Journal of Immunological Methods xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Immunological Methods

journal homepage: www.elsevier.com/locate/jim

In silico identification of potential epitopes present in human adenovirus proteins for vaccine design and of putative drugs for treatment against viral infection

Rafeka Hossain, Tahirah Yasmin, Md. Ismail Hosen, A.H.M. Nurun Nabi*

Laboratory of Population Genetics, Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka 1000, Bangladesh

ARTICLE INFO

Keywords:
Peptide vaccines
Epitopes
Drug targets
B and T cell mediated immunity
Adenovirus
Immunoinformatics

ABSTRACT

In silico approach using computational biology to design best probable epitopes and/or drug target(s) has given an edge to foresee active components for the treatment of many infectious diseases. This study aims to investigate the best probable epitopes from fiber, hexon and penton base proteins as well as probable drug targets to prevent and to cure adenovirus infection, respectively. After retrieving protein sequences, analysis of selection pressure; prediction of continuous/discontinuous B cell epitopes along with their antigenicity, immunogenicity, allergenicity; T cell epitopes along with their population coverage and echelon of conservancy were performed. Out of three proteins, fiber protein underwent the highest degree of selection pressure. Five peptides from fiber C-5, hexon C-5 and D-8, penton base B-3 and C-5 proteins were considered as the best potential B cell epitopes. Further analyses revealed that peptides present in fiber C-5, hexon C-5, penton base B-3 and C-5 proteins fulfilled the criteria of having surface accessibility, hydrophilicity, flexibility, antigenicity and beta turn. Several regions of proteins were identified as discontinuous B cell epitopes. Interestingly, a peptide present in 692–699 region of hexon C-5 and six amino acids at positions 100, 102, 105, 108, 112 and 114 of penton base B-3 proteins were recognized both as continuous and discontinuous B cell epitopes. Of all the predicted T cell epitopes, three nonamers from hexon C-5, D-8 and penton base C-5 proteins may elicit strong immune response by activating both humoral and cellular immunity as these were found to overlap with those of B cell epitopic peptides. Considering non-allergen, conservancy and population coverage properties, "SGYDPYYTY" of hexon protein C-5 was further validated using in silico docking study for its interaction with the HLA allele. This study also demonstrated the possibility of compounds like 3-(azepan-1-ium-1-yl) propane-1-sulfonate and E-5842 as the potential inhibitors of penton base and hexon proteins that could act as more effective drugs against the virus compared to the current ones. Therefore, further *in vitro* and animal model experiments using these predicted epitopes and compounds may pave the way for newer and more effective treatment approaches against adenovirus infection.

1. Introduction

Human adenovirus (HAdV) is a linear, non-segmented, double stranded DNA which is around 30–38 kbp (Rowe et al., 1953) and discovered in 1953 in the adenoid tissue–derived cell cultures, (Rowe et al., 1953). The main target for HAdV is the respiratory tract, but it also causes a wide variety of diseases in human including ocular infection, hemorrhagic cystitis, gastroenteritis, rash illness, keratoconjunctivitis, and febrile respiratory illness (Ray, 2004) due to involvement of varying HAdV serotypes. For instances, HAdV-B and C are mainly responsible for respiratory diseases while HAdV-B and D cause conjunctivitis (Robinson et al., 2011; Jones et al., 2007). The infection is transmitted through inhalation, direct contact with small droplet aerosols or the fecal-oral route. Although most adenovirus infections are mild, rarely these infections become fatal and there are a number of sequelae associated with the viral infection. Acute necrotizing bronchitis and bronchiolitis may develop in children and in debilitated and immunocompromised patients (Edwards et al., 1985). Furthermore, adenoviral infections in lung transplant recipients may produce a rapidly progressive course leading to premature death (Ohori et al., 1995; Simsir et al., 1998).

Being the third most common cause of viral gastroenteritis in children, HAdV infection can cause high mortality rates in both healthy individuals and immune-compromised patient. However, unlike other

* Corresponding author at: Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka 1000, Bangladesh.
E-mail address: nabi@du.ac.bd (A.H.M.N. Nabi).

https://doi.org/10.1016/j.jim.2018.01.005
Received 7 November 2017; Received in revised form 18 January 2018; Accepted 18 January 2018
0022-1759/ © 2018 Elsevier B.V. All rights reserved.

viral pneumonitis (e.g. herpes simplex virus, cytomegalovirus), no specific treatment for adenovirus pneumonitis exists. Although there is a vaccine against this virus which contains live adenovirus type 4 and type 7, it is only approved for military personnel 17 through 50 years of age (CDC, 2014). Besides, there is always the chance of a secondary mutation (a chance in the genetic material of the organism) which could lead to a transformation into a more contagious and virulent form of the microbe and furthermore they cannot be safely given to people who are immunocompromised (Hierholzer, 1992; Mandell et al., 2004).

Due to the several limitations of the current treatment such as incomplete efficiency of the drugs, limited availability of the vaccines only for U.S. Military and Coast Guards, there is a desperate need to develop additional approaches and drugs for the treatment of severe adenovirus infections (Lenaerts et al., 2008). Since prevention is always better than curative approaches, this study aims to design specific peptide vaccine against human adenovirus by targeting proteins like the fiber, hexon and penton base protein which are the major attachment proteins of the virus. These proteins perform the majority of functions facilitating the early stages of adenovirus infection that include initial cell-surface binding followed by receptor-mediated endocytosis, endosomal penetration and cytosolic entry, and intracellular trafficking toward the nucleus (Medina-Kauwe, 2013). Consequently, they are particularly important for vaccine development. Traditionally, vaccines are developed from live attenuated or inactivated viruses that elicit a strong humoral and cellular immune response leading to long-lasting memory that protects from reinfection. One of the major problems these vaccines brought, is crucial safety concern, because those pathogens being used for immunization may become activated and cause infection (Santana-Jorge et al., 2016). Moreover due to genetic variation of pathogen strains around the world, vaccines are likely to lose their efficacy in different regions or for a specific population. But novel vaccine approaches like epitope based peptide vaccines have the potential to overcome these barriers owing to their comparatively easy production and construction, chemical stability, and absence of infectious potential. At this time, the development of bioinformatics tools along with the knowledge on the host immune response and the availability of the complete genome sequences of the pathogens is playing a critical role on analyzing multiple genomes to select the protective epitopes in silico and it is conceived that cocktails of defined epitopes may provide a rationale design capable to elicit convenient humoral or cellular immune responses. And this concept is already being successfully pursued in a large number of researches especially to the development of vaccines targeting conserved epitopes in variable or rapidly mutating pathogens (Soria-Guerra et al., 2015).

Vaccines are mostly based on B cell immunity. However now-a-days, vaccine based on T cell epitope has been encouraged as the host can generate a strong immune response by CD8+ T cell against the infected cell. With time, due to antigenic drift, any foreign particle can escape the antibody mediated response; however, the T cell immune response often provides long-lasting immunity. Therefore, a multi epitope vaccine would be a good prospect that can trigger both cell mediated and humoral immune response. Cytotoxic T cell-mediated response is elicited by a pathway comprising intracellular antigen processing with linear epitopes as predominant targets. In this regard, the T cell epitopes selected for a vaccine must have binding affinity with more than one major histocompatibility complex (MHC) allele and must cover a major population and these were analyzed in this work using different bioinformatics tool. On the other hand, considering the fact that elicitation of humoral responses relies on the recognition of both linear epitopes and conformational epitopes, we focused on predicting both linear and discontinuous B cell epitopes. The latter is challenging since they must retain their native conformation to be functional (Soria-Guerra et al., 2015). Therefore, knowledge on the whole protein structure is crucial for predicting conformational B cell epitopes and bioinformatics databases of protein structures were utilized for retrieving the structures of the target proteins. At the same time, designing potential inhibitory molecules against these target proteins was another objective of the present study. As we know, vaccinations are by nature a pre-emptive treatment – they need to be given before an infection starts. Therefore, if a person gets infected due to unavailability of a vaccine, therapeutic treatment is essential for cure. The drug discovery process has been profoundly changed recently by the adoption of computational methods helping the design of new drug candidates more rapidly and at lower costs (Zoete et al., 2009).

In this study, protective and curative measures were suggested against adenovirus mediated infection by predicting probable epitopes and drugs via employing several in silico approaches to make significant contributions to the development of potential vaccine and drugs which would ultimately offer great advantages over the conventional time consuming and expensive process of therapeutic discovery. Major attachment proteins were analyzed and cluster of B- and T-cell epitopes have been identified on the basis of their physicochemical properties, interaction with HLA alleles, non-allergen property which may facilitate augmentation of both humoral and cell-mediated immunity. Further, potential inhibitors against penton base and hexon attachment proteins have also been proposed.

2. Materials and methods

The complete study plan is outlined in Fig. 1.

2.1. Retrieval of protein sequences

Sequences of fiber protein, hexon protein and penton base proteins were extracted in FASTA format from UniProtKB. More than 95% of the protein sequences provided by UniProtKB are derived from the coding sequences (CDS) which have been submitted to the public nucleic acid databases.

2.2. Phylogenetic analysis

The sequences were aligned using Clustal Omega (ClustalW) using default parameters. The Neighbor-Joining method was used for determining the evolutionary history of the taxa and the bootstrap (Felsenstein, 1985) consensus tree was taken to represent this history. The bootstrap value was set at 1000. The evolutionary distances were computed using the Maximum Composite Likelihood method. All the evolutionary analyses were conducted in MEGA6 software.

2.3. Selection pressure analysis

Selection pressure of each codon was calculated using values of dN/dS ratio. The codons with positive selection pressure ($dN/dS > 1$) were filtered and extracted in an excel file. The codons of threonine, tyrosine, serine, lysine, histidine, arginine, aspartate, glutamate and proline were identified and the numbers of these particular amino acids undergoing positive selection were determined. Afterwards the percentage of total codons which had shown positive selection was also measured and it was used to understand the selection pressure in each protein.

2.4. Epitope prediction

2.4.1. Continuous B cell epitope prediction

B cell epitopes are the sites of molecules that are recognized by antibodies of the immune system. In this study, continuous B cell epitopes were predicted by ABCpred server (Saha and Raghava, 2006) and BepiPred server (Larsen et al., 2006). To make the prediction more legitimate, overlapping epitopes were chosen as the most probable epitopes. Chou and Fasman beta-turn prediction (Chou and Fasman, 1978), Emini Surface accessibility prediction (Emini et al., 1985), Karplus and Schulz flexibility prediction (Karplus and Schulz, 1985), Kolaskar and Tongaonkar antigenicity prediction (Kolaskar and

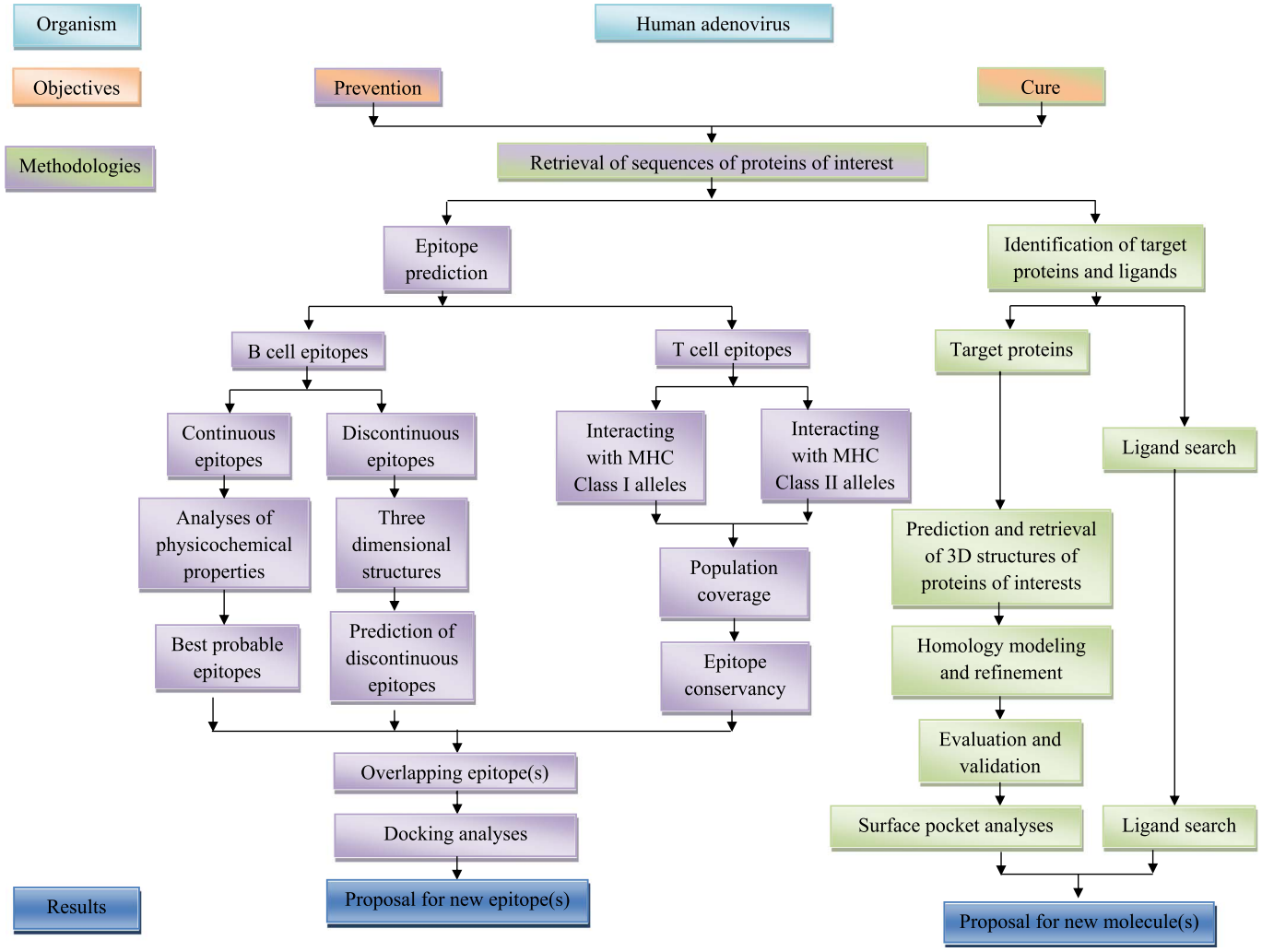


Fig. 1. Overall working flow chart.

Tongaonkar, 1990) and Parker hydrophilicity prediction (Parker et al., 1986) from Antigen Sequence Properties of IEDB server were applied for further analysis of these epitopes. In all the tools default settings were used with threshold value of 1.0.

2.4.2. *Discontinuous B cell epitope prediction*

Discontinuous B cell epitopes were predicted from the analyses of fiber, hexon and penton base proteins' 3D structures (PDB ID: 1KNB, 1P30 and 4AR2 respectively). For this prediction DiscoTope server was used that predicts discontinuous B cell epitopes from the three dimensional structures of respective proteins.

2.4.3. *T cell epitope prediction*

Tools available in Immune Epitope Database (IEDB) (tools.immuneepitope.org) were used to predict T cell epitopes. IEDB provides a dataset of experimentally characterized T cell epitopes, binding documentation of Major Histocompatibility Complex (MHC) and facts on MHC ligand elution experiments (Vita et al., 2010).

2.4.3.1. Prediction of MHC I binding, TAP transport efficiency and proteasomal cleavage efficiency. The ability of a peptide to become a potent T cell epitope was predicted using the overall scores obtained by the combined algorithm of MHC I binding, TAP transport efficiency and proteasomal cleavage efficiency in IEDB. The Stabilized Matrix Method (SMM) was taken into consideration for calculating IC_{50} values and the alleles having binding affinity (IC_{50} value) < 100 nM were chosen for further consideration. Peptide length was set as nonamer prior to the prediction.

2.4.3.2. Prediction of MHC class I and class II binding. The nonamers obtained from the combined algorithms were given as input in MHC Class I binding prediction tool available in the Immune Epitope Database (IEDB) server (http://tools.immuneepitope.org/mhci/).

The whole protein sequences were submitted in the IEDB MHC class II binding prediction tool (http://tools.immuneepitope.org/mhcii/) since MHC class II can accommodate much longer peptide. Similar to class I prediction method, SMM and IC_{50} value of ≤100 nM were set as parameters to predict best probable MHC Class II 15 mer epitopic peptides.

2.5. *Antigenicity prediction*

Antigenicity of the most common epitopes (100% overlapping) was checked from VaxiJen Server (Doytchinova and Flower, 2007). VaxiJen is the first server for alignment-independent prediction of protective antigens. It was developed to allow antigen classification solely based on the physicochemical properties of proteins without recourse to sequence alignment.

2.6. Analyses of population coverage, conservancy and allergenicity of the predicted T cell epitopes

MHC molecules are extremely polymorphic and over a thousand different human MHC (HLA) alleles are known. Selecting multiple peptides with different HLA binding specificities will allow increased coverage of the targeted population. Therefore, the population coverage analysis tool under default parameters (Bui et al., 2006) from IEDB Analysis resource was employed to determine the coverage of the epitopic peptides.

The epitope conservancy analysis tool (Bui et al., 2007) was used to observe how conserved the epitopes were among the different protein sequences used. All parameters were set at default. In addition, Allerdictor (Dang and Lawrence, 2014) and AllerHunter (Muh et al., 2009) were used for determining the allergenicity of the predicted epitopes.

2.7. Docking of the best selected peptides in the binding groove of HLA alleles

PEPFOLD server was used to predict the three dimensional structures of the selected peptides- "FRLSDSLAL" from hexon protein C5 and "NYFEGNLDM" from penton base protein B3 (Maupetit et al., 2010; Thevenet et al., 2012). The best models provided by the server were selected for the docking study.

On the other hand, the predicted three dimensional structure of the HLA-C*07:02 allele, as it was found to interact with both peptides, was obtained from a previous study (Sakib et al., 2014). The AutoDOCK tool from the MGL software package (version 1.5.6) was employed for the docking purpose (Morris et al., 1998; Morris et al., 2009).

These structures were then further prepared for running docking in AutoDock tool by adding polar hydrogen and water molecules. For the docking study both the HLA allele and ligand files were converted into PDBQT format. The grid/space box center was set at −10.839, 5.283 and 30.19 Å in the x-, y-, and z-axes, respectively to allow the epitopes to bind to the binding groove of the allele. The size was set at 28, 40 and 30 Å in the x, y, and z dimensions, respectively. All the analyses were done at 1.00 Å spacing. The thoroughness of global search algorithm i.e. the exhaustiveness parameter was kept at 8.00, while the number of outputs was set at 10. The docking was conducted based on the set parameters. The PDBQT files were converted in PDB format using OpenBabel (version 2.3.1) and visualized in PyMOL molecular Graphics system.

Since the prediction of 3D structure of HLA-C*07:02 allele was performed using template alpha chain of H-2 kb MHC class I molecule with PDB ID: 1KJ3 that covered 76% of query sequence modeled with 100% confidence, both the HLA and peptide were used as control in the docking study. Three dimensional structure of MHC class I H-2Kb molecule complexed with octapeptide PKB1 ("KVITFIDL") was retrieved from Protein Data Bank Database (ID: IKJ3) and visualized using PyMOL Graphics. The octapeptide was excluded before applying the structure of H-2Kb for comparing the validated data obtained for predicted structure of HLA-C*07:02.

Also, to assess HLA-C*07:02-epitope docking results, octapeptide PKB1 ('KVITFIDL') was used as the control. This peptide was docked with HLAs, HLA-C*07:02, and H-2Kb. The test epitope(s) and the control peptide were docked by setting similar parameters for each trial and successful binding of this peptide to these HLAs was demonstrated. Finally, H-2Kb – 'KVITFIDL' docking result was used as control to compare with the test docking results of HLA-C*07:02 complexed with selected epitopes.

2.8. Prediction of putative drug targets

2.8.1. Three-dimensional structure prediction of target proteins

Homology modeling uses a comparative protein structure prediction approach to predict three dimensional models of proteins. This method is based on the assumptions that proteins that possess similar sequences share similar three-dimensional structures, and only a limited number of protein folds exist in nature (Daga et al., 2010). In brief, this kind of modeling consists of i) suitable homologous template selection, optimal target-template sequence alignment, crude model building stage and ii) model evaluation, refinement and validation stage.

2.8.1.1. Homology modeling and refinement. Predictions of the three-dimensional structures of the selected proteins with higher confidence were performed using Phyre2 (Protein Homology/Analogy Recognition Engine) (Kelley et al., 2015) and RaptorX (Källberg et al., 2012) online tools. To refine the predicted structures, ModRefiner was used which can construct and refine protein structures from Cα traces based on a two-step, atomic-level energy minimization (Dong and Zhang, 2011).

2.8.1.2. Evaluation and validation of the predicted structures. To measure the accuracy and stereo-chemical quality of the predicted 3D structures the PROCHECK suite of programs (Laskowski et al., 1996) was applied. Ramachandran plot and QMEAN score were generated for each model. In addition, 3D profiling of the predicted proteins was done by Verify3D (Eisenberg et al., 1997) which is able to determine the compatibility of an atomic model with its own amino acid sequence.

2.8.2. Surface pocket analysis

Protein performs its function through interaction with other molecules and it is the three-dimensional structure of protein which provides the necessary shape and physicochemical texture to facilitate these interactions. CASTp server aims to provide a detailed quantitative characterization of interior voids and surface pockets of proteins, which are frequently associated with binding events (Laskowski et al., 1996). Therefore, in this present study this server was applied to find out the surface accessible pockets as well as the interior inaccessible cavities of the proteins and measure the area and volume of each pocket. The 3D structures were given as input.

2.8.3. Ligand search

The chemical structures of potential ligands and their structural analogs were retrieved from PubChem (https://pubchem.ncbi.nlm.nih.gov) and ZINC (http://zinc.docking.org/) database in .sdf format. Information on their physical properties was also collected from the same database.

2.8.4. Docking analyses between proteins of interests and ligands

The molecular docking study between ligand and protein was carried out using AutoDock tool from the MGL software. Both the protein and ligand files were at first converted into .pdbqt format to use them for the docking study by OpenBabel software. The protein structures were prepared by deleting water molecules and adding polar hydrogen atoms.

One compound called *N*-dodecyl-*N*,*N*-dimethyl-3-ammonio-1-propanesulfonate (Drugbank ID: DB02643) has been reported as an inhibitor of adenovirus which was found to be complexed with penton protein of serotype 2 (PDB ID: 1X9P) (Zubieta et al., 2005).To find out a more suitable inhibitor, structural analogues of the reported inhibitor were retrieved as well.

Structurally similar penton base protein of HAdV B3 (PDB ID: 4AR2) was selected for molecular docking studies to determine which of the inhibitor molecules interacted with the protein with minimum binding energy. Besides penton protein of HAdV C5 (PDB id: 3IZO) was also analyzed since C5 serotype is among the prevalent serotypes and 3IZO structure comprises the pentameric structure of penton base which is how they form complex in the virus. Docking between the experimental inhibitor and penton protein of Adenovirus 2 was used as a control.

On the other hand, citric acid (Drugbank ID: DB04272) has been proposed as a drug candidate against hexon protein. Citric acid was found to be complexed with hexon protein of HAdV C2 (PDB ID: 1P2Z)

(Rux et al., 2003). It was shown to interact with four different pockets of the protein. To find out the best inhibitor compounds among citric acid and its analogues, hexon protein C5 (PDB ID: 1P30) possessing similar structure as 1P2Z was chosen for performing docking studies. Again docking between the experimental ligand and hexon C2 was taken as control.

Against fiber protein, sialic acid (Drugbank ID: DB03721) has been reported as an effective inhibitor (Spjut et al., 2011). It was found to be complexed with fiber protein of HAdV D37 (PDB ID: 3QND). The experimental inhibitor and its derivatives were retrieved from similar ligand-protein complexed structures available in Protein Data Bank. In this study fiber protein of C5 was targeted due to its prevalence. So for docking study, its largest pocket as well as the pocket that matched with the control pocket of serotype D37 where sialic acid was found to bind was chosen to analyze. Meanwhile docking between the experimental ligand and fiber C37 was considered as control.

The grid box parameters were set in such a way that these ligands will bind to the binding site of the experimental inhibitor in their respective protein structure. To do so, the interaction between the experimental inhibitor and the protein was analyzed using Ligand Explorer tool. This region overlapped with the binding site residues of the protein, which were extracted using CASTp server (http://sts.bioe.uic.edu/castp/) and thus was defined as the center of grid box.

2.8.5. Pharmacophore analysis

Different properties of the selected ligands like mutagenicity, tumorigenicity, irritancy etc. were estimated by Data Warrior software (Sander et al., 2015).

3. Results

3.1. Retrieval of protein sequences and their antigenicity prediction

Fiber protein from 15 serotypes; hexon protein from 25 serotypes and penton base protein from 31 serotypes of adenovirus were retrieved for the study. Complete sequences of these proteins were collected from UniProt in FASTA format followed by determining antigenicity using VaxiJen by setting default parameters. All of them had shown antigenicity as they had crossed the threshold level (≥ 0.4).

3.2. Evolutionary analysis of selected proteins

The evolutionary history of fiber, hexon and penton base proteins were inferred using the Neighbor-Joining method (Saitou and Nei, 1987; Tamura et al., 2004) in MEGA6 (Tamura et al., 2013) software (Supplementary Figs. S1–S3). It was found that serotypes 1, 2, 5 and 6 which were among the major causative pathogens for respiratory illnesses, showed close association with respect to all three proteins.

3.3. Selection pressure analyses of fiber, hexon and penton base proteins

Penton protein exhibited the least number of positively selected codons with respect to positively and negatively charged amino acids, hydroxyl group containing amino acids and proline. These amino acids were selected for analyzing selection pressure since all of them have important roles in determining the structures of protein. Due to varying length of fiber protein in different serotypes of adenovirus, a complete analysis of selection pressure for fiber protein could not be performed. Therefore, conclusive analysis could not be performed from dN/dS ratio information of fiber protein. For instance, out of only 68 codons for which dN/dS ratio was calculated, 40 codons showed positive selection. On the other hand, in case of hexon protein out of 696 codons 58 codons, and in case of penton base protein out of 358 codons 28, showed positive selection. Fig. 2 represents selection pressure on different amino acids of these three different proteins of interests along with their percentage of positively selected codons.

3.4. Identification of B cell epitopes

Protein sequences from serotype B-3, serotype C-5 and serotype D-8 were chosen for predicting B cell epitope as they were considered to be among the most prevalent serotypes in respiratory tract disease and conjunctivitis (Ray, 2004).

From the analysis, it was found that the fiber proteins B-3, C-5 and D-8 had generated 4, 11 and 11 probable B cell epitopes with varying lengths ranging from 6 to 16 amino acid residues. Similarly, hexon proteins B-3, C-5 and D-8 had generated 14, 17 and 20 probable B cell epitopes and finally penton base protein had generated 9, 14 and 9 probable B cell epitopes. So, finally 109 epitopes were considered for further analysis to assess their efficacy as B cell epitopes.

3.5. Identification of the most probable continuous B cell epitopes

Most probable continuous B cell epitopes based on their physicochemical properties with respect to their surface accessibility, hydrophilicity, flexibility and beta turn characteristics, antigenic and immunogenic properties have been presented in Table 1 along with their extent of conservancy and allergenicity. Eighteen epitopes have been found from which five were considered as the best ones. The surface accessibilities of these B cell epitopes have clearly been shown in their respective three dimensional structures (Supplementary Fig. S4).

3.6. Identification of the most probable discontinuous B cell epitopes

Three dimensional (3D) structures of fiber protein, hexon protein and penton base protein were extracted from PDB server and given as input in Discotope tool. The resulting predicted non-linear B cell epitopes are marked in the following structures (Fig. 3).

From fiber protein C-5, 9 residues were identified as potential discontinuous B cell epitopes while 368 residues were identified from hexon protein C-5 and 134 residues were identified from penton base protein B-3. Amino acid sequence "^{692}SGYDPYYT^{699}" was found in both continuous and discontinuous B cell epitopes of hexon protein C-5 and in case of penton base protein B-3, both linear and nonlinear B cell epitopes were found in ^{100}D, ^{102}T, ^{105}E, ^{108}T, ^{112}N and ^{114}D amino acid residues.

3.7. T cell epitope prediction

3.7.1. MHC class I binding prediction

MHC- I binding prediction and MHC-I processing tool retrieved 254 nonamers from fiber C-5, 530 nonamers from hexon C-5, 570 from hexon D8, 292 from penton B-3 and 278 nonamers from penton C-5. On the other hand, MHC-II binding prediction tool generated 442 epitopes (15 mer peptides) from fiber C-5, 620 epitopes from hexon C-5, 679 from hexon D-8, 202 from penton B-3 and 270 epitopes from penton C-5.

3.7.2. MHC class II binding prediction

A good epitope should also interact with as many as MHC alleles. Thus, among the total peptides, only those peptides which interacted with highest number of HLA alleles (at least 6 MHC class I alleles that overlapped among class II alleles with IC_{50} value < 100 nM) were selected from each protein. From these studies, five, four, three, three and four epitopes were found respectively from fiber protein C-5, hexon protein C-5, hexon protein D-8, penton base proteins B-3 and C-5. These data have been presented in Supplementary Tables S1–S5.

From those, "IRFDENGVL" of fiber protein C-5, "MSMGALTDL" of hexon proteins C-5 and D-8, "YQNDHSNFL" and "NRFPENQIL" from penton base protein B3 and'YQNDHSNFL' from penton base protein C-5 were chosen for further analysis as they had shown the best result in population coverage, conservancy and allelic variation (as shown in Table 2).

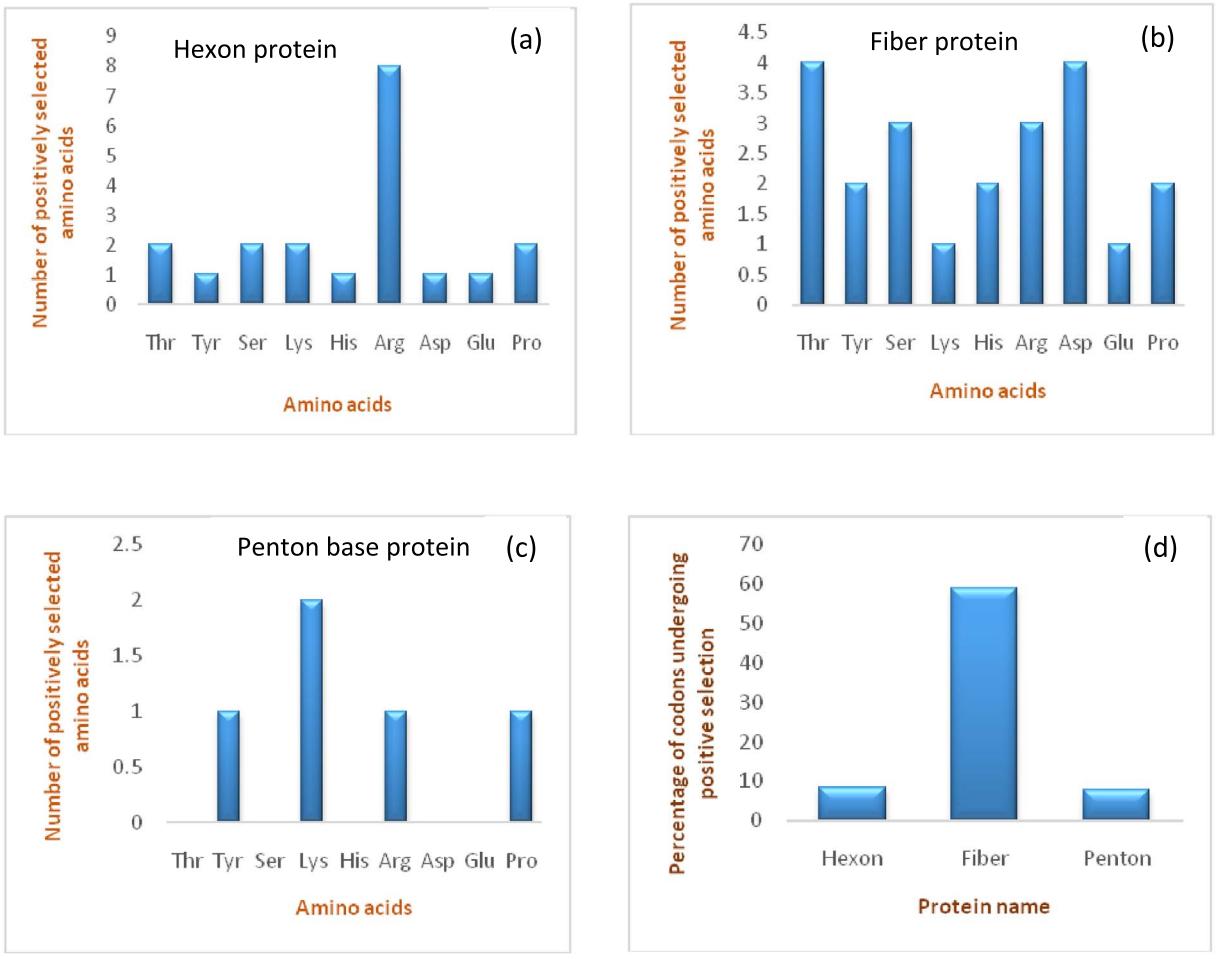


Fig. 2. Positive selection of different amino acids in hexon, fiber and penton base proteins of human adenovirus. Figures (a), (b) and (c) represent number of amino acids that had undergone positive selection in hexon, fiber and penton base proteins, respectively. The figures reflected that attachment proteins are under selection pressure by changing positively/negatively charged, cyclic, hydroxyl group containing amino acids at different numbers that may affect the functions of these respective proteins. Out of these proteins, fiber protein of HAdV is under highest selection pressure as shown in figure (d) considering the percentages of positively selected amino acids.

Table 1
Prediction of antigenicity, surface accessibility, flexibility, hydrophilicity, beta-turn of individual residues of each best probable continuous B cell epitopes along with their epitope conservancy and allergenicity.

Names of proteins, 'predicted peptides', (VaxiJen score)	Residue	Accessibility	Flexibility	Kolaskar antigenicity prediction	Hydrophilicity prediction	Beta-turn	Epitope conservancy analysis	Allergenicity
Fiber protein C5	T	1.333	1.069	0.948	4.871	1.191	100% conserved in	0.005
'PYDTETGPPTVPFL'	E	1.316	1.081	0.948	4.871	1.191	serotypes of group C	Non-allergen
(0.7435)	T	1.218	1.087	0.934	5.443	1.246		
	G	1.218	1.086	0.94	4.757	1.174		
Hexon protein C5	Y	1.223	1.045	1.002	3.743	1.401	100% conserved in all	0.004
"KETPS LGSGYDPYYTY"	D	1.43	1.034	1.043	2.657	1.341	available serotypes	Non-allergen
(0.6065)	P	2.086	1.01	1.028	2.471	1.274		
	Y	2.086	_	1.069	1.386	1.214		
Hexon protein D8	G	1.638	1.107	1.036	1.843	1.103	100% conserved in	0.005
'LRQGQPYPANFP'	Q	1.293	1.097	1.009	3.457	1.236	serotypes of group D	Non-allergen
0.7948								
Penton base protein B3	E	1.229	1.059	0.96	4.871	1.033	100% conserved in	0.005
'DFTPTEASTQTINFD'	A	1.377	1.058	0.975	4.986	1.036	serotypes of group B	Non-allergen
(1.0007)								
Penton base protein C5	E	1.471	1.055	0.956	2.771	1.117	100% conserved in	0.006
'MYEEGPPPSYESVVSA'	G	1.451	1.074	0.99	3.671	1.249	serotypes of group C	Non-allergen
(0.5860)	P	1.123	1.075	0.969	4.871	1.29		
	P	1.016	1.062	1.013	3.486	1.347		
	P	1.778	1.046	1.013	3.486	1.347		
	S	1.541	1.033	1.033	3.6	1.329		

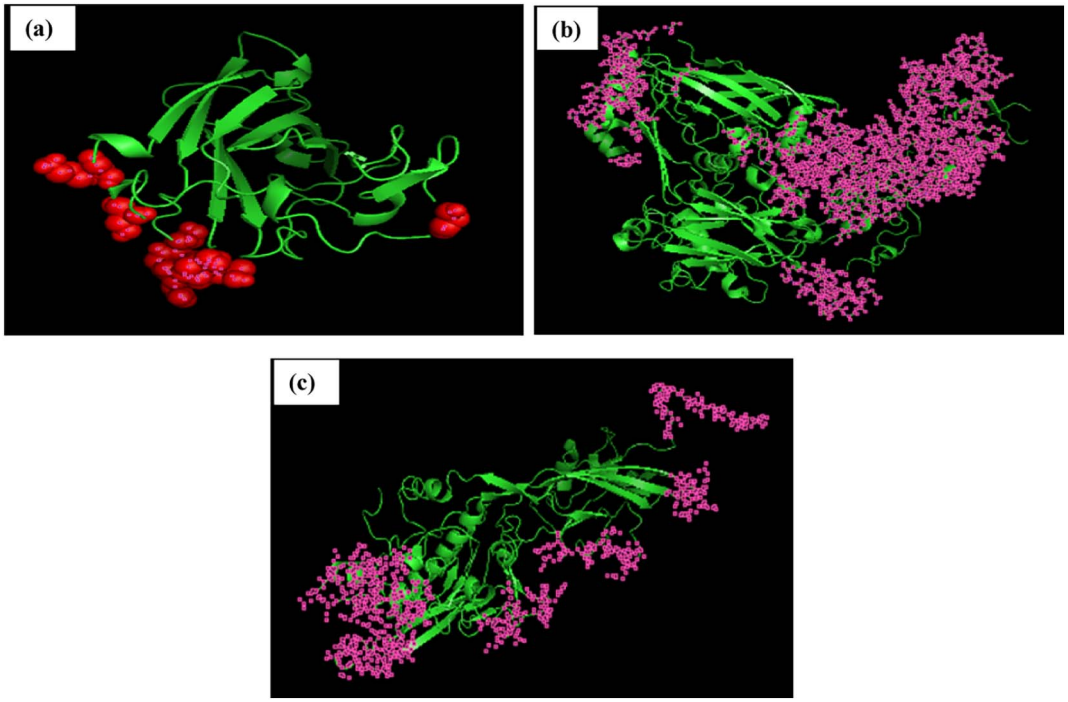


Fig. 3. Discontinuous B cell epitopes of fiber, hexon and penton base proteins. Discontinuous B cell epitopes predicted from 3D structures of (a) fiber protein C-5 (in red) (b) hexon protein C-5 (in pink) and (c) penton base protein B-3 (in pink). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.8. Molecular docking for HLA-epitope interaction

Using AutoDock Vina, binding models of predicted epitopes to their respective HLA molecules were generated. Binding energies indicate affinities of the epitopes toward the groove of the alleles i.e. the lower the binding energies, the higher the affinities. In case of class I HLA-C*07:02, epitope 'NRFPENQIL' bound to the binding groove with the binding energy −8.0 Kcal/mol [Fig. 4(a)]. Also, this epitope bound to the binding groove of H-2Kb with the binding energy of −6.7 Kcal/mol [Fig. 4(b)]. Control peptide 'KVITFIDL' bound to the groove of HLA-C*07:02 with the binding energy of −6.6 Kcal/mol [Fig. 4(c)].

Also, the binding affinities for the 'SGYDPYYTY' epitope to HLA-C*07:02 and H-2Kb were estimated to be −8.3 Kcal/mol [Fig. 4(d)] and −8.2 Kcal/mol [Fig. 4(e)], respectively. Control peptide 'KVITFIDL' bound to the groove of H-2Kb with the binding energy −7.3 Kcal/mol (Sakib et al., 2014).

3.9. Prediction of overlapping B and T cell epitopes

To find an epitope which would be able to elicit both B cell and T cell immune response, all the predicted B cell and T cell epitopes were compared and three epitopes were found to cover such criteria. The overlapping epitopes are given in Table 3.

3.10. Potential drug design against fiber, hexon and penton base proteins

In order to design anti-viral drugs along with vaccine against adenovirus, this study includes a protein-targeting drug discovery project. The aim is to identify a small-molecule compound that would bind to the target protein and modulate its function. When the protein's 3D structure is unknown, de novo structure determination of the protein target becomes the first step of the discovery project. The three-dimensional structure of the protein is then used to assess the energy of the interactions between the ligands and the binding site of the protein, which ultimately allows sorting of the inhibitory molecules identified during the in silico screening according to their binding affinity.

3.10.1. Three dimensional structure prediction of target proteins

The three dimensional structures of hexon protein C-5 (PDB ID: 1p30), penton base proteins B-3 (PDB ID: 4AR2) and C-5 (PDB ID: 3IZO, chains A, B, C, D and E) were retrieved from protein data bank.

Due to the unavailability of complete structure of fiber protein C-5 and hexon protein D-8, these two were predicted using two different 3D structure prediction tools: Phyre2 (Kelley et al., 2015) and RaptorX (Källberg et al., 2012). In case of fiber protein, phyre2 server could model 47% of residues at > 90% confidence whereas 57% of the sequence was predicted disordered. The overall confidence in the final model was considered too low (< 70%). On the other hand, RaptorX predicted structure showed the input had three domains (Fig. 5) which were modeled using the templates: 1knbA at p-value 1.93e−12; 1qiuA and 1v1iA at p-value: 1.39e−04 and 4rm6:A at p-value: 2.53e−09, respectively. A model with both good p-value and uGDT (GDT) is very likely to be of high quality. For alpha proteins, p-value $< 10^{-3}$ is a good indicator while for beta proteins, p-value $< 10^{-4}$ is a good indicator. For a protein with > 100 residues, uGDT > 50 is a good indicator and in this case overall uGDTscore was 341.

This model was further refined by ModRefiner tool (Xu and Zhang, 2011) (RMSD = 1.001, TM-score = 0.9914 to initial model) and this is presented in Fig. 6(a). Finally to measure its accuracy PROCHECK (Laskowski et al., 1993) performed an overall analysis of the model and delivered the Ramachandran plot shown in Fig. 6(b). The resultant structure had 86% of its residues laid out in the most favorable core region and 12.4% in additional allowed region with only 0.4% of the residues in generously allowed regions.QMEAN6 (Benkert et al., 2011) score was found 0.401 for the structure while *Z*-Score was −4.107. When analyzed by verify3D server (Bowie et al., 1991), the structure crossed the threshold since 81.07% of the residues had an averaged 3D-1D score ≥ 0.2.

In case of hexon protein D-8, RaptorX predicted model showed poor validation score in Ramachandran plot and only 78.9% residues were found in the most favored regions. On the other hand, the refined structure originally predicted by Phyre2 server showed almost 90% residues in the most favorable regions in Ramachandran plot which can be considered as a good quality model. Therefore, Phyre2 predicted

Table 2
Most probable T cell epitopes from different serotypes of human adenovirus.

Protein name	Predicted peptides	Interacting MHC-I alleles	Overlapping 15mer peptides	Interacting MHC-II alleles	Population coverage	Epitope conservancy	Allergenicity
Fiber protein C5	IRFDENGVL	HLA-C*03:03 HLA-C*06:02 HLA-C*07:01 HLA-C*07:02 HLA-C*12:03 HLA-C*14:02	SAHLIIRFDENGVLL AHLIIRFDENGVLLN HLIIRFDENGVLLNN LIIRFDENGVLLNNS QSAHLIIRFDENGVL	HLA-DRB1*13:02, HLA-DRB3*01:01	64.82%	25% conserved in serotypes of group C	0.006 Non-allergen
Hexon protein C5	MSMGALTDL	HLA-B*15:02 HLA-C*06:02 HLA-C*07:01 HLA-C*12:03 HLA-C*14:02 HLA-C*15:02	PFSSNFMSMGALTDL FSSNFMSMGALTDLG SSNFMSMGALTDLGQ SNFMSMGALTDLGQN NFMSMGALTDLGQNL FMSMGALTDLGQNLL	HLA-DRB1*04:04, HLA-DRB1*01:01, HLA-DRB1*04:01, HLA-DRB5*01:01, HLA-DRB1*04:05	54.28%	100% conserved in all available serotypes	0.005 Non-allergen
Hexon protein D8	MSMGALTDL	HLA-B*15:02 HLA-C*03:03 HLA-C*06:02 HLA-C*07:01 HLA-C*12:03 HLA-C*14:02 HLA-C*15:02	PFSSNFMSMGALTDL FSSNFMSMGALTDLG SSNFMSMGALTDLGQ SNFMSMGALTDLGQN NFMSMGALTDLGQNM FMSMGALTDLGQNML	HLA-DRB1*04:04, HLA-DRB1*01:01, HLA-DRB1*04:01, HLA-DRB1*04:05	54.28%	100% conserved in all available serotypes	0.005 Non-allergen
Penton base protein B3	YQNDHSNFL	HLA-A*02:06 HLA-B*15:02 HLA-B*39:01 HLA-C*03:03 HLA-C*05:01 HLA-C*06:02 HLA-C*12:03	SLNYQNDHSNFLTTV DIASLNYQNDHSNFL IASLNYQNDHSNFLT ASLNYQNDHSNFLTT LNYQNDHSNFLTTVV	HLA-DRB1*07:01, HLA-DRB3*01:01	42.75%	100% conserved in all available serotypes	0.005 Non-allergen
Penton base protein B3	NRFPENQIL	HLA-B*39:01 HLA-C*03:03 HLA-C*06:02 HLA-C*07:01 HLA-C*07:02 HLA-C*12:03	LTHVFNRFPENQILI THVFNRFPENQILIR HVFNRFPENQILIRP SLTHVFNRFPENQIL	HLA-DRB1*09:01	64.00%	100% conserved in serotypes of group B	0.006 Non-allergen
Penton base protein C5	YQNDHSNFL	HLA-A*02:06 HLA-B*15:02 HLA-B*39:01 HLA-C*03:03 HLA-C*05:01 HLA-C*06:02 HLA-C*12:03	SLNYQNDHSNFLTTV DVASLNYQNDHSNFL VASLNYQNDHSNFLT ASLNYQNDHSNFLTT LNYQNDHSNFLTTVI	HLA-DRB1*07:01, HLA-DRB3*01:01	42.75%	100% conserved in all available serotypes	0.006 Non-allergen

structure of hexon protein D-8 [Fig. 6(c)] was selected for further analysis.

In Ramachandran plot [Fig. 6(d)] the resultant structure showed 88.4% residues in the most favorable core region while 10.8% in additional allowed region and only 0.5% of the residues in generously allowed regions. QMEAN6 score and Z-Score were 0.505 and −2.857 respectively. The structure also passed verification check by Verify-3D tool since 80.79% of the residues had an averaged 3D-1D score ≥ 0.2.

3.10.2. Active site analysis

The three dimensional structures of the proteins were analyzed for pockets to find out the potential ligand binding sites by CASTp server (Dundas et al., 2006) and MetaPocket2.0 server (Zhang et al., 2011). A total of 147 pockets were identified as the most probable ligand binding sites in case of hexon protein C-5 while only 78 pockets of fiber protein C-5 were recognized. Similarly, 184 pockets from hexon D-8, 113 from penton B-3 and 580 pockets from penton C-5 were identified. The pockets with the largest area and volume from each protein were considered as the most potential ligand binding sites. Positions of these pocket residues (probable binding sites for ligands) along with their area and volume have been presented in Table 4 and shown in Fig. 7.

3.10.3. Ligand selection

The structures of *N*-dodecyl-*N*,*N*-dimethyl-3-ammonio-1-propanesulfonate, citric acid, sialic acid and their analogues were retrieved from PubChem (https://pubchem.ncbi.nlm.nih.gov) and ZINC database (http://zinc.docking.org/). A total of 18 molecules were recognized as the probable potential ligands for the binding sites of the retrieved as well as predicted 3D structures of the proteins of HAdV. The search for probable ligands was established based on the available ligand protein structures in PDB. All the molecules fulfilled Lipinski rule of 5 (Lipinski et al., 1997) which states that a compound is more likely to be membrane permeable and easily absorbed by the body if it matches the following criteria: its molecular weight is < 500; the compound's lipophilicity, expressed as a quantity known as logP (the logarithm of the partition coefficient between water and 1-octanol), is < 5; the number of groups in the molecule that can donate hydrogen atoms to hydrogen bonds (usually the sum of hydroxyl and amine groups in a drug molecule) is < 5; and the number of groups that can accept hydrogen atoms to form hydrogen bonds (estimated by the sum of oxygen and nitrogen atoms) is < 10. (Lipinski et al., 1997). These filters help in early assessing the 'drug-likeness' of compounds and also help in avoiding costly late-stage preclinical and clinical failures. In Table 5, the physical properties of the selected compounds are presented.

3.10.4. Docking study

Using AutoDock Vina (Trott and Olson, 2010), binding models of potential inhibitor molecules to their respective proteins were generated. The grid box parameters for the docking studies and the results are summarized in Table 6.

The most favorable docking is considered to be the conformation with the lowest binding energy. In case of penton base protein B-3, 3-(azepan-1-ium-1-yl) propane-1-sulfonate was found to bind with the protein with the lowest binding energy [Fig. 8(a)]. However with

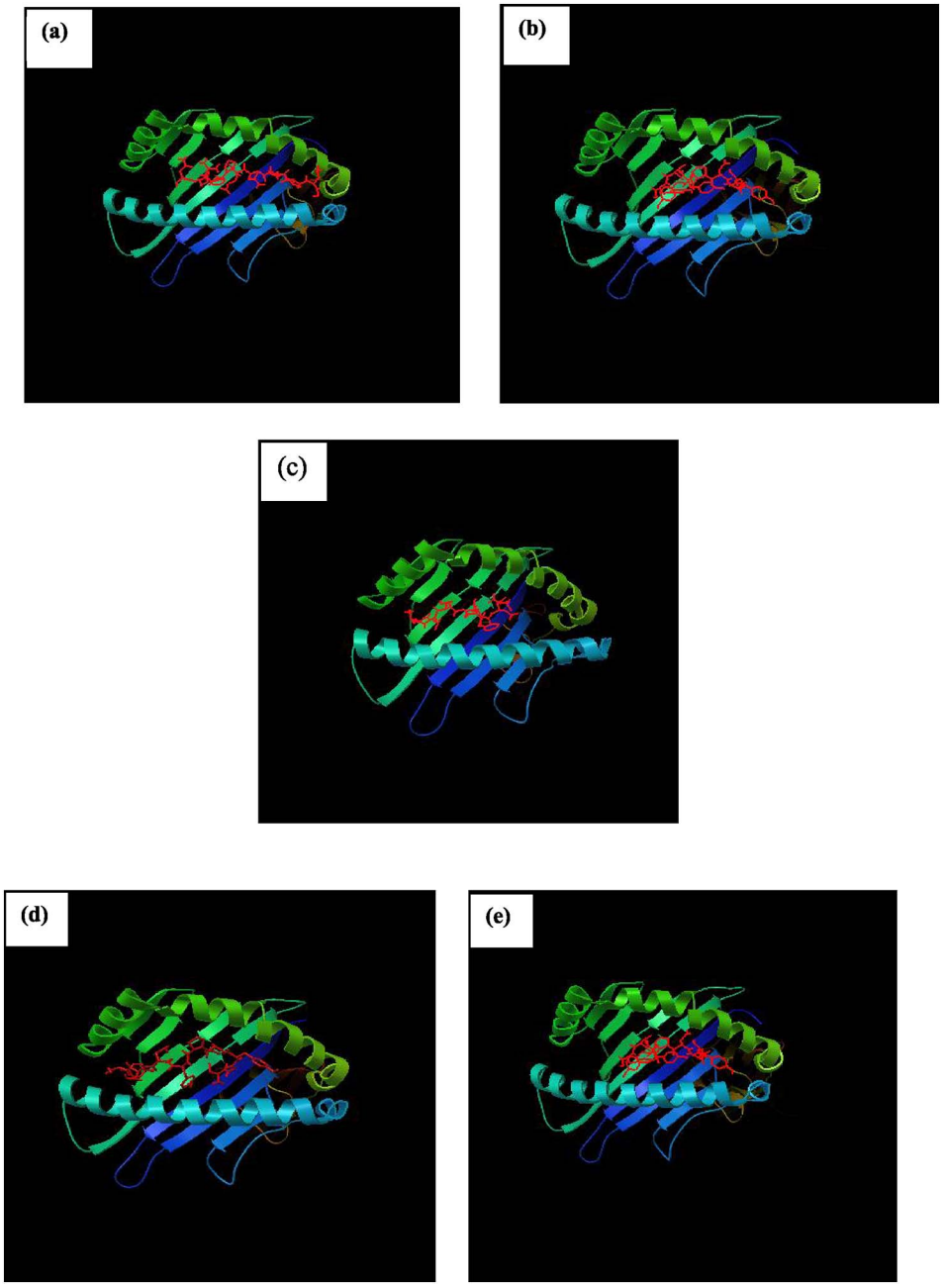


Fig. 4. Docking to predict the binding of predicted 'NRFPENQIL', 'SGYDPYYTY' and control epitopes to MHC class I molecule, HLA-C07:02 and H-2Kb. Binding of 'NRFPENQIL' to the binding grooves (a) of the predicted structure of HLA-C07:02 (binding energy: −8.0 Kcal/mol) and (b) of the 3D structure of H-2Kb (binding energy: −6.7 Kcal/mol); (c) binding of control peptide (PKB1, KVITFIDL) to the predicted 3D structure of HLA-C07:02 (−6.6 Kcal/mol). Binding of 'SGYDPYYTY' to the binding grooves (d) of the predicted structure of HLA-C07:02 (binding energy: −8.3 Kcal/mol) and (e) of the 3D structure of H-2Kb (binding energy: −8.2 Kcal/mol).

Table 3
Probable overlapping epitopes for eliciting both humoral and cell mediated immunity.

Protein name	Predicted peptides	Interacting MHC-I alleles	Overlapping 15mer peptides	Interacting MHC-II alleles	Predicted B cell epitopes	Population coverage	Epitope conservancy table	Allergenicity
Hexon protein C-5	SGYDPYYTY	HLA-A*29:02 HLA-B*15:02 HLA-C*03:03 HLA-C*07:02 HLA-C*12:03	GYDPYYTYSGSIPYL YDPYYTYSGSIPYLD SGYDPYYTYSGSIPY	HLA-DRB1*07:01 HLA-DRB1*07:01 HLA-DRB1*07:01	KETPS LGSGYDPYYTY	41.23%	100% conserved in 4 serotypes of group C	0.003 Non-allergen
Hexon protein D-8	YLAPTLRQG	HLA-C*03:03 HLA-C*07:02 HLA-C*12:03 HLA-C*14:02	SGFTGYLAPTLRQGQ GFTGYLAPTLRQGQP FTGYLAPTLRQGQPY	HLA-DRB1*01:01 HLA-DRB1*01:01 HLA-DRB1*01:01	LRQGQPYPANFP	39.71%	16.67% conserved in 4 serotypes of group D	0.005 Non-allergen
Penton base protein C-5	ESVVSAAPV	HLA-A*68:02 HLA-C*03:03 HLA-C*12:03 HLA-C*15:02	PPPSYESVVSAAPVA	HLA-DQA1*05:01/ DQB1*03:01 HLA-DRB1*01:01	MYEEGPPPSYESVVSA	23.94%	100% conserved in 4 serotypes of group C	0.011 Non-allergen

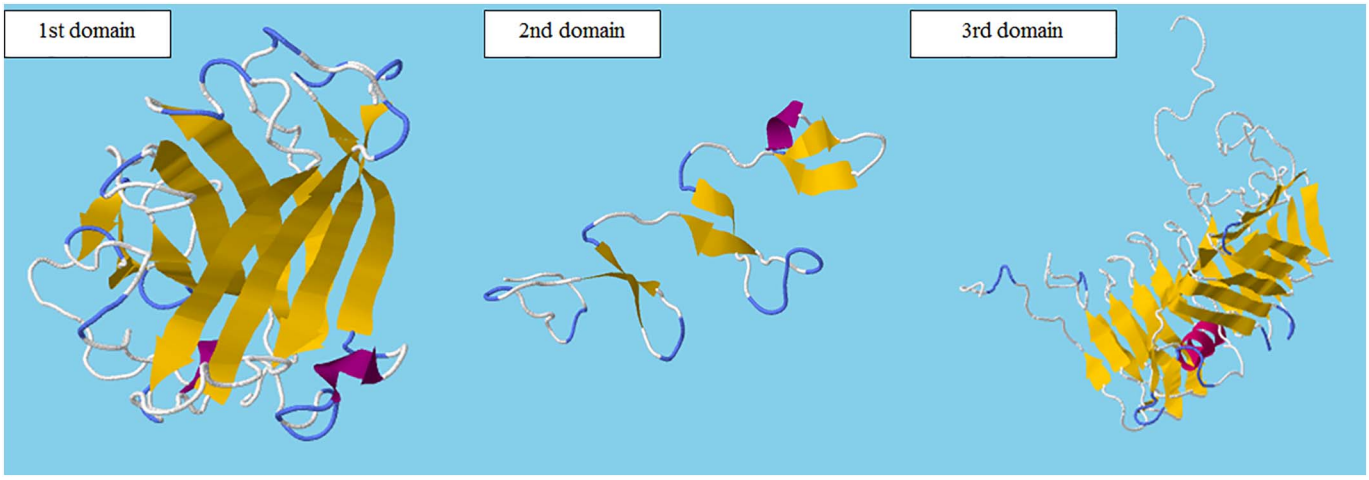


Fig. 5. Predicted 3 domains of fiber protein of HadV C-5 by RaptorX server. The templates used are 1knbA, 1qiuA, 1v1iA and 4rm6A. The visual images of the structures were generated in Jmol.

penton protein C-5, the experimental ligand: *N*-dodecyl-*N*,*N*-dimethyl-3-ammonio-1-propanesulfonate was shown to bind with lowest binding energy (−6.7 Kcal/mol) Fig. 8(c) although 3-(azepan-1-ium-1-yl) propane-1-sulfonate also showed comparatively high affinity (−6.3 Kcal/mol) Fig. 8(b).

In case of hexon protein, E-5842 (4-(4-fluorophenil)-1,2,3,6-tetrahydro-1-[4-(1,2,4-triazol-1-il) butyl] pyridine citrate) bound to the protein's three of the four pockets with lowest binding energy [Fig. 8(d), (e) and (f)].

During docking study between fiber protein and potential ligand molecules, compounds 4UO and 1P0 were found to bind with the largest pocket (determined by CastP server) of fiber protein C-5 with lowest binding energy (−4.8 Kcal/mol) [Fig. 8(g) and (h)]. However by comparing their physicochemical properties 1P0 was selected as the most potential inhibitor since 4UO compound violated more than one criteria of Lipinski's rule of 5 for becoming a drug candidate. The control docking complex was retrieved from HAdV D-37 serotype [Fig. 8(i)] of which binding groove didn't match with the largest pocket of C-5 serotype. Therefore, in this study the corresponding pocket in HAdV C-5 was also included and it was found 1P0 again showed the highest affinity for the groove (−6.6 Kcal/mol) [Fig. 8(j)].

3.10.5. Pharmacophore analysis

Although drug design techniques based on protein structure and protein-ligand interactions have become quite successful, there are many other physic-chemical properties, such as bioavailability, metabolic half-life, side effects etc. that first must be measured before a ligand can be considered a safe and efficacious drug. The polar surface area (PSA) is defined as the surface sum over all polar atoms, (oxygen, nitrogen, sulfur and phosphorus), including also attached hydrogen molecules. Molecules with a polar surface area of > 140 square angstrom are usually believed to be poor at permeating cell membranes. However it may also be noted that cellular membranes in eukaryotes are dynamic in nature which suggests that the presence of membrane proteins, receptors and carriers could overcome the rigidity of polar surface area requirements.

The logP value of a compound, which is the logarithm of its partition coefficient between n-octanol and water, is a well-established measure of the compound's hydrophilicity. Low hydrophilicity and therefore high logP values cause poor absorption or permeation. It has been shown, for compounds to have a reasonable probability of being well absorbed their logP value must not be > 5.0. Finally to find out whether these compounds will exert any undesirable or adverse effect on humans, animals, plants, or the environment, their toxicity was estimated which includes their mutagenic, tumorigenic and irritant properties (Table 7). It can be seen that E-5842 and 3-(azepan-1-ium-1-yl)propane-1-sulfonate would have comparatively more efficacious drug like properties. On the other hand, the polar surface area of 1P0 compound has crossed the threshold of 140 $Å^2$. It is known that the total sum of all polar regions of a molecule's surface correlates well with various bioavailability related properties, such as intestinal absorption, and blood brain barrier penetration. If the contributions of all polar atoms being exposed to the molecule's surface sum up to an area beyond 80 or 100 $Å^2$, then the molecules likelihood to pass membranes easily is significantly reduced.

4. Discussion

In this study, fiber, hexon and penton base proteins were selected as target sites which are the main attachment proteins of HAdV. The early stages of HAdV infection entail initial cell binding mediated by interaction of the fiber with primary cell-surface receptors, followed by secondary binding of the penton base to cell-surface integrins, triggering integrin receptor-mediated endocytosis (Medina-Kauwe, 2013). With the aim to predict epitopes that can be used in vaccine development to prevent viral entry or its interaction with host cell, our study focused on these proteins.

Phylogenetic tree of each of the target proteins was constructed to do their evolutionary analysis. From the evolutionary trees we found out comparatively close relation among the serotypes and also the portions where they share common ancestors in fiber, hexon and penton base proteins. During the selection pressure analysis, all the amino acids' dN/dS ratio was calculated and positively selected amino acids were determined. In this study, positively selected amino acids were identified to be such kind of amino acids that contribute to the protein structure and bioactivities e.g., histidine; to the stability of protein by forming ionic interaction and hydrogen bonds e.g. arginine and lysine; to the formation of salt bridge e.g., aspartic and glutamic acids; to the post-translational modification e.g. serine, threonine, tyrosine and thus, modulate/regulate functions of these attachment proteins. Regions with no positive selection can be considered as conserved. Phylogenetics and evolutionary pressure were analyzed to use this information later for digging out whether any of the amino acids within the predicted epitopes are under selection pressure or not as such pressure will probably reduce acceptability of the epitopes for the utility of vaccine development.

Different tools in IEDB server were used for identifying the most probable B cell epitopes according to their physicochemical properties that generated 18 epitopes. Out of these 18 epitopes, five were considered as the most potential B cell epitopes. Which include

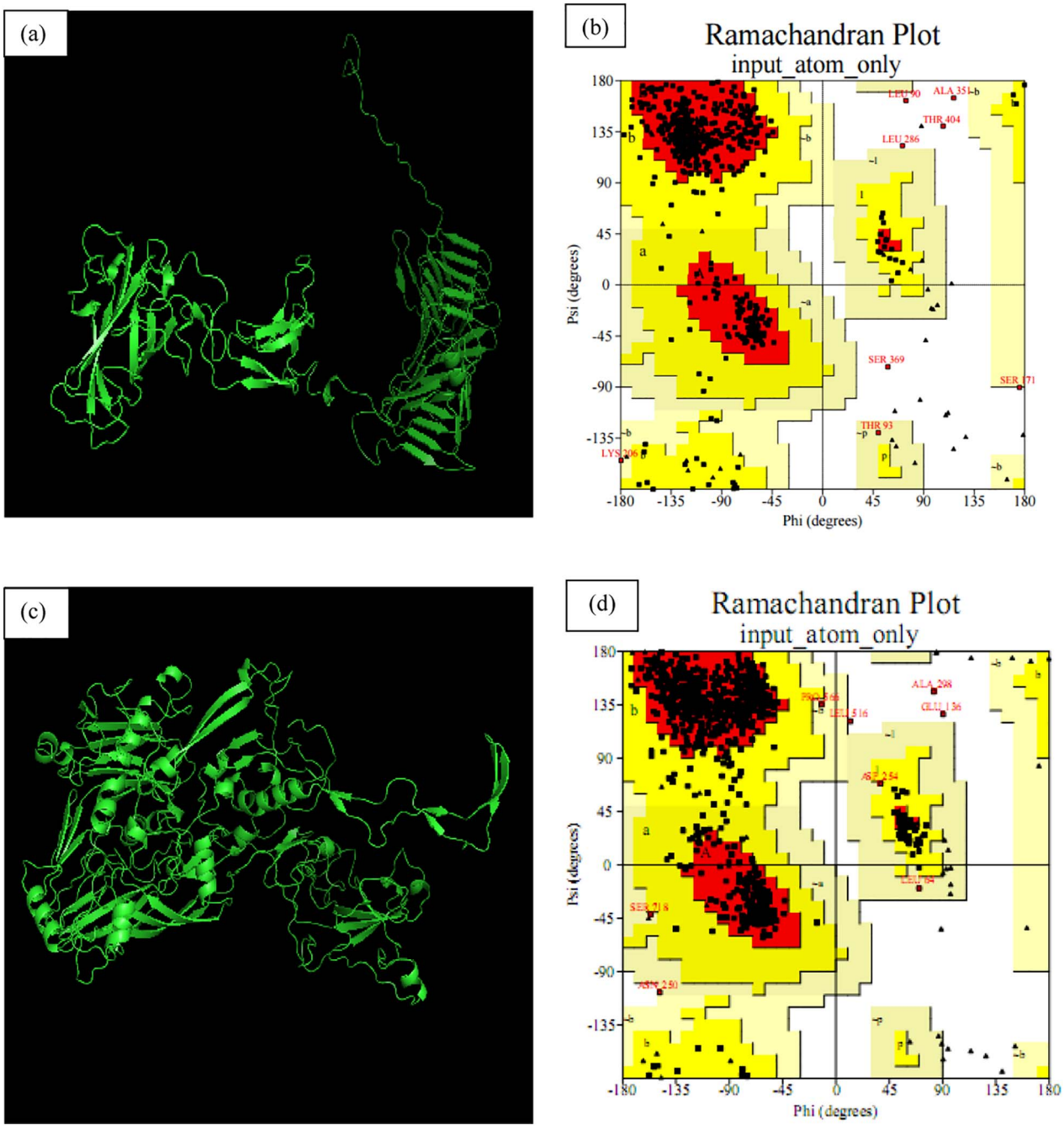


Fig. 6. a) Refined model of predicted three dimensional structure of fiber C-5 (a) and hexon D-8 proteins (c) and Ramachandran plot analyses of these protein models (b) and (d), respectively. Red region indicates favored regions, yellow region for allowed and light yellow shows generously allowed regions and white for disallowed regions. Phi and psi angels determine torsion angles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

"PYDTETGPPTVPFL" from fiber protein C-5, "KETPSLGSGYDPYYTY", from hexon protein C-5, "LRQGQPYPANFP" from hexon protein D-8, "DFTPTEASTQTINFD" from pentose base protein B-3 and "MYEEGPPPSYESVVSA" from pentone base protein C-5. It was found "TETG" from fiber Protein C-5, "YDPY" from hexon Protein C-5, "EA" from penton base protein B-3 and "EGPPPS" from penton base protein C5 - these regions fulfilled the criteria of having surface accessibility, hydrophilicity, flexibility, antigenicity and beta turn regions since they crossed the threshold level with maximum value in these analyses. In addition, experimentally solved 3D structures of fiber protein (1KNB), hexon protein (1P3O) and penton base protein (4AR2) were used to predict the discontinuous B cell epitopes. From these analyses, nine residues were identified from fiber protein C-5, 368 residues were identified from hexon protein C-5 and 134 residues were identified from fiber protein C-5 as potential discontinuous B cell epitopes. Interestingly amino acid sequence "^{692}SGYDPYYT^{699}" was recognized both as linear and discontinuous B cell epitopes; and from penton base protein B-3, amino acid residues ^{100}D, ^{102}T, ^{105}E, ^{108}T, ^{112}N and ^{114}D were recognized both as linear and discontinuous B cell epitopes.

Out of all the predicted T cell epitopes, 'IRFDENGVL' from fiber protein C-5, 'MSMGALTDL' from hexon protein C-5 and D-8, "YQNDHSNFL", and 'NRFPENQIL' from penton base protein B-3 and 'YQNDHSNFL' from penton base protein C-5 were found to interact with the maximum number of MHC class I and II alleles (Table 2). Furthermore, these nonamers were compared with the continuous B cell peptides and finally, 'SGYDPYYTY' from hexon protein C-5,

Table 4
Ligand binding sites predicted from different proteins of interests of human adenovirus.

Proteins	Area of pocket ($Å^2$)	Volume of pocket ($Å^3$)	Pocket residues
Hexon C-5	2765.5	7103.8	116, 117, 121–131, 168–174, 199, 201–207, 209–215, 218, 220–223, 232, 235–237, 243–246, 288, 290, 292, 298–302, 317–323, 413–417, 419, 420, 457–463, 465–467, 469, 470, 472–474, 477–479, 485, 486, 488, 489, 491, 503–508, 510, 513–515, 537, 824, 826–829, 832–836
Penton base B-3	505.4	992.9	148, 151, 163, 164, 166–170, 172, 211, 213, 214, 252, 253, 357, 358, 362–364, 366–369 and 375
Penton base C-5	6333.6	12,124	66, 68, 121, 122, 124–126, 465–479, 518, 520–528, 560, 562–564, 566.
Fiber C-5	2965.3	23,882	16, 18–25, 41–45, 81, 82, 109–111, 113, 132–134, 136, 138, 155, 156, 158, 160, 182, 184, 187–193, 195–200, 226–229, 231–234, 250–252, 254–258, 272, 273, 275, 276, 277, 279–283, 305–307, 309, 311–317, 319, 321–324, 327–329, 331–340, 344, 346, 349, 350, 353–359, 367, 369–371
Hexon D-8	5809.4	16,393	121–132, 134, 152–160, 190, 192–198, 200–206, 208–213, 222, 223, 225–227, 229, 233–236, 247–253, 255, 258, 260, 262, 263–268, 270–273, 275, 278, 280, 282, 290, 307–309, 312, 383, 384, 387, 388, 403–407, 409–414, 439–441, 448, 450–454, 456–458, 460, 461, 463, 464, 466, 468, 470–475, 488, 489, 493, 494, 497, 501–505, 507–510, 513–524, 526, 527, 529, 586–590, 649, 674, 676, 677, 690, 691, 706, 707, 791–793, 795–798, 804–809, 812–822, 824–831, 836–840, 846, 849, 851–853, 855, 898–902.

"YLAPTLRQG" from hexon protein D-8 and "ESVVSAAPV" from penton base protein C-5 were found overlapping with those of B cell epitopic peptides that would have the potential to elicit both B and T cell mediated immunity. Strong linkage disequilibrium between HLA-C and HLA-B alleles makes it difficult to distinguish HLA-C from HLA-B restricted responses. Sometimes HLA-C shares sequence homology with human class I HLA-A and HLA-B molecules. The role of HLA-C as a T-cell restriction element, particularly in HIV-1 infection, has been reported. HLA-C expression was found to continue even though HIV-1 Nef selectively downregulated HLA class I A and B molecules to minimize cytotoxic T lymphocyte scrutiny (Collins et al., 1998; Sakib et al., 2014). Moreover, HLA-C epitopes have been mistakenly identified as restricted by HLA-A or HLA-B (e.g., some B14 epitopes in HIV-1 p24 are now thought to be Cw8-restricted, Los Alamos Immunology Database) (Yusim et al., 2009). Though from the prediction process it was appeared that the CTL epitopes are tend to be biased toward HLA-C alleles, so it is indeed important to verify the rationale behind the identification of HLA-C restricted probable epitopes using animal models or *in vivo* studies. Also, HLA-C allele is no less important in eliciting immune response as these antigens can be recognized by cytotoxic T lymphocytes (Grunnet et al., 1976), can present peptides to T cells, can interact with class I receptors expressed on natural killer cells (Lanier et al., 1995) and regulate natural killer-mediated lysis of target cells (Collins et al., 1998; Colonna et al., 1992). HLA-C antigens function as transplantation determinants and that allele disparity at HLA-A and HLA-B is biologically relevant (Petersdorf et al., 1997).

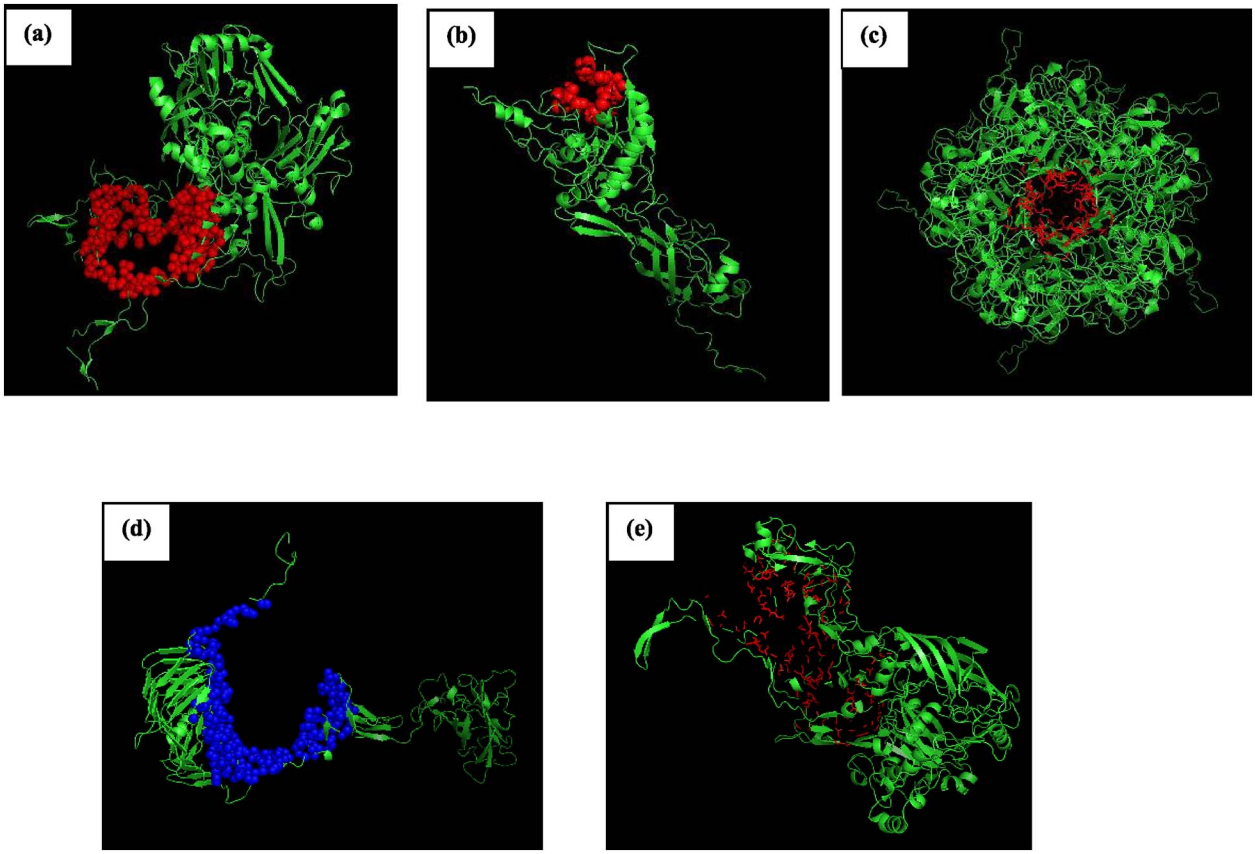


Fig. 7. Residues of largest pockets found in hexon, fiber and penton protein. a) hexon protein of Adenovirus C-5 (red), b) penton base of Adenovirus B3 (red), c) penton base of Adenovirus C-5 (red), d) fiber of Adenovirus C-5 (blue) and e) hexon of Adenovirus D-8 (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Physicochemical properties of potential ligand molecules.

Protein name	Compound	pH range	xlogP	H-bond donors	H-bond acceptors	Net charge	Molecular weight (g/mol)	Rotatable bonds
Penton base	*N*-Dodecyl-*N*,*N*-dimethyl-3-ammonio-1-propanesulfonate	pH 7	−0.80	0	4	0	335.554	15
	3-(Butylamino)propane-1-sulfonic	pH 7	−1.46	2	4	0	195.284	7
	3-(Azepan-1-ium-1-yl)propane-1-sulfonate	pH 7	−0.89	1	4	0	221.322	4
	3-(Piperidin-1-yl)propane-1-sulfonic acid	pH 7	−1.40	1	4	0	207.295	4
	3-(Triethylammonio)propane-1-sulfonate	pH 7	−4.14	0	4	0	223.338	7
	Ethyl dimethyl ammonio propane sulfonate	pH 7	−4.64	0	4	0	195.284	5
Hexon	Citric acid		−1.7	4	7	0	192.12	5
	Hydroxycitric acid		−2.6	5	8	0	208.12	5
	E-5842 (4-(4-fluorophenil)-1,2,3,6-tetrahydro-1-[4-(1,2,4-triazol-1-il)butyl] pyridine citrate	pH 7	2.18	1	4	1	301.389	6
	trans-Aconitic acid		−1.0	3	6	0	174.10	4
	2-Isopropyl-malic acid		−0.2	3	5	0	176.16	4
Fiber	Sialic acid	pH 7	−2.04	6	10	−1	308.263	5
	MNA		−3.4772	6	10		323.297	6
	18D		−3.4507	7	10		323.3	6
	42D		−3.7019	7	11		325.27	6
	4UO		−7.5793	13	20		633.55	11
	1P4		−3.2445	3	13		434.503	15
	1P0		−4.6077	3	13		392.423	12

Among the predicted epitopes 'NRFPENQIL' of penton base protein B-3 and 'SGYDPYYTY' of hexon protein C-5 were selected for docking study since both of them were non-allergen, showed 100% conservancy and highest population coverage among the epitopes which could induce either T cell immune response or both T cell and B cell immune response respectively. During docking analysis both showed good binding energy with respective HLA alleles compared to control peptides. The results are concluding that these epitopes can bind as efficiently as the controls considered in the study and thus confirm their affinity for the MHC-I molecules and uphold their position as potential vaccine candidates.

In addition, the selection pressure analysis showed that both hexon and penton protein had very low percentage of positive selection and so it could be considered that epitopes predicted from hexon protein and penton base protein would be more suitable candidates for vaccine designing.

However, for complete safeguard against adenovirus infection there is a vital need for the development of novel anti-adenovirus agents, including the prevention of fatal diseases from immuno-compromised patients, facilitating gene therapy trials that use adenovirus as a vector and for future development of anti-obesity schemes for those whose obesity might be linked to adenovirus. Moreover, if the effectiveness of

Table 6
Molecular docking parameters and resulting binding affinity between proteins (penton base protein and hexon protein) and selected compounds.

Names of proteins	Grid parameters			Ligand	Binding affinity (Kcal/mol)			
	Center dimensions (A°)	Size (A°)	Spacing (A°)					
Penton base	Penton (B-3)	52 × 20 × 64	1	*N*-Dodecyl-*N*,*N*-dimethyl-3-ammonio-1-propanesulfonate	Penton(B3)	Penton(C5)		
	50.32 × −50.484 × −10.278				−4.8	−6.7		
				3-(Butylamino)propane-1-sulfonic	−4.1	−5.5		
	Penton (C-5)	38 × 36 × 44	1	3-(Azepan-1-ium-1-yl)propane-1-sulfonate	−5.0	−6.3		
	217.156 × −0.004× 360.406			3-(Piperidin-1-yl)propane-1-sulfonic acid	−4.6	−5.8		
				3-(Triethylammonio)propane-1-sulfonate	−4.3	−5.3		
				Ethyl dimethyl ammoniopropane sulfonate	−4.1	−4.9		
Penton base C-2 (control)	50.32 × −50.484 × −10.278	52 × 20 × 64	1	*N*-Dodecyl-*N*,*N*-dimethyl-3-ammonio-1-propanesulfonate	−4.6			
Hexon	Pocket 1:	Pocket 1:	1	Citric acid	P1	P2	P3	P4
	30.785 × 40.386 × 4.803	44 × 50 × 52			−4.3	−2.4	−6.4	−5.5
				Hydroxycitric acid	−4.3	−2.3	−6.6	−5.6
	Pocket 2:	Pocket 2:		E-5842 (4-(4-fluorophenil)-1,2,3,6- tetrahydro-1-[4-(1,2,4-triazol-1-il)butyl] pyridine citrate)	−6.0	−2.2	−7.1	−6.8
	50 × 30 × −40	40 × 40 × 40		*trans*-Aconitic acid	−4.3	−2.1	−6.3	−5.6
	Pocket 3:	Pocket 3:		2-Isopropyl-malic acid	−4.2	−2.5	−6.1	−5.4
	15 × 30 × 45	44 × 54 × 40						
	Pocket 4:	Pocket 4:						
	70 × 40 × 55	46 × 48 × 50						
Hexon (control)				Citric acid	−4.2	−0.6	−6.2	−5.4
Fiber	−60.531 × 1.943 × 100.207	58 × 78 × 58	1	Sialic acid	−4.4			
				MNA	−4.5			
				18D	−4.4			
				42D	−4.5			
				4UO	−4.8			
				1P4	−4.6			
				1P0	−4.8			
Fiber (control)	16.93 5× 2.396 × −32.601	32 × 24 × 20	1	Sialic acid	−4.9			

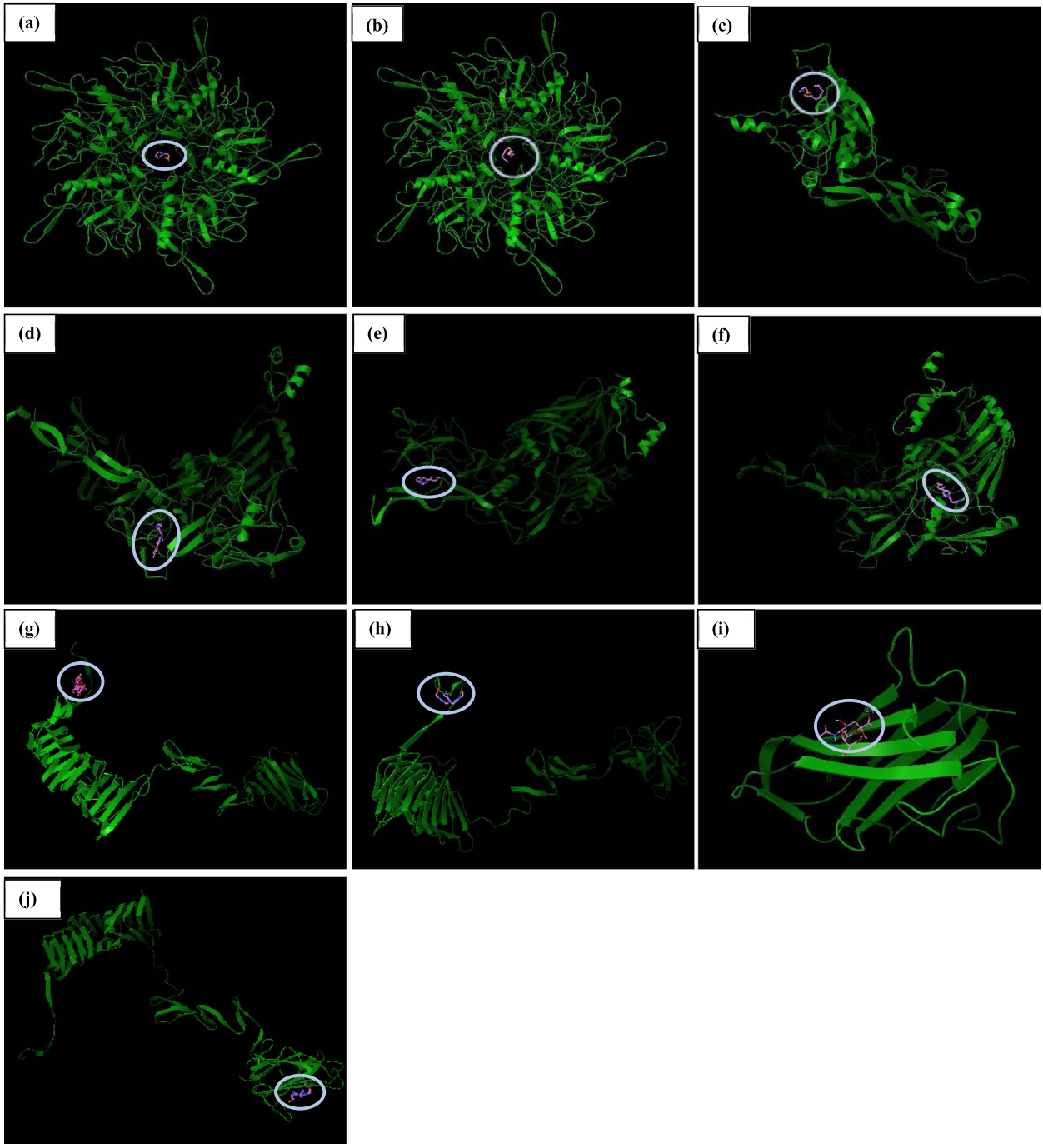


Fig. 8. Docking of penton, hexon and fiber proteins. Binding of 3-(azepan-1-ium-1-yl)propane-1-sulfonate to the binding pockets of penton base protein (binding energy: −5.0 Kcal/mol) (a), penton protein C-5 (binding energy: −6.3 Kcal/mol) (b) and binding of *N*-dodecyl-*N*,*N*-dimethyl-3-ammonio-1-propanesulfonate to the binding pocket of penton protein C-5 (binding energy: −6.7 Kcal/mol) (c); binding of 'E-5842' compound to the 1st, 3rd and 4th binding pockets of hexon protein (d), (e) and (f); bindings of two different 4UO (g) and 1P0 (h) compounds to the largest pocket of fiber protein C-5, (i) binding of sialic acid to the control protein (fiber protein C-5, PDB ID: 3QND) and (j) binding of 1P0 compound to the corresponding pocket (similar to control) of fiber protein C-5. Each Circle indicates the interaction of respective molecules with the respective binding pockets of proteins.

the vaccine is reduced due to mutation in the virus the drug may work alongside to diminish the symptoms faster and in case of any sudden outbreak in a non-vaccinated area vaccination would become obsolete (Hasan et al., 2015). Therefore, in the present project, we also predicted the active sites of the target proteins and subsequently predicted novel inhibitory candidate against them that must be assayed in vitro and in vivo to assess their real potential activity.

Three of the protein structures were available in protein data bank. However for the other two (fiber protein of adenovirus C-5 and hexon protein of adenovirus D-8), their three dimensional structures were modeled using homology modeling and the predicted structures scored satisfactorily well in structure validation analysis. From the binding site analysis, an insight could be gained on which particular sites of the selected proteins should be targeted for drug design. Docking study showed that 3-(azepan-1-ium-1-yl) propane-1-sulfonate could also be a better ligand for penton base protein as it bound the protein with the lowest binding energy (−5.0 Kcal/mol) compared to control (−4.6 Kcal/mol) and other analogues. Similarly, "E-5842" compound may be considered as a good inhibitor since it was found to interact with three of the four pockets of hexon protein with lower binding

Table 7
Pharmacological characteristics of selected ligand molecules.

Properties	E-5842	3-(Azepan-1-ium-1-yl) propane-1-sulfonate	1P0
Total surface area	283	192	342
Polar surface area	35.13	70.02	156.06
clogP	0.1604	−2.14	−4.60
Mutagenic	None	None	None
Tumorigenic	None	None	None
Irritant	None	None	None

energy (−6.0, −7.1 and −6.8 Kcal/mol) compared to control. Also, 1P0 compound was predicted to be more effective against fiber protein compared to current drug but its pharmacophore analysis was not promising. Therefore, this docking study may recommend 3-(azepan-1-ium-1-yl) propane-1-sulfonate and E-5842 compound to be more effective against penton base and hexon protein respectively through higher binding affinity and thus stronger inhibition. However, although both of them exhibited no tumorigenic, mutagenic and irritant properties during pharmacophore studies, their drug like properties, ADMET properties and toxicity should be fully analyzed and verified by biological experimentation before they can be established as potent drug candidates.

5. Conclusion

Immunoinformatics considers online immunological databases, tools and web servers which are immensely useful in reducing the time and cost involved in the traditional study of immunology. Here computational prediction of B cell and T cell epitopes and the epitopes eliciting both B cell and T cell immune response finally led us to some potential epitopes which would cover all the criteria to become a potential vaccine candidate against human adenovirus. However, the panel of CTL epitopes has been identified from the capsid proteins, which can be considered as the shortcoming of this study. This is because we aimed to identify such an epitope which will be able to elicit both humoral and cell mediated immunity if used as a vaccine candidate. As we know that surface proteins are the best choice for designing a probable B cell epitope(s), we preferred to deal with the fiber, hexon and penton base proteins of HAdV and analyses revealed epitopes which overlapped with those of B cell epitopes as reflected by the data presented in Table 3. We believe that the proposed predicted epitopes will be able to evoke better immune response when it will be tested in wet laboratory. It has also been shown in some cases previously that multiple epitopes may increase the specificity of immune responses but compromises the strength of the responses therefore testing the proposed epitopes' immunogenicity is imperative to rule out the possibility. At the same time we predicted potential novel target sites in viral proteins as well as drug candidates that might be tested to control HAdV. As a future course of action, to validate the findings of the current study extensive laboratory based in vivo trials are required for the ultimate development of pre therapeutic and therapeutic treatment against human adenovirus.

Conflict of interests

This work was carried out without any grant or other financial support. There is no conflict of interest regarding this work.

Contribution of authors

AHMNN conceived the idea and supervised the work. RH and TY carried out the analytical works. AHMNN, MIH and TY wrote the manuscript.

Acknowledgement

We would like to express our heartfelt gratitude to Professor Dr. AKM Mahbub Hasan and Assistant Professor Dr. Sajib Chakraborty of the Department of Biochemistry and Molecular Biology, University of Dhaka for their encouragement and constructive criticism about the work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jim.2018.01.005.

References

Benkert, P., Biasini, M., Schwede, T., 2011. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27 (3), 343–350.
Bowie, J.U., Lüthy, R., Eisenberg, D., 1991. A method to identify protein sequences that fold into a known three-dimensional structure. Science 253 (5016), 164–170.
Bui, H.H., Sidney, J., Dinh, K., Southwood, S., Newman, M.J., Sette, A., 2006. Population coverage calculation, predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics. 7, 153.
Bui, H.H., Sidney, J., Li, W., Fusseder, N., Sette, A., 2007. Epitope conservancy analysis-development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC Bioinformatics. 8 (1), 361.
Centers for Disease control and Prevention, 2014. Adenovirus VIS (Vaccine Information Statements). 6/11/. .
Chou, P.Y., Fasman, G.D., 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. Relat. Areas Mol. Biol. 47, 45–148.
Collins, K.L., Chen, B.K., Kalams, S.A., Walker, B.D., Baltimore, D., 1998. HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. Nature 391 (6665), 397–401 (Jan 22).
Colonna, M., Spies, T., Strominger, J.L., Ciccone, E., Moretta, A., Moretta, L., Pende, D., Viale, O., 1992. Alloantigen recognition by two human natural killer cell clones is associated with HLA-C or a closely linked gene. Proc. Natl. Acad. Sci. U. S. A. 89, 7983.
Daga, P.R., Duan, J., Doerksen, R.J., 2010. Computational model of hepatitis B virus DNA polymerase: Molecular dynamics and docking to understand resistant mutations. Protein Sci. Publ. Protein Soc. 19 (4), 796–807.
Dang, H.X., Lawrence, C.B., 2014. Allerdictor: fast allergen prediction using text classification techniques. Bioinformatics 30 (8), 1120–1128 btu004.
Dong, Xu, Zhang, Yang, 2011. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophys. J. 101, 2525–2534.
Doytchinova, I.A., Flower, Darren R., 2007. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 8, 4.
Dundas, Joe, Zheng, Ouyang, Tseng, Jeffery, Binkowski, Andrew, Turpaz, Yaron, Liang, Jie, 2006. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acid Res. 34, W116–W118.
Edwards, K.M., Thompson, J., Paolini, J., Wright, P.F., 1985. Adenovirus infections in young children. Pediatrics 76, 420–424.
Eisenberg, D., Luethy, R., James, B.W., 1997. VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol. 277, 396–404.
Emini, E.A., Hughes, J.V., Perlow, D.S., Boger, J., 1985. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J. Virol. 55, 836–839.
Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.
Grunnet, N., Kristensen, T., Kissmeyer-Nielsen, F., 1976. Cell mediated lympholysis in man. The impact of HLA-C antigens. Tissue Antigens 7 (301).
Hasan, M.A., Khan, M.A., Datta, A., Mazumder, M.H., Hossain, M.U., 2015. A comprehensive immunoinformatics and target site study revealed the corner-stone toward Chikungunya virus treatment. Mol. Immunol. 65 (1), 189–204. http://dx.doi.org/10.1016/j.molimm.2014.12.013. (Epub 2015 Feb 14).
Hierholzer, J.C., 1992. Adenoviruses in the immunocompromised host. Clin. Microbiol. Rev. 5, 262–274.
Jones, M.S., Harrach, B., Ganac, R.D., Gozum, M.M., Dela Cruz, W.P., Riedel, B., Pan, C., Delwart, E.L., Schnurr, D.P., 2007. New adenovirus species found in a patient presenting with gastroenteritis. J. Virol. 81, 5978–5984.
Källberg, Morten, Wang, Haipeng, Wang, Sheng, Peng, Jian, Wang, Zhiyong, Lu, Hui, Xu, Jinbo, 2012. Template-based protein structure modeling using the RaptorX web server. Nat. Protoc. 7, 1511–1522.
Karplus, P.A., Schulz, G.E., 1985. Prediction of chain flexibility in proteins. Naturwissenschaften 72, 212–213.
Kelley, L.A., et al., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10, 845–858.
Kolaskar, A.S., Tongaonkar, P.C., 1990. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett. 276, 172–174.
Lanier, L.L., Gumperz, J.E., Parham, P., Melero, I., Lopez-Botet, M., Phillips, J.H., 1995. The NKB1 and HP-3E4 NK cell receptors are structurally distinct glycoproteins and independently recognize polymorphic HLA-B and HLA-C molecules. J. Immunol. 154,

3320.
Larsen, Jens Erik Pontoppidan, Lund, Ole, Nielsen, Morten, 2006. BepiPred-improved method for predicting linear B-cell epitopes. Immun. Res. 2, 2.
Laskowski, R.A., MacArthur, M.W., Moss, D., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 283–291.
Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R., Thornton, J.M., 1996. Aqua AND Procheck-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR. 8, 477–486.
Lenaerts, L., De Clercq, E., Naesens, L., 2008. Clinical features and treatment of adenovirus infections. Rev. Med. Virol. 18, 357–374.
Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J., 1997. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev. 23, 3–25.
Mandell, G.L., Bennett, J.E., Dolin, R., 2004. Adenovirus — Principles and Practice of Infectious Disease, 6th edition. .
Maupetit, J., Derreumaux, P., Tuffery, P., 2010 Mar. A fast and accurate method for large-scale de novo peptide structure prediction. J. Comput. Chem. 31 (4), 726–738.
Medina-Kauwe, L.K., 2013. Development of adenovirus capsid proteins for targeted therapeutic delivery. Ther. Deliv. 4 (2), 267–277. http://dx.doi.org/10.4155/tde.12.155.
Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., Olson, A.J., 1998. Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. J. Comput. Chem. 19, 1639–1662.
Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J., 2009. Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. J. Comput. Chem. 16, 2785–2791.
Muh, H.C., Tong, J.C., Tamm, M.T., 2009. AllerHunter: a SVM-pairwise system for assessment of allergenicity and allergic cross-reactivity in proteins. PLoS ONE 4 (6).
Ohori, N.P., Michaels, M.G., Jaffe, R., Williams, P., Yousem, S.A., 1995. Adenovirus pneumonia in lung transplant recipients. Hum. Pathol. 26, 1073–1079.
Parker, J.M., Guo, D., Hodges, R.S., 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, 5425–5432.
Petersdorf, E.W., Longton, G.M., Anasetti, C., Mickelson, E.M., McKinney, S.K., Smith, A.G., Martin, P.J., Hansen, J.A., 1997. Association of HLA-C disparity with graft failure after marrow transplantation from unrelated donors. Blood 89 (5), 1818–1823.
Ray, C.G., 2004. Influenza, respiratory syncytial virus, adenovirus, and other respiratory viruses. In: Ryan, K.J., Ray, C.G. (Eds.), Sherris Medical Microbiology — An Introduction to Infectious Diseases, 4th ed. The McGraw-Hill Companies, Inc., USA, pp. 495–512.
Robinson, C.M., Seto, D., Jones, M.S., Dyer, D.W., Chodosh, J., 2011. Molecular evolution of human species D adenoviruses. Infect. Genet. Evol. 11, 1208–1217. http://dx.doi.org/10.1016/j.meegid.2011.04.031.
Rowe, W.P., Huebner, R.J., Gilmore, L.K., Parrott, R.H., Ward, T.G., 1953. Isolation of a cytopathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. Proc. Soc. Exp. Biol. Med. 84, 570–573.
Rux, J.J., Kuser, P.R., Burnett, R.M., 2003. Structural and phylogenetic analysis of adenovirus hexons by use of high-resolution x-ray crystallographic, molecular modeling, and sequence-based methods. J. Virol. 77, 9553–9566.
Saha, S., Raghava, G.P.S., 2006. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 65 (1), 40–48.
Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.
Sakib, M.S., Islam, M.R., Hasan, A.K.M.M., Nabi, A.H.M.N., 2014. Prediction of epitope-based peptides for the utility of vaccine development from fusion and glycoprotein of Nipah virus using in silico approach. Adv. Bioinforma. 2014http://dx.doi.org/10.1155/2014/402492. (17 pages).
Sander, Thomas, Freyss, Joel, von Korff, Modest, Rufener, Christian, 2015. DataWarrior: an open-source program for chemistry aware data visualization and analysis. J. Chem. Inf. Model. 55 (2), 460–473.
Santana-Jorge, K.T.O., Santos, T.M., Tartaglia, N.R., Aguiar, E.L., Souza, R.F.S., Mariutti, R.B., Azevedo, V., 2016. Putative virulence factors of *Corynebacterium pseudotuberculosis* FRC41: vaccine potential and protein expression. Microb. Cell Factories 15 (83). http://dx.doi.org/10.1186/s12934-016-0479-6.
Simsir, A., Greenebaum, E., Nuovo, G., Schulman, L.L., 1998. Late fatal adenovirus pneumonitis in a lung transplant recipient. Transplantation 65, 592–594.
Soria-Guerra, R.E., Nieto-Gomez, R., Govea-Alonso, D.O., Rosales-Mendoza, S., 2015. An overview of bioinformatics tools for epitope prediction: implications on vaccine development. J. Biomed. Inform. 53, 405–414. http://dx.doi.org/10.1016/j.jbi.2014.11.003.
Spjut, S., Qian, W., Bauer, J., Storm, R., Frängsmyr, L., Stehle, T., Arnberg, N., Elofsson, M., 2011. A potent trivalent sialic acid inhibitor of adenovirus type 37 infection of human corneal cells. Angew. Chem. Int. Ed. Engl. 50 (29), 6519–6521.
Tamura, K., Nei, M., Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences (USA). 101, 11030–11035.
Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.
Thevenet, P., Tufféry, P., Lamiable, A., 2012. A critical assessment of HMM taboo sampling strategies applied to the generation of peptide 3D models (in preparation) 27 (12), 1715–1716.
Trott, O., Olson, A.J., 2010. AutoDockVina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J. Comput. Chem. 31, 455–461.
Vita, R., Zarebski, L., Greenbaum, J.A., Emami, H., Hoof, I., Salimi, N., Damle, R., Sette, A., Peters, B., 2010. The immune epitope Database 2.0. Nucleic Acids Res. 38 (Database issue), D854–D862.
Xu, D., Zhang, Y., 2011. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophys. J. 101, 2525–2534.
Yusim, K., BTM, Korber, Christian, B. (Eds.), 2009. HIV Molecular Immunology. Los Alamos National Laboratory, Los Alamos, NM, USA.
Zhang, Zengming, Li, Yu, Lin, Biaoyang, Schroeder, Michael, Huang, Bingding, 2011. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. Bioinformatics 27 (15), 2083–2088.
Zoete, Vincent, Grosdidier, Aurélien, Michielin, Olivier, 2009. Docking, virtual high throughput screening and in silico fragment-based drug design. J. Cell. Mol. Med. 13 (2), 238–248.
Zubieta, C., Schoehn, G., Chroboczek, J., Cusack, S., 2005. The structure of the human adenovirus 2 penton. Mol. Cell 17, 121–135.

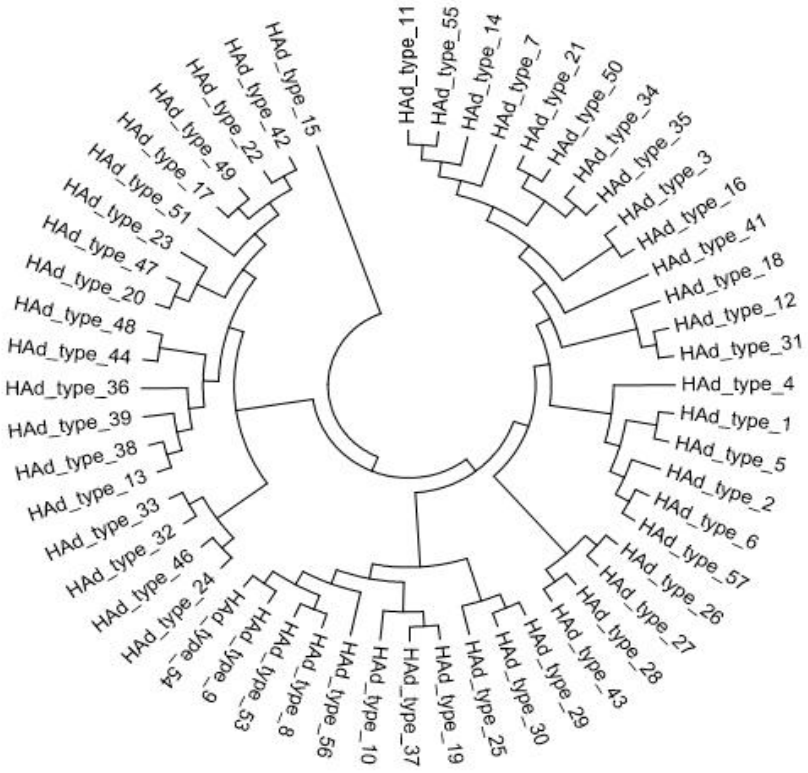


Figure S1. Evolutionary relationships of Fiber Protein of Human Adenovirus. The dendogram is presented by the number of different adenovirus serotypes.The analysis involved 54 nucleotide sequences. There were a total of 458 positions in the final dataset. From the tree, it can be inferred that, HAdV type 7,11,14,55 are comparatively closelyrelated; HAdV types 21,34,35,50 are comparatively closely related and they altogether share a common ancestor. Again, from the figure it can be observed that, HAdV types 8, 9, 53, 54, 56 are relatively closely related and they share an ancestor with HAdV types 25,29,30 and with 10,19,37.

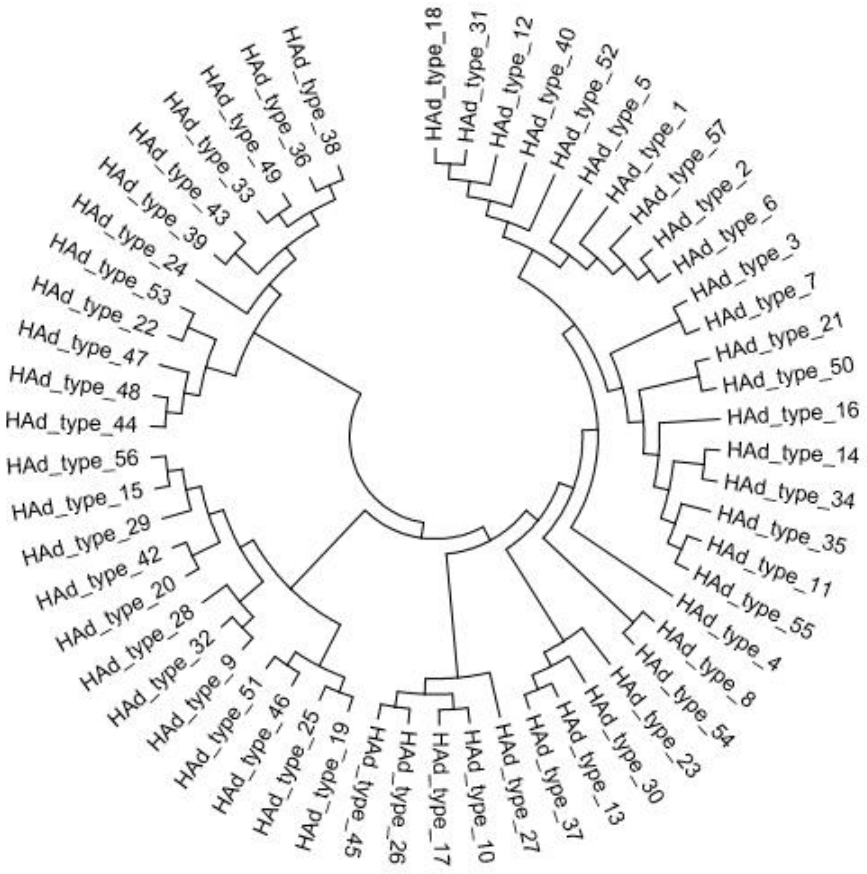


Figure S2. Evolutionary relationships of Hexon Protein of Human Adenovirus.

The analysis involved 56 nucleotide sequences. There were a total of 2613 positions in the final dataset. From this tree, it can be observed that, HAdV types 12, 18, 31,40, 42 are comparatively closely related; HAdV types 1,2,5,6,57 are relatively closely related and they share a common ancestor. Again, HAdV types 19,25,46,51 are evolutionarily closely related; HAdV types 9,15, 20, 28, 29, 32, 42, 56 are comparatively closely related and they share a common ancestor as well. It can also be said that, HAdV types 33,36,38,49 are phylogenetically more close; HAdV types 24,39,43 are comparatively closely related; HAdV types 22,44,47,48,53 again related and they altogether share a common ancestor.

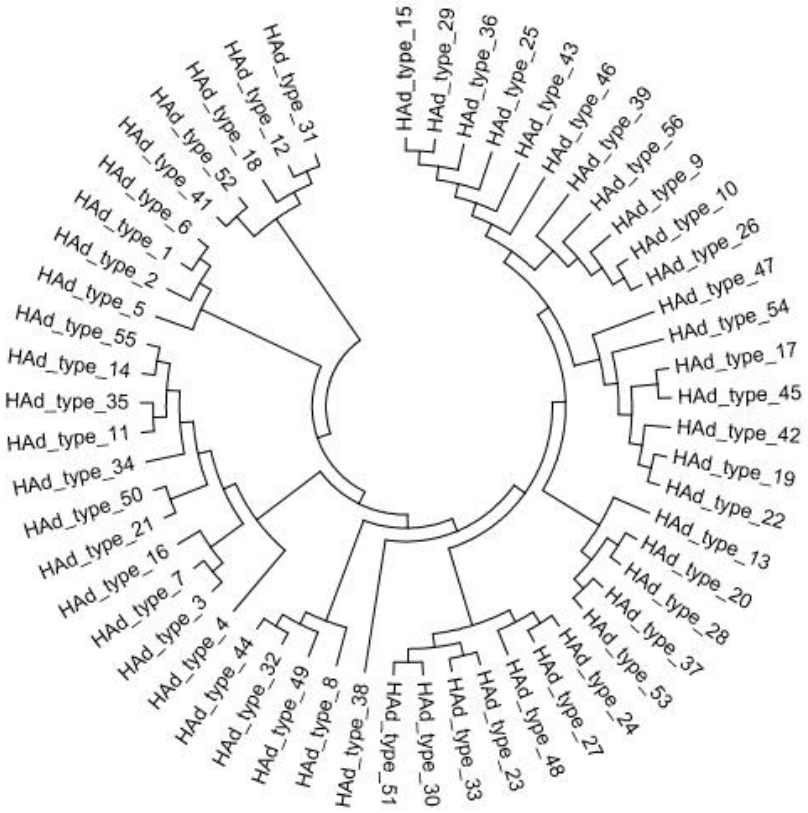


Figure S3. Evolutionary relationships of Penton Base Protein of Human Adenovirus.

The analysis involved 55 nucleotide sequences. There were a total of 1423 positions in the final dataset. From the tree, it can be found that, HAdV types 15, 25, 29, 36, 43, 46 are comparatively closely related; HAdV types 9, 10, 26, 39, 56 are relatively closely related and together they share a common ancestor. Again, HAdV types 17,19,22,42,45,47,54 are comparatively closely related. It can also be shown that, HAdV types 27, 28, 48 are comparatively closely related; HAdV types 23,30,33,51 are comparatively closely related and they altogether share a common ancestor. HAdV types 3,4,7,16,21,34,50 are comparatively closely related, HAdV types 11,14,35,55 are related and finally HAdV types 12,18,31,41,52 are more close.

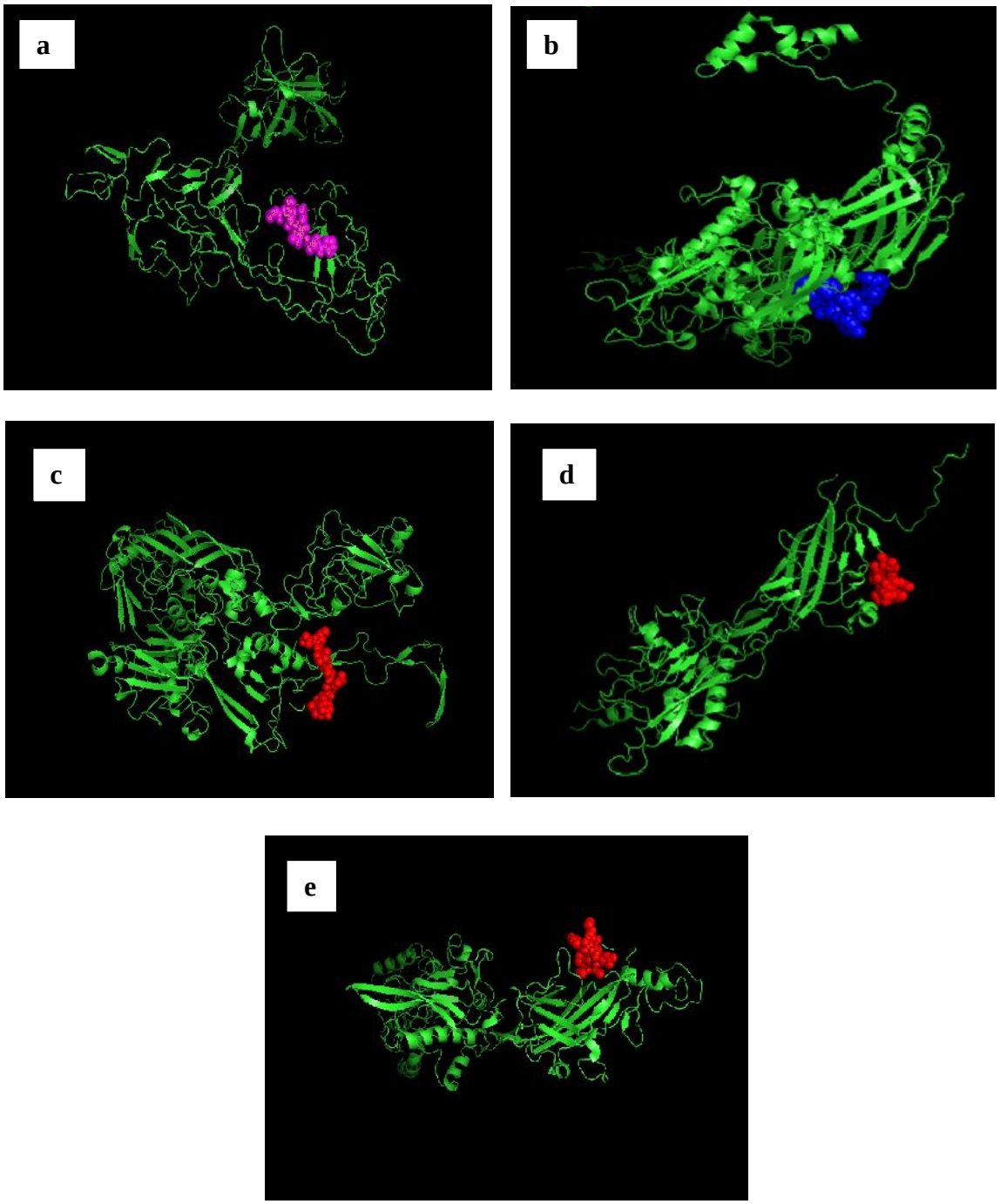


Figure S4. Three dimensional structures of B cell epitopes of hexon, fiber and penton base protein. B cell epitopes from (a) fiber protein of adenovirus C-5 (in pink), (b)hexon protein of adenovirus C-5 serotype (in blue), (c)hexon protein of adenovirus D-8 (in red), (d)penton base protein of adenovirus B-3 (in red) and (e) penton base protein of adenovirusC-5 (in red).

Table S1. T- cell epitope prediction of Fiber protein C-5 of Human Adenovirus.

Peptide	Interacting MHC-I Alleles	Overlapping 15 mer peptides	Interacting MHC-II Alleles	Population Coverage	Epitope Conservancy Table
NYNK GLYLF	HLA-A*23:01	AHNLDINY NKGLYLF	HLA-DRB1*13:02	50.49%	25% conserved inserotypes of group C
	HLA-C*14:02		HLA-DRB1*07:01		
	HLA-A*24:02	HNLDINYN KGLYLFT	HLA-DRB1*13:02		
	HLA-C*12:03		HLA-DRB1*07:01		
	HLA-B*15:02	NLDINYNK GLYLFTA	HLA-DRB1*13:02		
	HLA-C*07:01		HLA-DRB1*07:01		
			HLA-DRB1*01:01		
		LDINYNKG LYLFTAS	HLA-DRB1*13:02		
			HLA-DRB1*07:01		
			HLA-DRB1*01:01		
		NYNKGLYL FTASNNS	HLA-DRB1*07:01		
			HLA-DRB1*01:01		
			HLA-DRB1*04:01		
			HLA-DRB1*04:04		
		DINYNKGL YLFTASN	HLA-DRB1*13:02		
		INYNKGLY LFTASNN	HLA-DRB1*13:02		
SAYSM SFSW	HLA-B*58:01	SAYSMSFS WDWSGHN	HLA-DRB3*01:01	32.79%	25% conserved inserotypes of group C
	HLA-C*12:03				
	HLA-C*03:03				
	HLA-B*53:01				
	HLA-B*57:01				
	HLA-A*32:01				
	HLA-C*15:02				
NEIFAT SSY	HLA-C*12:03	HNYINEIFA TSSYTF	HLA-DRB1*07:01	36.53%	25% conserved inserotypes of group C
	HLA-B*18:01		HLA-DRB1*01:01		
	HLA-B*44:03	NYINEIFAT	HLA-DRB1*07:01		

		SSYTFS			
	HLA-C*03:03		HLA-DRB1*01:01		
	HLA-B*44:02	INEIFATSS YTFSYI	HLA-DRB1*07:01		
	HLA-B*15:02	YINEIFATS SYTFSY	HLA-DRB1*07:01		
		NEIFATSSY TFSYIA	HLA-DRB1*07:01		
		GHNYINEIF ATSSYT	HLA-DRB1*01:01		
		SGHNYINEI FATSSY	HLA-DRB1*01:01		
YSMSF SWDW	HLA-B*58:01	YSMSFSWD WSGHNYI	HLA-DRB3*01:01	44.95%	25% conserved inserotypes of group C
	HLA-C*12:03	SAYSMSFS WDWSGHN	HLA-DRB3*01:01		
	HLA-C*07:01	AYSMSFSW DWSGHNY	HLA-DRB3*01:01		
	HLA-B*57:01				
	HLA-C*15:02				
	HLA-C*03:03				
	HLA-B*53:01				
IRFDE NGVL	HLA-C*07:01	SAHLIIRFD ENGVLL	HLA-DRB1*13:02	64.82%	25% conserved inserotypes of group C
	HLA-C*12:03		HLA-DRB3*01:01		
	HLA-C*06:02	AHLIIRFDE NGVLLN	HLA-DRB1*13:02		
	HLA-C*07:02		HLA-DRB3*01:01		
	HLA-C*03:03	HLIIRFDEN GVLLNN	HLA-DRB1*13:02		
	HLA-C*14:02		HLA-DRB3*01:01		
		LIIRFDENG VLLNNS	HLA-DRB1*13:02		
			HLA-DRB3*01:01		
		QSAHLIIRF DENGVL	HLA-DRB1*13:02		
			HLA-DRB3*01:01		

Table S2.T-cell epitope prediction of Hexon protein C-5 of Human Adenovirus.

Peptide	Interacting MHC-I	Overlapping 15 mer peptides	Interacting MHC-II Alleles	Population Coverage	Epitope Conservancy Table
FLYSNIALY	HLA-B*15:01	NLWRNFLYSNIALYL	HLA-DRB1*01:01	32.05%	100% conserved in serotypes of group C
	HLA-C*14:02		HLA-DPA1*02:01/DPB1*01:01		
	HLA-C*12:03		HLA-DRB1*07:01		
	HLA-B*15:02		HLA-DPA1*03:01/DPB1*04:02		
	HLA-A*29:02		HLA-DPA1*03:01/DPB1*04:02		
	HLA-C*03:03		HLA-DRB1*13:02		
			HLA-DRB1*13:02		
		LWRNFLYSNIALYLP	HLA-DRB1*01:01		
			HLA-DPA1*02:01/DPB1*01:01		
			HLA-DRB1*07:01		
		WRNFLYSNIALYLPD	HLA-DRB1*01:01		
			HLA-DPA1*02:01/DPB1*01:01		
			HLA-DPA1*03:01/DPB1*04:02		

			HLA-DRB1*07:01		
			HLA-DRB1*13:02		
		NFLYSNIALYLPDKL	HLA-DRB1*01:01		
		RNFLYSNIALYLPDK	HLA-DPA1*02:01/DPB1*01:01		
			HLA-DRB1*01:01		
			HLA-DPA1*03:01/DPB1*04:02		
			HLA-DRB1*07:01		
			HLA-DRB1*13:02		
CLYATFFPM	HLA-C*03:03	KFDSICLYATFFPMA	HLA-DPA1*01:03/DPB1*02:01	58.01%	100% conserved in serotypes of group C
	HLA-C*14:02	FDSICLYATFFPMAH	HLA-DPA1*01:03/DPB1*02:01		
	HLA-A*02:06	SICLYATFFPMAHNT	HLA-DPA1*01:03/DPB1*02:01		
	HLA-C*12:03	DSICLYATFFPMAHN	HLA-DPA1*01:03/DPB1*02:01		
	HLA-A*02:01	IKFDSICLYATFFPM	HLA-DPA1*01:03/DPB1*02:01		
	HLA-B*15:01	CLYATFFPMAHNTAS	HLA-DRB1*01:01		
	HLA-B*15:02	ICLYATFFPMAHNTA	HLA-DPA1*01:03/DPB1		

			*02:01		
		KFDSICLY ATFFPMA	HLA-DPA1*01/DPB1*04:01		
		FDSICLYAT FFPMAH	HLA-DPA1*01/DPB1*04:01		
		DSICLYATF FPMAHN	HLA-DPA1*01/DPB1*04:01		
		SICLYATFF PMAHNT	HLA-DPA1*01/DPB1*04:01		
		ICLYATFFP MAHNTA	HLA-DRB1*01:01		
		ICLYATFFP MAHNTA	HLA-DPA1*01/DPB1*04:01		
		CLYATFFP MAHNTAS	HLA-DPA1*01:03/DPB1*02:01		
		CLYATFFP MAHNTAS	HLA-DRB1*04:04		
		KFDSICLY ATFFPMA	HLA-DPA1*02:01/DPB1*01:01		
		FDSICLYAT FFPMAH	HLA-DPA1*02:01/DPB1*01:01		
		DSICLYATF FPMAHN	HLA-DPA1*02:01/DPB1*01:01		
		SICLYATFF	HLA-DPA1*02:01/DPB1		

		PMAHNT	*01:01		
FRKDV NMVL	HLA-C*12:03	YTYEWNFR KDVNMVL	HLA-DRB3*01:01	64.89%	100% conserved in serotypes of group C
	HLA-C*07:01	TYEWNFRK DVNMVLQ	HLA-DRB3*01:01		
	HLA-B*15:02	YEWNFRK DVNMVLQ S	HLA-DRB3*01:01		
	HLA-C*06:02	EWNFRKD VNMVLQSS	HLA-DRB3*01:01		
	HLA-C*07:02	WNFRKDV NMVLQSSL	HLA-DRB3*01:01		
	HLA-C*03:03	YTYEWNFR KDVNMVL	HLA-DRB1*03:01		
	HLA-B*39:01	TYEWNFRK DVNMVLQ	HLA-DRB1*03:01		
		YEWNFRK DVNMVLQ S	HLA-DRB1*03:01		
		EWNFRKD VNMVLQSS	HLA-DRB1*03:01		
		WNFRKDV NMVLQSSL	HLA-DRB1*03:01		
		TYEWNFRK DVNMVLQ	HLA-DRB1*13:02		
		YEWNFRK DVNMVLQ S	HLA-DRB1*13:02		
		EWNFRKD VNMVLQSS	HLA-DRB1*13:02		
		YTYEWNFR KDVNMVL	HLA-DRB1*13:02		

		WNFRKDV NMVLQSSL	HLA-DRB1*13:02		
		WNFRKDV NMVLQSSL	HLA-DRB1*01:01		
		FRKDVNM VLQSSLGN	HLA-DRB1*01:01		
		NFRKDVN MVLQSSLG	HLA-DRB3*01:01		
		FRKDVNM VLQSSLGN	HLA-DRB3*01:01		
		NFRKDVN MVLQSSLG	HLA-DRB1*01:01		
		YTYEWNFR KDVNMVL	HLA-DRB5*01:01		
		EWNFRKD VNMVLQSS	HLA-DRB1*01:01		
MSMG ALTDL	HLA-C*15:02	PFSSNFMS MGALTDL	HLA-DRB1*04:04	54.28%	100% conserved in all available serotypes
	HLA-B*15:02	FSSNFMSM GALTDLG	HLA-DRB1*04:04		
	HLA-C*06:02	SSNFMSMG ALTDLGQ	HLA-DRB1*04:04		
	HLA-C*07:01	SNFMSMG ALTDLGQN	HLA-DRB1*04:04		
	HLA-C*12:03	PFSSNFMS MGALTDL	HLA-DRB1*01:01		
	HLA-C*03:03	FSSNFMSM GALTDLG	HLA-DRB1*01:01		
	HLA-C*14:02	SSNFMSMG ALTDLGQ	HLA-DRB1*01:01		
		SNFMSMG	HLA-DRB1*01:01		

		ALTDLGQN			
		NFMSMGA LTDLGQNL	HLA-DRB1*04:04		
		FMSMGALT DLGQNLL	HLA-DRB1*04:04		
		FMSMGALT DLGQNLL	HLA-DRB1*01:01		
		NFMSMGA LTDLGQNL	HLA-DRB1*01:01		
		SSNFMSMG ALTDLGQ	HLA-DRB1*04:01		
		FSSNFMSM GALTDLG	HLA-DRB1*04:01		
		PFSSNFMS MGALTDL	HLA-DRB1*04:01		
		SNFMSMG ALTDLGQN	HLA-DRB1*04:01		
		PFSSNFMS MGALTDL	HLA-DRB5*01:01		
		FSSNFMSM GALTDLG	HLA-DRB5*01:01		
		SSNFMSMG ALTDLGQ	HLA-DRB5*01:01		
		SNFMSMG ALTDLGQN	HLA-DRB5*01:01		
		PFSSNFMS MGALTDL	HLA-DRB1*04:05		
		SSNFMSMG ALTDLGQ	HLA-DRB1*04:05		
		FSSNFMSM GALTDLG	HLA-DRB1*04:05		
		SNFMSMG	HLA-DRB1*04:05		

		ALTDLGQN			

Table S3.T-cell epitope prediction of Hexon protein D-8 of Human Adenovirus.

Peptide	**Interacting MHC-I Alleles**	**Overlapping 15 mer peptides**	**Interacting MHC-II Alleles**	**Population Coverage**	**Epitope Conservancy Table**
FLYSN VALY	HLA-C*03:03	NLWKSFLY SNVALYL	HLA-DRB1*01:01	38.09%	100% conserved in serotypes of group D
	HLA-A*29:02	LWKSFLYS NVALYLP	HLA-DRB1*01:01		73.33% conserved in all available serotypes
	HLA-B*15:02	KSFLYSNV ALYLPDS	HLA-DRB1*01:01		
	HLA-C*12:03	WKSFLYSN VALYLPD	HLA-DRB1*01:01		
	HLA-C*14:02	SFLYSNVA LYLPDSF	HLA-DRB1*01:01		
	HLA-B*15:01	NLWKSFLY SNVALYL	HLA-DRB1*07:01		
	HLA-B*35:01	NLWKSFLY SNVALYL	HLA-DRB1*04:04		
		NLWKSFLY SNVALYL	HLA-DRB1*04:05		
		LWKSFLYS NVALYLP	HLA-DRB1*07:01		
		ANLWKSFL YSNVALY	HLA-DRB1*01:01		
		ANLWKSFL YSNVALY	HLA-DRB1*04:04		
		KSFLYSNV ALYLPDS	HLA-DRB1*07:01		
		NLWKSFLY SNVALYL	HLA-DRB1*13:02		
		WKSFLYSN VALYLPD	HLA-DRB1*07:01		
		ANLWKSFL YSNVALY	HLA-DRB1*04:05		
		KSFLYSNV ALYLPDS	HLA-DRB1*13:02		
		LWKSFLYS NVALYLP	HLA-DRB1*13:02		

		WKSFLYSN VALYLPD	HLA-DRB1*13:02		
		LWKSFLYS NVALYLP	HLA-DRB1*04:04		
		WKSFLYSN VALYLPD	HLA-DRB1*04:04		
		SFLYSNVA LYLPDSF	HLA-DRB1*13:02		
		NLWKSFLY SNVALYL	HLA-DPA1*02:01/DPB1*01:01		
		LWKSFLYS NVALYLP	HLA-DPA1*02:01/DPB1*01:01		
		KSFLYSNV ALYLPDS	HLA-DRB1*04:04		
		WKSFLYSN VALYLPD	HLA-DPA1*02:01/DPB1*01:01		
		KSFLYSNV ALYLPDS	HLA-DPA1*02:01/DPB1*01:01		
		NLWKSFLY SNVALYL	HLA-DRB1*03:01		
		KSFLYSNV ALYLPDS	HLA-DRB1*03:01		
		NLWKSFLY SNVALYL	HLA-DPA1*03:01/DPB1*04:02		
		SFLYSNVA LYLPDSF	HLA-DRB1*07:01		
		WKSFLYSN VALYLPD	HLA-DRB1*03:01		
		LWKSFLYS NVALYLP	HLA-DRB1*03:01		
		KSFLYSNV ALYLPDS	HLA-DPA1*03:01/DPB1*04:02		
		ANLWKSFL YSNVALY	HLA-DRB1*03:01		
FVYSG TIPY	HLA-B*35:01	FDPYFVYS GTIPYLD	HLA-DRB1*07:01	38.09%	16.67% conserved in serotypes of group D
	HLA-C*03:03	GFDPYFVY SGTIPYL	HLA-DRB1*07:01		
	HLA-A*29:02	SGFDPYFV YSGTIPY	HLA-DRB1*07:01		

	HLA-C*14:02	DPYFVYSG TIPYLDG	HLA-DRB1*07:01		
	HLA-C*12:03	PYFVYSGTI PYLDGT	HLA-DRB1*07:01		
	HLA-B*15:02				
	HLA-B*15:01				
MSMG ALTDL	HLA-C*15:02	PFSSNFMS MGALTDL	HLA-DRB1*04:04	54.28%	100% conserved in all available serotypes
	HLA-B*15:02	FSSNFMSM GALTDLG	HLA-DRB1*04:04		
	HLA-C*06:02	SSNFMSMG ALTDLGQ	HLA-DRB1*04:04		
	HLA-C*07:01	SNFMSMG ALTDLGQN	HLA-DRB1*04:04		
	HLA-C*12:03	PFSSNFMS MGALTDL	HLA-DRB1*01:01		
	HLA-C*03:03	FSSNFMSM GALTDLG	HLA-DRB1*01:01		
	HLA-C*14:02	SSNFMSMG ALTDLGQ	HLA-DRB1*01:01		
		SNFMSMG ALTDLGQN	HLA-DRB1*01:01		
		NFMSMGA LTDLGQN M	HLA-DRB1*04:04		
		FMSMGALT DLGQNML	HLA-DRB1*04:04		
		FMSMGALT DLGQNML	HLA-DRB1*01:01		
		NFMSMGA LTDLGQN M	HLA-DRB1*01:01		
		SSNFMSMG ALTDLGQ	HLA-DRB1*04:01		
		FSSNFMSM GALTDLG	HLA-DRB1*04:01		
		PFSSNFMS MGALTDL	HLA-DRB1*04:01		
		SNFMSMG ALTDLGQN	HLA-DRB1*04:01		
		PFSSNFMS MGALTDL	HLA-DRB5*01:01		
		FSSNFMSM GALTDLG	HLA-DRB5*01:01		
		SSNFMSMG ALTDLGQ	HLA-DRB5*01:01		
		SNFMSMG	HLA-DRB5*01:01		

		ALTDLGQN			
		PFSSNFMS MGALTDL	HLA-DRB1*04:05		
		SSNFMSMG ALTDLGQ	HLA-DRB1*04:05		
		FSSNFMSM GALTDLG	HLA-DRB1*04:05		
		SNFMSMG ALTDLGQN	HLA-DRB1*04:05		

Table S4.T-cell epitope prediction of Penton Base Protein B-3 of Human Adenovirus.

Peptide	**Interacting MHC-I Alleles**	**Overlapping 15 mer peptides**	**Interacting MHC-II Alleles**	**Population Coverage**	**Epitope Conservancy Table**
YQND HSNFL	HLA-A*02:06	SLNYQNDH SNFLTTV	HLA-DRB3*01:01	42.75%	100% conserved in all available serotypes
	HLA-C*06:02	DIASLNYQ NDHSNFL	HLA-DRB3*01:01		
	HLA-B*15:02	IASLNYQN DHSNFLT	HLA-DRB3*01:01		
	HLA-B*39:01	ASLNYQND HSNFLTT	HLA-DRB3*01:01		
	HLA-C*03:03	LNYQNDHS NFLTTVV	HLA-DRB3*01:01		
	HLA-C*05:01	SLNYQNDH SNFLTTV	HLA-DRB1*07:01		
	HLA-C*12:03	LNYQNDHS NFLTTVV	HLA-DRB1*07:01		
		ASLNYQND HSNFLTT	HLA-DRB1*07:01		
		IASLNYQN DHSNFLT	HLA-DRB1*07:01		
		DIASLNYQ NDHSNFL	HLA-DRB1*07:01		
YTYEA FHPD	HLA-C*14:02	GVYTYEAF HPDIVLL	HLA-DRB1*09:01	28.00%	100% conserved in serotypes of group B
	HLA-C*03:03	VYTYEAFH PDIVLLP	HLA-DRB1*09:01		90.32% conserved in all available serotypes

	HLA-A*02:06	YTYEAFHP DIVLLPG	HLA-DRB1*09:01		
	HLA-C*12:03	PGVYTYEA FHPDIVL	HLA-DRB1*09:01		
	HLA-C*15:02	MPGVYTYE AFHPDIV	HLA-DRB1*09:01		
	HLA-A*68:02				
NRFPE NQIL	HLA-C*06:02	LTHVFNRF PENQILI	HLA-DRB1*09:01	64.00%	100% conserved in serotypes of group B
	HLA-C*07:01	THVFNRFP ENQILIR	HLA-DRB1*09:01		
	HLA-C*07:02	HVFNRFPE NQILIRP	HLA-DRB1*09:01		
	HLA-C*12:03	SLTHVFNR FPENQIL	HLA-DRB1*09:01		
	HLA-B*39:01				
	HLA-C*03:03				

Table S5. T-cell epitope prediction of Penton Base Protein C-5 of Human Adenovirus.

Peptide	**Interacting MHC-I Alleles**	**Overlapping 15 mer peptides**	**Interacting MHC-II Alleles**	**Population Coverage**	**Epitope Conservancy Table**
YQND HSNFL	HLA-A*02:06	SLNYQNDH SNFLTTV	HLA-DRB3*01:01	42.75%	100% conserved in all available serotypes
	HLA-C*06:02	DVASLNYQ NDHSNFL	HLA-DRB3*01:01		
	HLA-B*15:02	VASLNYQN DHSNFLT	HLA-DRB3*01:01		
	HLA-B*39:01	ASLNYQND HSNFLTT	HLA-DRB3*01:01		
	HLA-C*03:03	LNYQNDHS NFLTTVI	HLA-DRB3*01:01		
	HLA-C*05:01	SLNYQNDH SNFLTTV	HLA-DRB1*07:01		

	HLA-C*12:03	LNYQNDHS NFLTTVI	HLA-DRB1*07:01		
		ASLNYQND HSNFLTT	HLA-DRB1*07:01		
		VASLNYQN DHSNFLT	HLA-DRB1*07:01		
		DVASLNYQ NDHSNFL	HLA-DRB1*07:01		
STFTQ YRSW	HLA-C*12:03	ISNDSTFTQ YRSWYL	HLA-DRB1*07:01	29.88%	100% conserved in serotypes of group C
	HLA-B*57:01	SNDSTFTQ YRSWYLA	HLA-DRB1*07:01		
	HLA-A*32:01	NDSTFTQY RSWYLAY	HLA-DRB1*07:01		
	HLA-B*58:01	DSTFTQYR SWYLAYN	HLA-DRB1*07:01		
	HLA-C*03:03	STFTQYRS WYLAYNY	HLA-DRB1*07:01		
	HLA-C*14:02	SNDSTFTQ YRSWYLA	HLA-DRB1*15:01		
		NDSTFTQY RSWYLAY	HLA-DRB1*15:01		
		DSTFTQYR SWYLAYN	HLA-DRB1*15:01		
		STFTQYRS WYLAYNY	HLA-DRB1*15:01		
		ISNDSTFTQ YRSWYL	HLA-DRB1*15:01		
		STFTQYRS WYLAYNY	HLA-DRB1*01:01		
		SNDSTFTQ YRSWYLA	HLA-DRB1*01:01		
		DSTFTQYR SWYLAYN	HLA-DRB1*01:01		
		NDSTFTQY RSWYLAY	HLA-DRB1*01:01		
		ISNDSTFTQ YRSWYL	HLA-DRB1*01:01		
RSWYL AYNY	HLA-A*32:01	STFTQYRS WYLAYNY	HLA-DRB1*07:01	26.28%	100% conserved in serotypes of group C
	HLA-A*30:02	STFTQYRS WYLAYNY	HLA-DRB1*15:01		
	HLA-A*29:02	TFTQYRSW YLAYNYG	HLA-DRB1*15:01		
	HLA-C*12:03	FTQYRSWY LAYNYGD	HLA-DRB1*15:01		
	HLA-C*15:02	TFTQYRSW	HLA-DRB1*07:01		

		YLAYNYG			
	HLA-B*58:01	FTQYRSWY LAYNYGD	HLA-DRB1*07:01		
		STFTQYRS WYLAYNY	HLA-DRB1*01:01		
NSNAA AAAM	HLA-C*03:03	AEENSNAA AAAMQPV	HLA-DQA1*05:01/DQB 1*03:01	51.04%	100% conserved in serotypes of group C
	HLA-C*15:02	EENSNAAA AAMQPVE	HLA-DQA1*05:01/DQB 1*03:01		
	HLA-C*05:01	GAEENSNA AAAAMQP	HLA-DQA1*05:01/DQB 1*03:01		
	HLA-B*35:01	SGAEENSN AAAAAMQ	HLA-DQA1*05:01/DQB 1*03:01		
	HLA-C*12:03	ENSNAAAA AMQPVED	HLA-DQA1*05:01/DQB 1*03:01		
	HLA-B*15:02	AEENSNAA AAAMQPV	HLA-DQA1*01:02/DQB 1*06:02		
	HLA-C*07:01	EENSNAAA AAMQPVE	HLA-DQA1*01:02/DQB 1*06:02		
		NSNAAAA AMQPVED M	HLA-DQA1*05:01/DQB 1*03:01		
		GAEENSNA AAAAMQP	HLA-DQA1*01:02/DQB 1*06:02		
		GSGAEENS NAAAAAM	HLA-DQA1*05:01/DQB 1*03:01		
		ENSNAAAA AMQPVED	HLA-DQA1*01:02/DQB 1*06:02		
		NSNAAAA AMQPVED M	HLA-DQA1*01:02/DQB 1*06:02		
		SGAEENSN AAAAAMQ	HLA-DQA1*01:02/DQB 1*06:02		