Hepatitis C Virus Subtype 3a Envelope Protein 1 Binding with Human Leukocyte Antigen Class I Types of Pakistani Population: Candidate Epitopes for Synthetic Peptide Vaccine

Hamid Nawaz Tipu

Combined Military Hospital, Khuzdar Cantt, Pakistan

ABSTRACT

The object of this cross sectional study was to determine the HCV subtype 3a envelope protein binding affinity with Human Leukocyte Antigen. **Envelope 1 (E1)** protein is one of the structural proteins responsible for entering the cells through the receptors. The binding affinity of E1 protein epitopes to the selected **Human Leukocyte Antigen (HLA)** class I alleles was investigated using the computer-based tools. These prediction tools were also used to design the synthetic vaccine's candidate epitopes and to identify the individuals/populations who are likely to be responder to those vaccines.

The mean frequency of **HLA** I antigens in Pakistani population was calculated. Three alleles each from HLA A and B were selected. E1 protein sequence extracted from HCV 3a isolates was retrieved and twenty-four sequences of it were selected. NetMHCcons 1.0 server was used to determine the binding affinities of HLA alleles to the epitope sequences of 10 amino acids in length.

A02, A03, A11, A24, A33, B08, B13, B15, B35 and B40 were the first five antigens more prevalent in Pakistan each from HLA A and HLA B.. We did not find any binding affinity between HLA A*201, B*1501 and B*4001 and epitopes from E1 sequences in a threshold of 50 nM. Totally five various epitopes derived from different isolates were characterized.

The prediction of HLA-E1 epitope specific bindings and the forthcoming response can be a useful bioinformatics tool to uncover the right synthetic peptides for vaccine design purposes.

Keywords: Bioinformatics; Hepatitis C virus envelope 1 (E1) protein; Human Leukocyte Antigen

INTRODUCTION

Hepatitis C virus (HCV) is known as a leading cause of chronic liver diseases and has infected over 10 million people in Pakistan. HCV genotype 3a accounts for over 80% of cases in our country.[1,2] It is classified within the Flaviviridae family and has a positive single stranded RNA. The genome length is about 9.6 kb, including the

Corresponding Author: Hamid Nawaz Tipu,
Bachelor of Medicine, Bachelor of Surgery (MBBS), Fellow of College of Physicians & Surgeons Pakistan in Immunology (FCPS),

Classified Pathologist & Consultant Immunologist, Combined Military Hospital, Khuzdar Cantt, Pakistan. Tel: (+92 33) 3547 8215, E-mail: hnt1779@yahoo.com

genes encoding over 3000 amino acids long polyprotein and non-translated regions at 5' and 3', termini.[3] Following transcription, this polyprotein is then cleaved into four structural proteins (core, E1, E2 and P7) and six non-structural proteins (NS2, NS3, NS4A NS4B, NS5A and NS5B).[4] The envelope proteins mediate cell entry by recognition of cellular membrane receptors. E1 is fusogenic subunit and E2 is receptor binding subunit of HCV. The highest sequence variability is found in E1 and E2 glycoproteins while the lowest variable sites are located in 5' and 3' UTRs.[5] Therefore it is important to target viral envelope glycoproteins to stop viral entering into the cells. A vaccine containing the recombinant glycoprotein (gp) E1 and E2 along with MF59 as an adjuvant is under phase I clinical trials. The vaccine's efficacy in preventing the infection as well as reducing the rate of chronicity have been shown in chimpanzee models.[6,7] Most neutralizing antibodies recognize conserved regions in E1 and E2 proteins. Besides glycoprotein-based vaccines have a capacity to induce the cross-genotype neutralizing activity.[8,9]

In order to generate a successful immune response, a viral antigen needs to be processed and presented in the presence of HLA class I molecules to T lymphocytes.Up to July 2014, 2884 HLA-A, 3590 HLA-B, 2375 HLA-C and 1649 HLA-DR alleles (besides other HLA class I and II loci) have been identified.[10] Binding affinity of HLA molecules to different antigens (both self and non self) depends on the types of HLA alleles.This is the basis of protective responses against foreign antigens and also development of autoimmune diseases.[11] This differential binding is due to allele specific amino acid composition and thus distinct polarity and stereochemistry of antigen binding pockets on HLA molecule.[12]

HCV vaccine development has not met much of a success because epitopes that induce antibody formation are yet poorly defined. Synthetic peptide vaccines derived from the virus's proteins can mount sufficient immune response while reducing the side effects. But an important factor is whether the considred epitopes can bind HLA or not.[11,13] Due to this differential HLA alleles and peptide binding affinities, we hypothesized that it would be possible to determine high affinity binding epitopes from E1 protein of HCV using online databases, softwares and tools. We further narrowed down our analysis

according to HLA types of Pakistani population and RNA sequence of E1 protein determined in HCV isolates from them. This will help to identify specific epitopes of the E1 protein with the strong binding activity to HLA alleles. 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[14], however, it must be emphasized that ultimately such predictive *in silico* work needs to be confirmed through *in vivo* experiments[15].

MATERIALS AND METHODS

This cross sectional study was carried out in August 2014, in Combined Military Hospital, Khuzdar Cantonment. It consisted of sequential steps mentioned as A through D. Data collection and results sections of each step have been mentioned together as results of one step serve as data in next step.

A: HLA Alleles Selection Based on Mean Allele Frequency and Dissimilar Binding Preferences
Data Collection and Analysis

HLA A and B alleles found in Pakistani populations were determined from HLA allele frequencies website.[16] Initial search for HLA A and B with 4 digit level of resolution revealed very few alleles (especially for HLA B locus). This is understandable as little sequence based HLA typing has yet been done in Pakistan. So we had to resort to 2 digit level of resolution.

Allele frequencies for all HLA A and B loci were entered in Microsoft excel, for different Pakistani populations in database. From this, mean allele frequencies were calculated for all HLA A and B loci.

Five commonest loci each from HLA A and HLA B were selected for further analysis. These were fed into MHCcluster 2.0 server from Technical University of Denmark to select 3 most dissimilar alleles that should have different binding preferences.[17] Three alleles from each locus were selected for further analysis.

Results

At 2 digit level of resolution, 17 HLA A and 15 HLA B antigens were identified in different Pakistani populations.

HLA A and B antigens mean frequencies are shown

in Figure 1. It is seen that A02, A03, A11, A24, A33, B08, B13, B15, B35 and B40 are first five antigens each form HLA A and HLA B, more prevalent in Pakistan. HLA A*0201, 0301, 1101, 2402, 3301 and HLA B*0801, 1501, 3501 and 4001 were taken as representative of their corresponding antigens.

Results from MHCcluster 2.0 server are shown in Figure 2. It shows MHC specificity tree and MHC specificity heat map for analyzed HLA alleles. HLA A*0201, 0301, 2402 and HLA B*0801, 1501, 4001 were selected for further progress into the study.

B: Hepatitis C Virus (HCV) Envelope Protein (E1) Sequences Inclusion and Alignment
Data Collection and Analysis

HCV genotype 3a was searched in National center for biotechnology information (NCBI) nucleotide database that has been sequenced in Pakistani isolates of virus.

The results were manually screened for E1 sequences. Manual screening was required since many results included core protein and envelope protein E2 sequences as well. Twenty seven E1 sequences were copied in textpad file in FASTA format.

These 27 nucleotide sequences were translated into amino acid sequences in Molecular Evolutionary Genetic Analysis (MEGA) 6 software. Then these 27 sequences were aligned with one another using same software. 24 were further selected as 3 were misaligned.

Results

It revealed 223 results for HCV genotype 3a E1. Manual screening for E1 protein shortened the results to 27 isolates from Pakistan.

Following translation of nucleotides into amino acids, alignment was carried out. Three sequences were further removed from analysis as these were misaligned with the rest of our sequences, leaving 24 amino acids sequences. Figure 3 shows aligned translated amino acid sequences of isolates. Conserved sites are marked by asterix at the top.

C: Binding Affinities of Selected HLA Alleles with E1 Amino Acid Sequences and 3D Model Generation
Data Collection and Analysis

This was carried out at NetMHCcons 1.0 server from Technical University of Denmark. It predicts peptide binding to known HLA I alleles, integrating NetMHC, NetMHCpan and PickPocket methods. The first two methods are known as artificial neural
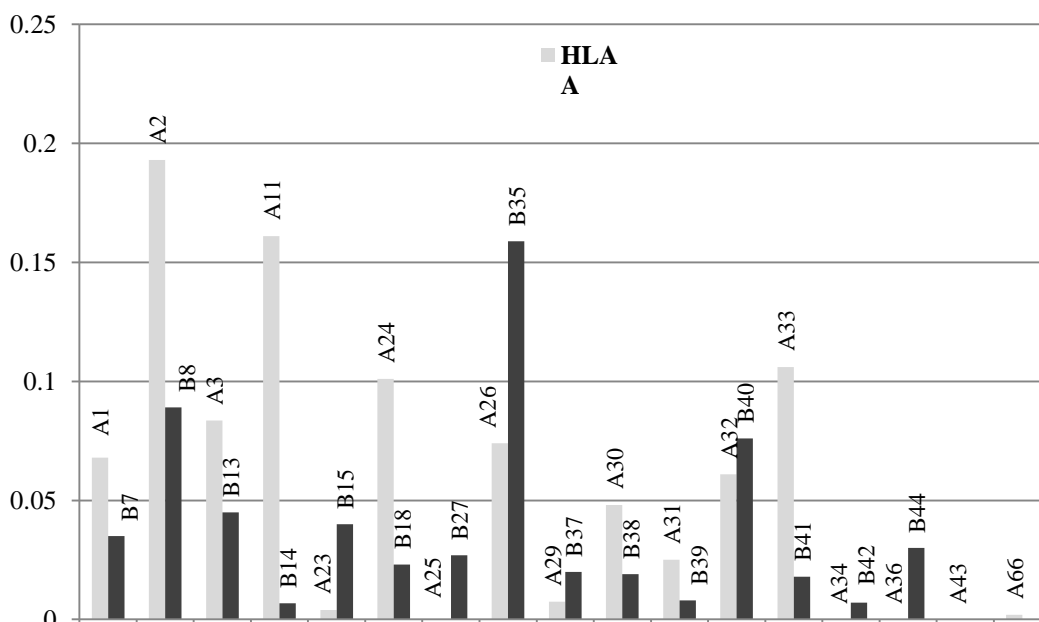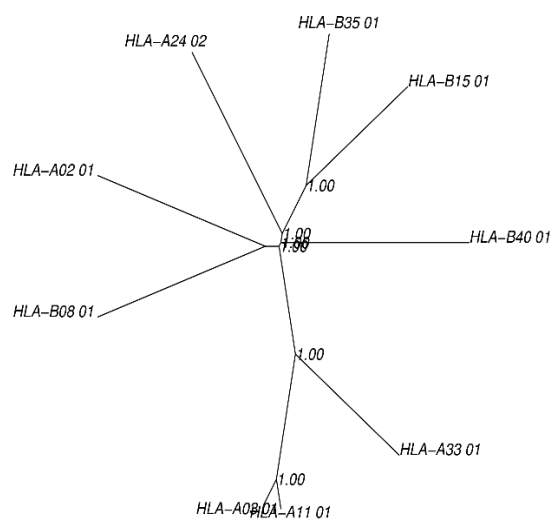

Figure 1. HLA A and B antigens with their mean frequency in Pakistani population

MHC specificity tree
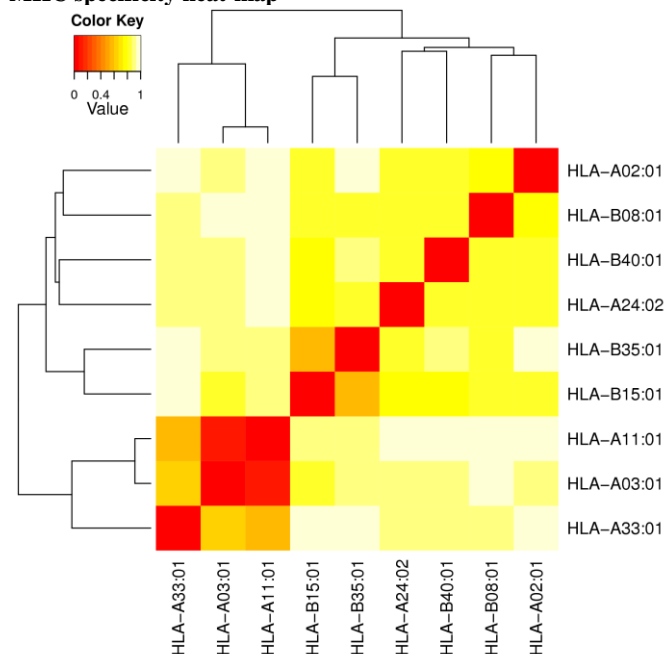
MHC specificity heat-map

Figure 2. HLA A*0201, 0301, 2402 and HLA B*0801, 1501, 4001 were selected for further progression into the study based on dissimilar binding preferences
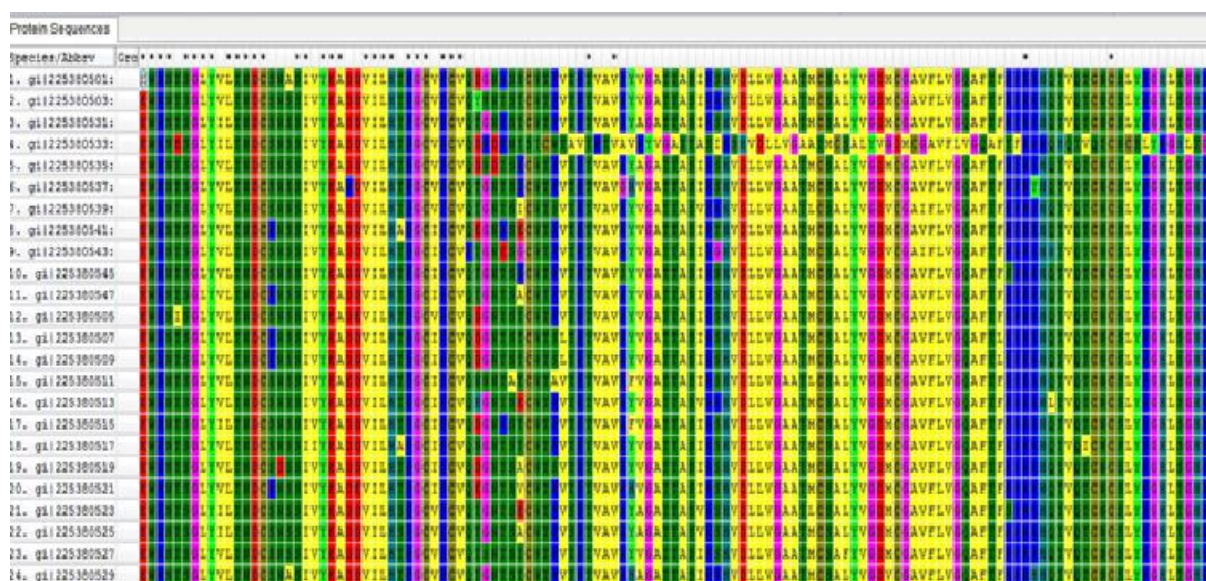

Figure 3. Translated amino acid sequences of E1 protein of isolates from Pakistan. Conserved sites are marked with asterix at the top.

network (ANN) based while the third one is called matrix based method.[18] Peptide length was fixed at 10 amino acids regarding its convenient binding capacity to the majority of HLA I molecules. Threshold of strong binders was set at less than 50 nM. FASTA amino acid sequence of all 24 isolates was copied into the server and analyzed for binding with selected HLA A and B alleles, at 10 amino acids sequence length.
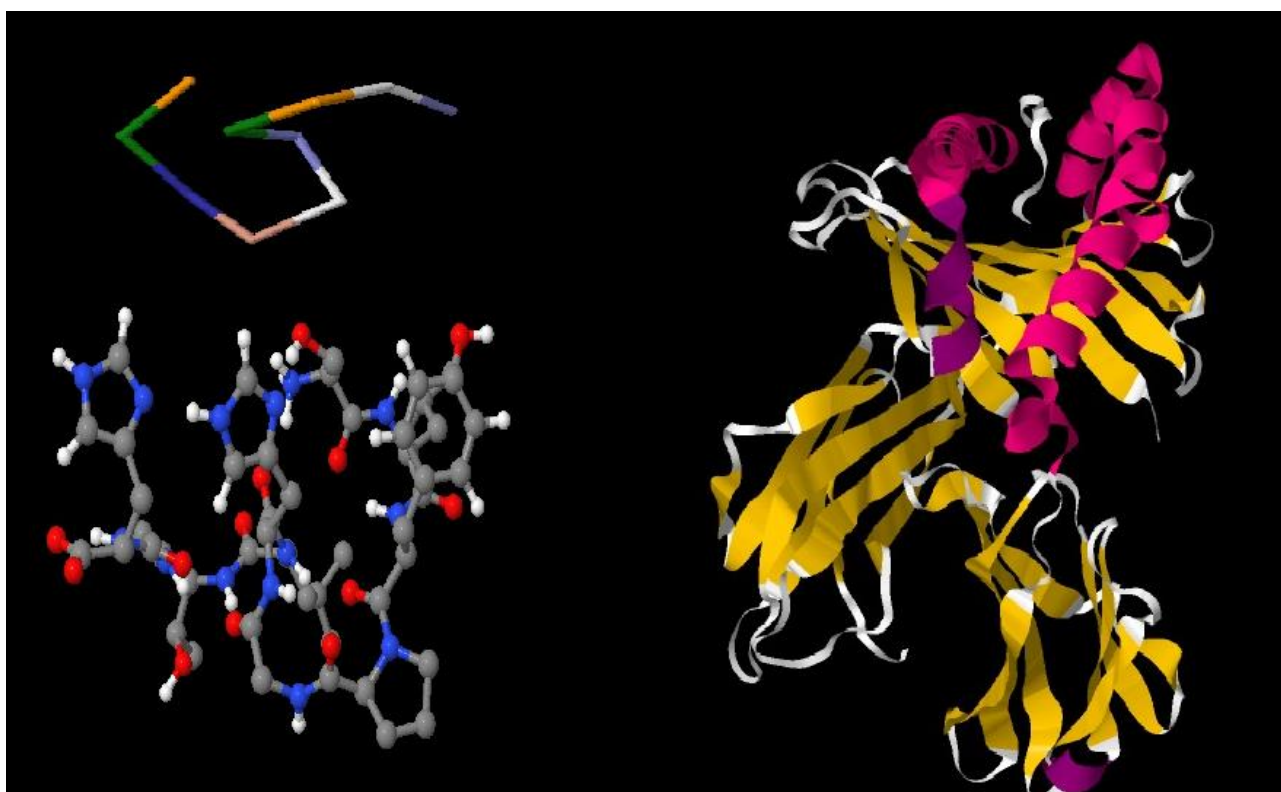
Figure 4. Left side shows linear and ball and stick models of peptide 1A (SLYPGHLTGH) generated in Pepfold. Right side shows HLA A*0301 in complex with peptide (PDB ID 2XPG)

The resulting strong binder peptide epitopes were fed into Pepfold server available from Mobyle@RPBS to generate 3D structure of the peptide.

Results

Number of strongly binding epitopes from 24 amino acid sequences against selected six HLA alleles (three each of HLA A and HLA B) are shown in table 1. HLA A*201, B*1501 and B*4001 do not bind any of the epitopes from E1 sequences analyzed at 50 nM threshold. HLA B*801 recognizes epitopes from 21 isolates. HLA A*301 and A*2402 recognize epitopes from all 24 isolates.

In total five different epitopes were identified, with variations within them according to isolates. We hypothetically divided these epitopes as shown in table 2.

Figure 4 shows 3D structure generation of peptide 1A followed by its loading onto HLA*0301 (PDB ID 2xpg).

Several weak binders (at threshold 50-500 nM) were also predicted by the program, including 25 epitopes for HLA A*0201, 72 for HLA B*1501 and only 2 for HLA B*4001 (data not shown).

D: Comparison with Reference Sequence and Epitope Mapping
Data Collection and Analysis

For verification of our computed epitopes, we took the reference sequence of HCV genotype 3 E1 protein (NC_009824.1 and YP_001491550.1) from NCBI reference sequence database and determined if any of our epitopes (or variation thereof) were also found in reference sequence annotated at NCBI.

Results

While comparing with NCBI reference sequence (YP_001491550.1), we found that all our computed epitopes (or variations thereof) were present in reference sequence.

Hepatitis C Virus Subtype 3a Envelope Protein 1

Table 1. Epitope sequences bound by representative HLA alleles from 24 isolates of HCV subtype 3a E1 protein. Total number of isolates recognized is also shown in the last row

No	Accession numbers	HCV Isolate	HLA A*201	HLA A*301	HLA A*2402	HLA B*801	HLA B*1501	HLA B*4001
1	225380501	HCV 32	0	SLYPGHLTGH	RYVGATTASI VFLVGQAFTF	TFRPRRHQTV	0	0
2	225380503	HCV 31	0	SLYPGHLSGH	RYVGATTASI VFLVGQAFTF	TFRPRRHQTV	0	0
3	225380531	HCV47	0	SLYPGHLTGH	RYAGATTASI VFLVGQAFTF	TFRPRRHQTV	0	0
4	225380533	HCV K40	0	SLYPGHLTGH	RYVGATTASI VFLVGQAFTF	0	0	0
5	225380535	HCV5	0	SLYPGHLTGH	RYAGATTASI VFLVGQAFTF	TFRPRRHQTV	0	0
6	225380537	HCV37	0	SLYPGHLTGH	VFLVGQAFTF	TFRPRYHQTV	0	0
7	225380539	HCV K60	0	SLYPGHLTGH	IFLVGQAFTF	TFRPRRHQTV	0	0
8	225380541	HCV9	0	SLYPGHISGH	RYVGATTASI VFLVGQAFTF	TFRPRRHQTV	0	0
9	225380543	HCV23	0	SLYPGHLTGH	RYVGATTASI IFLVGQAFTF	TFRPRRHHTV	0	0
10	225380545	HCV43	0	SLYPGHLTGH	SYVGATTASI VFLVGQAFTF	0	0	0
11	225380547	HCV16	0	SLYPGHLTGH	KYVGATTASI VFLVGQAFTF	TFRPRRHQTV	0	0
12	225380505	HCV27	0	SLYPGHLTGH	KYVGATTASI VFLVGQAFTF	TFRPRRHQTV	0	0
13	225380507	HCV28	0	SLYPGHLSGH	RYVGATTASI	TLRPRRHQTV	0	0
14	225380509	HCV41	0	SLYPGHLSGH	RYVGATTASI	TLRPRRHQTV	0	0
15	225380511	HCV61	0	SLYPGHLSGH	VFLVGQAFTF	TFRPRRHQTV	0	0
16	225380513	HCV K3	0	SLYPGHLSGH	VFLVGQAFTF	TFRPRRHLTV	0	0
17	225380515	HCV38	0	SLYPGHLSGH	VFLVGQAFTF	TFRPRRHQTV	0	0
18	225380517	HCV44	0	SLYPGHLSGH	KYVGATTASI VFLVGQAFTF	TFRPRRHQTV RPRRHQTVQI	0	0
19	225380519	HCV51	0	SLYPGHLSGH	RYVGATTASI VFLVGQAFTF	TFRPRRHQTV	0	0
20	225380521	HCV54	0	SLYPGHLTGH	VFLVGQAFTF	TFRPRRHQTV	0	0
21	225380523	HCV19	0	SLYPGHLTGH	VFLVGQAFTF	0	0	0
22	225380525	HCV13	0	SLYPGHLTGH	VFLVGQAFTF	TFRPRRHQTV	0	0
23	225380527	HCV17	0	SLYPGHLTGH	RYVGATTASI VFLVGQAFTF	TFRPRRHQTV	0	0
24	225380529	HCV30	0	SLYPGHLTGH	VFLVGQAFTF	TFRPRHHQTV	0	0
	Total number of epitopes bound		0	24	37	22	0	0
	HCV isolates identified (out of total 24)		0	24	24	21	0	0

Table 2: Hypothetically divided five different epitopes from E1 protein of HCV, classified as strong HLA binders. Highlighted are the variations within epitopes according to different HCV isolates.

S No.	Hypothetical Epitope Number	Starting Positions in E1 Protein Sequence from Different Isolates	Amino Acid Sequence
1	1A	114-116	SLYPGHLTGH
	1B		SLYPGHLSGH
2	2A	56-58	RYVGATTASI
	2B		RYAGATTAS
	2C		SYVGATTASI
	2D		KYVGATTASI
3	3A	91-93	VFLVGQAFTF
	3B		IFLVGQAFTF
4	4A	99	TFRPRRHQTV
	4B		TFRPRYHQTV
	4C		TFRPRRHHTV
	4D		TLRPRRHQTV
	4E		TFRPRRHLTV
5	5A	101	RPRRHQTVQI

DISCUSSION

Currently three prophylactic HCV vaccines are under clinical trials, one containing recombinant glycoproteins E1 and E2 as immunogens, by National Institute of Allergy and Infectious Diseases (NIAID), USA. It is yet to enter the phase II clinical trials.[7] E proteins of HCV offer a unique target for vaccine development as it has previously been shown in chimpanzees that developed antibodies (in HCV genotype 1) are cross reactive against E proteins of other genotypes as well.[8] We have identified five epitopes from E1 protein of HCV 3a isolates, the most prevalent subtype in Pakistan.[19] Mapping of these epitopes to reference sequence has shown that these epitopes are also present in it, thus qualifying these epitopes as candidates for peptide vaccine not only in Pakistan but also worldwide.

In addition to being immunogenic, designing a vaccine must take into account target population. An HCV vaccine has to consider HCV genomic sequences and distribution with regard to Pakistani population.[20,21] In the present study, we have first identified commonest HLA types expressed in Pakistani population. Thus, we were able to cover most of them using this approach with regards to their HLA types (taking into account publicly available data). Then we selected the most dissimilar HLA antigens for further analysis in order to high coverage of E1 protein epitopes.

Similar bioinformatics approach was used by other scientists in an effort to identify immunogenic epitopes from HCV. Ikram A et al have identified both B and T cell epitopes of HCV subtype 3a from a different envelope protein (E2) than ours. In addition they have analyzed isolates as well as HLA types from all over the world. In contrast we have restricted our analysis to HCV isolates from Pakistan and HLA types of Pakistani population to identify the best possible epitopes suitable for our country.[22] Similar study was conducted by Idrees S et al for analyzing partial sequence of E1 protein (NCBI ID ACN92051) and considering all available HLA types in bioinformatics servers employed.[23] The difference in epitope prediction is expected to be based on different analysis of HLA alleles and different bioprediction servers and tools that we used. Shehzadi et al analyzed entire viral proteome for similar investigation. In their study on extracting E1 epitopes, they predicted the binding activity of our two epitopes (number 2 and 3) to HLA class II.[21]

Three alleles included in our analysis A*0201,

B*1501 and B*4001 were not predicted to be bound to any of the epitopes. Extrapolating its implication is that people carrying these alleles will not be able to respond to a vaccine containing these epitopes only. However, it is due to the fact that epitope binding threshold was set at 50nm to predict strong binders only. Several weak binders (at threshold 50-500 nM) were also predicted by the program, including 25 epitopes for HLA A*0201, 72 for HLA B*1501 and only 2 for HLA B*4001. The amino acid sequences of these epitopes have not been shown here. However, adding these weak binders to a peptide vaccine can cover most of the population.

CONCLUSION

By identifying HCV subtype 3a E1 protein epitopes as binders of HLA alleles in dry lab, this article introduces "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. It 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.

REFERENCES

1. Ur Rehman I, Vaughan G, Purdy MA, Xia GL, Forbi JC, Rossi LM, et al. Genetic history of hepatitis C virus in Pakisan. Infect Genet Evol 2014; 318:24.

2. Saleha S, Kamal A, Ullah F, Khan N, Mahmood A, Khan S. Prevalence of hepatitis C virus genotypes in district Bannu, Khyber Pakhtunkhwa, Pakistan. Hepat Res Treat 2014; 2014:165826.

3. Rehman I, Idress M, Ali M, Ali L, Butt S, Hussain, A et al. Hepatitis C virus genotype 3a with phylogenetically distinct origin is circulating in Pakistan. Genet Vaccines Ther 2011; 9(1):2.

4. Afzal S, Idrees M, Ali M, Ilyas M, Hussain A, Akram M, et al. Envelope 2 protein phosphorylation sites S75 and 277 of hepatitis C genotype 1a and interferon resistance: A sequence alignment approach. Virol J 2011; 8:71.

5. Ahfaq UA, Javed T, Rehman S, Nawaz Z, Riazuddin S. An overview of HCV molecular biology, replication and immune responses. Virol J 2011; 8:161.

6. Frey SE, Houghton M, Coates S, Abrignani S, Chien D, Rosa D, et al. Safety and immunogenicity of HCV E1E2 vaccine adjuvanated with MF59 administered to healthy adults. Vaccine 2010; 28(38):6367-73.

7. Law LMJ, Landi A, Magee WC, Tyrrell DL, Houghton M. Progress towards a hepatitis C virus vaccine. Emerging microbes and infections 2013; 2: e79.

8. Meunier JC, Gottwein JM, Houghton M, Russell RS, Emerson SU, Bukh J, et al. Vaccine induced cross genotype reactive neutralizing antibodies against hepatitis C virus. J Infect Dis 2011; 204(8):1186-90.

9. Stamataki Z, Coates S, Abrignani S, Houghton M, McKeating JA. Immunization of human volunteers with hepatitis C virus envelope glycoproteins elicits antibodies that cross neutralize heterologous virus strains. J Infect Dis 2011; 204(5):811-3.

10. HLA Nomenclature: HLA Alleles Numbers. [cited 2014, August 25] Available from: http://hla.alleles.org/nomenclature/stats.html

11. Tipu HN, Ahmed D. Differential binding of Aspergillus fumigatus allergenic proteins to human leukocyte antigen (HLA) alleles in dry lab: potential for immunotherapy. Cell Biol: Res Ther 2014; 3:1.

12. Hillen N, Mesteri G, Lemmel C, Weinzierl AO, Muller M, Wernet D, et al. Essential Differences in Ligand Presentation and T Cell Epitope Recognition among HLA Molecules of the HLA B44 Supertypes. Eur J Immunol 2008; 38(11):2993-3003.

13. 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.

14. Whelan FJ, Yap NVL, Surette MG, Golding GB, Bowdish DM. A Guide to Bioinformatics for Immunologists. Front Immunol 2013; 4:416.

15. Sirskyj D, Diaz-Mitoma F, Golshani A, Kumar A, Azizi A. Innovative Bioinformatic Approaches for Developing Peptide-based Vaccines Against Hypervariable Viruses. Immunol Cell Biol 2011; 89(1):81-9.

16. Allele frequencies in worldwide populations [online] [cited 2014, August 22] Available from: http://www.allelefrequencies.net/hla6006a.asp

17. Thomsen M, Lundegaard C, Buus S, Rasmussen M, Lund O, Nielsen M. MHCcluster, a method for functional clustering of MHC molecules. Immunogenetics 2013; 65(9):655-65.

18. Karosiene E, Lundegaard C, Lund O, Nielsen M. NetMHCcons: a consensus method for the major

histocompatibility complex class I predictions. Immunogenetics 2012; 64(3):177-86.

19. Attaullah S, Khan S, Ali I. Hepatitis C virus genotypes in Pakistan: A systemic review. Virol J 2011; 8:433.

20. Groot DSA, Sbai H, Aubin SC, Mcmurry J, Martin W. Immuno-informatics: Mining genomes for vaccine development. Immunol Cell Biol 2002; 80(30):255-69.

21. Shehzadi A, Ur Rehman S, Idrees M. Promiscuous prediction and conservancy analysis of CTL binding epitopes of HCV 3a viral proteome from Punjab Pakistan: an In Silico Approach. Virol J 2011; 8:55.

22. Ikram A, Anjum S, Tahir M. In silico identification and conservation analysis of B-cell and T-cell epitopes of hepatitis C virus 3a genotype enveloped glycoprotein 2 from Pakistan: A step towards heterologous vaccine design. Hepat Mon 2014; 14(6):e9832.

23. Idrees S, Ashfaq UA. Structural analysis and epitope prediction of HCV E1 protein isolated in Pakistan: an in-silico approach. Virol J 2013; 10:113.