. **2010**, *9(3),* 203–214.

9 Tufts PR. CSDD 2014 Cost Study. Cost to Develop and Win Marketing Approval for a New Drug Is $2.6 Billion. http://csdd.tufts.edu/news/complete_story/pr_tufts_csdd_2014_cost_study . (Accessed 10th July, 2018).

10 Rappuoli R. Reverse vaccinology, a genome-based approach to vaccine development. *Vaccine* **2001,** *19*, 2688–2691.

11 Poland GA, Kennedy RB, Ovsyannikova IG. Vaccinomics and personalized vaccinology: Is science leading us toward a new path of directed vaccine development and discovery? *PLoS Pathog*, **2011,** *7,* 1002344.

12 Poland GA, Whitaker JA, Poland CM, Ovsyannikova IG, Kennedy RB. Vaccinology in the third millennium: Scientific and social challenges. *Curr. Opin. Virol* **2016**, *17*, 116–125.

13 Nandy A, Dey S, Roy P, Basak SC. Epidemics and Peptide Vaccine response – A brief review. *Curr Topics Med Chem* **2018,** *18,* 2202-2208. doi: 10.2174/1568026618666181112144745.

14 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.

15 Sarkar T. Das S, De A, Nandy P, Chattopadhyay S, Chawla-Sarkar M, Nandy A. H7N9 influenza outbreak in China 2013: In silico analyses of conserved segments of the hemagglutinin as a basis for the selection of peptide vaccine targets. *Comput. Biol. Chem* **2015,** *59,* 8–15.

16 Ghosh A, Chattopadhyay S, Chawla-Sarkar M, Nandy P, Nandy A. In silico study of rotavirus VP7 surface accessible conserved regions for antiviral drug/vaccine design. *PLoS ONE* **2012,** *7,* 40749.

17 Dey S, De A, Nandy A. Rational Design of Peptide Vaccines Against Multiple Types of Human Papillomavirus. *Cancer Inform*. **2016,** *15(Suppl 1),* 1–16. doi:10.4137/CIN.S39071.

18 Dey S, Nandy A, Basak SC, Nandy P, Das S. A Bioinformatics approach to designing a Zika virus vaccine. *Comp Biol Chem* **2017,** *68*, 143-152.

19 Vracko M, Basak SC, Sen D, Nandy A. Clustering of Zika viruses originating from different geographical regions using computational sequence descriptors. *Curr Comp Aided Drug Des.* (communicated)

20 Ebola Virus disease - Democratic Republic of Congo; World Health Organization – Ebola health update – DRC **2019** https://www.who.int/emergencies/diseases/ebola/drc-2019

21 WHO – Ebola virus disease - Ebola treatments approved for compassionate use in current outbreak – WHO 6th June **2018**. https://www.who.int/ebola/drc-2018/treatments-approved-for-compassionate-use/en/ (Accessed 15th November 2019)

22 "Synthetic Peptide Vaccines" World Health Organization. Retrieved 24 July, **2015**. https://www.who.int/biologicals/vaccines/synthetic_peptide_vaccines/en

23 Nanbo A, Watanabe S, Halfmann P, Kawaoka Y. The spatio-temporal distribution dynamics of Ebola virus proteins and RNA in infected cells. *Scientific Reports* **2013,** *3*, 1206.

24 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*, 77-87.