Binding Affinities of Dengue Virus Envelope Glycoprotein Residues with Human Leukocyte Antigen Alleles: Dry Lab Candidates for Synthetic Vaccines

Hamid Nawaz Tipu[a], Masroor. E. Babar*[b], Tanveer Hussain[c]

[a]Combined Military Hospital, KhuzdarCantt. Pakistan
[b]Virtual University of Pakistan, Lahore.
[c]Department of Livestock Production, University of Veterinary and Animal Sciences, Lahore, Pakistan.
Emails: HNT: hnt1779@yahoo.com, MEB: masroorbabar@hotmail.com, TH: tanveer.hussain@uvas.edu.pk

Abstract: Determining dengue virus four serotypes E protein nanomeric epitopes that bind strongly with HLA alleles were identified through computer generated databases and prediction tools to identify individuals/populations who are likely to respond to vaccines. FASTA sequences of E proteins retrieved from NCBI proteins database were fed into MEGA6 software for conserved sequence identification. Binding of conserved sequence as well as nanomeric peptides from entire E protein length with HLA I and II alleles were determined using NetMHC 3.4 and NetMHCII pan 3.0 servers respectively. Conserved sequence among four serotypes did not bind HLA alleles. Individual E protein analysis identified total 97 nanomeric epitopes from four serotypes that qualified as strong binders. HLA specific E protein epitope prediction can help identify synthetic peptide vaccine candidates and predict responses as well.

[Hamid Nawaz Tipu, Masroor. E. Babar, Tanveer Hussain. **Binding Affinities of Dengue Virus Envelope Glycoprotein Residues with Human Leukocyte Antigen Alleles: Dry Lab Candidates for Synthetic Vaccines.** *Life Sci J* 2014;11(8s):552-557]. (ISSN:1097-8135). http://www.lifesciencesite.com. 115

Keywords Dengue virus envelope (E) protein, Human Leukocyte Antigen, Bioinformatics, Vaccine Candidate

Introduction

Dengue virus, belonging to flaviviridae family, is transmitted to humans by infective females of mosquitoes *Aedes aegypti* and *Aedes albopictus* [1-2].World Health Organization (WHO) has declared dengue infection endemic in South Asia. About 50-100 million cases of dengue fever and 500,000 cases of dengue hemorrhagic fever are reported annually, resulting in around 24,000 deaths [3]. Pakistan is at very high risk of large epidemics owing to overcrowded cities, unsafe drinking water, inadequate sanitation and large number of refugees [3]. First outbreak was officially reported in 1994, though virus is believed to have existed in the country since long ago[4]. Since then, existence of virus has been reported from various parts of the country. In 2011, infection rapidly assumed proportion of epidemic; more than 250 people were reported dead in provincial capital city Lahore [5].

Dengue virus is an RNA virus with 11 Kilobases genome that encodes three structural proteins, nucleocapsid or core protein (C), a membrane associated protein (M), an envelope protein (E) and seven nonstructural proteins [6]. E protein binds cellular receptors and mediates fusion of viral and cellular membranes during viral entry into the cells [7]. The native E protein forms a homodimer, with each subunit consisting of three beta barrel domains, named domain I (EDI), domain II (EDII) and domain III (EDIII). The hydrophobic viral fusion peptide is located at the tip of EDII [8] and is well conserved among different serotypes. Figure 1 illustrates structure of E protein. There are four antigenically distinct serotypes of the virus (DENV 1-4) [9]. Antibody epitopes are both unique to each serotype as well as shared among them. Previous infection results in antibody formation that despite cross reactivity with all four serotypes (and other flaviviruses), prevents against re infection with same serotype only. In fact, infection with a new serotype may actually exacerbate the disease [8]. Following infection, patients develop antibodies against structural as well as nonstructural proteins [10]. Since E protein is the main antigen on the surface and is involved in fusion, it is the main target of neutralizing antibodies. However, exact serotype specific epitopes against which antibodies arise remain poorly defined [8, 11].

For a vaccine to be successful it must induce antibodies against all four serotypes but dengue vaccine development has been very slow due to these ill-defined antibodies inducing epitopes. Development of epitopic vaccines using small segments of viral proteome provides effective and controlled immune response reducing lethal effects of live vaccine [12]. However, it must take into account binding of epitopes with human leukocyte antigen (HLA), the complexity of which is compounded by extreme HLA polymorphism. T cells recognize an antigen when presented with human leukocyte antigen complex on the surface of antigen presenting cells. HLA alleles how differential binding affinities for various antigens, both foreign and self, and this determines

basis of protective response against foreign antigens by the individual as well as development of autoimmune diseases. This differential binding is due to allele specific amino acid composition and thus distinct polarity and stereochemistry of antigen binding pockets on HLA molecule [13]. As of January 2014, 2579 HLA-A, 3285 HLA-B, 2133 HLA-C and 1512 HLA-DR alleles (besides other HLA class I and II loci) have been identified [14].Owing to differential binding affinities of HLA alleles with the antigen (in this case, dengue virus all four serotypes envelope glycoproteins E), we hypothesized that it is possible to determine, using computer generated models, tools, algorithms, databases and servers, specific binding affinity of a specific protein sequence with specific HLA allele. This will help identify specific epitope of the E proteins to which individual's HLA alleles bind strongly. This approach in future can aid in designing synthetic peptide vaccines for wet lab testing and clinical trials. In fact, this 'vaccinomics' approach is already being employed in designing bacterial and viral vaccines[15].However it must be emphasized that ultimately such predictive *in silico* work needs to be confirmed through *in vivo* experiments[16].

Materials and Methods

Public database of National Centre of Biotechnology Information (NCBI) was searched for dengue virus envelope glycoprotein, specifically protein database at site: http://www.ncbi.nlm.nih.gov/protein. Although protein sequences in different countries have been identified, in South Asian countries, E protein sequences from all four serotypes have not yet been reported. So we included top hit as representative of E protein sequence for that particular serotype. These sequences in FASTA format were saved in text pad file for further analysis.

This file was imported in Molecular Evolutionary Genetic Analysis (MEGA) software [17] and all four sequences were aligned against one another to identify conserved sequences of minimum 9 amino acid in length, since this is the minimum length bound by HLA alleles and is used in most HLA binding prediction softwares.

The FASTA sequence of conserved amino acids epitopes was fed into NetMHC 3.4 server and NetMHCII pan 3.0server from Technical University of Denmark [18, 19]. Theseserverspredict binding of each peptide sequence to different HLA class I and HLA class II loci respectively, using artificial neural networks (ANN). For our study we analyzed HLA A, B and DRB1 loci. These servers give affinity of peptides with HLA alleles in the form of 1-log50k with higher value indicating stronger binding.

The affinity of conserved epitope with HLA class I was tested at default threshold values of NetMHC 3.4 server. HLA loci tested were HLA A 01, 02, 03, 24 and 26 with HLA A*0101, 0201, 0301, 2402 and 2601 as representative alleles. HLA B loci included were B 07, 08, 27, 39, 40, 58 and 15 with HLA B*0702, 0801, 2705, 3901, 4001, 5801 and 1501 as representative alleles.

The affinity of conserved epitope with HLA class II was tested at default threshold values of NetMHCII pan 3.0 server. HLA loci tested were HLA DRB1*01, 03, 04, 07, 08, 09, 10, 11, 13, 15 and 16 with HLA DRB1*0101, 0301, 0401, 0701, 0802, 0901, 1001, 1101, 1302, 1501 and 1601 as representative alleles respectively.

Following inability of conserved residues peptide to bind HLA alleles (as will be evident in results section), we took the entire FASTA sequence of E proteins of all four serotypes separately and tested their binding affinities with HLA class I and II alleles, as mentioned in steps 4 and 5 above.

Results

Dengue virus E protein from four serotypes identified from NCBI database were gi|111380034, gi|58223370, gi|5007020, gi|55275822 respectively, with all proteins 495 amino acids in length except for serotype 3 which was 493 amino acids long.

Alignment in MEGA6 and conserved amino acids identification revealed only one sequence more than 9 amino acids in length. This sequence spanned 97-111 residue positions, sequenced as VDRGWGNGCGLFGKG. This sequence is highlighted in yellow in figure 1.

Binding affinity analysis of this sequence with HLA class I alleles at NetMHC 3.4 server revealed that none of the six nanomers from this 15 amino acids peptide bound with any of the alleles with sufficient strong affinity to qualify as binder, at default threshold values. Similar were the findings with HLA class II alleles at NetMHCII pan 3.0 server. None of the nanomers qualified as binder with any of the representative alleles tested.

Strong binder peptides of 9 amino acids length from entire sequence of E protein (from 4 serotypes) with their respective HLA class I binding alleles are shown in table I.

Strong binder peptides of 9 amino acids length from entire sequence of E protein (from 4 serotypes) with their respective HLA class II binding alleles are shown in table II.

Only three of eleven HLA class II alleles were able to bind one nanomer each (from all four serotypes), in marked contrast to HLA class I alleles where eleven of twelve analyzed alleles were able to bind nanomers, more than one in most cases. This is not unexpected as viruses usually follow HLA class I antigen presentation pathway.

Chart I shows number of bound vs unbound HLA class I alleles by E proteins of individual serotypes. Such chart is omitted for HLA class II alleles since E protein showed negligible binding, as is evident in table II.

Chart II shows total number of epitopes from E protein of all four serotypes, bound by analyzed HLA class I alleles. Such chart is omitted for HLA class II alleles since the numbers were negligible, as is evident in table II.

Table I: Strongly binding nanomers of E protein with their respective HLA class I binding alleles, with lower affinity indicating stronger binding

HLA class I alleles	DENV 1 E nanomers with binding affinity (amino acid position)	DENV 2 E nanomers with binding affinity (amino acid position)	DENV 3 E nanomers with binding affinity (amino acid position)	DENV 4 E nanomers with binding affinity (amino acid position)	Total nanomers bound
A*0101	TSEIQLTDY 49(169)	ITEAELTGY 37(169)	None	None	**2**
A*0201	TMKEKSWLV 38(199) WLVHKQWFL 10(205) KQWFLDLPL 38(209) LTLKGISYV 34(291) SLSMTCIAV 27(473) GMVTLYLGV 30(482)	WLVHRQWFL 11(205) FLDLPLPWL 9(212) RLITVNPIV 12(349) SLSVSLVLV 26(473)	RQWFFDLPL 39(207) YAMCLNTFV 6(296) RLITANPVV 14(347) ALFSGVSWI 12(443) WIMKIGIGV 39(450)	KMKKKTWLV 36(199) WLVHKQWFL 10(205) KQWFLDLPL 38(209) SLGKAVHQV 31(430) TMFGGVSWM 46(445) SMAMSCIAV 13(473)	**21**
A*0301	SLITCAKFK 39(111) VTFKTAHAK 39(237) KALKLSWFK 47(384) ALKLSWFKK 42(385)	VTFKNPHAK 34(237) MSYSMCTGK 19(296)	VTFKNAHAK 37(235) CLNTFVLKK 18(299) ITANPVVTK 38(349)	VTFKVPHAK 30(237) MSYTMCSGK 15(296) ALTLHWFRK 34(385)	**12**
A*2402	None	None	SYAMCLNTF 5(295)	None	**1**
A*2601	None	None	None	None	**nil**
B*0702	TPQAPTSEI 25(164) LPWTSGAST 41(217)	None	LPEYGTLGL 28(172) TPTWNRKEL 14(225)	TPRSPSVEV 17(164) VPHAKRQDV 27(241)	**6**
B*0801	None	MRGAKRMAI 29(405) GMNSRSTSL 46(466)	None	MKMKKKTWL 36(198)	**3**
B*2705	None	KRFVCKHSM 30(87) RQWFLDLPL 22(209) KRHALGRLI 35(343)	RQWFFDLPL 26(207)	None	**4**
B*3901	MKIGIGILL 18(454)	IQMSSGNLL 34(269)	MKIGIGVLL 16(452)	VHNGDTHAV 22(142)	**4**
B*4001	TEVTNPAVL 22(47) TEHGTIATI 37(155) QETWNRQDL 33(227) QEGAMHTAL 22(255) AETQHGTVL 6(312)	TEAKQPATL 36(47) AELTGYGTV 27(172) QEGAMHTAL 22(255) AETQHGTIV 34(312)	TEATQLATL 19(47) QEGAMHTAL 22(253) SETQHGTIL 8(310) GESNIVIGI 6(371)	LEYTVVVTV 30(134) VELPDYGEL 39(172) QEGAMHSAL 36(255) AETQHGTTV 26(312)	**17**
B*5801	KSWLVHKQW 5(203) GASTSQETW 7(222) QTSGTTTIF 46(270) VSIEAEPPF 24(364) MAILGDTAW 6(411) IGIGILLTW 10(456) IAVGMVTLY 40(479)	KAWLVHRQW 6(203) MAILGDTAW 6(411) GAAFSGVSW 24(444) ILIGAVITW 41(456)	KAWMVHRQW 5(201) GATTETPTW 9(220) MAILGDTAW 6(409) TALFSGVSW 20(442) IGIGVLLTW 7(454) IAIGIITLY 33(477)	KTWLVHKQW 6(203) GADTSEVHW 24(222) MAILGETAW 6(411) STMFGGVSW 13(444)	**21**
B*1501	VSIEAEPPF 45(364)	IQKETLVTF 19(231) QMSSGNLLF 28(270)	IQNSGGTSI 41(267) NSKNTSMSF 44(466)	LGKAVHQVF 38(431)	**6**
Total nanomers	**27**	**23**	**25**	**22**	**97**
Total alleles bound	**8**	**9**	**9**	**8**	

Table II: Strongly binding nanomers of E protein with their respective HLA class II binding alleles, with lower level indicating stronger binding

HLA class II alleles	DENV 1 E nanomers with binding level (amino acid position)	DENV 2 E nanomers with binding level (amino acid position)	DENV 3 E nanomers with binding level acid position)	DENV 4 E nanomers with binding level (amino acid position)	Total nanomers bound
DRB1*0101	None	None	None	None	**Nil**
DRB1*0301	None	LRMDKLQLK 0.4(286)	None	None	**1**
DRB1*0401	None	None	None	None	**Nil**
DRB1*0701	None	None	None	None	**Nil**
DRB1*0802	None	None	None	MILMKMKKK 0.17(195)	**1**
DRB1*0901	None	None	None	None	**Nil**
DRB1*1001	None	None	None	None	**Nil**
DRB1*1101	None	None	None	MILMKMKKK 0.4(195)	**1**
DRB1*1302	None	None	None	None	**Nil**
DRB1*1501	None	None	None	None	**Nil**
DRB1*1601	None	None	None	None	**Nil**
Total nanomers	**Nil**	**1**	**Nil**	**2**	**3**
Total alleles bound	**Nil**	**1**	**Nil**	**2**	

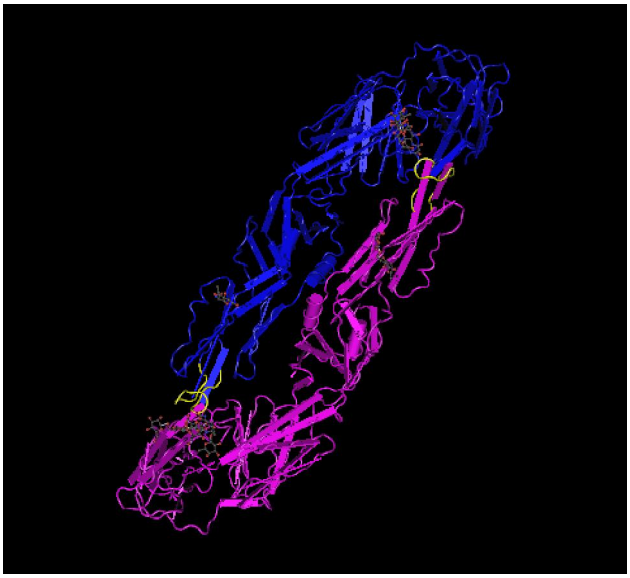

Fig 1: Structure of dengue virus envelope glycoprotein: E protein is a homodimer (right and left), each subunit has three domains, conserved sequence is indicated in yellow (Adapted and modified from Protein Data Bank ID 1UZG)

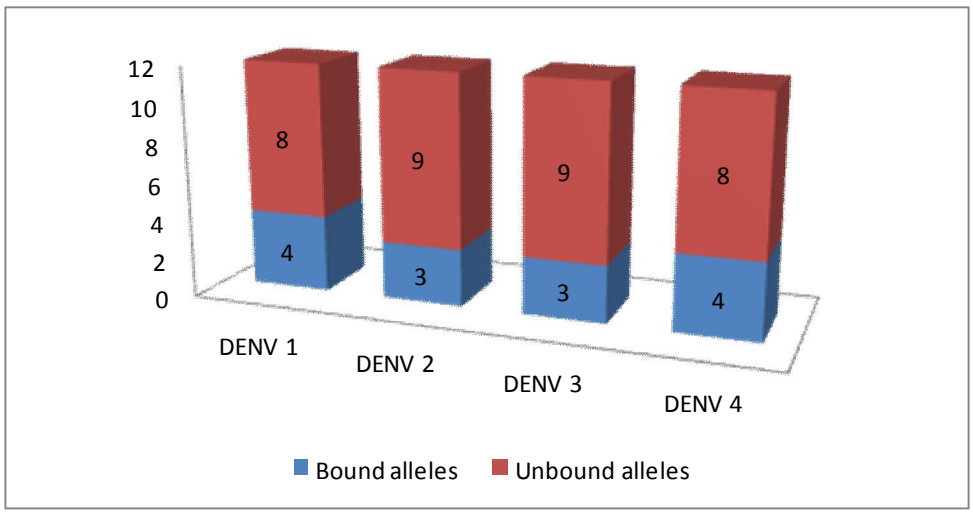

Fig 2: Number of bound Vs unbound analyzed HLA class I alleles by E protein of dengue virus four serotypes

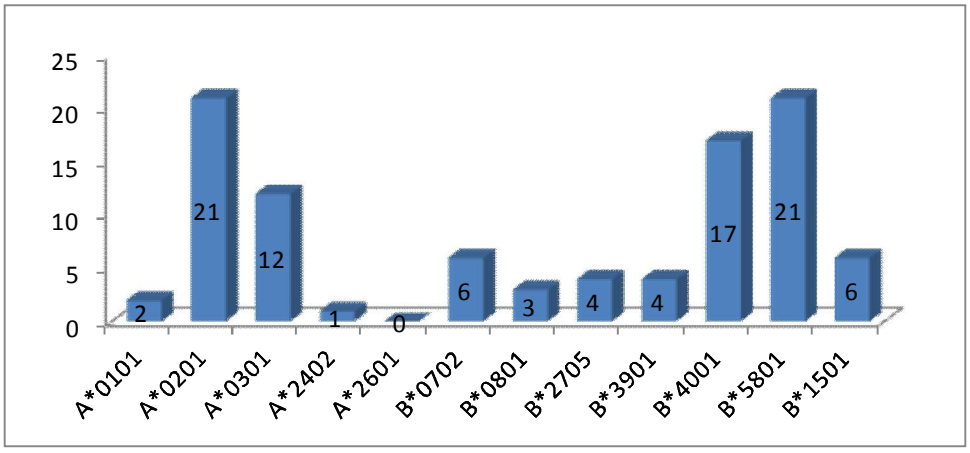

Fig 3: Number of epitopes (total 97 nanomers) from E protein of all four serotypes bound by analyzed HLA class I alleles

Discussion

Conserved sequence identification and binding with HLA alleles

Firstly, in our analysis, we identified conserved amino acids sequence in all four serotypes that must be of sufficient length to bind HLA and hence be a candidate for synthetic peptide vaccine from all four serotypes. Only one such sequence (highlighted in yellow in figure I) was identified spanning 97-111 positions in E protein sequence of all four serotypes. However, predicting binding affinities of this epitope proved that it did not bind any of the analyzed HLA class I or II alleles at nanomers length, and hence was a poor candidate for vaccine. This explains difficulty in preparing a single vaccine for all serotypes; firstly, only one conserved peptide sequence of sufficient length is available, secondly, even this sequence did not bind any of the analyzed HLA alleles.

Epitopes identification from E protein of four serotypes

This demanded the use of peptides from all four serotypes separately. So we determined binding of nanomer peptides from E proteins (of all four serotypes) with HLA alleles to identify potential candidates for vaccines. This bioinformatics approach for predicting and designing vaccine candidates is known as "reverse vaccinology" and has arisen because conventional experimental approaches are extremely laborious, expensive and time consuming. "Reverse vaccinology" involves computational methods to identify all potential candidate immunogens from genome of a pathogen/allergen. Once appropriate vaccine candidates have been identified, genes of interest can be cloned to produce corresponding protein. Subsequent in vivo and in vitro testing can further validate potential use in specific population[16, 20, 21].Determining binding of E peptides revealed negligible binding with HLA class II alleles as is evident in table II, it was expected since most of the viruses follow HLA class I antigen presentation pathway, for being presented to T lymphocytes. Binding determination with HLA class I alleles identified 27, 23, 25 and 22 (total 97) nanomers from E protein of DENV 1, 2, 3, and 4 respectively that showed significant binding affinities, only nanomers that qualified as "strong binders" under default threshold values were included. We propose that these epitopes can be effectively used singly or fused together as polytopes and used as vaccines. Another advantage is that use of specific peptides instead of whole viral proteome can eliminate several unwanted side effects [12]. However it must be emphasized that ultimately such predictive *in silico* work needs to be confirmed through *in vivo* experiments [16].

Population specific synthetic peptide vaccine candidates

These nanomers bind most of the representative HLA class I alleles as shown in chart I. Hence these should be expected to be presented to CD8 T lymphocytes and elicit an immune response. Separate population HLA allele frequency data (and table I) can be used to predict which nanomers are better vaccine candidates for that specific population.

Individualized vaccine therapy

Lastly, since this analysis also identifies HLA class I alleles that bind maximum/minimum/no epitopes from E protein, it can be used to predict individuals with specific HLA genotype if they will mount an immune response against the E protein. It will help identifying individuals who are better candidates for vaccine inoculation.

Conclusion

By identifying dengue virus four serotypes E protein nanomeric epitopes as binders of HLA alleles in dry lab, this article highlights the introduction of "Immunoinformatics" and "vaccinomics" in identification of synthetic peptide vaccine candidate epitopes. In addition, it also shows how to identify individuals likely to respond to synthetic/recombinant vaccines and to what epitopes of proteins they are likely to respond, directing their tailored therapy. This must be emphasized at the end that this predictive work although can significantly reduce cost and labor, ultimately, it has to be confirmed through in vivo experiments/clinical trials.

Conflict of Interest

The author declares no conflict of interest.

***Corresponding author**
Masroor. E. Babar
Tel: +92-334-4261313
E-mail: masroorbabar@hotmail.com

References

1. Batty ME, Stone A, Fitzsimons DW, Hanna JN, Lam SK, Vong S, Guzman MG, Mendez-Galvan JF, Halstead SB, Letson GW, Kuritsky J, Mahoney R, Margolis HS. Best practices in deugue surveillance: a report from the Asia-Pacific and Americas dengue prevention boards. PLoSNegl Trop Dis 2010; 4(11): e890.
2. Fatima Z, Idrees M, Bajwa MA, Tahir Z, Ullah O, Zia MQ, Hussain A, Akram M, Khubaib B, Afzal S, Munir S, Saleem S, Rauff B, Badar S, Naudhani M, Butt S, Aftab M, Ali L, Ali M. Serotype and genotype analysis of dengue virus by sequencing followed by phylogenetic analysis

using samples from three mini outbreaks 2007-2009 in Pakistan. BMC Microbiol 2010; 11(200): 1-8.
3. Jahan F. Dengue fever (DF) in Pakistan. Asia Pacific Family Medicine 2011; 10:1.
4. Chan YC, Salahuddin NI, Khan J, Tan HC, Seah CL, Li J, Chow VT. Dengue hemorrhagic fever outbreak in Karachi, Pakistan, 1994. Trans R Soc Trop Med Hyg 1995; 89: 619-20.
5. Idrees S, Ashfaq UA. A brief review of dengue molecular virology, diagnosis, treatment and prevalence in Pakistan. Genetic Vaccines and Therapy 2012; 10: 6.
6. Umareddy I, Pluquet O, Wang QY, Vasudevan SG, Chevet E, Gu F. Dengue virus serotype infection specifies the activation of the unfolded protein response. Virology Journal 2007; 4: 91.
7. Modis Y, Ogata S, Clements D, Harrison SC. Structure of the dengue virus envelope protein after membrane fusion. Nature 2004; 427(22): 313-9.
8. Wahala WMPB, de Silva AM. The human antibody response to dengue virus infection. Viruses 2011; 3: 2374-95.
9. Ross TM. Dengue virus. Clin Lab Med 2010; 30: 149-160.
10. Lai CY, Tsai WY, Lin SR, Kao CL, Hu HP, King CC, Wu HC, Chang GJ, Wang WK. Antibodies to envelope glycoprotein of dengue virus during the natural course of infection are predominantly cross-reactive and recognize epitopes containing highly conserved residues at the fusion loop of domain II. J Virol 2008; 82: 6631-43.
11. Dowd KA, Pierson TC. Antibody mediated neutralization of flaviviruses: A reductionist view. Virology 2011; 411(2): 306-15.
12. Shehzadi A, Shakoori FR, Shakoori AR. Immunoinformatics study of promiscuous HLA binding antigenic nanomers of surface glycoproteins of Influenza A (H1N1) pdm09 of Asian origin. Pakistan J Zool 2013; 45(3): 733-40.
13. Hillen N, Mesteri G, Lemmel C, Weinzierl AO, Muller M, Wernet D, Hennenlotter J, Stenzl A, Rammensee HG, Stevanovic S. Essential Differences in Ligand Presentation and T CelllEpitipe Recognition among HLA Molecules of the HLA B44 Supertypes. Eur J Immunol 2008; 38: 2993-3003.
14. HLA Nomenclature: HLA Alleles Numbers. [online] [cited 2014, April 11] Available from: http://hla.alleles.org/nomenclature/stats.html
15. Whelan FJ, Yap NVL, Surette MG, Golding GB and Bowdish DME. A Guide to Bioinformatics for Immunologists. Frontiers in Immunology 2013; 4: 1-16.
16. Sirskyj D, Diaz-Mitoma F, Golshani A, Kumar A, Azizi A. Innovative Bioinformatic Approaches for Developing Peptide-based Vaccines Against Hypervariable Viruses. Immunology and Cell Biology 2011; 89: 81-9.
17. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetic analysis version 6. Molecular biology and evolution 2013; 30: 2725-9.
18. Nielsen M, Lund O. NN-align. An Artificial Neural Network-based Alignment Algorithm for MHC class II Peptide Binding Prediction. BMC Bioinformatics 2009; 10: 296.
19. Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II Binding Affinity Using SMM-align, a Novel Stabilization Matrix Alignment Method. BMC Bioinformatics 2007; 8: 238.
20. Tomar N, Rajat K De. Immunoinformatics: An Integrated Scenario. Immunology 2010; 131: 153-68.
21. Whelan FJ, Yap NVL, Surette MG, Golding GB and Bowdish DME. A Guide to Bioinformatics for Immunologists. Frontiers in Immunology 2013; 4: 1-16.

8/15/2014