



Reverse vaccinology approach to design a novel multi-epitope subunit vaccine against avian influenza A (H7N9) virus



Mahmudul Hasan^{a,b,*}, Progga Paromita Ghosh^a, Kazi Faizul Azim^a, Shamsunnahar Mukta^{a,c}, Ruhshan Ahmed Abir^d, Jannatun Nahar^d, Mohammad Mehedi Hasan Khan^{a,e}

^a Faculty of Biotechnology and Genetic Engineering, Sylhet Agricultural University, Sylhet, 3100, Bangladesh

^b Department of Pharmaceuticals and Industrial Biotechnology, Sylhet Agricultural University, Sylhet, 3100, Bangladesh

^c Department of Plant and Environmental Biotechnology, Sylhet Agricultural University, Sylhet, 3100, Bangladesh

^d Department of Genetic Engineering and Biotechnology, Shahjalal University of Science and Technology, Sylhet, 3114, Bangladesh

^e Department of Biochemistry and Chemistry, Sylhet Agricultural University, Sylhet, 3100, Bangladesh

ARTICLE INFO

Keywords:

Reverse vaccinology
Epitope based vaccine
Avian influenza A(H7N9)
B-cell immunity
T-cell immunity

ABSTRACT

H7N9, a novel strain of avian origin influenza was the first recorded incidence where a human was transited by a N9 type influenza virus. Effective vaccination against influenza A (H7N9) is a major concern, since it has emerged as a life threatening viral pathogen. Here, an *in silico* reverse vaccinology strategy was adopted to design a unique chimeric subunit vaccine against avian influenza A (H7N9). Induction of humoral and cell-mediated immunity is the prime concerned characteristics for a peptide vaccine candidate, hence both T cell and B cell immunity of viral proteins were screened. Antigenicity testing, transmembrane topology screening, allergenicity and toxicity assessment, population coverage analysis and molecular docking approach were adopted to generate the most antigenic epitopes of avian influenza A (H7N9) proteome. Further, a novel subunit vaccine was designed by the combination of highly immunogenic epitopes along with suitable adjuvant and linkers. Physicochemical properties and secondary structure of the designed vaccine were assessed to ensure its thermostability, hydrophilicity, theoretical PI and structural behavior. Homology modeling, refinement and validation of the designed vaccine allowed to construct a three dimensional structure of the predicted vaccine, further employed to molecular docking analysis with different MHC molecules and human immune TLR8 receptor present on lymphocyte cells. Moreover, disulfide engineering was employed to lessen the high mobility region of the designed vaccine in order to extend its stability. Furthermore, we investigated the molecular dynamic simulation of the modeled subunit vaccine and TLR8 complexed molecule to strengthen our prediction. Finally, the suggested vaccine was reverse transcribed and adapted for *E. coli* strain K12 prior to insertion within pET28a(+) vector for checking translational potency and microbial expression.

1. Introduction

Influenza is a highly infectious respiratory disease and major health concerned issue throughout the world. On an average, it causes 3 to 5 million cases of severe illness and up to 650,000 respiratory deaths each year [1]. Majority of influenza infections occur due to H1, H2 and H3 subtypes belonging to the Orthomyxoviridae family [2,3]. Influenza A(H7N9), an emerging virus of novel avian origin was first identified to cause human infection in China in 2013 [4]. Although, the strain has been classified as “low pathogenic avian influenza” (LPAI) by the Centers for Disease Control and Prevention (CDC) due to not being exceptionally lethal to infected poultry, it does not mean that the

lethality is also decreased in humans. On an average, one third of all instances of infections have ended in death [5]. Epidemiology of the outbreaks indicated live bird markets as the main origin of human infections (75% of the occurrences resulting from direct contact with poultry). Literature studies suggest that the virus was mutated from its actual form in the wild birds to a strain capable of infecting humans [6].

The influenza A H7N9 strain was reported to cause fever and lower respiratory tract infections of the lung epithelium, resulting in acute respiratory distress syndrome and even death in extreme cases [4]. The average time for onset of lung failure was between 3 and 14 days and those patients that did die, died 1–3 weeks after the initiation of symptoms. Clinical investigations implied that the strain tends to cause

* Corresponding author. Department of Pharmaceuticals and Industrial Biotechnology, Faculty of Biotechnology and Genetic Engineering, Sylhet Agricultural University, Sylhet, 3100, Bangladesh.

E-mail addresses: mhasan.pib@sau.ac.bd, hasan_sust@yahoo.com (M. Hasan).

more dreadful and long lasting signs in aged patients (approximately 41.7 days) than subtype H5N1 [7]. The ability of subtype H7N9 virus to cause a pandemic among people raises major human health concern [8,9]. Occurrences of H7N9 are now accumulating at a rate which is five times faster than avian influenza A H5N1 [5]. At this pace, it will exceed the burden of avian influenza A H5N1 very soon (860 incidences and 454 deaths, as of December 2018) [10]. If the virus becomes able to transmit from human to human, it will attain prevalent proportions easily and result in pandemic outbreak.

Exploitation of vaccination as a strategy in combating wide spread infections has resulted in considerable advances in the fight against many infectious diseases including influenza, pertussis, diphtheria, varicella, hepatitis, smallpox, rotavirus, tetanus and polio. The outburst of human infection by influenza H7N9 strain has accentuated the significance of producing rapid and more efficient influenza vaccines than the currently accessible ones. Researchers have also studied the live attenuated H7N9 virus vaccine candidate and other H7 subtype virus vaccines for their potential to protect from H7N9 virus infection [11–14]. However, making these types of H7N9 influenza vaccines often has several obstacles [15]. The conventional vaccines include attenuated or killed agents which may take more than 15 years to develop while relying on adequate antigen expression from in vitro culture models. Although these types of vaccines have saved innumerable lives, it can lead to undesirable consequences too [16]. Sometimes microorganisms are difficult to cultivate and in some instances to attenuate resulting in cross contamination or adverse immune responses [17]. Again, majority of the methods so far utilized to acquire and purge the target antigen were unsuccessful, demonstrating that these approaches are not always feasible [18]. Subunit vaccines consist of only antigenic part of the pathogen with the potential to induce a protective immune response inside host while overcoming the problems caused by live attenuated vaccines such as risk of emergence of virulence property.

Reverse vaccinology or vaccinomics is a novel approach that has been used tremendously to introduce new vaccines. The strategy aims to combine immunogenomics and immunogenetics with bioinformatics for the development of novel vaccine targets [19]. This rapid *in silico* based method has acquired great acceptance with the recent development in the genome and protein sequence databases [20]. The possibility of the vaccine candidates is reinforced through different assessments. Detection of T cell epitopes and HLA (human leukocyte antigen) ligands aids in the process of vaccine discovery [21]. Vaccines may be detected as a foreign particle when introduced in the body. So, allergenicity prediction is a vital step in the development of a peptide vaccine. Hydrophilicity is also an important criterion for the prediction of a B-cell epitope. Our study was conducted to design a non-allergic, immunogenic and thermostable chimeric vaccine against avian influenza A(H7N9) virus utilizing vaccinomics approach while the wet lab researchers are anticipated to authenticate our prediction.

2. Materials and methods

In the present study, a computational method was adopted for predicting vaccine candidates against avian influenza A (H7N9) strain. The flow chart summarizing the protocol over reverse vaccinology approach for designing an epitope-based vaccine has been illustrated in Fig. 1.

2.1. Viral strain selection

National Center for Biotechnology Information (NCBI), the store house of biomedical and genomic information advances science and health in a great deal. It was used for the selection of avian influenza A (H7N9) strains and analysis of associated information including genus, family, host, transmission, disease, genome and proteome.

2.2. Protein sequence retrieval

The entire viral proteome of influenza A (H7N9) strain was retrieved by using NCBI database (<https://www.ncbi.nlm.nih.gov/protein/?term=avian+influenza+H7N9>).

2.3. Antigenic protein screening and structure analysis

Antigenicity indicates to the ability of antigens to be recognized by the immune system. The VaxiJen v2.0 server (<http://www.ddg-pharmfac.net/vaxijen/>) was used for analyzing the whole protein antigenicity and to determine the most potent antigenic proteins [22]. Surface proteins or outer membrane proteins are always preferred to target as vaccine candidates, though few studies also suggested that proteins located inside the viral particle can also be used to develop vaccine candidates to confer immunity [23–25]. Here, the study was focused on the immunogenicity of membrane proteins (outer membrane and inner membrane) of influenza A (H7N9) while considering the antigenic score above threshold value (0.4). The secondary structure parameters (solvent accessibility, transmembrane helices, globular regions and coiled coil regions) of the target proteins were predicted using the ExPASy's (<https://web.expasy.org>) secondary structure prediction tool ProtParam (<https://web.expasy.org/protparam/>) [26,27].

2.4. T-cell epitope prediction

The IEDB facilitates optimum access of investigated data characterizing antibody and T cell epitopes studied in different organisms, namely humans, non-human primates and other animal species. A tool from the Immune Epitope Database (<http://tools.iedb.org/mhci/>) was used to predict the MHC-I binding [28]. MHC-I restricted CD8⁺ cytotoxic T lymphocytes (CTLs) have potential roles in controlling virus infection. Hence, identification of T cell epitopes is critical for understanding the mechanism of T cell activation and epitope based vaccine design. The protein sequences of hemagglutinin and matrix protein1 were added to the query box at different time. All of the alleles were selected for the binding analysis.

2.5. Transmembrane topology prediction and antigenicity analysis of epitopes

TMHMM server (<http://www.cbs.dtu.dk/services/TMHMM/>) was used for the prediction of transmembrane helices in proteins. Here, the topology is given as the position of the transmembrane helices differentiated by 'i' and 'o' when the loop is on the inside and outside respectively [29]. VaxiJen v2.0 server (<http://www.ddg-pharmfac.net/vaxijen/>) was used for screening epitope antigenicity and determining the best antigenic epitopes.

2.6. Population coverage analysis

Population coverage of proposed epitopes was analyzed by the IEDB population coverage calculation tool analysis resource (<http://tools.iedb.org/population/>). HLA distribution pattern varies among various ethnic groups and geographic areas around the world. Thus, population coverage is considered a potential screening parameter when designing an effective vaccine to cover as much as possible populations.

2.7. Conservancy analysis

The epitope conservancy analysis is a crucial step in reverse vaccinology approach as it determines the degree of desired epitope distribution in the homologous protein set. To ensure the conservancy pattern, homologous sequences of the selected antigenic proteins for Influenza A (H7N9) strains were retrieved from the NCBI database by using BLASTp tool. Further, the epitope conservancy analysis tool

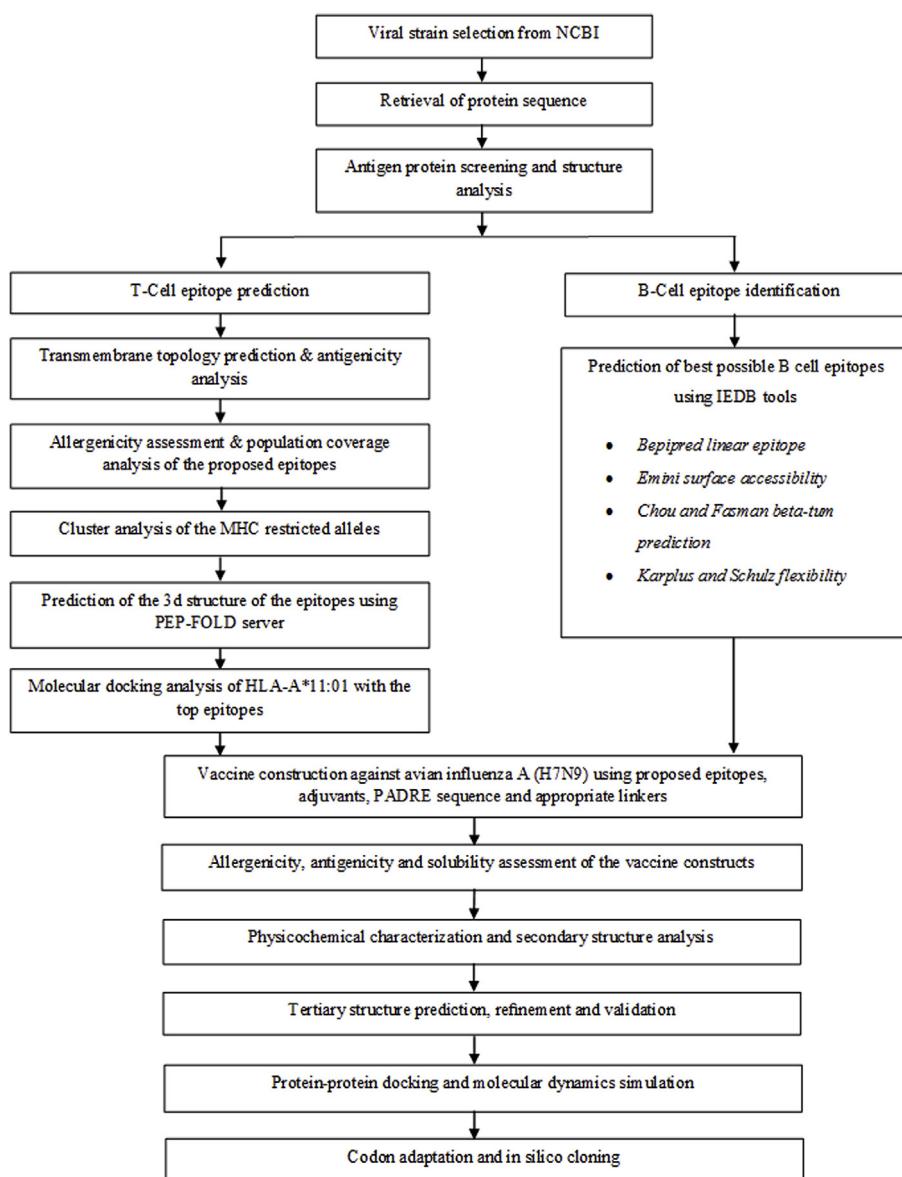


Fig. 1. Flow chart summarizing the protocols over multi-epitope subunit vaccine development against influenza A (H7N9) through reverse vaccinology approach.

(<http://tools.iedb.org/conservancy/>) at the IEDB was selected for analysis of conservancy pattern. The conservancy levels were determined by focusing on the identities of hemagglutinin and matrix protein 1.

2.8. Allergenicity assessment of predicted epitopes

Various bioinformatics tools are now frequently used for the prediction of allergens. Four servers, AllerTOP (<http://www.ddg-pharmfac.net/AllerTop/>) [30], AllergenFP (<http://www.ddg-pharmfac.net/AllergenFP/>) [31], PA³P (<http://lpa.saogabriel.unipampa.edu.br:8080/pa3p/pa3p.jsp>) [32] and Allermatch (HYPERLINK \o "http://Error! Hyperlink reference not valid. www.allermatch.org/allermatch.py/form") [33] were utilized to predict the allergenicity of our proposed epitopes for vaccine development.

2.9. Cluster analysis of the MHC restricted alleles

The distinct specificity of the MHC alleles remains uncharacterized in most cases due to their vast polymorphic nature among species. Structure based clustering techniques have been proven efficient to identify

the superfamilies of MHC proteins with similar binding specificities. In this study, we used a tool from MHCcluster v2.0 [34] server that provides pictorial tree-based visualizations and highly instinctive heat-map of the functional alliance between MHC variants.

2.10. Designing three dimensional (3D) epitope structure

After different bioinformatics analysis, top ranked epitopes were subjected for the docking study. PEP-FOLD is a *de novo* approach aimed at predicting peptide structures from amino acid sequences [35]. It offers new possibilities to refine pre-existing models and also comes with the possibility to synthesize candidate conformations of peptide-protein complexes, by folding peptides on a user specified patch of a protein [36,37].

2.11. Molecular docking analysis

MGLTools, designed at the Molecular Graphics Laboratory (MGL) of The Scripps Research Institute which usually visualize and analyze the molecular structures of biological compounds [38]. It includes AutoDock which is an automated docking software developed to predict how

small molecules, for instance, substrates, drug or vaccine candidate bind to a receptor of known 3D structure [39]. AutoDockTools or ADT is the free GUI for AutoDock developed by the same laboratory that developed AutoDock [40]. All the analyses were done at 1.00- Å spacing. The exhaustiveness parameter was kept at 8.00 while the number of outputs was set at 10. These parameters were performed in AutoDOCK tool. The docking was conducted using AutoDOCKVina program based on parameters mentioned above. All the output PDBQT files were converted in PDB format using OpenBabel (version 2.3.1). The best output was selected on the basis of lower binding energy. The docking interaction was visualized with the PyMOL molecular graphics system, version 1.5.0.4 (<https://www.pymol.org/>).

2.12. B-cell epitope identification

To find the potential antigen that would interact with B lymphocytes and initiate an immune response is the prime concerned of the B cell epitope prediction. From the experimental study, it was found that the flexibility of the peptide is linked to its antigenicity. Proper surface accessibility is also prerequisite for being a potential B cell epitope. B cell epitope prediction Tools from IEDB can be used to identify the B cell antigenicity depending on few algorithms including the Kolaskar and Tongaonkar antigenicity scale [41], Emini surface accessibility prediction [42], Karplus and Schulz flexibility prediction [43], Beppied linear epitope prediction analysis [44], Chou and Fasman beta turn prediction [45] and Parker hydrophilicity prediction [46].

2.13. Vaccine construction

Subunit vaccines comprise antigenic components of a pathogenic organism to induce an immunogenic reaction in the host. In this study, the predicted T cell and B cell epitopes were combined in a sequential manner to construct the final vaccine proteins. A total six vaccine sequences were constructed each associated with different adjuvants including beta defensin (a 45 mer peptide), L7/L12 ribosomal protein and HABA protein (*M. tuberculosis*, accession number: AGV15514.1). Interactions of adjuvants with toll like receptors (TLRs) polarize CTL responses and stimulate robust immune-reaction [47]. Beta defensin adjuvant can act as an agonist to TLR1, TLR2 and TLR4 receptor while L7/L12 ribosomal protein and HBHA protein are agonists to only TLR4 [48]. PADRE sequence was also incorporated along with the adjuvant peptides to overcome the problems caused by highly polymorphic HLA alleles [49].

The first three vaccine constructs (i.e. V1, V2, V3) were designed where each epitope was included separately divided by flexible linkers. The vaccine proteins started with an adjuvant followed by top CTL epitopes for both matrix protein 1 and hemagglutinin and then by top BCL (B-cell Lymphoma) epitopes in the similar fashion. EAAAK linkers were used to join the adjuvant and CTL epitopes [48,50]. Similarly, GGGS and KK linkers were used to conjugate the CTL and BCL epitopes respectively. Thus utilized linkers ensured effective separation of individual epitopes in vivo [51,52]. Although AAY linkers enhance epitope presentation and remove junctional epitopes [53,54], Yang et al. proved the superiority of GGGS linkers than AAY for epitope-based vaccines [55]. In addition, three more constructs (i.e. V4, V5, V6) were designed using larger cassettes that contain multiple epitopes (reference). Epitope cluster analysis tool from IEDB (ref) was used to identify the epitope clusters with overlapping peptides for both proteins using the top 12 CTL epitopes and top 10 BCL epitopes as input. The identified clusters and singletons were further utilized to design construct V4, V5 and V6 using the same adjuvants described above.

2.14. Allergenicity, antigenicity and solubility prediction of different vaccine constructs

AllerTOP v.2.0 [30] sever was used to predict the non-allergic

behavior of the vaccine constructs. It developed an algorithm by taking an account of the auto cross covariance transformation of proteins into uniform vectors of similar length. The server predicts with an accuracy ranging from 70% to 89% depending on species. We further used VaxiJen v2.0 server [22] to determine the probable antigenicity of the vaccine proteins for suggesting the superior vaccine candidate. The server evaluates the immunogenic potential of the given proteins through an alignment independent algorithm. Proso II server was used to predict the solubility of the vaccine proteins upon over expression in *E. coli* [56]. The server employs a sequence based approach by exploiting subtle differences among soluble and insoluble proteins from a range of database and predicts with an overall accuracy of 71%. The solubility of the proposed vaccines was further analyzed via Protein-sol software [57,58]. The program calculated the surface distribution of charge, hydrophobicity and the stability at 91 different combinations of pH and ionic strength.

2.15. Physicochemical characterization and secondary structure analysis of vaccine protein

ProtParam is a tool provided by Expasy server (<http://expasy.org/cgi-bin/protpraram>) [26] which was used to functionally characterize our vaccine constructs. Various physicochemical properties of the vaccine proteins were analyzed including molecular weight, isoelectric pH, aliphatic index, GRAVY values, hydropathicity, instability index and estimated half-life. By comparing the pK values of different amino acids, the server computes these parameters of a given protein sequence. The volume occupied by the aliphatic side chains is known as the aliphatic index of protein. Grand average of hydropathicity was calculated by the sum of hydropathicity obtained for all of the amino acid residues divided by the total number of amino acid residues present in the protein. The PSIPRED v3.3 [59] and NetTurnP 1.0 program [60,61] were used to predict the alpha helix, beta sheet and coil structure of the vaccine protein.

2.16. Vaccine tertiary structure prediction, refinement and validation

The RaptorX server [62,63] predicted the 3D structure of our modeled vaccine based on the degree of similarity between target protein and available template structure from PDB. To improve the accuracy of the predicted 3D modeled structure, refinement was performed using ModRefiner [64] followed by FG-MD refinement server [65]. ModRefiner draws the initial model closer to its native state in terms of hydrogen bonds, side-chain positioning as well as backbone topology and thereby resulting in significant improvement in physical quality of the local structure. FG-MD is another molecular dynamics based algorithm for protein structure refinement at the atomic level. The refined protein structure was further validated through Ramachandran plot assessment at RAMPAGE [66,67].

2.17. Vaccine protein disulfide engineering

The stability of the model vaccine was enhanced by adopting a logical approach to disulfide bond formation. Disulfide bonds strengthen the geometric conformation of proteins and are provide significant stability. DbD2, an online tool was used to design such bonds for the predicted vaccine protein [68]. The server can detect the pair of residues having the capacity to form a disulfide bond if the individual amino acids mutated to cysteine. It provides a list of residue pairs with proper geometry and ability to form disulfide bonds.

2.18. Protein-protein docking

Molecular docking is an *in silico* approach which aims to evaluate the binding affinity between a receptor molecule and a ligand [48]. Inflammation mediated by ssRNA virus are involved with immune

Table 1
ProtParam analysis of the retrieved viral proteins.

Viral Proteins	Accession ID	Molecular Weight	Instability Index	Aliphatic Index	Theoretical pI	No. of Amino acids	Extinction Co-Efficient	Gravy
Polymerase PB2	AGL44433.1	85905.34	45.04	86.77	9.56	759	79090	-0.32
Polymerase PB1	AGL44434.1	86383.13	40.17	72.27	9.31	757	86750	-0.50
Polymerase PA	AGL44436.1	82734.49	49.83	77.51	5.43	716	94310	-0.50
Hemagglutinin	AGL44438.1	62108.36	32.64	80.11	6.25	560	67350	-0.30
Nucleocapsid protein	AGL44439.1	56316.75	43.33	69.20	9.48	498	52745	-0.58
Neuraminidase	AGL44440.1	51808.08	43.59	69.61	6.74	465	10805	-0.45
Matrix protein 1	AGL44441.1	27716.96	33.98	84.09	9.15	252	13075	-0.24
Nonstructural protein 1	AGL44443.1	24718.21	59.60	88.99	5.54	217	23615	-0.36

receptors present over the immune cells, mainly by TLR-7 and TLR-8 [69,70]. An approach for protein-protein docking was made to determine the binding affinity of our designed subunit vaccine with the TLR-8 immune receptor by using ClusPro 2.0 [71], hdoc [72,73] and PatchDock server [74]. The 3D structure of human TLR-8 immune receptor was retrieved from RCSB protein data bank. PDB files of both TLR-8 receptor and vaccine protein was uploaded to the above mentioned servers to obtain the desirable complexes in terms of better electrostatic interaction and free binding energy. The given solutions from PatchDock server were again subjected to the FireDock server for refining the complexes.

2.19. Molecular dynamics simulation

Molecular dynamics study was performed to demonstrate the stability of protein-protein complex for further strengthening our prediction. Essential dynamics can be compared to the normal modes of proteins to determine their stability [75,76]. It is a powerful tool and alternative to costly atomistic simulation [77,78]. iMODS is an online server that explained the collective motion of proteins via analysis of normal modes (NMA) in internal coordinates [79]. The server was used to investigate the structural dynamics of protein complex due to its much faster and effective assessments than the typical molecular dynamics (MD) simulations tools [80,81]. It predicted the direction and magnitude of the immanent motions of the complex in terms of deformability, B-factors, eigenvalues and covariance. For a given molecule, the deformability of the main chain depends on the ability to deform at each of its residues. The eigenvalue associated to each normal mode represents the motion stiffness. This value is directly related to the energy required to deform the structure. The lower the eigenvalue, the easier the deformation is [82].

2.20. Codon adaptation and in silico cloning

Codon adaptation tools are commonly used to adapt the codon usage to the well characterized prokaryotic organisms for accelerating the expression rate in them. We selected *E. coli* strain K12 as host for the cloning purpose of our vaccine construct V1. Due to the dissimilarity between the codon usage of human and *E. coli* strain, this approach was adopted with an objective of achieving a higher expression rate of the vaccine protein V1 in the selected host. We avoided rho independent transcription termination, prokaryote ribosome binding site and cleavage site of restriction enzyme BgIII and ApaI while using (JCAT) server [83]. For confirming the restriction cloning, the optimized sequence of vaccine protein V1 was reversed followed by conjugation with BamHI and ApaI restriction site at the N-terminal and C-terminal sites respectively. The SnapGene [48,50] restriction cloning module was used to insert the adapted sequence into pET28a(+) vector between the BamHI (198) and ApaI(1334).

3. Results

3.1. Protein sequence retrieval

A total of 4 influenza strains (i.e. H1N1, H3N2, H5N1, H7N9) were analyzed and strain H7N9 was selected due to its severity and burden in the recent years. The whole viral proteome of avian influenza A (H7N9) was extracted from NCBI database. All the sequences were from Shanghai, China. Among 8 viral proteins, we selected two structural proteins, hemagglutinin (NCBI Accession ID: AGL44438.1) and matrix protein 1 (NCBI Accession ID: AGL44441.1) for further analysis.

3.2. Antigenic protein prediction and structure analysis

The antigenicity scoring (Supplementary Table 1) and literature study revealed that three membrane proteins, i.e., hemagglutinin, neuraminidase and matrix protein 1 could be prioritized. Hemagglutinin, an outer surface protein was selected for its highest VaxiJen score (0.5294) while matrix protein 1, an inner membrane protein was selected due to its good antibody response in the previous studies [84,85]. The total prediction score was 0.4517 for matrix protein 1 provided via VaxiJen server which was well above the threshold value (0.40), revealing its immunogenic potential to stimulate protective response in the host. Various physiochemical properties of the proteins were also analyzed by ProtParam tools as shown in Table 1.

3.3. T-cell epitope prediction

The results were analyzed according to epitope-HLA cell binding. Numerous immunogenic epitopes from hemagglutinin and matrix protein 1 were identified to be T cell epitopes that can bind a large number of HLA-A and HLA-B alleles using MHC-I binding prediction tool of IEDB.

3.4. Transmembrane topology prediction and antigenicity analysis

Top 10 epitopes of both matrix protein 1 and hemagglutinin, binding the maximum number of HLA alleles were selected as putative T cell epitope candidates based on their transmembrane topology screening by TMHMM and antigenic scoring by Vaxijen (Table 2). Here, highest Vaxijen score was found for GALGLVCAT (1.6997) of matrix protein 1 and RIDFHWLML (2.807) of hemagglutinin. Epitope candidates with a positive score of immunogenicity showed potential to elicit strong T cell response.

3.5. Population coverage analysis

The IEDB population coverage calculation tool analysis resource was used for analyzing the population coverage pattern of proposed epitopes found from previous step. In the present study, all the indicated alleles in supplementary data were found to be optimum binders of the predicted epitopes and used to demonstrate the population coverage for both matrix protein 1 and hemagglutinin (Fig. 2).

Table 2

Predicted T-cell epitopes of Matrix protein 1 and Hemagglutinin.

Matrix Protein 1				Hemagglutinin					
Epitopes	Start	End	No. of HLAs Binding with Epitope	Antigenic Score (AS)	Epitopes	Start	End	No. of HLAs Binding with Epitope	Antigenic Score (AS)
GALGLVCAT	2	10	81	1.6997	RIDFHWML	28	36	81	2.807
ILGFVFVFTLT	59	68	81	1.5511	SGRIDFHWL	26	34	81	2.6637
QRLEDVFAG	25	34	81	1.5221	LSGRIDFHW	25	33	81	2.5448
GAKEVALSY	41	49	81	1.4158	IDFHWMLN	29	37	81	2.3793
VFAGKNADL	31	39	81	1.4046	GRIDFHWLM	27	35	81	2.2043
DVFAGKNAD	30	38	81	1.3868	NGLSGRIDF	23	31	81	1.8385
LEDVFAGKGN	28	36	81	1.2732	WFSFGASC	37	45	81	1.8337
GKNADLEAL	34	42	81	1.2324	FLRGKSMGI	58	66	81	1.7331
FAGKNADLE	32	40	81	1.1915	DFHWMLNP	30	38	81	1.7075
ILSPLTKGI	51	59	81	1.1641	GLSGRIDFH	24	32	81	1.6021

3.6. Conservancy analysis

Conservancy pattern ensures a broad spectrum vaccine efficacy for any epitope candidates. Here, after getting positive result from population coverage, putative epitopes generated from hemagglutinin and matrix protein 1 were allowed for conservancy analysis. Two sets of homologous protein sequences (100 strains) of matrix 1 ([Supplementary Table 2](#)) and hemagglutinin ([Supplementary Table 3](#)) were retrieved for the input of IEDB conservancy analysis server. Putative epitopes from matrix 1 were found highly conserved with 81–100% conservancy level, where epitopes from hemagglutinin showed insignificant conservancy level ([Table 3](#)).

3.7. Allergenicity assessment of T-Cell epitopes

Based on the allergenicity assessment by 4 servers (AllerTOP, AllergenFP, PA³P, Allermatch), 7 epitopes of Matrix Protein 1 and 6 epitopes of hemagglutinin were found to be non-allergen for human ([Table 3](#)).

3.8. Cluster analysis of the MHC restricted alleles

MHCcluster v2.0 generated a clustering of 27 class I HLA molecules which were identified to interact with our predicted epitopes. The output was generated through conventional phylogenetic method on the basis of sequence data available for different HLA-A and HLA-B alleles. [Fig. 3](#) is illustrating the function based clustering of HLA alleles (heat map) with red zones indicating strong correlation and yellow zone showing weaker interaction.

3.9. Molecular docking analysis and HLA allele interaction

For the docking analysis, the six epitopes from hemagglutinin and six epitopes from matrix protein 1 were subjected to PEP-FOLD3 web-based server for 3D structure conversion to analyze the interactions with HLA molecule. This server modeled five 3D structures of each of the proposed epitope and the best one was identified for further docking analysis. MGLTools was used for performing molecular docking study in order to investigate the binding pattern of HLA molecules and our predicted epitope. The HLA-A*11:01 was selected for docking on the basis of the available Protein Data Bank (PDB) structure deposited in the database which interacted with our proposed epitopes. The Protein Data Bank structure of 4UQ2 (Epstein-Barr virus complexed with human HLA-A*11:01) was retrieved from the Research Collaboratory for Structural Bioinformatics (RCSB) protein database. All of these epitopes were allowed for docking study with HLA-A*11:01 using AutoDock and analyzed their binding energies ([Table 4](#)).

The result showed that 'DFHWMLNP' epitope of hemagglutinin bound in the groove of the HLA-A*11:01 with an energy of -8.7 kcal/mol ([Fig. 4:A](#)). The demonstrated energy was -6.5 kcal/mol for epitope 'DVFAGKNAD' of matrix protein 1 ([Fig. 4:B](#)). The docking interaction was visualized with the PyMOL molecular graphics system, version 1.5.0.4.

3.10. B-cell epitope identification

In this study, B-cell epitopes of both matrix protein 1 and hemagglutinin were generated using four different algorithms. Bepipred prediction method showed that the peptide sequences from 80 to 90

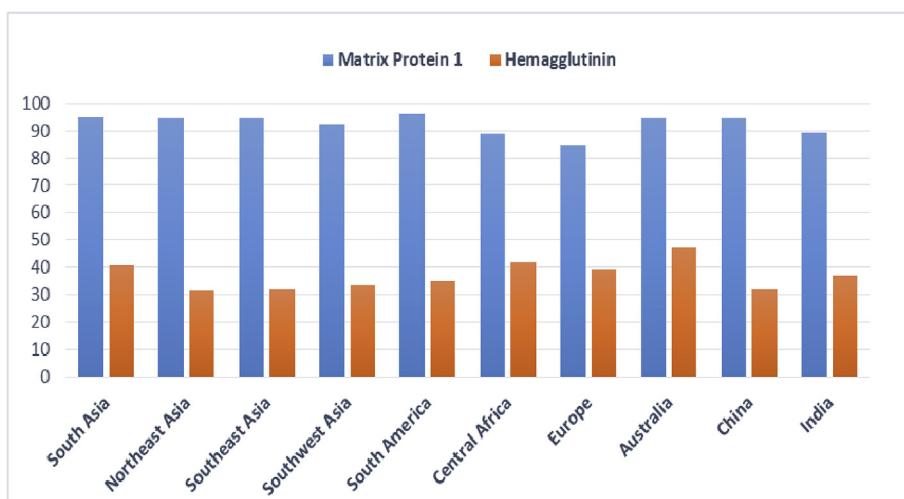


Fig. 2. Population coverage analysis of matrix protein 1 and hemagglutinin.

Table 3

Conservancy level and allergenicity pattern of putative T cell epitopes generated from matrix protein 1 and hemagglutinin.

Matrix Protein 1					Hemagglutinin				
Epitope	Start	End	Conservancy Level	Allergenicity	Epitope	Start	End	Conservancy Level	Allergenicity
GALGLVCAT	2	10	97.85%	Non Allergen	RIDFHWLML	28	36	98.77%	Non Allergen
ILGFVFTLT	59	68	100%	Non Allergen	SGRIDFWL	26	34	98.77%	Allergen
QRLEDVFAG	25	34	97.85%	Non Allergen	LSGRIDFWH	25	33	95.06%	Non Allergen
GAKEVALSY	41	49	93.55%	Allergen	IDFHWLMLN	29	37	98.77%	Allergen
VFAGKNADL	31	39	83.87	Non Allergen	GRIDFHWM	27	35	98.77%	Allergen
DVFAGKNAD	30	38	83.87	Non Allergen	NGLSGRIDF	23	31	95.06%	Non Allergen
LEDVFAGKN	28	36	96.77	Allergen	WFSFGASC	37	45	100%	Non Allergen
GKNADLEAL	34	42	81.72	Non Allergen	FLRGKSMGI	58	66	98.77%	Non Allergen
FAGKNADLE	32	40	83.87	Allergen	DFHWLMLNP	30	38	98.77%	Allergen
ILSPLTKGI	51	59	97.85	Non Allergen	GLSGRIDFH	24	32	95.06%	Allergen

and 215–230 amino acids were able to induce preferred immune responses as B cell epitopes (Fig. 5:A). Emini surface accessibility prediction predicted 99–110, 155–170 and 235–245 amino acid residues to be more accessible (Fig. 5:B). Chou and Fasman beta-turn prediction method indicated the regions from 80 to 90 and 215–235 as potential Beta-turn regions (Fig. 5:C). Karplus and Schulz flexibility prediction method displayed that the region of 75–90 and 210–235 amino acid residues are the most flexible regions (Fig. 5:D).

In case of hemagglutinin, bipred prediction method indicated that the peptide sequences from 208 to 236 and 446–456 amino acids are able to induce the preferred immune responses as B cell epitopes (Fig. 6:A). The regions from 168 to 182, 393–400 and 490–507 amino acid residues were more accessible according to Emini surface accessibility prediction algorithm (Fig. 6:B). Chou and Fasman beta-turn prediction method displayed the regions from 78 to 84 and 392–400 as potential Beta-turn regions (Fig. 6:C). Karplus and Schulz flexibility prediction method indicated the region of 148–157 and 385–400 amino acid residues as a most flexible region (Fig. 6:D). Allergenicity pattern of the predicted B-cell epitopes is shown in Table 5.

3.11. Vaccine construction

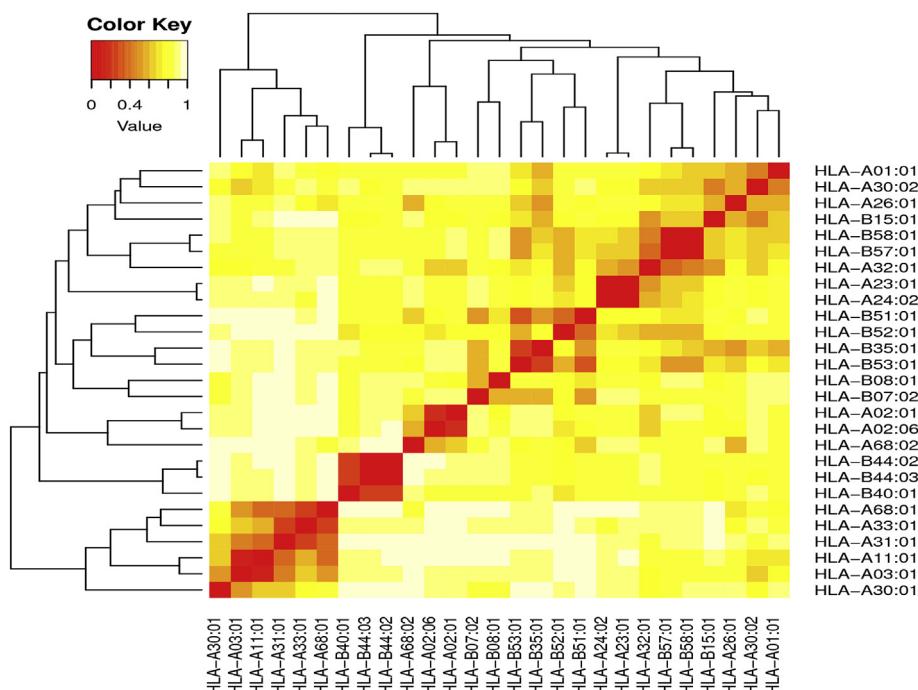
Each construct comprised a protein adjuvant and PADRE peptide

Table 4

Binding energy of predicted epitopes and MHC class I molecule, HLA-A*11:01 generated from molecular docking by AutoDock.

Matrix Protein 1				Hemagglutinin			
Epitopes	Start	End	Binding Energy (kcal/mol)	Epitopes	Start	End	Binding Energy (kcal/mol)
GALGLVCAT	2	10	-5.9	RIDFHWLML	28	36	-7.0
ILGFVFTLT	59	68	-6.2	LSGRIDFWH	25	33	-7.4
QRLEDVFAG	25	34	-6.4	NGLSGRIDF	23	31	-6.7
DVFAGKNAD	30	38	-6.5	WFSFGASC	37	45	-8.6
GKNADLEAL	34	42	-6.0	FLRGKSMGI	58	66	-6.9
ILSPLTKGI	51	59	-6.3	DFHWLMLNP	30	38	-8.7

sequence, while the rest was occupied by the T-cell and B-cell epitopes and their respective linkers. The first three designed vaccines (i.e. V1, V2, V3) consist of 12 CTL epitopes and 10 BCL epitopes where CTL and BCL epitopes were conjugated together via GGGG and KK linkers (Table 6). Constructs V1, V2 and V3 were 401, 487 and 516 residues long respectively. The predicted epitopes were separated by suitable linker with a view to ensuring maximal immunity in the body. PADRE sequence was used to enhance the potency and efficacy of the peptide

**Fig. 3.** Cluster analysis of the HLA alleles in general (red color in the heat map indicates strong interaction while the yellow zone is indicating weaker interaction).

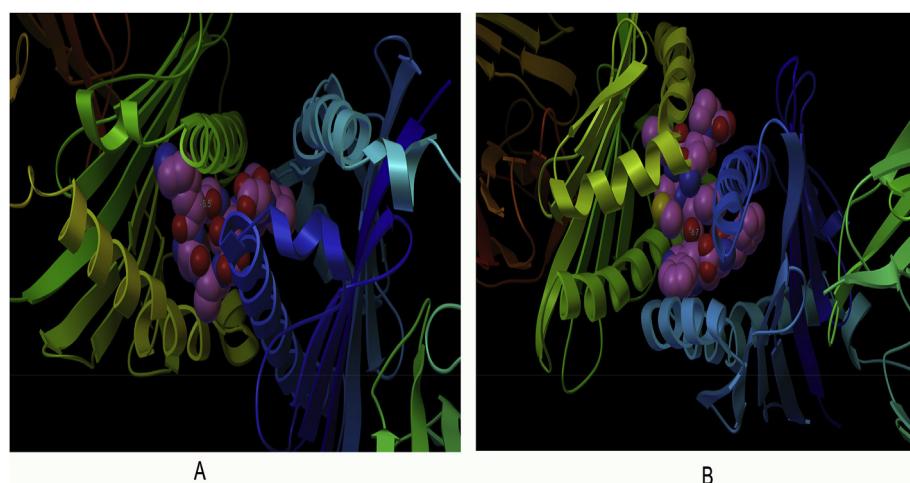


Fig. 4. Docking of predicted hemagglutinin epitope 'DFHWLMLNP' and matrix protein 1 epitope 'DVFAGKNAD' to HLA-A*11:01 with a binding energy of -8.7 kcal/mol and -6.5 kcal/mol, respectively.

vaccine. Construction of vaccine protein V4, V5 and V6 were based on identifying larger cassettes containing multiple epitopes. Epitope cluster analysis tool from IEDB predicted the epitope clusters among the top CTL epitopes proposed in Table 4. Cluster QRLEDVFAGKNADLEAL for matrix protein 1 and NGLSGRIDFHWLMLNP for hemagglutinin were identified comprising three (i.e. QRLEDVFAG, DVFAGKNAD, GKNADLEAL) and four (i.e. NGLSGRIDF, LSGRIDFHW, RIDFHWLML, DFHWLMLNP) overlapping peptides respectively. Again, a total five singletons were identified for matrix protein 1 (GALGLVCAT, ILGFVF-TLT, ILSPLTKGI) and hemagglutinin (WFSFGASC, FLRGKSMGI). Both the clusters and singletons were used along with top BCL epitopes to design construct V4, V5 and V6 which were 351, 436 and 465 residues long respectively. No cluster was found among the top BCL epitopes used to design the final vaccine constructs.

3.12. Allergenicity, antigenicity and solubility prediction of different vaccine constructs

Allertop v2.0 was used to predict the non-allergic behavior of vaccine constructs. All three vaccine constructs were found to be non-allergic in nature. Antigenicity of shortlisted three vaccine constructs were further predicted using VaxiJen 2.0 server. Results indicated V1 construct as most potent vaccine candidate with a good antigenic nature (0.70) that can stimulate a strong immune response (Table 6). However, Constructs designed using larger cassettes (i.e. V4, V5 and V6) showed immunogenicity significantly lower than the ones with individually separated epitopes (i.e. V1, V2 and V3). Solubility prediction was performed via Proso II server to find out the vaccine protein that had the best chance to be soluble upon heterologous expression. All three constructs showed solubility more than the threshold value 0.6 (0.832, 0.652 and 0.81 for V1, V2 and V3 respectively). The solubility was further investigated via Protein-sol server [Fig. 7]. It predicted the solubility of proteins in terms of QuerySol (scaled solubility value). The experimental dataset (PopAvrSol) had a population average of 0.45. So, any solubility scores larger than 0.45 were expected to have a higher solubility than the average soluble *E. coli* proteins from the experimental solubility dataset and vice versa [86]. A QuerySol of 0.66 for construct V1 ensured its higher solubility during heterologous expression in the *E. coli*.

3.13. Physicochemical characterization of vaccine protein

The ProtParam tool from Expasy server was used to characterize the physical and chemical parameters of the vaccine proteins. The

molecular weight of the vaccine construct V1 was found to be 41.99 kDa which indicated its good antigenic potential. The theoretical pI 10.09 explained that the protein will have net negative charge above the pI and vice versa. At 0.1% absorption, the extinction coefficient was 41940, assuming all cysteine residues are reduced. The estimated half-life of the designed construct V1 was predicted to be 1h in mammalian reticulocytes *in vitro* while more than 10 h in *E. coli* *in vivo*. Thermostability and hydrophilic nature of the protein was represented by the aliphatic index and GRAVY value which were 70.62 and -0.291 respectively. The instability index was computed to be 30.34 which classified the protein as a stable one and ensured that the construct possesses good characteristics to initialize an immunogenic reaction in the body.

3.14. Secondary and tertiary structure prediction

We used the PSIPRED and NetTurnP 1.0 server to obtain the secondary structure of the designed constructs. The predicted structure of the vaccine protein V1 confirmed to have 16.70% alpha helix, 10.1% sheet and 73.37% coil regions (Fig. 8). RaptorX predicted the tertiary structure of the construct V1 consisting 4 domains (Fig. 9:A). The server performed homology modeling by detecting and using 5tg8A from protein data bank (PDB) as the best suited template for Vaccine V1. All 401 amino acid residues were modeled with only 8% residues in the disordered region. The quality of the 3D model was defined by P value which was $1.37e^{-6}$ for our vaccine construct. A low P value confirmed the good model quality of the predicted structure. The Ramachandran plot analysis was also performed to validate the model (Fig. 9:B).

3.15. Tertiary structure refinement and validation

To improve the quality of predicted 3D modeled structure beyond the accuracy, refinement was performed using ModRefiner followed by FG-MD refinement server. Upon evaluation of the vaccine model on the Ramachandran plot, we found 92.4% residues in the favored region, 6.4% residues in the allowed and 1.3% in the outlier region before refinement. However, after refinement 94% and 6% residues were in the favored and allowed region respectively, while no residues were present in the outlier region (Fig. 9:B). Modeled tertiary structures of vaccine protein V2 and V3 have been shown in Fig. 10.

3.16. Vaccine protein disulfide engineering

Residues in the highly mobile region of the protein sequence were

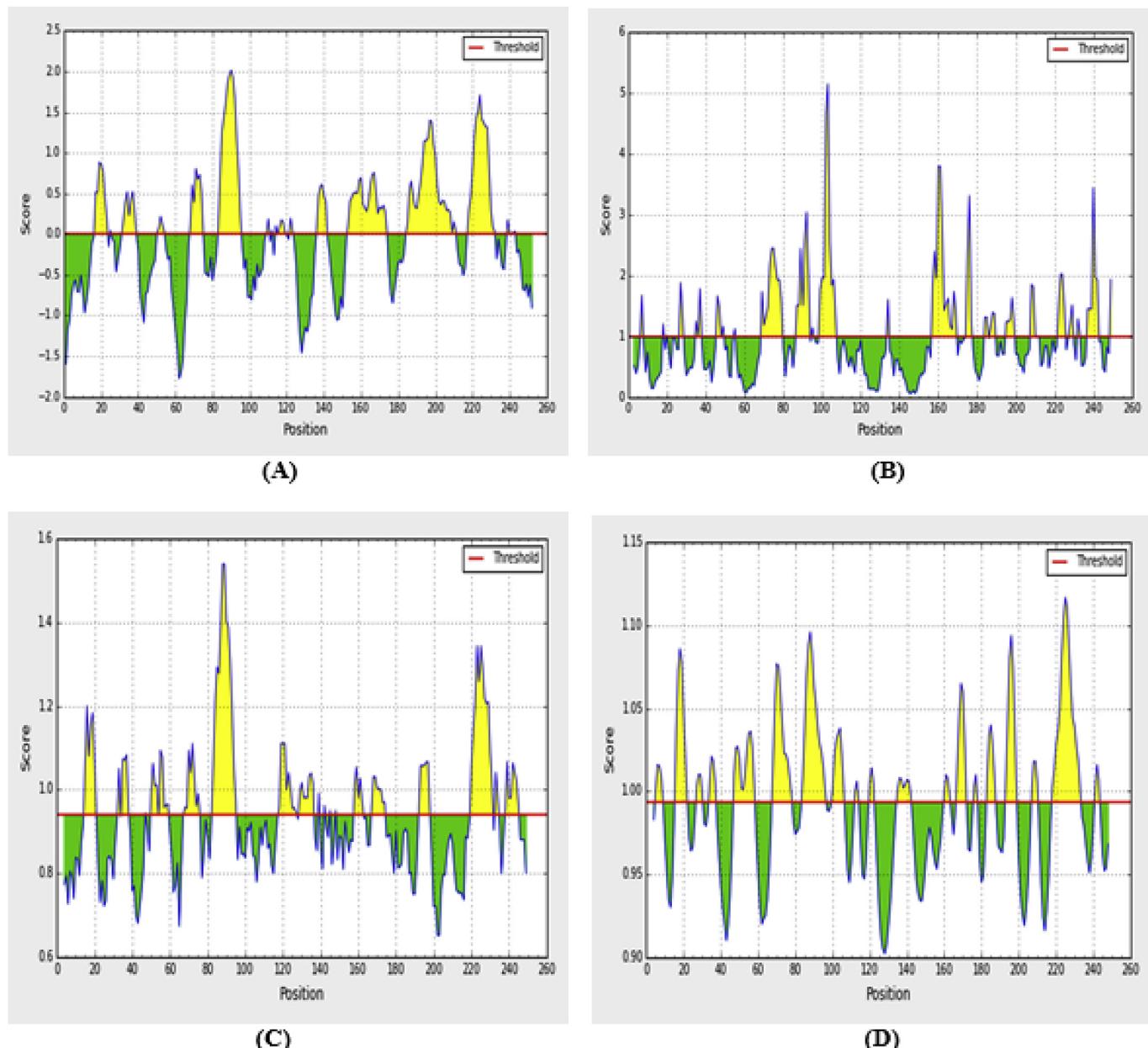


Fig. 5. B-cell epitope prediction of matrix protein 1 showing the most potent regions in yellow color, above the threshold value. (A: Bepipred Linear Epitope prediction of B cell with threshold value 0.00. B: Emini surface accessibility prediction with threshold value 1.000. C: Chou & fasman beta-turn prediction with 0.9 threshold value. D: Karplus & schulz flexibility prediction of B-cell epitope with threshold value 1.0.

mutated with cysteine to perform Disulfide engineering. A total 47 pairs of amino acid residue were identified having the capability to form disulfide bond by DbD2 server. After evaluation of the residue pairs in terms of energy, chi3 and B-factor parameter, only 2 pair satisfied the disulfide bond formation. Those residue pairs were PRO 30 - ALA 390 and ILE 324 - LYS 384 (Fig. 11). All these 4 residues were replaced with cysteine residue. The value of chi3 considered for the residue screening was between –87 and +97 while the energy value was less than 2.5.

3.17. Protein-protein docking

Docking analysis was performed between vaccine constructs and different HLA alleles (Table 7), where construct V1 showed biologically significant results and found to be superior in terms of free binding energy. Besides, docking was also conducted to evaluate the binding affinity of designed vaccine construct with human TLR8 receptor using

ClusPro, hdoc and PatchDock web servers (Fig. 12). The ClusPro server generated 30 protein-ligand complexes as output along with respective free binding energy. The lowest energy of –1247.4 was obtained for the complex named cluster 1. The hdoc server predicted the binding energy for the protein-protein complex was –362.99, while FireDock output refinement of PatchDock server showed the lowest global energy of –29.55 for solution 4. The lowest binding energy of the complexes indicates the highest binding affinity between TLR-8 and vaccine construct.

3.18. Molecular dynamics simulation

NMA (Normal mode analysis) was performed to describe the stability of proteins as well as their large scale mobility. The iMODS server rendered such analysis by taking an account of the internal coordinates of the complex molecule (Fig. 13:A). The direction of each residue was

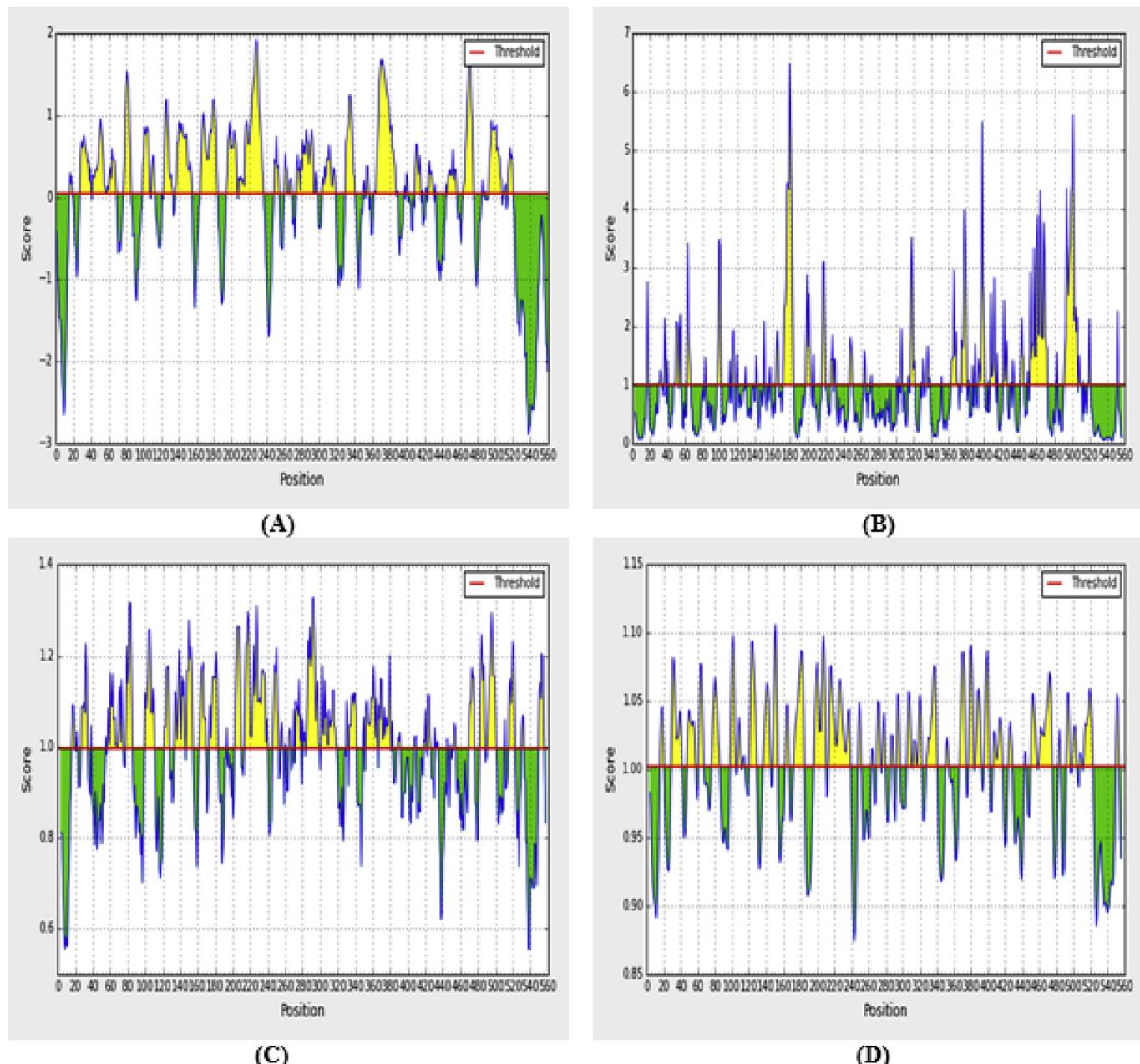


Fig. 6. B-cell epitope prediction of hemagglutinin showing the most potent regions in yellow color, above the threshold value. (A: Bepipred Linear Epitope prediction with threshold value 0.00. B: Emini surface accessibility prediction with threshold value 1.000. C: Chou & Fasman beta-turn prediction with 1.041 threshold value. D: Karplus & Schulz flexibility prediction of B-cell epitope with threshold value 1.0).

Table 5

Allergenicity assessment of the predicted B-cell epitopes generated from matrix protein 1 and hemagglutinin.

Protein	Start	End	Length	Peptide	Allergenicity
Matrix protein 1	218	232	15	TVGTHPNSSTGLKDD	Non Allergen
	84	94	11	LNGNGDPNNMD	Non Allergen
	99	107	9	LYKKLKREM	Non Allergen
	157	169	13	AQHRSHRQMATT	Non Allergen
	143	154	12	ALGLVCATCEQI	Allergen
	4	19	16	LTEVETVVLISIPSGP	Non Allergen
Hemagglutinin	172	183	12	QMTKSYKNTRK	Non Allergen
	491	506	16	IRNNTYDHSKYREEAM	Allergen
	4	15	12	QILVFALIAIP	Non Allergen
	531	550	20	GASCFILLAIIVMGLVFICVK	Non Allergen
	466	475	10	RENAEEDGTG	Non Allergen
	215	234	20	SSNYQQSFVPSPGARPQVNG	Non Allergen

Table 6

Allergenicity prediction and antigenicity analysis of the constructed vaccines.

Vaccine Constructs	Composition	Complete Sequence of Vaccine Constructs	Allergenicity	VaxiJen score (Threshold 0.4)
V1	Selected T cell & B cell epitopes of matrix protein 1 and hemagglutinin with β defensin adjuvant & PADRE sequence	EAAAKGIINTLQKYCRVRRGRCAVLSCLPKEEQI IGKCSTRGRKCCRKEAAAKAKFVAAWTLKAAA GGGSDFVAGKNADGGGSQRLEDVFAGGGGSILSP LTKGIGGGSILGFVFTLTGGGSCKNADLEALGGGS ALGLVCATGGGSDFHWMLNPGGGSWFSFGAS CFGGGSGLSGRIDFHWWGGSRIDFHWMLGGGSF LRGKSMGIGGGSNGLSSGRIDFKKTVGTHPNSTGL KDDKKLNNGDPPNNMDKKLYKKLREMKAQH RSHRQMATTKKLITEVETYVLSIIPSGPKQMTKS YKNTRKSKKQILVFALIAIIPKKGASCFLAIVMGL VFICVKKKRENAAEDGTGKSSNYQQSFVPSPGAR PQVNNGKKAKFVAAWTLKAAAGGGS	Non Allergen	0.7083
V2	Selected T cell & B cell epitopes of matrix protein 1 and hemagglutinin with L7/L12 ribosomal protein adjuvant & PADRE sequence	EAAAKMAKLSTDDELLDAFKEMTLLSDFVKFEE TEFVTAAPVAVAAGAAPAGAAVEAAEQQSEFD VILEAAGDKKIVKVKVREIVSGLGLKEAKDLVDGA PKPLLEKVAKEADEAKAKLEAAGATVTVKEAAAK AKFVAAWTLKAAAGGGSDVAGKNADGGGSQRL EDVFAAGGGGSILSPLTKGIGGGSILGFVFTLTGGGS KNADLEALGGGSALGLVCATGGGSDFHWMLNLP GGGSWFSFGASCFFGGGSLSGRIDFHWWGGSLRIDFH WLMGGGSFLRGKSMGIGGGSNGLSSGRIDFKKTVG THPNSTGLKDDKKLNNGDPPNNMDKKLYKKLKR EMKKAQHRSRQMATTKKLITEVETYVLSIIPSGPKK QMTKSYKNTRKSKKQILVFALIAIIPKKGASCFLAIV MGLVFICVKKKRENAAEDGTGKSSNYQQSFVPSPGA RPQVNNGKKAKFVAAWTLKAAAGGGS	Non Allergen	0.6268
V3	Selected T cell & B cell epitopes of matrix protein 1 and hemagglutinin with HABA adjuvant & PADRE sequence	EAAAKMAENPNIDDLPAPlLAALGAADALALATVNDL IANLRERAETTRAETRTRVERRARLTKFQEDLPEQFI ELRDKFTEELRKAAEGYLEAATNRYNELVERGEAA LQRRLSQTAFEDASARAEGYVQDQAELTQEALGTV SQTRAVGERAAKLVGIELEAAAKAKFVAAWTLKAA AGGGSDVAGKNADGGGSQRLEDVFAGGGGSILSP LTKGIGGGSILGFVFTLTGGGSCKNADLEALGGGS ALGLVCATGGGSDFHWMLNPGGGSWFSFGASC GGGSLSGRIDFHWWGGSLRIDFHWMLGGGSFLRG KSMGIGGGSNGLSSGRIDFKKTVGTHPNSTGLKDD KKLNNGDPPNNMDKKLYKKLREMKAQHRSRQ MATTKKLITEVETYVLSIIPSGPKQMTKSYKNTRKS KKQILVFALIAIIPKKGASCFLAIVMGLVFICVKKKR ENAAEDGTGKSSNYQQSFVPSPGARPQVNNGKKAKF VAAWTLKAAAGGGS	Non Allergen	0.6457
V4	Selected T cell & B cell epitopes of matrix protein 1 and hemagglutinin with β defensin adjuvant & PADRE sequence	EAAAKGIINTLQKYCRVRRGRCAVLSCLPKEEQ IGKCSTRGRKCCRKEAAAKAKFVAAWTLKAAA AGGGSQRLEDVFAGKNADLEALGGGSILSP IGGGSILGFVFTLTGGGSCKNADLEALGGGS GKSMGIKTVGTHPNSTGLKDDKKLNNGDPP NMDKKLYKKLREMKAQHRSRQMATT LTEVETYVLSIIPSGPKQMTKSYKNTRKS VFAFIAIIPKKGASCFLAIVMGLVFICVKKKREN AEEDGTGKSSNYQQSFVPSPGARPQVNNGKKAKF VAAWTLKAAAGGGS	Non Allergen	0.6150
V5	Selected T cell & B cell epitopes of matrix protein 1 and hemagglutinin with L7/L12 ribosomal protein adjuvant & PADRE sequence	EAAAKMAKLSTDDELLDAFKEMTLLSDFVKFEE EETFEVTAAPVAVAAGAAPAGAAVEAAEQQS EFDVILEAAGDKKIVKVKVREIVSGLGLKEAKDL VDGAPKPLLEKVAKEADEAKAKLEAAGATVTV KEAAAKAKFVAAWTLKAAAGGGSQRLEDVFAG KNADLEALGGGSILSP GGSALGLVCATGGGSNGLSSGRIDFHWML NPGGGSWFSFGASCFFGGGSFLRGKSMGIKKTVG THPNSTGLKDDKKLNNGDPPNNMDKKLYKKL KREMKAQHRSRQMATTKKLITEVETYVLSIIP SGPKQMTKSYKNTRKS VFAFIAIIPKKGASCFLAIVMGLVFICVKKKREN AEEDGTGKSSNYQQSFVPSPGARPQVNNGKKAKF AGGGS	Non Allergen	0.5581

(continued on next page)

Table 6 (continued)

Vaccine Constructs	Composition	Complete Sequence of Vaccine Constructs	Allergenicity	VaxiJen score (Threshold 0.4)
V6	Selected T cell & B cell epitopes of matrix protein 1 and hemagglutinin with HABA adjuvant & PADRE sequence	EAAKMAENPNIDDLPPAPLLAALGAADLALATV NDLIANLRERAETRAETRTRVEERRARLTKFQE DLPEQFIELRDKFTTEELRKAAEGLYEAATNRYNE LVERGEAALQRRLRSQTAFEDASARAEGYVDQAV ELTQEALGTVASQTRAVGERAALKVGIELEAAAK AKFVAAWTLKAAAGGGSQRLEDVFAGKNADLE ALGGGSILSPLTKGIGGGSILGFVFTLTGGGSGAL GLVCATGGGSNLGSGRIDFHWLMLNPAGGS WFSFGASCFCGGGSFLRGKSMGIKKTVGTHPNSST GLKDDKKLNNGNDPNMDKKLYKKLKREMKK AQHRSRQMATTKKLTVETVVLISIPSPKKQ MTKSYKNTRKSQQILVFALIAIIPKKGASCFLLA IVMGLVIFCVKKRKENAEEDGTGKSSNYQQSFV PSPGARPQVNGKKAKFVAAWTLKAAAGGGS	Non Allergen	0.5833

given by arrows and the length of the line represented the degree of mobility in the 3D model. Results showed that the mobility of vaccine protein V1 and TLR8 were oriented towards each other. The B-factor values inferred via NMA, was equivalent to RMS (Fig. 13:C). The probable deformability of the complex was linked with the distortion of the individual residues, indicated by hinges in the chain (Fig. 13:D). The eigenvalue found for the complex was $2.8978e^{-05}$ (Fig. 13:B). The variance associated to each normal mode was inversely related to the eigenvalue [81,87]. The covariance matrix indicated coupling between pairs of residues, where they may be associated with correlated, uncorrelated or anti-correlated motions, indicated by red, white and blue colors respectively (Fig. 13:E). The result also generated an elastic network model (Fig. 13:F). It identified the pairs of atoms those are connected via springs. Each dot in the diagram showed one spring between the corresponding pair of atoms and was colored based on extent of stiffness. The darker the grays, the stiffer the springs is.

3.19. Codon adaptation and in silico cloning

The regulatory systems for protein expression in human and E. coli K12 strains are different. Therefore, codon adaptation was performed

considering the expression system of the host. Vaccine protein V1 was reverse transcribed and codon adaptation index of the adapted codons was found to be 0.974 indicating the higher proportion of most abundant codons. The GC content of the optimized codons (49.45%) was also significant. The construct also lacked restriction sites for BglII and ApaI ensuring its safety for the cloning purpose. The optimized codons were then inserted into pET28a(+) vector along with BglII and ApaI restriction sites. A clone of 5649 base pair was obtained including the 1213 bp desired sequence and the rest belonging to the vector. The desired region was shown in red color in between the pET28a(+) vector sequence (Fig. 14).

4. Discussion

To combat against the ever rising global burden of virus driven diseases, design of a new vaccine in a timely fashion is an age demanding scientific challenge. With the blessing of *in silico* studies and the progress of sequence-based technology, we have enough data about the genomics and proteomics of different viral pathogens. Therefore, it is possible to design peptide vaccines based on a neutralizing epitope using various bioinformatics tools. The “vaccinomics” approach has

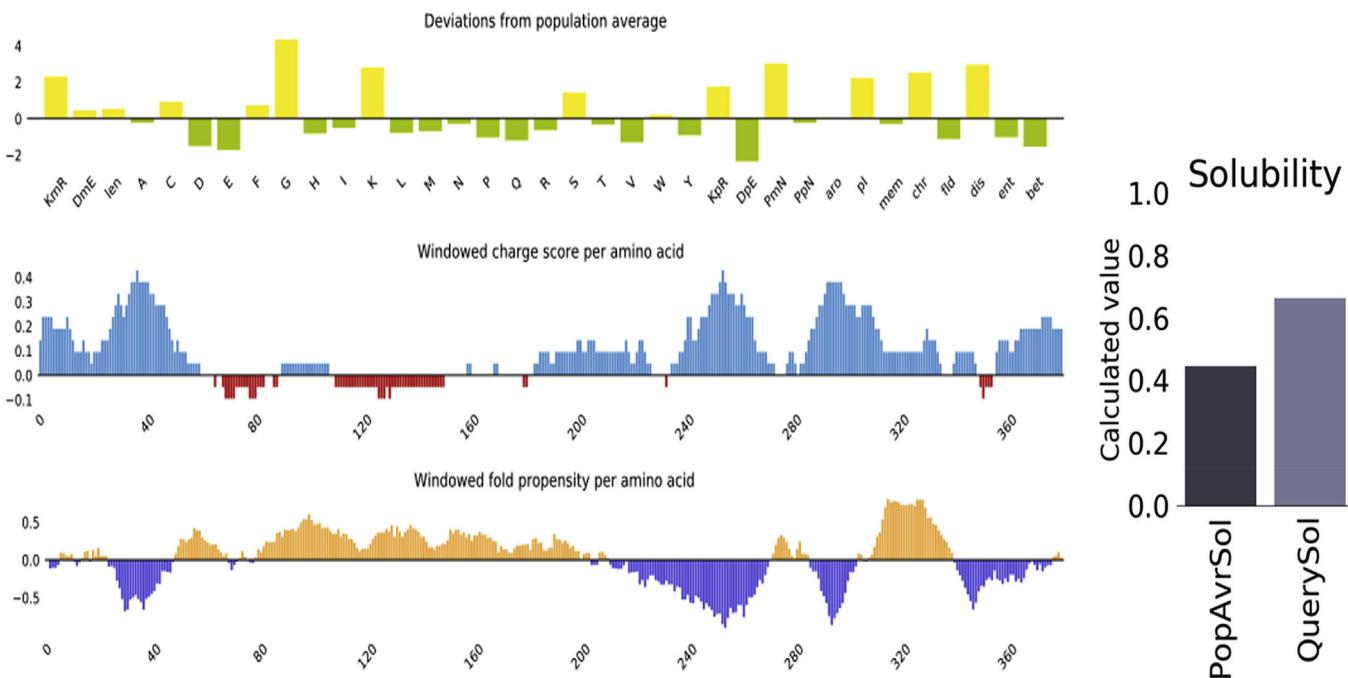


Fig. 7. Solubility prediction of designed vaccine construct V1 using via Protein-sol server.

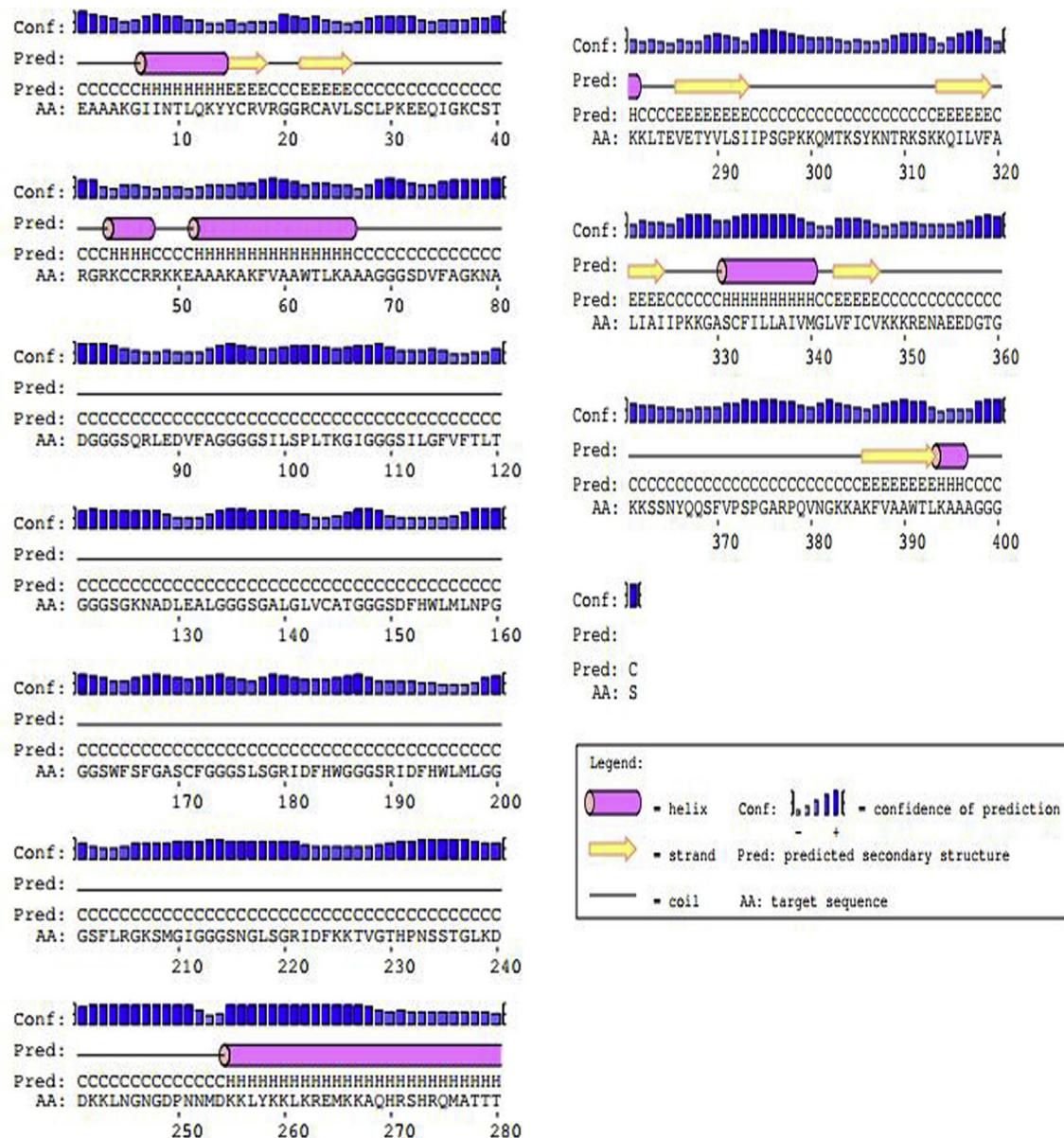


Fig. 8. Secondary structure prediction of constructed vaccine protein V1 using PESIPRED server.

already been proved to be promising for defending diseases such as malaria [88], multiple sclerosis [89] and tumors [90]. In the present study, we retrieved the entire viral proteome of avian influenza A(H7N9) strains by using National Center for Biotechnology Information (NCBI) database. The physicochemical properties of viral proteins were analyzed through ProtParam server [26,91]. The VaxiJen server assessed all of the retrieved protein sequences in order to find out the most potent antigenic protein. Among the eight viral proteins, a combination of outer and inner membrane proteins was chosen based on their antigenicity scoring and literature study. Matrix protein 1 (Accession ID: AGL44441.1) & hemagglutinin (Accession ID: AGL44438.1) were selected to design the final vaccine constructs based on their ability to confer immunity and allowed for further analysis. Hemagglutinin is a surface glycoprotein responsible for binding to host cell [92] while matrix protein 1 (inner membrane protein) is one of the most abundant proteins of influenza A virus playing essential structural and functional roles in the virus life cycle [93,94]. Other structural proteins for example neuraminidase (NA) undergoes periodic antigenic shift when NA genes re-assort with a virus of a different subtype, thus

evasive antibodies [85,95,96]. However, previous studies confirmed good antibody response in the host induced via M1 protein [84,85,97].

Most antigens and vaccines trigger not only B cell response but also T cell response. Vaccine induces production of antibodies that are synthesized by B cells and mediates effector functions by interacting specifically to a pathogen or toxin [98]. However, humoral response from memory B cells can easily be overcome over time by surge of antigens while cell mediated immunity often elicits long lasting immunity [99,100]. Cytotoxic CD8+ T lymphocytes (CTL) restrict the spread of infectious agents by recognizing and killing infected cells or secreting specific antiviral cytokines [101]. Thus, T cell epitope-based vaccination is a unique process to induce strong immune response against infectious agents [102]. 8–10 amino acids long peptide antigens are usually involved in either MHC or TCR binding or both [103]. In the current study, peptide lengths were set at 9 before making software based class I T-cell epitope prediction.

MHC-I binding predictions of the IEDB with recommended methods were used, where approximately 26576 immunogenic epitopes of hemagglutinin and 13340 immunogenic epitopes of matrix protein 1 were

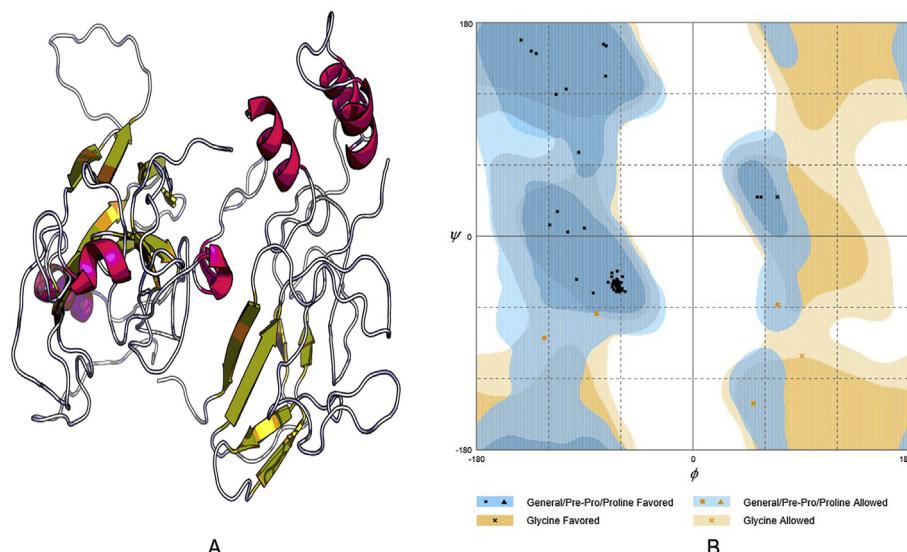


Fig. 9. Tertiary structure prediction and validation of vaccine protein V1. (A) Tertiary structure of modeled construct V1, (B) Ramachandran plot analysis of vaccine protein V1.

generated to be T cell epitopes that can bind a large number of HLA-A and HLA-B alleles with high binding affinity. Top ten epitopes, bound with the highest number of HLA alleles were selected as putative T cell epitope candidates based on their protein transmembrane topology screening and VaxiJen score. Allergenicity, a prominent deterrent in vaccine development was found in most vaccines candidate as “allergic” reaction in immune stimulation process. There is a formula of allergenicity prediction by the FAO or WHO and it states that a sequence could be a potentially allergenic if it has an identity of at least six contiguous amino acids over a window of 80 amino acids when compared to known allergens [104]. In this study, AllerTOP, AllerGenFP, PA3P, Allermatch were used for the allergenicity assessment of the T-cell epitopes of both proteins. However, 6 epitopes of Matrix protein 1 and 6 epitopes of hemagglutinin were found to be non-allergenic to humans. Population coverage is another potential parameter in reverse vaccinology approach. Results indicated that population from most geographic regions of the world can be covered by all predicted T-cell epitopes of Matrix protein 1 (more than 90%) and Hemagglutinin (more than 40%). In Asia, mostly in the People's Republic of China, more than 1500 people have been affected and at least 600 people have died as a result of H7N9 bird flu virus since 2013 [5,7]. As the MHC superfamilies play a vital role in vaccine design and drug development,

MHC cluster analysis was also performed to determine the functional relationship between MHC variants. To ensure the binding between HLA molecules and our predicted epitope, a docking study was performed using MGLTools. The 13 epitopes of hemagglutinin and matrix protein 1 were subjected to PEP-FOLD3 web-based server for 3D structure conversion, in order to analyze the interactions with different HLAs. The HLA-A*11:01 was selected for docking analysis. Epitope 'DFHWLMLNP' from hemagglutinin was found to be best as it bound in the groove of the HLA-A*11:01 with an energy of -8.7 kcal/mol. Again, 'DVFAGKNAD' was found best for matrix protein 1 and its binding energy was -6.5 kcal/mol. However, in this study we developed a multi-epitope subunit vaccine to ensure better immune protection. All the finalized epitopes showed a lower binding energy which was biologically significant.

For B-cell epitope prediction, we predicted amino acid scale-based methods for the identification of potential B-cell epitopes using four algorithms from IEDB server. Bepipred prediction method predicted the peptide sequence from 215 to 230 regions for matrix protein 1 and 386–400 regions for hemagglutinin with the ability to induce the preferred immune responses. Emini surface accessibility prediction, Chou & Fasman beta-turn prediction and Schulz & Karplus flexibility prediction method was used to identify the easily accessible, most potent beta-

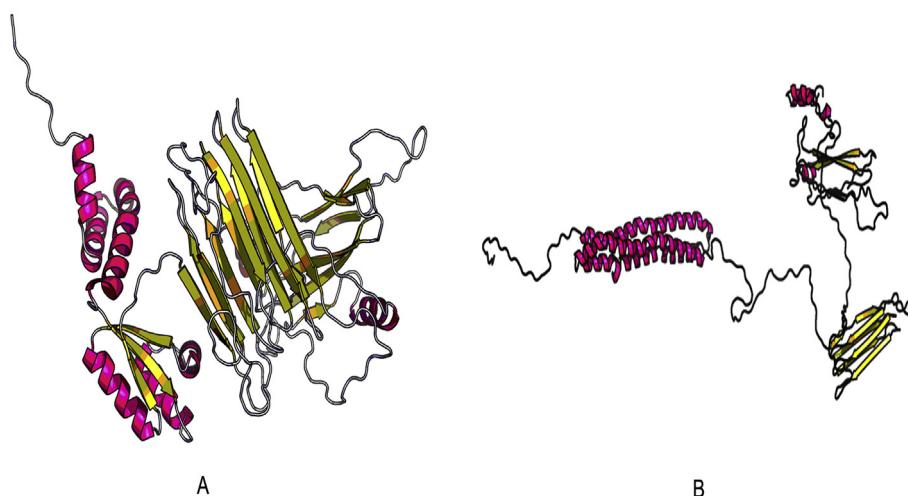


Fig. 10. 3D modeled structure of vaccine protein V2 and V3 generated via RaptorX server.

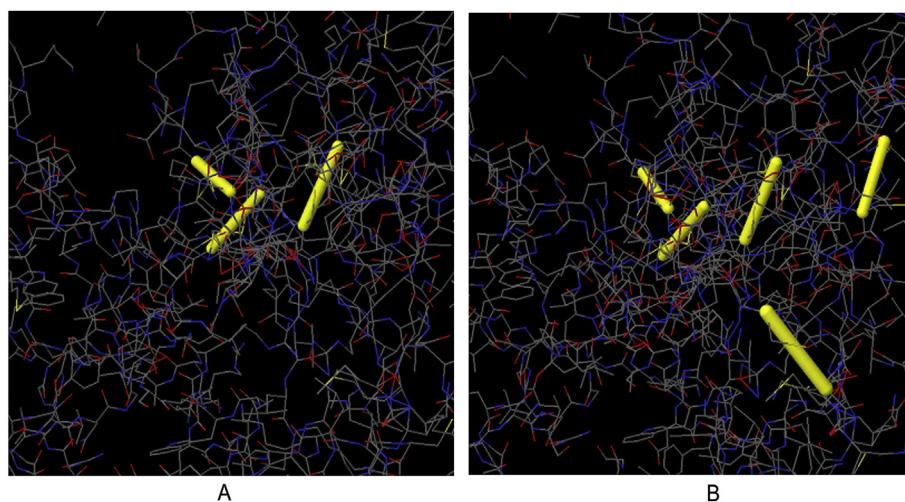


Fig. 11. Disulfide engineering of vaccine protein V1. (A) Initial form, (B) Mutant form.

Table 7

Docking score of vaccine construct V1 with different HLA alleles including HLA-DRB1*03:01 (1A6A), (HLA-DRB5*01:01 (1H15), HLA-DRB1*01:01 (2FSE), HLA-DRB3*01:01 (2Q6W), HLA-DRB1*04:01 (2SEB) and HLA-DRB3*02:02 (3C5J).

Vaccine Construct	PDB ID of the HLA Alleles	Global Energy	Hydrogen Bond Energy	ACE	Score	Area
V1	1A6A	-39.35	-0.56	-0.19	18396	2859.1
	1H15	-38.14	-2.24	-6.64	18554	2598.5
	2FSE	-48.38	-7.57	6.13	16774	2607.6
	2Q6W	-11.43	-2.71	15.93	17628	2477.4
	2SEB	-29.33	-5.19	1.05	14434	2062.7
	3C5J	-11.38	-2.12	3.06	12262	1718.8
V2	1A6A	-9.34	-3.36	10.20	16352	2754.90
	1H15	-10.36	-3.86	8.65	18040	2603.20
	2FSE	-11.58	-1.58	-0.10	16718	2336.20
	2Q6W	-50.57	-0.97	-8.13	17552	2711.10
	2SEB	1.60	-3.27	-0.14	18380	2537.00
V3	1A6A	-24.43	-2.67	14.52	17402	2534.40
	1H15	-29.48	-3.60	15.78	16664	2158.00
	2FSE	-37.31	-3.06	4.25	14598	2050.70
	2Q6W	-11.81	-5.50	10.98	17868	2500.80
	2SEB	9.64	0.00	1.12	14958	2942.50
	3C5J	34.18	-0.99	16.84	15626	2168.10

turn regions and most flexible regions respectively. From all above, the most potent B cell epitopes for both proteins were identified as vaccine candidates against avian influenza A(H7N9) strain. The final vaccine proteins were constructed using the promiscuous epitopes and protein adjuvants along with PADRE peptide sequence. Adjuvants were previously used as immunomodulator to enhance the activity of several vaccine candidates [105,106]. Beta defensin adjuvant has already been proved as an effective immune-stimulator against different organisms in a number of experiments [105,107–109]. In this study, a total six vaccines were designed following two different strategies. In case of first three constructs (V1, V2, V3) each epitopes were separately linked via suitable linkers while the remaining constructs (V4, V5, V6) were designed using larger cassettes containing multiple epitopes. Cluster QRLEDVFAGKNADLEAL comprising 3 CTL epitopes and NGLSGRIDFHWLMLNP comprising 4 CTL epitopes were identified for matrix protein 1 and hemagglutinin respectively. Such combination of overlapping peptides was also described in literature [110]. Studies also revealed that PADRE containing vaccine constructs showed better CTL responses than the vaccines lacked it [111]. The constructed vaccines were further checked for their non-allergic behavior and immunogenic potential. However, results showed that inclusion of individual epitopes separately in the vaccine construct enhanced the antigenicity score by a wide margin than incorporating larger cassettes such in case of V4, V5 and V6. Construct V1 was found superior in terms of antigenicity,

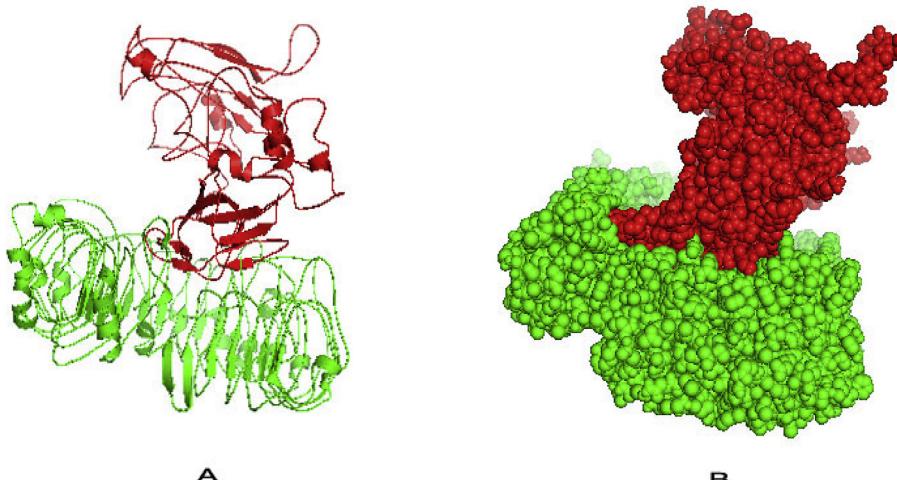


Fig. 12. Docked complex of vaccine construct V1 with human TLR8; A: Cartoon format and B: Ball structure.

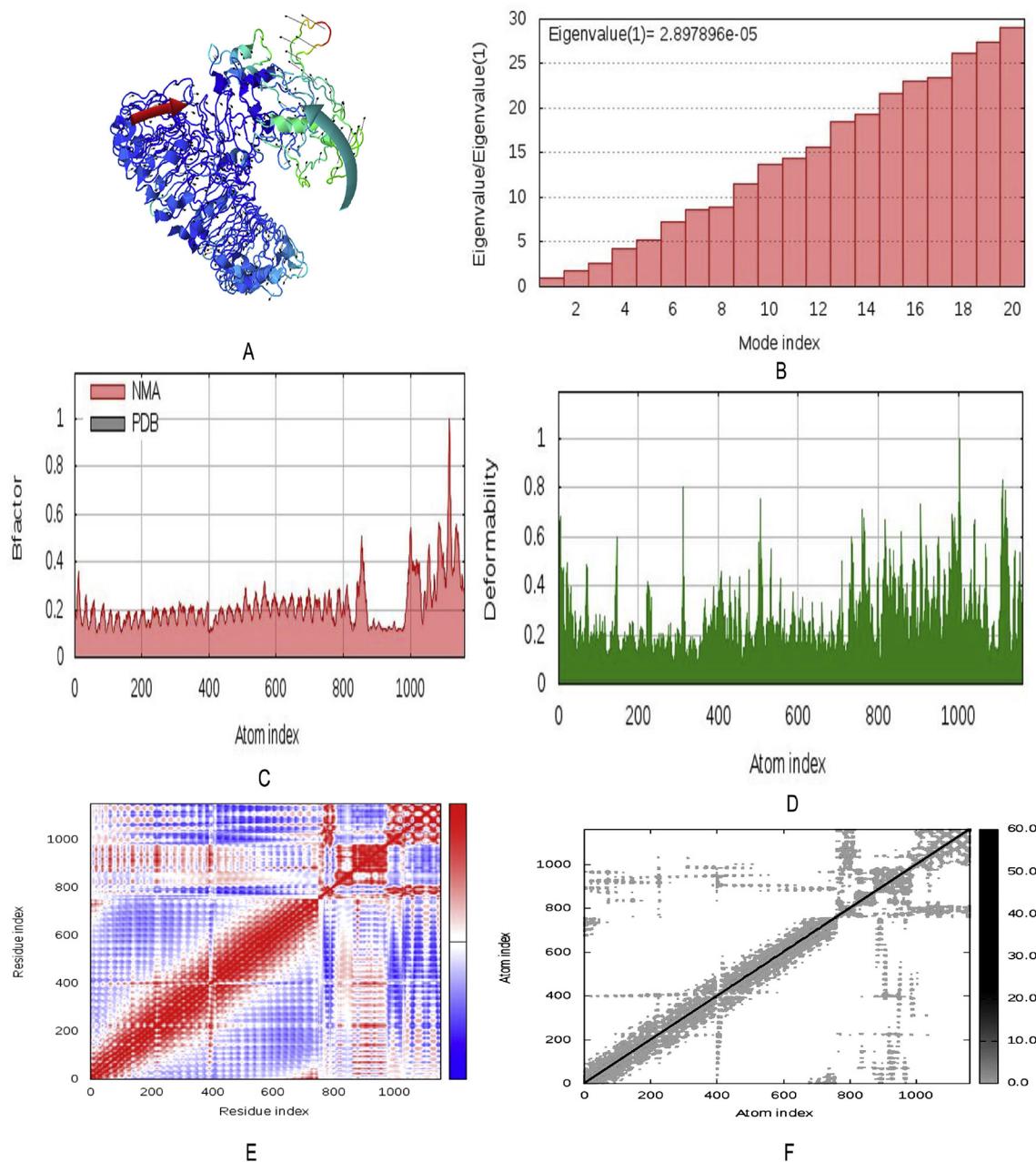


Fig. 13. Molecular dynamics simulation of vaccine protein V1-TLR8 complex. Stability of the protein-protein complex was investigated through mobility (A), eigenvalue (B), B-factor (C), deformability (D), covariance (E) and elastic network (F) analysis.

solubility and allergenicity. The physicochemical properties and secondary structure of V1 was also analyzed before tertiary structure prediction and refinement of 3D model.

To strengthen our prediction, we checked the interaction between our vaccine construct with different HLA molecules (i.e. DRB1*0101, DRB3*0202, DRB5*0101, DRB3*0101, DRB1*0401, and DRB1*0301). Again construct V1 was found to be best considering the free binding energy. Moreover, docking analysis was also performed to explore the binding affinity of vaccine protein V1 and human TLR8 receptor to evaluate the efficacy of used adjuvant. Molecular dynamics study was conducted to determine the complex stability as well. Structural dynamics had been investigated previously using subsets of atoms and covariance analysis [75]. Literature studies linked the stability of macromolecules with correlated fluctuations of atoms [112,113]. In the present study, essential dynamics was compared to the normal modes of proteins to determine its stability through iMODS server. The analysis

revealed negligible chance of deformability for each individual residues, as location of hinges in the chain was not significant and thereby strengthening our prediction. Finally, the designed vaccine construct V1 was reverse transcribed and adapted for *E. coli* strain K12 prior to insertion within pET28a(+) vector for its heterologous cloning and expression. However, our predicted *in silico* results were generated through different computational analysis of sequences and using various immune databases. We suggest further wet lab based research using model animals for experimental validation of our predicted vaccine candidates.

5. Conclusion

In this study, we developed a unique chimeric subunit vaccine (V1) against influenza A H7N9 using the finalized epitopes generated from the selected proteins via various bioinformatics tools. Our study

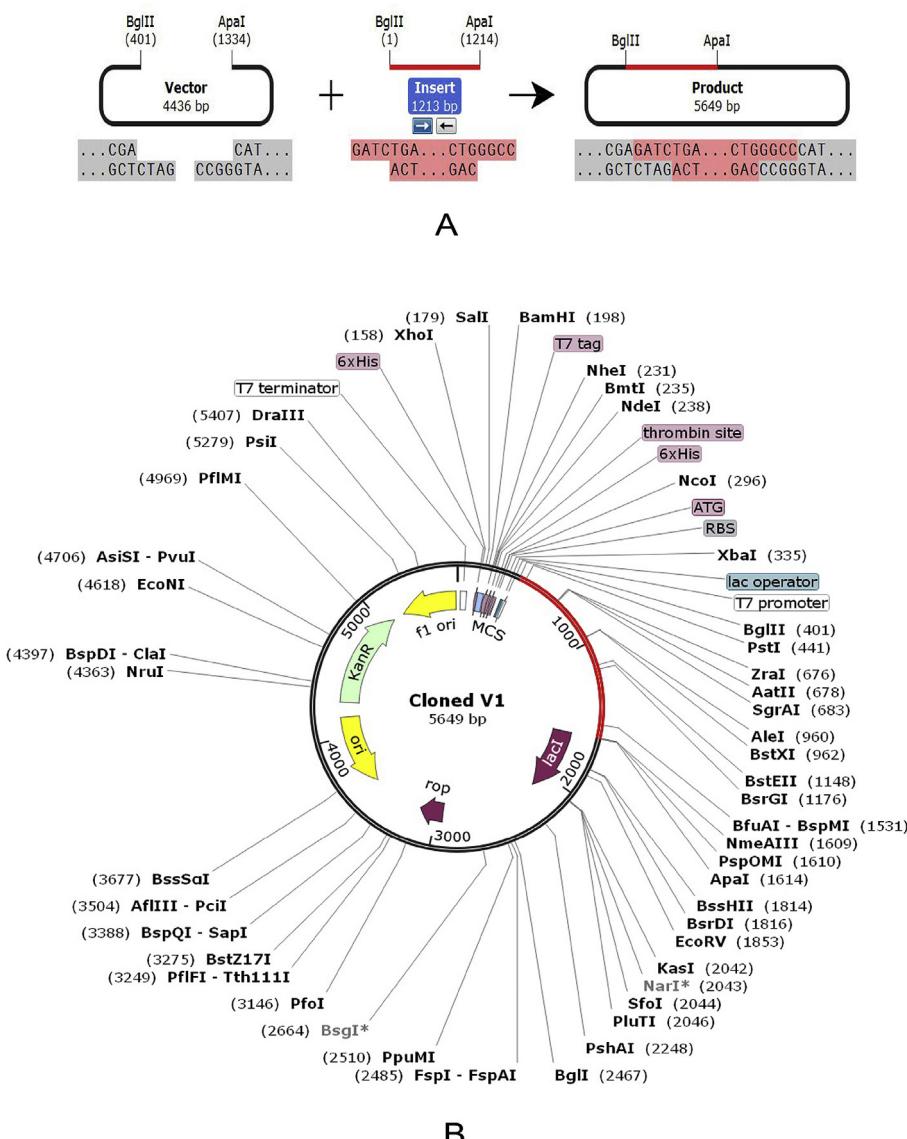


Fig. 14. In silico restriction cloning of the gene sequence of final vaccine construct V1 into pET28a(+) expression vector; (A) Restriction digestion of the vector pET28a(+) and construct V1 with BglII and ApaI (B) Inserted desired fragment (V1 Construct) between BglII (401) and ApaI (1334) indicated in red color.

revealed that reverse vaccinology approach could be used successfully for predicting vaccine candidates against viral pathogens. *In silico* studies can save both time and costs for researchers and help to find out the potential solution with fewer trials and error repeats of assays. Therefore, our present study will help to develop a suitable therapeutics and prompts the future vaccine development against avian influenza A(H7N9) virus and other emerging infectious diseases.

Glossary

LPAI, Low pathogenic avian influenza; CDC, Disease Control and prevention; NCBI, National center for Biotechnology information; IEDB, Immune epitope database and analysis resource; CTLs, Cytotoxic T lymphocytes; MGL, Molecular graphics laboratory; MHC, Major histocompatibility complex.

Conflicts of interest

Authors declare no conflict of interests.

Author's contribution

MH: Conceptualization, Supervision, Project administration and Reviewing. PPG: Experiment Design, Data Handling and Data Analysis. KFA: Data Handling, Data analysis, Draft Preparation and Manuscript writing. SM: Data Handling and Data Analysis. RAA: Data Handling and Data Analysis. JN: Data Handling and Data Analysis. MMH: Reviewing. Mahmudul Hasan (MH), Progga Paromita Ghosh (PPG), Kazi Faizul Azim (KFA), Shamsunnahar Mukta (SM), Ruhshan Ahmed Abir (RAA), Jannatun Nahar (JN), Mohammad Mehedi Hasan Khan (MMH).

Funding information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

Authors like to acknowledge the Department of Pharmaceuticals and Industrial of Sylhet Agricultural University for the technical support of the project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.micpath.2019.02.023>.

References

- [1] World Health Organization Factsheet, Influenza (Seasonal). 6 November, (2018) [https://www.who.int/news-room/fact-sheets/detail/influenza-\(seasonal\)](https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)).
- [2] K. Yuen, P. Chan, M. Peiris, D. Tsang, T. Que, K. Shortridge, et al., Clinical features and rapid viral diagnosis of human disease associated with avian influenza A H5N1 virus, *Lancet* 351 (1998) 467–471.
- [3] M. Garcia, J.M. Crawford, J.W. Latimer, E. Rivera-Cruz, M.L. Perdue, Heterogeneity in the haemagglutinin gene and emergence of the highly pathogenic phenotype among recent H5N2 avian influenza viruses from Mexico, *J. Gen. Virol.* 77 (1996) 1493–1504.
- [4] Y. Chen, W. Liang, S. Yang, N. Wu, H. Gao, J. Sheng, et al., Human infections with the emerging avian influenza A H7N9 virus from wet market poultry: clinical analysis and characterisation of viral genome, *Lancet* 381 (2013) 1916–1925.
- [5] R. Gao, B. Cao, Y. Hu, Z. Feng, D. Wang, W. Hu, et al., Human infection with a novel avian-origin influenza A (H7N9) virus, *N. Engl. J. Med.* 368 (2013) 1888–1897.
- [6] D. Liu, W. Shi, Y. Shi, D. Wang, H. Xiao, W. Li, et al., Origin and diversity of novel avian influenza A H7N9 viruses causing human infection: phylogenetic, structural, and coalescent analyses, *Lancet* 381 (2013) 1926–1932.
- [7] B.J. Cowling, L. Jin, E.H. Lau, Q. Liao, P. Wu, H. Jiang, et al., Comparative epidemiology of human infections with avian influenza A H7N9 and H5N1 viruses in China: a population-based study of laboratory-confirmed cases, *Lancet* 382 (2013) 129–137.
- [8] Y. Shi, W. Zhang, F. Wang, J. Qi, Y. Wu, H. Song, et al., Structures and receptor binding of hemagglutinins from human-infecting H7N9 influenza viruses, *Science* 342 (2013) 243–247.
- [9] T.T.-Y. Lam, J. Wang, Y. Shen, B. Zhou, L. Duan, C.-L. Cheung, et al., The genesis and source of the H7N9 influenza viruses causing human infections in China, *Nature* 502 (2013) 241.
- [10] Cumulative number of confirmed human cases of avian influenza A(H5N1) reported to WHO. 13 December, (2018) https://www.who.int/influenza/human_animal_interface/H5N1_cumulative_table_archives/en/.
- [11] Q. Xu, Z. Chen, X. Cheng, L. Xu, H. Jin, Evaluation of live attenuated H7N3 and H7N7 vaccine viruses for their receptor binding preferences, immunogenicity in ferrets and cross reactivity to the novel H7N9 virus, *PLoS One* 8 (2013) e76884.
- [12] I. Isakova-Sivak, D. Korenkov, T. Smolnogina, T. Tretiak, S. Donina, A. Rekstkin, A. Naykin, S. Shcherbik, N. Pearce, L.M. Chen, T. Bousse, Comparative studies of infectivity, immunogenicity and cross-protective efficacy of live attenuated influenza vaccines containing nucleoprotein from cold-adapted or wild-type influenza virus in a mouse model, *Virology* 500 (2017) 209–217.
- [13] L. Rudenko, I. Kiseleva, E. Krutikova, E. Stepanova, I. Isakova-Sivak, S. Donina, A. Rekstkin, M. Pisareva, E. Bazhenova, T. Kotomina, A. Katelnikova, Two Live Attenuated Vaccines against Recent Low-and Highly Pathogenic H7N9 Influenza Viruses Are Safe and Immunogenic in Ferrets, *Vaccines* 6 (2018) 74.
- [14] D.M. Carter, C.E. Bloom, G.A. Kirchenbaum, V. Tsvetnitsky, I. Isakova-Sivak, L. Rudenko, et al., Cross-protection against H7N9 influenza strains using a live-attenuated H7N3 virus vaccine, *Vaccine* 33 (2015) 108–116.
- [15] T.M. Babu, M. Levine, T. Fitzgerald, C. Luke, M.Y. Sangster, H. Jin, et al., Live attenuated H7N7 influenza vaccine primes for a vigorous antibody response to inactivated H7N7 influenza vaccine, *Vaccine* 32 (2014) 6798–6804.
- [16] A.W. Purcell, J. McCluskey, J. Rossjohn, More than one reason to rethink the use of peptides in vaccine design, *Nat. Rev. Drug Discov.* 6 (2007) 404–414.
- [17] R. Rappuoli, Reverse vaccinology, *Curr. Opin. Microbiol.* 3 (2000) 445–450.
- [18] S.S. Wong, T. Jeevan, L. Kercher, S.W. Yoon, A.M. Petkova, J.C. Crumpton, et al., A single dose of whole inactivated H7N9 influenza vaccine confers protection from severe disease but not infection in ferrets, *Vaccine* 32 (2014) 4571–4577.
- [19] G.A. Poland, I.G. Ovsyannikova, R.M. Jacobson, Application of pharmacogenomics to vaccines, *Pharmacogenomics* 10 (5) (2009) 837–852.
- [20] D.R. Flower, Bioinformatics for vaccinology, first ed., John Wiley & Sons, 2008, p. 302.
- [21] N.B.V. Petrovsky, Computational immunology: The coming of age, *Immunol. Cell Biol.* 80 (2012) 248–254.
- [22] I.A. Doytchinova, D.R. Flower, VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, *BMC Bioinf.* 8 (2007) 4.
- [23] D.M. Post, B. Slutter, B. Schilling, A.T. Chande, J.A. Rasmussen, B.D. Jones, A.K. D’Souza, L.M. Reinders, J.T. Harty, B.W. Gibson, M.A. Apicella, Characterization of inner and outer membrane proteins from Francisellatularensis strains LVS and Schu S4 and identification of potential subunit vaccine candidates, *mBio* 8 (2017) e01592-17.
- [24] R. Sharmin, A.B. Islam, A highly conserved WDYPKCDRA epitope in the RNA directed RNA polymerase of human coronaviruses can be used as epitope-based universal vaccine design, *BMC Bioinf.* 15 (2014) 161.
- [25] L. Ravichandran, A. Venkatesan, J. Febin Prabhu Dass, Epitope-based immunoinformatics approach on RNA-dependent RNA polymerase (RdRp) protein complex of Nipah virus (NiV), *J. Cell. Biochem.* (2018) 1–14.
- [26] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein Identification and Analysis Tools on the ExPASy Server, *Nucleic Acids Res.* 31 (2003) 3784–3788.
- [27] M. Hasan, Z.F. Joy, E.H. Bhuiyan, M.S. Islam, In Silico Characterization and Motif Election of Neurotoxins from Snake Venom, *Am. J. Biochem. Biotechnol.* 11 (2015) 84.
- [28] R. Vita, J.A. Overton, J.A. Greenbaum, J. Ponomarenko, J.D. Clark, J.R. Cantrell, et al., The immune epitope database (IEDB) 3.0, *Nucleic Acids Res.* 43 (2014) D405–D412.
- [29] A. Krogh, B. Larsson, G. Von Heijne, E.L. Sonnhammer, Predicting transmembrane protein topology with a hidden markov model: application to complete genomes1, *J. mol. biol.* 305 (2001) 567–580.
- [30] I. Dimitrov, I. Bangov, D.R. Flower, I. Doytchinova, v AllerTOP, 2- a server for in silico prediction of allergens, *J. Mol. Model.* 20 (2013) 2278.
- [31] I. Dimitrov, L. Naneva, I. Doytchinova, I. Bangov, AllergenPP: allergenicity prediction by descriptor fingerprints, *Bioinformatics* 30 (2014) 846–851.
- [32] C. Chrysostomou, H. Seker, Prediction of protein allergenicity based on signal-processing bioinformatics approach, 36th annual international conference of the IEEE engineering in medicine and biology society, (2014).
- [33] M.W.E.J. Fiers, G.A. Kleter, H. Nijland, A. Peijnenburg, J.P. Nap, R. Ham, Allermatch™, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines, *BMC Bioinf.* 5 (2004) 133.
- [34] M. Thomsen, C. Lundsgaard, S. Buus, O. Lund, Nielsen M. MHCcluster, a method for functional clustering of MHC molecules, *Immunogenetics* 65 (2013) 655–665.
- [35] J. Maupetit, P. Derreumaux, P. Tuffery, A fast method for large scale De Novo peptide and miniprotein structure prediction, *J. Comput. Chem.* 31 (2010) 726–738.
- [36] H. Kaur, A. Garg, G. Raghava, PEPstr: a de novo method for tertiary structure prediction of small bioactive peptides, *Protein Pept. Lett.* 14 (2007) 626–631.
- [37] Z. Wang, J. Eickholt, J. Cheng, APOLLO: a quality assessment service for single and multiple protein models, *Bioinformatics* 27 (2011) 1715–1716.
- [38] M.F. Sanner, Python: A Programming Language for Software Integration and Development, *J. Mol. Graph. Model.* 17 (1) (1999) 57–61.
- [39] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, et al., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, *J. Comput. Chem.* 30 (2009) 2785–2791.
- [40] M.F. Sanner, D. Stoffler, A.J. Olson, VIPER, a visual programming environment for Python, *Proceedings of the 10th International Python conference* (2002) 103–115.
- [41] A. Kolaskar, P.C. Tongaonkar, A semi-empirical method for prediction of antigenic determinants on protein antigens, *FEBS Lett.* 276 (1990) 172–174.
- [42] E.A. Emini, J.V. Hughes, D. Perlrow, J. Boger, Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide, *J. Virol.* 55 (1985) 836–839.
- [43] P. Karplus, G. Schulz, Prediction of chain flexibility in proteins, *Naturwissenschaften* 72 (1985) 212–213.
- [44] M.C. Jespersen, B. Peters, M. Nielsen, P. Marcatili, BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes, *Nucleic Acids Res.* 45 (2017) W24–W29.
- [45] P. Chou, G. Fasman, Prediction of the secondary structure of proteins from their amino acid sequence, *Adv. Enzymol.* 47 (1978) 45–148.
- [46] J. Parker, D. Guo, R. Hodges, 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 (1986) 5425–5432.
- [47] A. Rana, Y. Akhter, A multi-subunit based, thermodynamically stable model vaccine using combined immunoinformatics and protein structure based approach, *Immunobiology* 221 (2016) 544–557.
- [48] V. Solanki, V. Tiwari, Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against *Acinetobacter baumannii*, *Sci. Rep.* 8 (1) (2018) 9044.
- [49] H. Ghaffari-Nazari, J. Tavakkol-Afshari, M.R. Jaafari, S. Tahaghoghi-Hajghorbani, E. Masoumi, S.A. Jalali, Improving multi-epitope long peptide vaccine potency by using a strategy that enhances CD4+ T help in BALB/c mice, *PLoS One* 10 (2015) e0142563.
- [50] R.K. Pandey, R. Ojha, V.S. Aathmanathan, M. Krishnan, V.K. Prajapati, Immunoinformatics approaches to design a novel multi-epitope subunit vaccine against HIV infection, *Vaccine* 36 (2018) 2262–2272.
- [51] N. Hajighahramani, N. Nezafat, M. Eslami, M. Negahdaripour, S.S. Rahmatabadi, Y. Ghasemi, Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against *Staphylococcus aureus*, *Infect. Genet. Evol.* 48 (2017) 83–94.
- [52] R.K. Pandey, S. Sundar, V.K. Prajapati, Differential expression of miRNA regulates T cell differentiation and plasticity during visceral leishmaniasis infection, *Front. Microbiol.* 7 (2016) 206.
- [53] T. Farhad, N. Nezafat, Y. Ghasemi, Z. Karimi, S. Hemmati, N. Erfani, Designing of complex multi-epitope peptide vaccine based on omps of *Klebsiella pneumoniae*: an in silico approach, *Int. J. Pept. Res. Ther.* 21 (2015) 325–341.
- [54] N. Nezafat, Y. Ghasemi, G. Javadi, M.J. Khoshnoud, E. Omidinia, A novel multi-epitope peptide vaccine against cancer: an in silico approach, *J. Theor. Biol.* 349 (2014) 121–134.
- [55] Y. Yang, W. Sun, J. Guo, G. Zhao, S. Sun, H. Yu, Y. Guo, J. Li, X. Jin, L. Du, S. Jiang, In silico design of a DNA-based HIV-1 multi-epitope vaccine for Chinese populations, *Hum. Vaccines Immunother.* 11 (2015) 795–805.
- [56] P. Smialowski, A.J. Martin-Galiano, A. Mikolajka, T. Girschick, T.A. Holak, D. Frishman, Protein solubility: sequence based prediction and experimental verification, *Bioinformatics* 23 (2006) 2536–2542.
- [57] M. Hebditch, M.A. Carballo-Amador, S. Charonis, R. Curtis, J. Warwick, Protein-Sol: a web tool for predicting protein solubility from sequence, *Bioinformatics* 33 (2017) 3098–3100.

- [58] P. Chan, R.A. Curtis, J. Warwicker, Soluble expression of proteins correlates with a lack of positively-charged surface, *Sci. Rep.* 3 (2013) 3333.
- [59] T. Kosciolek, D.T. Jones, De novo structure prediction of globular proteins aided by sequence variation-derived contacts, *PLoS One* 9 (3) (2014) e92197.
- [60] B. Petersen, C. Lundegaard, T.N. Petersen, NetTurnP-neural network prediction of beta-turns by use of evolutionary information and predicted protein sequence features, *PLoS One* 5 (11) (2010) e15079.
- [61] M. Thayesen-Andersen, N.H. Packer, Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching, *Glycobiology* 22 (11) (2012) 1440–1452.
- [62] M. Karlberg, G. Margaryan, S. Wang, J. Ma, J. Xu, RaptorX server: a resource for template-based protein structure modeling, *InProtein Structure Prediction* 1137 (2014) 17–27.
- [63] J. Peng, J. Xu, RaptorX: exploiting structure information for protein alignment by statistical inference, *Proteins: Structure, Function, and Bioinformatics* 79 (S10) (2011) 161–171.
- [64] D. Xu, Y. Zhang, Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-step Atomic-level Energy Minimization, *Biophys. J.* 101 (2011) 2525–2534.
- [65] J. Zhang, Y. Liang, Y. Zhang, Atomic-Level Protein Structure Refinement Using Fragment-Guided Molecular Dynamics Conformation Sampling, *Structure* 19 (2011) 1784–1795.
- [66] S.C. Lovell, I.W. Davis, W.B. Arendall, P.I.W. Bakker, J.M. Word, M.G. Prisant, Richardson JS and Richardson DCStructure validation by Calpha geometry: phi, psi and Cbeta deviation, *Protein Struct. Funct. Genet.* 50 (2002) 437–450.
- [67] Al-Hakim, M. Hasan, R. Ali, M.F. Rabbee, Joy ZF. Marufatuzzahan, In-silico characterization and homology modeling of catechol 1,2 dioxygenase involved in processing of catechol- an intermediate of aromatic compound degradation pathway, *Glob. J. Sci. Front. Res. G Bio-Tech Genet.* 15 (2015) 1–13.
- [68] D.B. Craig, A.A. Dombrowski, Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins, *BMC Bioinf.* 14 (2013) 346.
- [69] F. Heil, H. Hemmi, H. Hochrein, F. Ampenberger, C. Kirschning, S. Akira, G. Lipford, H. Wagner, S. Bauer, Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8, *Science* 303 (5663) (2004) 1526–1529.
- [70] J. Cros, N. Cagnard, K. Woollard, N. Patey, S.Y. Zhang, B. Senechal, A. Puel, S.K. Biswas, D. Mosboush, C. Picard, J.P. Jais, Human CD14dim monocytes patrol and sense nucleic acids and viruses via TLR7 and TLR8 receptors, *Immunity* 33 (3) (2010) 375–386.
- [71] S.R. Comeau, D.W. Gatchell, S. Vajda, C.J. Camacho, ClusPro: a fully automated algorithm for protein-protein docking, *Nucleic Acids Res.* 32 (2004) W96–W99.
- [72] S. Macalino, S. Basith, N. Clavio, H. Chang, S. Kang, S. Choi, Evolution of In Silico Strategies for Protein-Protein Interaction Drug Discovery, *Molecules* 23 (8) (2018) 1963.
- [73] P. Extra-Kangueane, C. Nilofer, Protein-Protein Docking: Methods and Tools. In *Protein-Protein and Domain-Domain Interactions*, (2018), pp. 161–168.
- [74] D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, H.J. Wolfson, PatchDock and SymmDock: servers for rigid and symmetric docking, *Nucleic Acids Res.* 33 (2005) W363–W367.
- [75] D.M.F. Aalten, B.L. Groot, J.B.C. Findlay, H.J.C. Berendsen, A. Amadei, A Comparison of Techniques for Calculating Protein Essential Dynamics, *J. Comput. Chem.* 18 (2) (1997) 169–181.
- [76] K. Wuthrich, G. Wagner, Rene Richarz, Werner Braun, Correlations between internal mobility and stability of globular proteins, *Biophys. J.* (1980) 549–558.
- [77] F. Tama, C.L. Brooks, Symmetry, form, and shape: guiding principles for robustness in macromolecular machines, *Annu. Rev. Biophys. Biomol. Struct.* 35 (2006) 115–133.
- [78] Q. Cui, I. Bahar, normal mode analysis theoretical and applications to biological and chemical systems, *Briefings Bioinf.* 8 (5) (2007) 378–379.
- [79] J.R. Lopez-Blanco, J.I. Aliaga, E.S. Quintana-Orti, P. Chacon, iMODS: internal coordinates normal mode analysis server, *Nucleic Acids Res.* 42 (2014) W271–W276.
- [80] F.M. Awan, A. Obaid, A. Ikram, H.A. Janjua, Mutation-structure function relationship based integrated strategy reveals the potential impact of deleterious missense mutations in autophagy related proteins on hepatocellular carcinoma (HCC): a comprehensive informatics approach, *Int. J. Mol. Sci.* 18 (1) (2017) 139.
- [81] P.K. Prabhakar, A. Srivastava, K.K. Rao, P.V. Balaji, Monomerization alters the dynamics of the lid region in campylobacter jejuni CstII: an MD simulation study, *J. Biomol. Struct. Dyn.* 34 (4) (2016) 778–779.
- [82] J.R. Lopez-Blanco, J.I. Garzon, Chacon P. iMod, multipurpose normal mode analysis in internal coordinates, *Bioinformatics* 27 (20) (2011) 2843–2850.
- [83] A. Grote, et al., JCat: a novel tool to adapt codon usage of a target gene to its potential expression host, *Nucleic Acids Res.* 33 (2005) W526–W531.
- [84] N. Lohia, M. Baranwal, Immune responses to highly conserved influenza A virus matrix 1 peptides, *Microbiol. Immunol.* 61 (6) (2017) 225–231.
- [85] M. Terajima, J. Cruz, A.M. Leporati, L. Orphian, J.A. Babon, P. Pazoles, J. Jameson, F.A. Ennis, Influenza A virus matrix protein 1-specific human CD8+ T-cell response induced in trivalent inactivated vaccine recipients, *J. Virol.* 82 (18) (2008) 9283–9287.
- [86] T. Niwa, B.W. Ying, K. Saito, W. Jin, S. Takada, T. Ueda, H. Taguchi, Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins, *Proc. Natl. Acad. Sci. Unit. States Am.* 106 (2009) 4201–4206.
- [87] J. Kovacs, P. Chacon, R. Abagyan, Predictions of Protein Flexibility: First Order Measures. PROTEINS: Structure, Function and Bioinformatics, *Proteins* 56 (4) (2004) 661–668.
- [88] J.A. Lopez, C. Weilenman, R. Audran, et al., A synthetic malaria vaccine elicits a potent CD8(+) and CD4(+) T lymphocyte immune response in humans. Implications for vaccination strategies, *Eur. J. Immunol.* 3 (2001) 1989–1998.
- [89] K.L. Knutson, K. Schiffman, M.L. Disis, Immunization with a HER-2/neu helper peptide vaccine generates HER-2/neu CD8 T-cell immunity in cancer patients, *J. Clin. Invest.* 107 (2001) 477–484.
- [90] D.N.E.E. Bourdette, C. Smith, A highly immunogenic trivalent T cell receptor peptide vaccine for Knutson KL, Schiffman K, Disis ML (2001) Immunization with a HER-2/neu helper peptide vaccine generates HER-2/neu CD8 T-cell immunity in cancer patients, *J. Clin. Invest.* 107 (2005) 477–484.
- [91] Das, S. Chakraborty, M. Hasan, A.M. Rahman, In silico analysis to elect superior bacterial alkaline protease for detergent and leather industries, *Journal of Advances in Biotechnology* 5 (2015) 685–698.
- [92] R.J. Russell, P.S. Kerry, D.J. Stevens, D.A. Steinhauer, S.R. Martin, S.J. Gamblin, J.J. Skehel, Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion, *Proc. Natl. Acad. Sci. U. S. A* 105 (46) (2008) 17736–17741.
- [93] E.V. Shtykova, L.A. Baratova, N.V. Fedorova, V.A. Radyukhin, A.L. Ksenofontov, V.V. Volkov, A.V. Shishkov, A.A. Dolgov, L.A. Shilova, O.V. Batishchev, C.M. Jeffries, Structural analysis of influenza A virus matrix protein M1 and its self-assemblies at low pH, *PLoS One* 8 (12) (2013) e82431.
- [94] R. Lamb, V. Krug, Orthomyxoviridae: the viruses and their replication, in: D. Knipe, P. Howley (Eds.), *Fundamental virology*, fourth ed., Lippincott Williams & Wilkins, Philadelphia, Pa, 2001, pp. 725–770.
- [95] E.D. Kilbourne, N.H. Arden, Inactivated influenza vaccines, in: S.A. Plotkin, W.A. Orenstein (Eds.), *Vaccines*, third ed., W.B. Saunders Company, Philadelphia, PA, 1999, pp. 531–551.
- [96] N.M. Bouvier, P. Palese, The biology of influenza viruses, *Vaccine* 26 (2008) D49–D53.
- [97] A. Naskalska, E. Szolajska, L. Chaperot, J. Angel, J. Plumas, J. Chroboczek, Influenza recombinant vaccine: Matrix protein M1 on the platform of the adenovirus dodecahedron, *Vaccine* 27 (2009) 7385–7393.
- [98] N.R. Cooper, G.R. Nemerow, The role of antibody and complement in the control of viral infections, *J. Invest. Dermatol.* 83 (1984) 121–127.
- [99] R. Bacchetta, S. Gregori, M.G. Roncarolo, CD4+ regulatory T cells: mechanisms of induction and effector function, *Autoimmun. Rev.* 4 (2005) 491–496.
- [100] J. Igietseme, F. Eko, Q. He, C.M. Black, Antibody regulation of T-cell immunity: implications for vaccine strategies against intracellular pathogens, *Expert Rev. Vaccines* 3 (2014) 23–34.
- [101] L. Garcia KC Teyston, I.A. Wilson, Structural basis of T cell recognition, *Annu. Rev. Immunol.* 17 (1999) 369–397.
- [102] B. Shrestha, Role of CD8+ T cells in control of West Nile virus infection, *J. Virol.* 12 (2004) 8312–8321.
- [103] H.G. Rammensee, K. Falk, O. Rotzschke, Peptides naturally presented by MHC class I molecules, *Annu. Rev. Immunol.* 11 (1993) 213–244.
- [104] T.M. McKeever, S. Lewis, C. Smith, R. Hubbard, Vaccination and allergic disease: a birth cohort study, *Am. J. Public Health* 94 (6) (2004) 985–989.
- [105] T. Mohan, P. Verma, D.N. Rao, Novel adjuvants & delivery vehicles for vaccines development: a road ahead, *Indian J. Med. Res.* 138 (5) (2013) 779.
- [106] V. Solanki, V. Tiwari, Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against *Acinetobacter baumannii*, *Scientific report* 8 (1) (2018) 9044.
- [107] T. Mohan, C. Sharma, A.A. Bhat, D.N. Rao, Modulation of HIV peptide antigen specific cellular immune response by synthetic α - and β -defensin peptides, *Vaccine* 31 (2013) 1707–1716.
- [108] D. Yang, A. Biragyn, L.W. Kwak, J.J. Oppenheim, Mammalian defensins in immunity: more than just microbial, *Trends Immunol.* 23 (2002) 291–296.
- [109] T. Mohan, D. Mitra, D.N. Rao, Nasal delivery of PLG microparticle encapsulated defensin peptides adjuvanted gp41 antigen confers strong and long-lasting immunoprotective response against HIV-1, *Immunol. Res.* 58 (2013) 139–153.
- [110] N. Lohia, M. Baranwal, Identification of Conserved Peptides Comprising Multiple T Cell Epitopes of Matrix 1 Protein in H1N1 Influenza Virus, *Viral Immunol.* 28 (10) (2015) 570–579.
- [111] C.Y. Wu, A. Monie, X. Pang, C.F. Hung, T.C. Wu, Improving therapeutic HPV peptide-based vaccine potency by enhancing CD4+ T help and dendritic cell activation, *J. Biomed. Sci.* 17 (2010) 88.
- [112] J.B. Clarage, T. Romo, B.K. Andrews, B.M. Pettitt, G.N. Phillips, A sampling problem in molecular dynamics simulations of macromolecules, *Proc. Natl. Acad. Sci. U.S.A.* 92 (1995) 3288–3292.
- [113] D.L.D. Caspar, Problems in simulating macromolecular movements, *Structure* 3 (1995) 327–329.