Bangladesh J Microbiol, Volume 35, Number 1, June 2018, pp 45-55

Original Article

In Silico screening of T-cell and B-cell Epitopes of Rotavirus VP7 and VP4 proteins for Effective Vaccine Design

Md. Sadikur Rahman Shuvo, Sanjoy Kumar Mukharjee and Firoz Ahmed*

Department of Microbiology, Noakhali Science and Technology University, Noakhali-3814, Bangladesh.

Rotavirus is one of the deadliest causative agents of childhood diarrhea which causes half a million child death across the globe, mostly in developing countries. However, effective vaccine strategies against rotavirus are yet to be established to prevent these unwanted premature deaths. In this regard, *in silico* vaccine design for rotavirus could be a promising alternative for developing countries due to its efficiency in shortening valuable time and cost. The present study described an epitope-based peptide vaccine design against rotavirus, using a combination of T-cell and B-cell epitope predictions and molecular docking approach. To perform this, sequences of rotavirus VP7 and VP4 proteins were retrieved from the NCBI database and subjected to different bioinformatics tools to predict most immunogenic T-cell and B-cell epitopes. From the identified epitopes, the sequence VMSKRSRSL of VP7 and TQFTDFVSL of VP4 was identified as the most potential epitopes based on their antigenicity, conservancy and interaction with major histocompatability complex I (MHC-I) alleles. Moreover, the peptide VMSKRSRSL interacted with human leukocyte antigen, HLA-B*08:01 and TQFTDFVSL interacted with HLA-A*02:06 with considerable binding energy and affinity score. Combined population coverage for our identified epitopes was found 70.53% and 45.64% for world population and South Asian population respectively. All these results suggest that, the epitopes identified in this study could be a very good vaccine candidate for the strains of rotavirus circulating in Bangladesh. However, as this study is completely dependent on computational prediction algorithms, further *in vivo* screening is required to come up in a precise conclusion about these epitopes for effective rotavirus vaccination.

Key words: Rotavirus, Epitope, VP4, VP7, Vaccine design.

Introduction

Group A rotavirus is the most widely found pathogenic rotavirus and it is the major cause of gastroenteritis in infants and young children worldwide[1]. In spite of recent progress in controlling the global burden of rotavirus, the mortality rate due to this virus is still high. An estimated 215,000 child deaths has been documented in 2013 in resource-poor settings of Asia and sub Saharan Africa countries. In Bangladesh, the number of death was documented as many as 2700 in 2013[2, 3]. Along with these early life deaths, other health and economic consequences are worsening the rotavirus disease episodes in Bangladesh.

Rotavirus, a double stranded RNA virus of *Reoviridiae* family, contains a genome of 11 segments that encode for 6 structural proteins (VP1, VP2, VP3, VP4, VP6, VP7) and 6 nonstructural proteins (NSP1, NSP2, NSP3, NSP4, NSP5, NSP6) [4]. Among all these proteins, the outer capsid proteins, VP7 and VP4 grab special attentions due to their important role in rotavirus characteristics. These two proteins can elicit neutralizing antibody responses that are serotype specific and serotype crossreactive[5, 6, 7]. Current serotype dependent rotavirus classification system is based on these two proteins. The G type is based on glycoprotein VP7 and the P type is based on protease sensitive protein VP4. Moreover, these two proteins have been selected as a target of vaccine preparation for many years due to their surface exposed structure and considerably large size.

In group A rotavirus, the frequency of reassortment and interspecies transmission to form novel serotypes is high due to its multiple host range along with the presence of segmented genome[8]. Therefore, the effectiveness of a monovalent vaccine to give protection against all G and P types as well as against any emerged unusual strains in case of outbreak is of great question[9]. In this situation, highly conserved sequence specific vaccine target should be made to find out effective vaccine strategies for a specific geographic region.

At present, epitope-based vaccine design using *in silico* methods is a very promising approach of vaccine development due to its time and cost effectiveness. Some recent studies show that, epitope based vaccination strategies can efficiently elicit defensive immune responses against diverse infective agents[10, 11, 12, 13]. In this context, this *in silico* study attempted to find out effective epitopes (T-cell and B-cell epitope) to design vaccine candidate against group A rotavirus based on the available VP7 and VP4 sequences circulating in Bangladesh.

Materials and Methods

A graphical representation of methods applied to predict T cell and B cell epitopes for rotavirus vaccine design is shown in Figure 1. Stepwise details of methods are as follows:

Corresponding author:
Dr. Firoz Ahmed, Associate Professor, Department of Microbiology, Noakhali Science & Technology University, Mobile: +8801713195658, Email: firoz19701016@gmail.com.

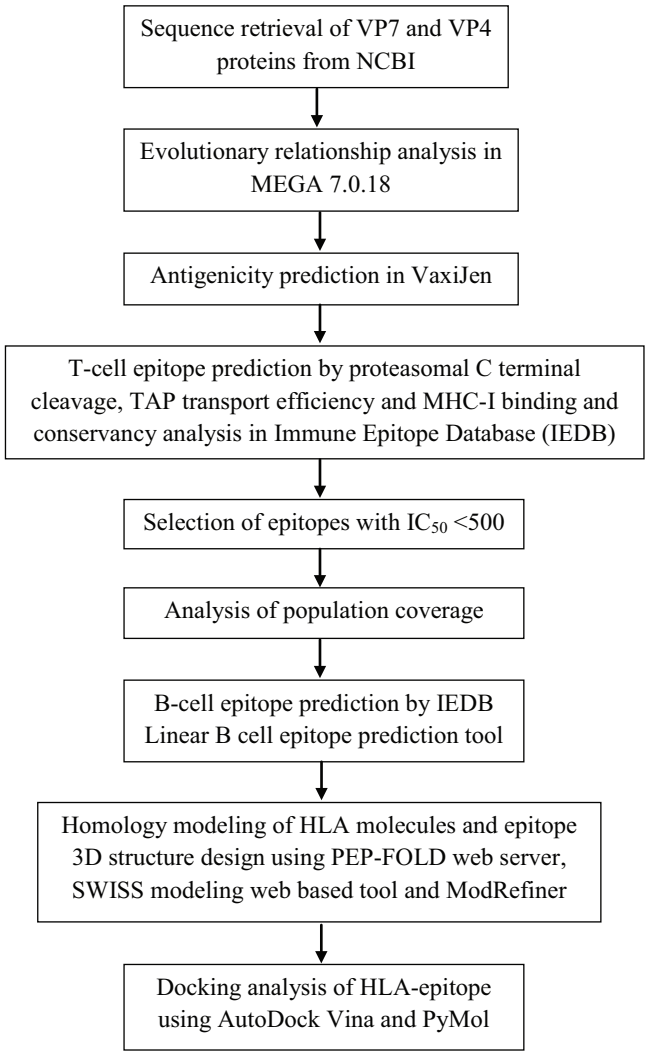

Figure 1: *Graphical representation of methods applied to predict T cell and B cell epitopes for rotavirus vaccine design.*

Database preparation and evolutionary relationship analysis

VP7 and VP4 protein sequences of rotavirus A circulating in Bangladesh were retrieved from protein database of NCBI (https://www.ncbi.nlm.nih.gov/protein). Advanced filtering system of NCBI was used to filter the sequences of Bangladeshi source and human specific rotavirus A sequences. The retrieved sequences of VP7 and VP4 were assembled in separate files for multiple sequence alignment using MEGA 7.0.18 software[14]. ClustalW algorithm was used for multiple sequence alignment. Phylogenetic tree was constructed by neighbor joining method with a bootstrap value of 1000. Representative sequences from each cluster of phylogenetic tree were used for further analysis.

Prediction of Antigenicity

The representative VP7 and VP4 protein sequences were tested for antigenicity prediction using VaxiJen v2.0 online server[15] at threshold value of 0.5 and 0.4 for VP7 and VP4 protein sequences respectively. This server predicted the level of antigenicity for each protein sequence. The most antigenic sequences were analyzed for probable vaccine target identification.

Identification of T-cell epitope

The most antigenic VP7 and VP4 proteins were subjected to T-cell epitope identification by NetCTL-1.2 web tool[16]. Interacting major histocompatability complex I (MHC-I), transport efficiency, proteosomal cleavage prediction and transporter of antigenic peptide (TAP) values are involved to calculate the summed score. Threshold score value was set to 0.75 for moderate sensitivity and specificity. Epitopes were selected for all 12 different super types of human leukocyte antigen (A1, A2, A3, A24, A26 B7, B8, B27, B39, B44, B58, and B62). The best epitopes were tested for conservancy analysis in Epitope Conservancy Analysis tool[17]. The epitopes possessing at least 80% conservancy were elected for finding their half maximal inhibitory concentration (IC_{50}) value followed by binding to human leukocyte antigen (HLA) by stabilized matrix method (SMM)[18]. Here the reference HLA dataset was used for binding prediction of epitopes and the peptide length was set to 9. Epitopes of IC_{50} value less than 50nM has higher affinity and IC_{50} values less than 500nM has moderate affinity to HLA[19]. In this study, we considered epitopes of highest conservancy to find their respective HLA binding at IC_{50} value <500nM.

Population coverage prediction

The potential epitopes that showed interaction with various major histocompatability complex (MHC) alleles were used for calculating the human population coverage for each of the epitopes. Population coverage tool of IEDB[20] was used to predict at what percentage the world population has the MHC alleles that bind the selected epitopes. This tool calculates average number of HLA/epitope combination conceded by the population of different geographical distributions. We focused specifically on the South Asian and Asian population for the population coverage study.

B cell epitope prediction

IEDB Linear B cell epitope prediction tool[21] was used for predicting probable B cell epitope of VP7 and VP4 protein sequences which gave higher vaxijen value. B cell epitope can be of various lengths ranging from 2 to 85. Hydrophobicity of the protein was checked by Parker Hydrophobicity Prediction tool[22]. Kolaskar and Tongaonkar antigenicity [23] tool detected potential antigenicity throughout the sequences. Emini surface accessibility prediction detected accessible regions in the proteins[24].

Designing 3D structure of epitopes and HLA homology modeling

Three dimensional structures of the best epitopes of VP7 and VP4 datasets were predicted using PEP-FOLD web server[25]. The tool predicted all possible structures. Best structures were selected based on lowest energy model which were used for further analysis. Structural patterns of the epitopes were verified by structure alignment with the original VP7 and VP4 proteins by PyMol tool[26]. Homology modeling of HLA proteins was done by SWISS modeling[27] web based tool. ModRefiner[28] was used

to correct the hypothetical structure. The predicted structures were validated using QMEAN[29], ERRAT[30] and PROCHECK[31]. Ramachandran plot was constructed by PROCHECK evaluated backbone conformation by checking non-GLY residues in the excluded regions. Overall models were validated in ERRAT server for accuracy.

Docking analysis of HLA-Epitope

AutoDock Tool 1.5.6[32] was employed in docking analysis of HLA-Epitope interaction. HLA molecules were considered as protein and epitopes as ligands. Proteins were added with polar hydrogen. The parameters were used as default of AutoDock Tool for docking analysis. The grid box was prepared according to the protein volume and docking site. Energy range was kept at 4 as default. Binding affinity of ligands was recorded from log file. Docking was visualized using PyMol visualizing tool.

Results

Sequence retrieval and evolutionary analysis of VP4 and VP7 proteins

Total 44 VP7 and 42 VP4 protein sequences were retrieved from NCBI database. After performing multiple sequence alignment, phylogenetic trees were constructed (Figure 2). The trees show closer relationships among VP7 and VP4 proteins. Analyzing the phylogenetic trees, 4 clusters were observed among all VP7 and VP4 protein sequences. One representative protein sequence from each cluster was selected which results in 4 VP7 and 4 VP4 proteins for further analysis.

Prediction of antigenic proteins

All the selected proteins were subjected to antigenicity test in VaxiJen web tool in order to identify the most potential antigenic VP7 and VP4 proteins. The proteins which satisfied the condition, threshold level > 0.5 for VP7 and > 0.4 for VP4 were considered as antigenic [33]. The proteins that showed higher degree of antigenicity than others were selected. Our selected VP7 protein (AAT78645.1) and VP4 protein (ADI59479.1) showed the highest VaxiJen value 0.5247 and 0.4897 respectively (Table 1). These two sequences were used for further analysis.

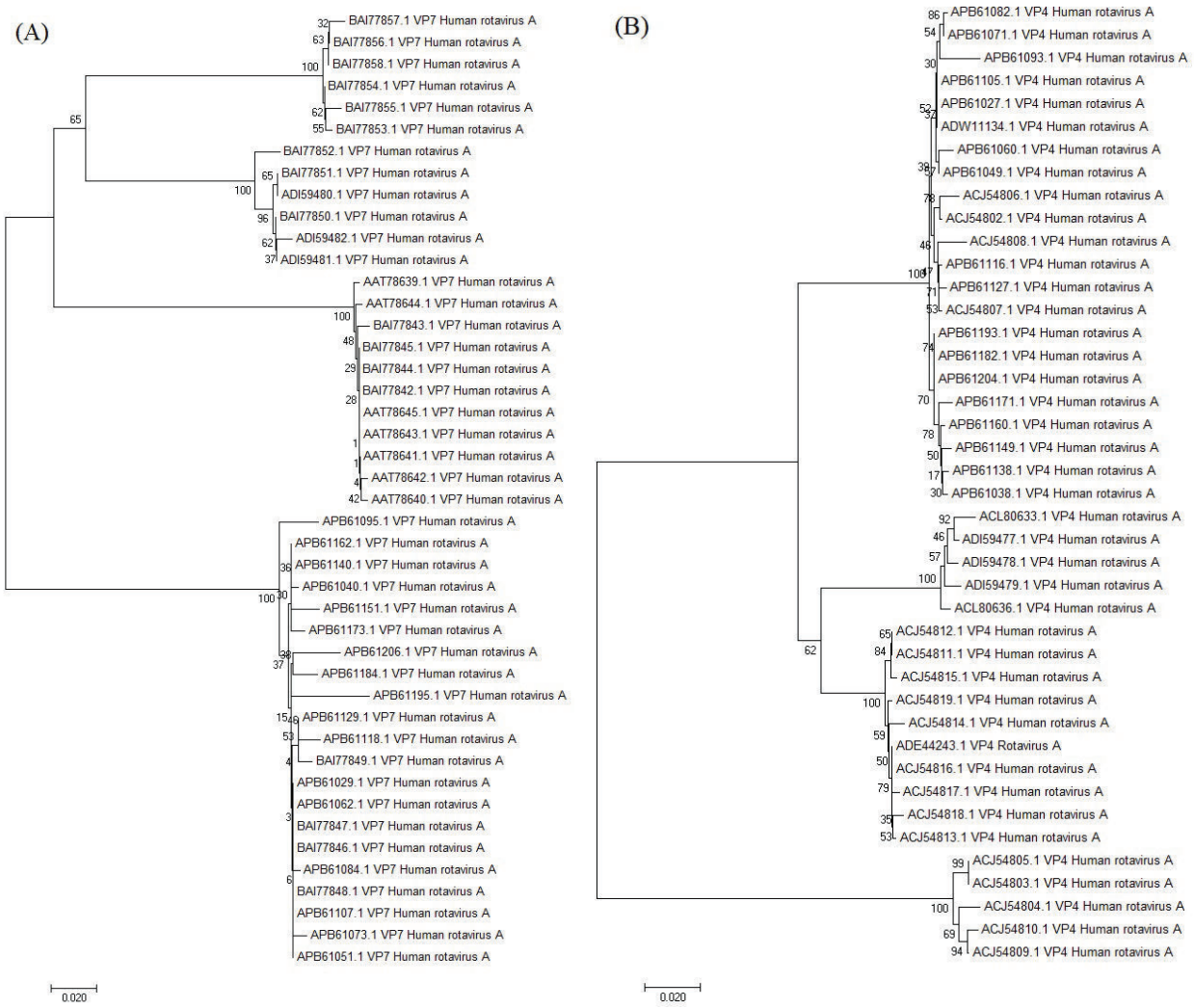

Figure 2: *Evolutionary relationship of available VP7 (A) and VP4 (B) proteins of rotavirus represented in a phylogenetic tree.*

Table 1. *Antigenic protein prediction by Vaxijen server. Threshold values were 0.5 and 0.4 for rotavirus VP7 and VP4 proteins respectively.*

Protein	Protein accession number	Vaxijen Score
VP7	AAT78645.1	0.5247
	BAI77854.1	0.5040
	BAI77850.1	0.5031
	APB61062.1	0.4697
VP4	ADI59479.1	0.4897
	APB61138.1	0.4818
	ACJ54804.1	0.4645
	ACJ54814.1	0.4626

Potential T-cell epitope prediction

Selected VP7 and VP4 proteins were then analyzed in NetCTL web tool for epitope identification at threshold value 0.75. A total of 92 VP7 and 182 VP4 epitopes were found. Among them 6 VP7 and 6 VP4 epitopes were selected based on their NetCTL score, conservancy among the sequences and antigenicity by Vaxijen score (Table 2). The NetCTL value comprises of transporter associated with antigen processing (TAP)/proteasomal cleavage/MHC-I combined predictor. These selected T-cell epitopes were subjected to MHC-I binding prediction. Then stabilized matrix method (SMM) was used to predict IC_{50} for the epitopes with MHC-I in binding condition. Epitopes with IC_{50} < 500nM shows moderate affinity with MHC-I and those elicited comparatively higher affinity (IC_{50} < 500nM) were selected for next study (Table 3). Among the VP7 epitopes, VMSKRSRSL was identified to interact with most number of MHC-I alleles (HLA-B*08:01, HLA-A*02:03, HLA-B*15:01) with higher affinity as well as diverse MHC-I types. In case of VP4, TQFTDFVSL was considered as the best epitope which interact with five diverse MHC-I alleles (HLA-A*02:06, HLA-B-15:01, HLA-A*32:01, HLA-B*40:01, HLA-A*02:01) (Table 3). Moreover, these two epitopes showed very good conservancy and antigenicity scores (VMSKRSRSL conservancy: 95.45%, VaxiJen score: 1.3703; TQFTDFVSL conservancy: 100%, VaxiJen score: 0.9837) (Table 2).

Population coverage

Cumulative population coverage percentage was obtained for epitopes VMSKRSRSL and TQFTDFVSL. Combined world population coverage was found 70.53% at average hit 1.03 for MHC-I. On the other hand, combined population coverage for South Asian ethnic region was calculated 45.64% at average hit 0.58 (Table 4, Figure 3).

B-cell epitope identification

IEDB linear B-cell epitope identification tool analyzed the VP7 and VP4 sequences for potential epitope identification. According to Kolaskar and Tongaonkar's antigenicity method, for VP7 protein, the average antigenicity score was 1.038 with maximum and minimum score were 1.213 and 0.896 respectively. For VP4 protein, the average antigenicity score was 1.015 while maximum and minimum scores were 1.193 and 0.842 respectively. In Kolaskar and Tongaonkar's method, values >1.00 are denoted as potential antigen depending on the physicochemical properties of the protein residue [23]. Highest surface accessibility for VP7 was found in 66-73 regions and for VP4 in 666-679 regions. Parker Hydrophilicity index for VP7 was average 0.694, minimum -6.629, maximum 6.500 and for VP4 average 1.603, minimum -4.671 and maximum 7.143 at threshold value 1.603. Predicted epitope list are given in Table 5.

Validation of predicted HLA structures

Three dimensional structures of all the MHC-I alleles that interacts with the potential VP7 and VP4 epitopes were generated using SWISS homology modeling. In this study, the structural validation properties of HLA-B*08:01 and HLA-A*02:06 are discussed as these two MHC-I alleles interacted with epitope VMSKRSRSL (VP7) and TQFTDFVSL (VP4) with considerable binding energy and affinity score. Ramachandran plot for HLA-B*08:01 3D

Table 2: *Most potential epitopes of VP7 and VP4 proteins according to the overall score predicted by the NetCTL server, conservancy of epitopes and their antigenicity by Vaxijen score.*

Protein	Epitopes	NetCTLScore (nM)	Conservancy	VaxijenScore
VP7(AAT78645.1)	LISIILLNY	1.8567	84.09% (37/44)	1.2721
	MYGIEYTTI	1.6403	97.73% (43/44)	1.0924
	VMSKRSRSL	1.6348	95.45% (42/44)	1.3703
	GWPTGSVYF	1.4572	86.36% (38/44)	0.4816
	LADLILNEW	1.4316	100% (44/44)	0.2509
	YGIEYTTIL	1.3855	97.73% (43/44)	0.8176
VP4(ADI59479.1)	SIIIHSEFY	2.0953	64.29% (27/42)	0.2863
	ISIIIHSEF	1.9571	64.29% (27/42)	0.5370
	TQFTDFVSL	1.4200	100% (42/42)	0.9837
	FIPKRSYRV	1.1427	64.29% (27/42)	0.7320
	SLVPTNDDY	1.0701	85.71% (36/42)	1.0581
	TLKNLNDNY	1.0473	95.24% (40/42)	0.5909

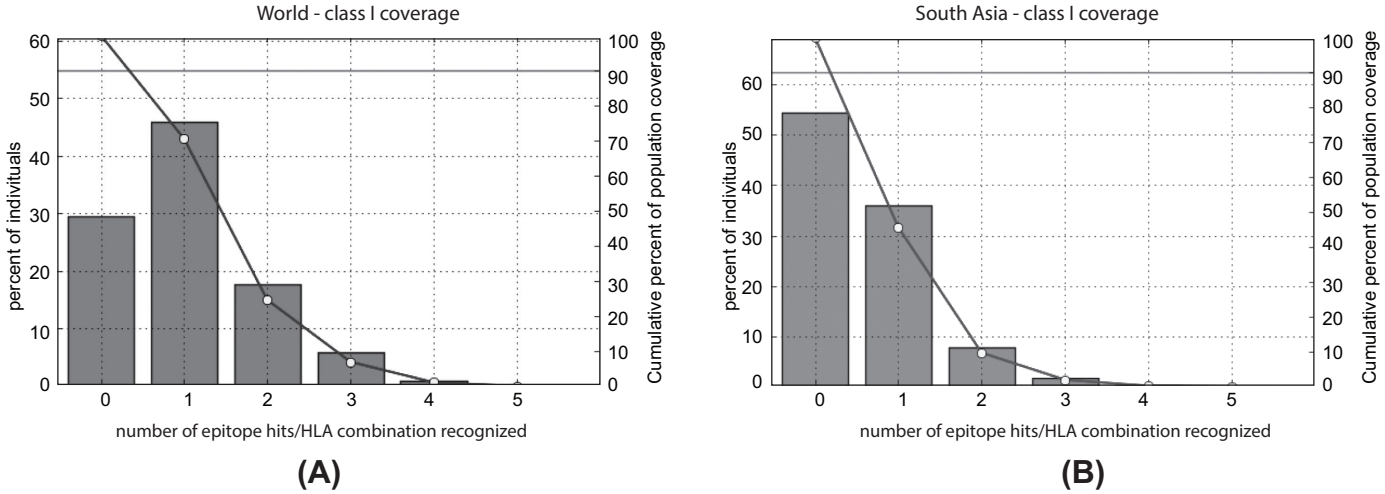

Figure 3: *Combined population coverage for the world (A) and South Asian ethnic region (B) based on MHC restriction data using the Immune Epitope Database analysis resource.*

Table 3: *Potential T-cell epitopes, along with their interacting HLA alleles, inhibitory concentration values (SMM IC_{50}), TAP (transporter associated antigen processing) score and proteasome score.*

Protein	Epitope	Interacting MHC-I allele	SMM IC_{50}	TAP score	Proteasome score
VP7	LISIILLNY	HLA-A*30:02	326.18	1.30	1.38
(AAT78645.1)		HLA-B*15:01	347.54		
	MYGIEYTTI	HLA-B*15:01HLA-A*23:01	69.25	0.20	1.25
		HLA-A*24:02	77.74		
	VMSKRSRSL	HLA-B*08:01	42.22	0.45	1.44
		HLA-A*02:03	143.86468.81		
	GWPTGSVYF	HLA-A*24:02	340.9	1.10	1.38
	LADLILNEW	HLA-B*58:01	42.2	0.33	1.46
		HLA-B*53:01	159.54		
		HLA-B*57:01	179.4		
	YGIEYTTIL	HLA-A*02:06	43.81	0.37	1.45
		HLA-A*02:01	461.43		
VP4	SIIIHSEFY	HLA-A*30:02	39.13	1.34	1.41
(ADI59479.1)		HLA-B*58:01	56.92		
		HLA-A*32:01	243.75		
		HLA-B*35:01	478.17		
	ISIIIHSEF	HLA-B*58:01HLA-B*15:01	67.5227.51	1.14	1.22
		HLA-B*57:01	335.6		
		HLA-B*35:01	360.23		
		HLA-A*68:01	387.69		
		HLA-A*23:01	476.9		
	TQFTDFVSL	HLA-A*02:06	21.51	0.50	1.64
		HLA-B*15:01	115.08		
		HLA-A*32:01	225.39		
		HLA-B*40:01	428.45		
		HLA-A*02:01	446.8		
	FIPKRSYRV	HLA-A*02:06	36.77	0.10	1.04
		HLA-A*02:03	42.55		
		HLA-A*02:01	90.39		
		HLA-A*68:02	143.08		
		HLA-B*08:01	195.23		
	SLVPTNDDY	HLA-A*30:02	329.2	1.31	1.09
		HLA-B*15:01	478.63		
	TLKNLNDNY	HLA-A*30:02	434.97	1.26	1.30
		HLA-B*15:01	487.53		

structure generated from PROCHECK states that it has 93.3% residues in most favored core regions, 5.8% residues in additional allowed regions and 0% residues in disallowed regions (Figure 6). The protein model that has > 90% residues in the core region and allowed is considered to be a high quality model [34]. The quality factor of the model analyzed by ERRAT tool is 97.3684 where a good structure model should have a quality factor > 80.00 [35]. The model has a QMEAN value of -0.72 (Figure 6). In case of Ramachandran plot for HLA-A*02:06, 3D structure states that it has 93.0% residues in most favored core regions, 6.6% residues in additional allowed regions and 0% residues in disallowed regions (Figure 7). The quality factor of the model analyzed by ERRAT tool is 93.2584. The model has a QMEAN value of 0.09 (Figure 7).

Molecular docking of HLA-Epitope interaction

AutoDock Vina was used to dock the epitopes with the predicted HLA molecules. All the docking results are interpreted in Figure 5. Binding energies of the epitopes with HLA molecules are shown in Table 6.

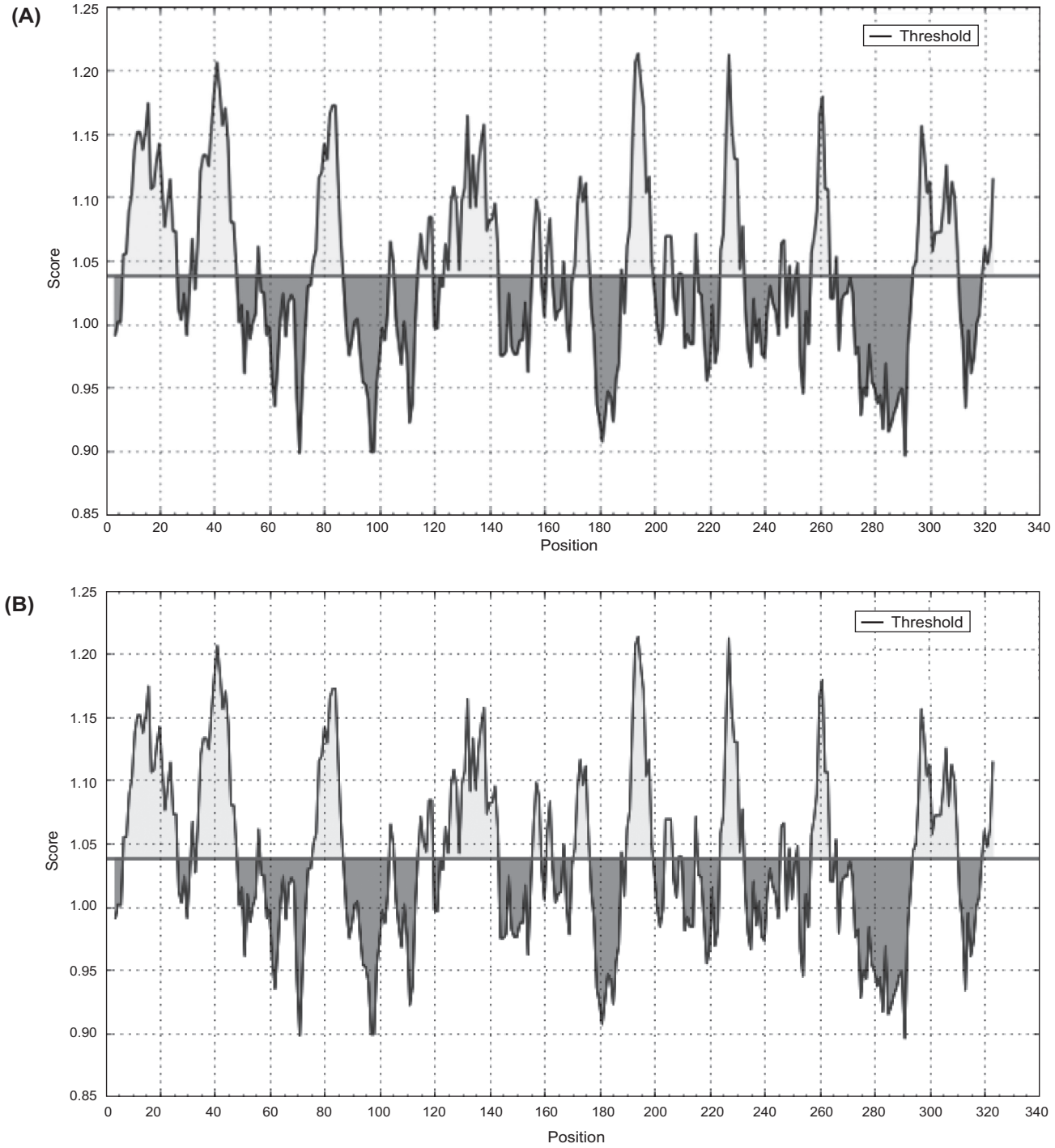

Figure 4: *Kolaskar and Tongaonkar antigenicity prediction of rotavirus protein VP7 (A) and VP4 (B).*

Table 4: *Combined population coverage for epitopes of both VP7 and VP4 proteins.*

Area	Coverage	pc90	Average hit
World population coverage	70.53%	0.34	1.03
South Asian population coverage	45.64%	0.18	0.58

Table 5: *Prediction of potential antigenic B-cell epitope region of VP7 and VP4 proteins by Kolaskar and Tongaonkar antigenicity prediction method.*

Protein	Start	End	Epitope sequences	Length
VP7	50	54	AQNYG	5
	62	76	SMDTVYSNSTREEVF	15
	86	102	PTEASTQISDGEWKDSL	17
	105	131	MFLTKGWPTGSVYFKEYSNIVDFSVDP	27
	145	155	DQNLELDMSEL	11
	175	186	YQQSGESNKWIS	12
	197	222	PLNTQTLGIGCQTTNVDSFETVAENE	26
	231	238	DGINHKIN	8
	247	256	RNCKKLGPRE	10
	266	292	ANILDITADPTTNPQIERMMRVNWKRW	27
	308	323	QVMSKRSRSLNSAAFY	16
VP4	14	77	YSVDLHDEIEQIGSEKTQNVTVNPGPFAQTRYAPVNWGHGEINDSTTVEPVLDGPYQPTTFKPP	64
	112	116	VSPTN	5
	126	135	KQFNVENSSD	10
	144	155	KGSSQSDFSNRR	12
	180	195	TPRATTDSSNTADLNN	16
	211	230	QESKCNEYINNGLPPIQNTR	20
	240	252	RSIQYRRAQVNED	13
	258	266	TSLWKEMQY	9
	281	307	IRLGGLGYKWSEISYKAANYQYNYSRD	27
	322	346	VNNFSYNGGSLPTDFSISRYEVIKE	25
	355	361	WDDSKAF	7
	376	384	SVKCVGGSY	9
	391	398	GEWPIMNG	8
	434	460	PSFSIIRTRTVNLYGLPAANPNNGNEY	27
	477	492	DDYQTPIMNSVTVRQD	16
	498	508	NDLREEFNSLS	11
	553	557	KSKLA	5
	572	582	SSASRSASIRS	11
	588	633	SNWSDASKSVLNVTDSVNDISTQTSTISKKLRLKEMITQTEGISFD	46
	643	654	KIDMSTQIGKNT	12
	663	673	SEKFIPKRSYR	11
	694	711	VDTLNEIPFDINKFAELV	18
	726	739	LKNLNDNYGITRIE	14

Table 6: *Binding energies of the epitopes with HLA molecules.*

Protein	Epitope	MHC-I alleles	Binding affinity (kcal/mol)
VP7	VMSKRSRSL	HLA-B*08:01	-8.5
		HLA-A*02:03	-7.5
		HLA-B*15:01	-8.8
VP4	TQFTDFVSL	HLA-A*02:06	-7.4
		HLA-B-15:01	-9.3
		HLA-A*32:01	-8.1
		HLA-B*40:01	-8.8
		HLA-A*02:01	-7.6

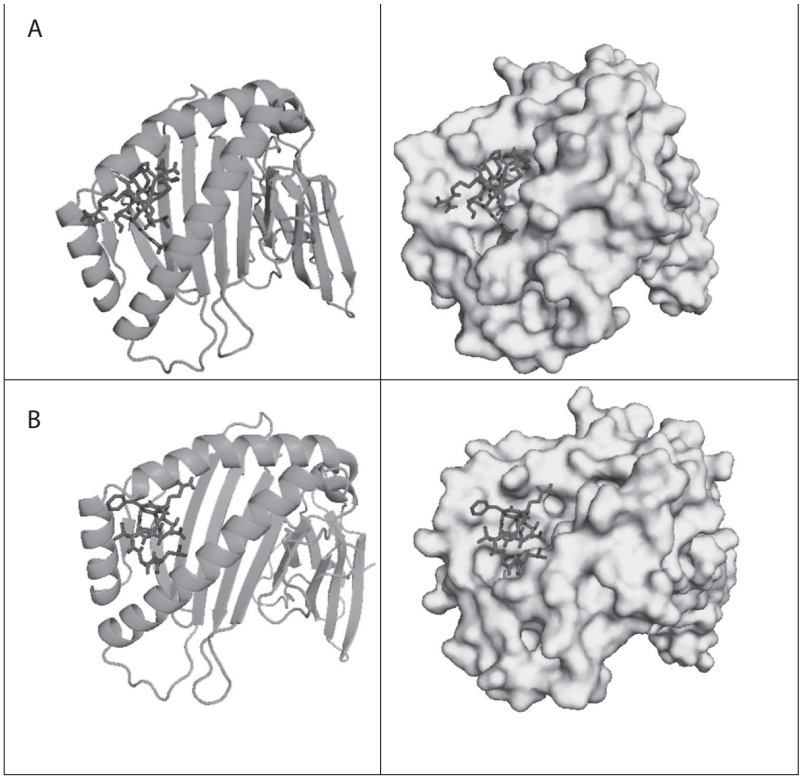

Figure 5: *Molecular interaction analysis of (A) HLA-B*08:01 and epitope VMSKRSRSL (VP7 protein) complex, (B) HLA-A*02:06 and epitope TQFTDFVSL (VP4 protein) complex, where both the epitopes bind in the groove of their respective MHC-I molecules. The models are generated by Autodock Vina.*

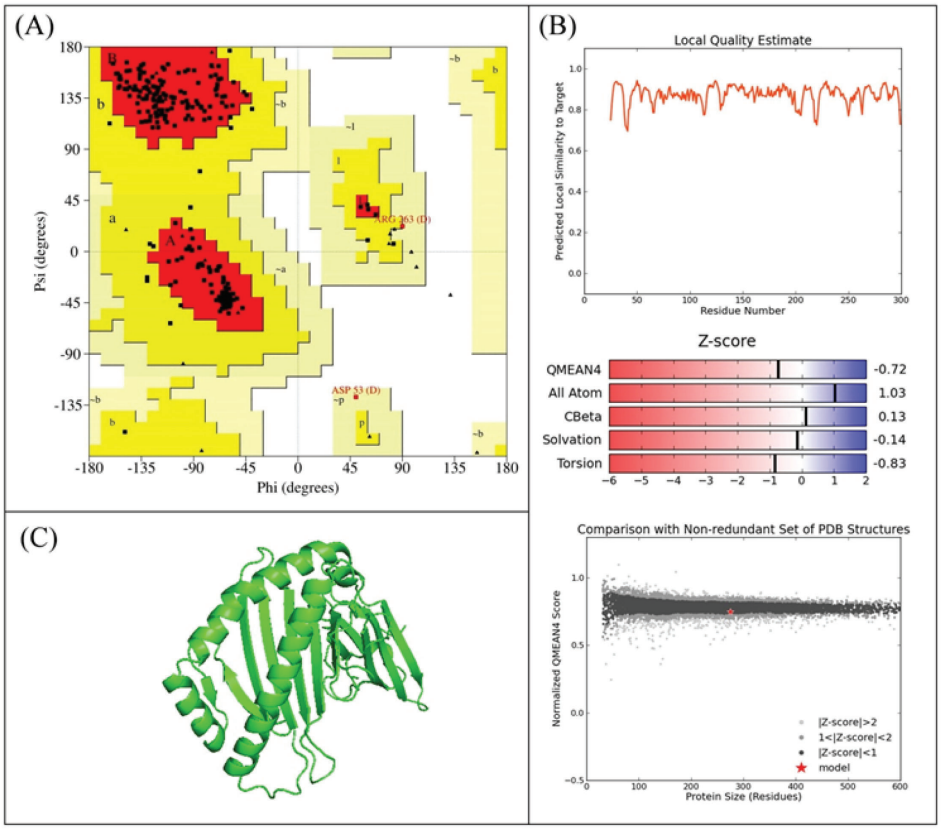

Figure 6: *Evaluation of structural superiority of HLA-B*08:01 by (A) Ramachandran plot, (B) QMEAN assessment, and (C) three-dimensional structure of final model of HLA-B*08:01.*

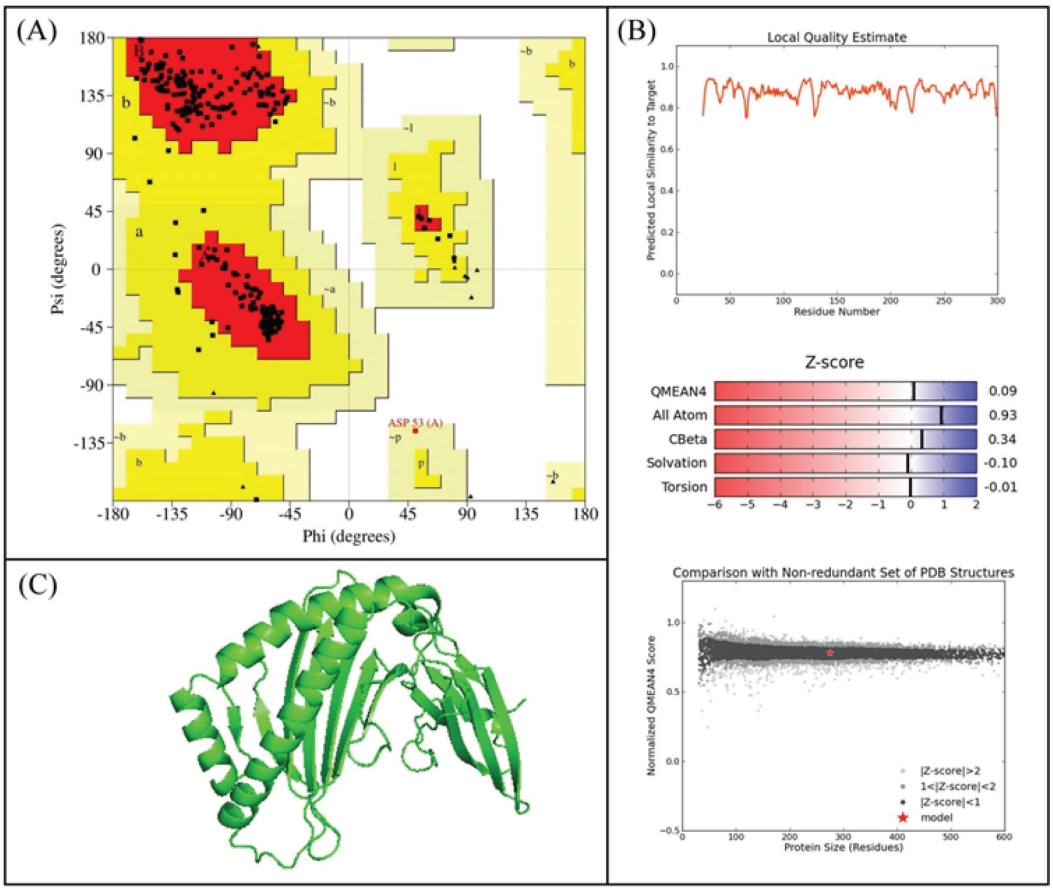

Figure 7: *Evaluation of structural superiority of HLA-A*02:06 by (A) Ramachandran plot, (B) QMEAN assessment, and (C) three-dimensional structure of final model of HLA-A*02:06.*

Discussion

In the past, vaccine development required pathogenic viruses to be attenuated or inactivated which had ample of effects on the expansion of common viral diseases. This procedure played a remarkable role in public health sector. In the twenty first century, invention of diverse groups of viruses, specially the RNA viruses urges the extensive study to discover molecular based vaccines [36]. In this case computational vaccine design is comparatively time efficient and accurate before lab scale vaccine development. *In silico* epitope screening against MHC-I molecules is much safe, specific and economic.

Bangladesh is at high risk of diarrheal epidemics. Rotavirus is the second most commonly identified pathogen causing diarrheal diseases in Bangladesh [37]. There is scarcity of epitope based vaccine design study in Bangladesh. In our study we focused on the strains of human rotavirus A which are commonly circulating in Bangladesh. Here we worked with 44 VP7 and 42 VP4 protein sequences of human rotavirus A from Bangladesh origin. Evolutionary analysis of the retrieved sequences helped to find the representative sequences having a probable vaccine target. Having antigenicity is the prime criteria of a molecule to be a vaccine. We verified the antigenicity in two phases, at first the whole proteins were tested for antigenicity and in second round, the antigenicity of the probable epitopes were calculated. This gave a precise notion of an epitope's antigenicity and its efficiency as a vaccine agent. From numerous predicted epitopes, we dropped down to 6 VP7 and 6 VP4 potential epitopes based on their NetCTL score which represent both proteasomal processing score and transporter associated antigen processing score. Among them, we found the most potential VP7 and VP4 epitopes based on the number of HLA molecules they are capable of binding as well as how conserved they are among the retrieved sequences. So the epitope selection was involved in two way verification system that helped to predict prospective epitopes more precisely.

We studied population coverage of epitope-HLA combination to predict at what percentage the potential epitopes can provoke the MHC-I. It was estimated that world's 70% population has the predicted epitope-HLA combination on the other hand 45.64% Indian Asian population covers the same type of condition. As our study is focused in developing vaccine agent in Bangladesh, we predicted Indian Asian ethnic group for coverage study. 90% population coverage of some ethnic group can be achieved by at least eleven unique epitopes [38]. In this study, we found 45.64% population coverage for Indian Asian ethnic group by only two unique epitopes. This indicates better population coverage.

Binding affinity of epitopes and HLA molecules were predicted by IC_{50} value. IC_{50} < 50 nM renders higher affinity, IC_{50} < 500 nM for intermediate affinity and IC_{50} < 5000 nM for lower affinity [33]. In our study we considered IC_{50} < 500 which indicates intermediate affinity. At this range we found 3 (three) HLA molecules binding to VMSKRSRSL epitope and 5 (five) HLA molecules to binding to TQFTDFVSL epitope. Although IC_{50} < 500 nM was used for finding the epitopes showing at least intermediate affinity, most of the epitopes showed IC_{50} < 250 nM which is significant in the sense that they have higher affinity than average [33, 39]. Those HLA molecules which showed lowest IC_{50} values were considered for docking analysis.

VP7 epitope VMSKRSRSL and VP4 epitope TQFTDFVSL were our most potential epitopes based on various physicochemical properties. They were validated using *in silico* molecular docking by AutoDock Vina. All three dimensional structures of HLA molecules were generated by homology modeling and the structures were verified. The verified structures were used in docking. The docked epitopes showed fair binding affinity ranging from -7.4 kcal/mol to 9.3 kcal/mol. Along with this, B-cell epitopes were predicted for both VP7 and VP4 proteins that can be used for analyzing antibody-epitope binding study. In this case, molecular dynamics study is the best option to calculate energy scoring adding valuable information to sequence based results.

Conclusion

In this study, the determined VP7 and VP4 epitopes showed higher MCH-I binding affinity, very good conservancy and acceptable population coverage. Moreover, they have significant number of MHC-I alleles interaction. They showed proper binding with HLA molecules in the docking study. These properties can make the epitopes significant vaccine agent which needs to be verified by wet laboratory procedures.

References

1. Walker CLF, Rudan I, Liu L, Nair H, Theodoratou E, Bhutta ZA, O'Brien KL, Campbell H and Black RE. 2013. Global burden of childhood pneumonia and diarrhoea. *The Lancet*. **381**(9875): 1405-1416.
2. Tate JE, Burton AH, Boschi-Pinto C and Parashar UD. 2016. Global, regional, and national estimates of rotavirus mortality in children <5 years of age, 2000–2013. *Clinical Infectious Diseases*. **62**: S96-S105.
3. World Health Organization. Rotavirus deaths by country 2000–2013. Available at: <http://www.who.int/immunization/monitoring_surveillance/rotavirus_deaths_by_country_2000-2013.xlsx?ua=1> [accessed on 17.05.2018].
4. Estes MK. 2001. Rotaviruses and their replication. In: Knipe DM, Howley PM eds. Fields virology. 4th ed. Vol 2. Philadelphia, PA: Lippincott Williams & Wilkins, pp 1747-1786.
5. Coulson BS and Kirkwood C. 1991. Relation of VP7 amino acid sequence to monoclonal antibody neutralization of rotavirus and rotavirus monotype. *Journal of virology.* **65**(11): 5968-5974.
6. Taniguchi K, Hoshino Y, Nishikawa K, Green KY, Maloy WL, Morita Y, Urasawa S, Kapikian AZ, Chanock RM and Gorziglia M. 1988. Cross-reactive and serotype-specific neutralization epitopes on VP7 of human rotavirus: nucleotide sequence analysis of antigenic mutants selected with monoclonal antibodies. *Journal of virology.* **62**(6):1870-1874.
7. Kirkwood CD, Bishop RF and Coulson BS. 1996. Human rotavirus VP4 contains strain-specific, serotype-specific and cross-reactive neutralization sites. *Archives of virology.* **141**(34): 587-600.
8. Martella V, Banyai K, Matthijnssens J, Buonavoglia C and Ciarlet M. 2010. Zoonotic aspects of rotaviruses. *Veterinary microbiology.* **140**(3-4): 246-255.
9. Dennehy PH. 2008. Rotavirus vaccines: an overview. *Clinical Microbiology Reviews.* **21**(1): 198-208.
10. Firbas C, Jilma B, Tauber E, Buerger V, Jelovcan S, Lingnau K, Buschle M, Frisch J and Klade CS. 2006. Immunogenicity and safety of a novel therapeutic hepatitis C virus (HCV) peptide vaccine: a randomized, placebo controlled trial for dose optimization in 128 healthy subjects. *Vaccine.* **24**(20): 4343-4353.
11. Stanekova Z and Vareckova E. 2010. Conserved epitopes of influenza A virus inducing protective immunity and their prospects for universal vaccine development. *Virology J.* **7**: 351.
12. He L, Cheng Y, Kong L, Azadnia P, Giang E, Kim J, Wood MR, Wilson IA, Law M and Zhu J. 2015. Approaching rational epitope vaccine design for hepatitis C virus with meta-server and multivalent scaffolding. *Scientific reports.* **5**: 12501.
13. Sominskaya I, Skrastina D, Dislers A, Vasiljev D, Mihailova M, Ose V, Dreilina D and Pumpens P. 2010. Construction and immunological evaluation of multivalent hepatitis B virus (HBV) core virus-like particles carrying HBV and HCV epitopes. *Clinical and Vaccine Immunology.* **17**(6): 1027-1033.
14. Kumar S, Stecher G and Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*. **33**(7): 1870-1874.
15. Doytchinova IA and Flower DR. 2007. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC bioinformatics*. **8**(1): 4.
16. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O and Nielsen M. 2007. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC bioinformatics*. **8**(1): 424.
17. Bui HH, Sidney J, Li W, Fusseder N and Sette A. 2007. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC bioinformatics*. **8**(1): 361.
18. Peters B and Sette A. 2005. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC bioinformatics*. **6**(1): 132.
19. Sette A, Vitiello A, Reherman B, Fowler P, Nayersina R, Kast WM, Melief CJ, Oseroff C, Yuan L, Ruppert J, Sidney J. 1994. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *The Journal of Immunology*. **153**(12): 5586-5592.
20. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ and Sette A. 2006. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC bioinformatics*. **7**(1): 153.
21. Larsen JEP, Lund O and Nielsen M. 2006. Improved method for predicting linear B-cell epitopes. *Immunome research*. **2**(1): 2.
22. Parker JMR, Guo D and Hodges RS. 1986. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry*. **25**(19): 5425-5432.
23. Kolaskar AS and Tongaonkar PC. 1990. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS letters*. **276**(1-2): 172-174.
24. Emini EA, Hughes JV, Perlow D and Boger J. 1985. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *Journal of virology*. **55**(3): 836-839.
25. Shen Y, Maupetit J, Derreumaux P and Tuffery P. 2014. Improved PEP-FOLD approach for peptide and miniprotein structure prediction. *Journal of chemical theory and computation*. **10**(10): 4745-4758.
26. DeLano WL. 2002. The PyMOL molecular graphics system. *http://pymol. org*.

27. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L and Schwede T. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research*. **42**(Web Server issue): W252-W258.

28. Yang J and Zhang Y. 2015. I-TASSER server: new development for protein structure and function predictions. *Nucleic acids research*. **43**(Web Server issue): W174-W181.

29. Benkert P, Künzli M and Schwede T. 2009. QMEAN server for protein model quality estimation. *Nucleic acids research*. **37**(Web Server issue): W510-W514.

30. Colovos C and Yeates TO. 1993. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein science*. **2**(9): 1511-1519.

31. Laskowski RA, MacArthur MW, Moss DS and Thornton JM. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography*. **26**(2): 283-291.

32. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS and Olson AJ. 2009. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*. **30**(16): 2785-2791.

33. Dash R, Das R, Junaid M, Akash MFC, Islam A and Hosen SZ. 2017. In silico-based vaccine design against Ebola virus glycoprotein. *Advances and Applications in Bioinformatics and Chemistry*. **10**: 11-28.

34. Dash R, Hosen SZ, Sultana T, Junaid M, Majumder M, Ishat IA and Uddin MMN. 2016. Computational Analysis and Binding Site Identification of Type III Secretion System ATPase from *Pseudomonas aeruginosa*. *Interdisciplinary Sciences: Computational Life Sciences*. **8**(4): 403-411.

35. Monterrubio-López GP and Ribas-Aparicio RM. 2015. Identification of novel potential vaccine candidates against tuberculosis based on reverse vaccinology. *BioMed Research International*. **2015**:16.

36. Graham BS, Ledgerwood JE and Nabel GJ. 2009. Vaccine development in the twenty first century: changing paradigms for elusive viruses. *Clinical Pharmacology & Therapeutics.* **86**(3): 234-236.

37. Schwartz BS, Harris JB, Khan AI, Larocque RC, Sack DA, Malek MA, Faruque AS, Qadri F, Calderwood SB, Luby SP and Ryan ET. 2006. Diarrheal epidemics in Dhaka, Bangladesh, during three consecutive floods: 1988, 1998, and 2004. *The American journal of Tropical Medicine and Hygiene*. **74**(6): 1067-1073.

38. Longmate J, York J, La Rosa C, Krishnan R, Zhang M, Senitzer D, Diamond DJ. 2001. Population coverage by HLA class-I restricted cytotoxic T-lymphocyte epitopes. *Immunogenetics*. **52**(3-4): 165-173.

39. Mostafa SM, Islam AB. 2016. An in silico approach predicted potential therapeutics that can confer protection from maximum pathogenic Hantaviruses. *Future Virology*. **11**(6): 411-428.