



Immunoinformatics approach for a novel multi-epitope subunit vaccine design against various subtypes of Influenza A virus

Shipra Sharma^{a,1}, Vibha Kumari^{a,c,1}, Bajarang Vasant Kumbhar^b, Amrita Mukherjee^a, Ruchika Pandey^a, Kiran Kondabagil^{a,*}

^a Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai, India

^b HaystackAnalytics Private Limited, Society for Innovation and Entrepreneurship (SINE), Indian Institute of Technology Bombay, Powai, Mumbai, India

^c Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria 3800, Australia

ARTICLE INFO

Keywords:

Influenza A virus
Innate immunity
Cytotoxic T cell epitopes
Helper T cell epitopes
Vaccine
Immunoinformatics

ABSTRACT

Vaccination is the best strategy for the control and prevention of contagious diseases caused by Influenza A viruses. Extraordinary genetic variability and continual evolvability are responsible for the virus having survival and adaptation to host cell immune response, thus rendering the current influenza vaccines with suboptimal effectiveness. Therefore, in the present study, using a novel immunoinformatics approach, we have designed a universal influenza subunit vaccine based on the highly conserved epitopic sequences of rapidly evolving (HA), a moderately evolving (NP) and slow evolving (M1) proteins of the virus. The vaccine design includes 2 peptide adjuvants, 26 CTL epitopes, 9 HTL epitopes, and 7 linear BCL epitopes to induce innate, cellular, and humoral immune responses against Influenza A viruses. We also analyzed the physicochemical properties of the designed construct to validate its thermodynamic stability, hydrophilicity, PI, antigenicity, and allergenicity. Furthermore, we predicted a highly stable tertiary model of the designed subunit vaccine, wherein additional disulfide bonds were incorporated to enhance its stability. The molecular docking and molecular dynamics simulations of the refined vaccine model with TLR3, TLR7, TLR8, MHC-I and MHC-II showed stable vaccine and receptors complexes, thus confirming the immunogenicity of the designed vaccine. Collectively, these findings suggest that our multi-epitope vaccine construct may confer protection against various strains of influenza A virus subtypes, which could prevent the need for annual reformulation of vaccine and alleviate disease burden.

1. Introduction

For more than a century, Influenza A virus (IAV), a highly infectious pathogen, has posed a major global health concern with significant mortality and morbidity. IAV, a single-stranded segmented RNA virus, invariably outsmarts its host and causes acute, contagious respiratory illness (Taubenberger and Morens, 2010). The World Health Organisation, in 2017, estimated that seasonal influenza may cause 290,000–650,000 deaths annually increasing the global social and economic burden of Influenza-triggered respiratory diseases (http://www.who.int/influenza/surveillance_monitoring/bod/en/). Owing to its error-prone polymerase activity, IAV is highly mutable and a rapidly evolving pathogen, and has acquired resistance to commercially available drugs like Amantadine, Oseltamivir, and Zanamivir (Bright et al., 2006; Hurt et al., 2009; Moscona, 2005). Moreover,

influenza viruses can quickly cross the species barrier by recombination of HA genes, thus creating a novel influenza strain to which a majority of the population lacks pre-existing immunity.

Vaccination is the most appropriate practice to check the spread of IAV. However, the lack of an efficacious method to predict the circulating strains often leads to a mismatch between the annual vaccine and circulating viruses. To illustrate, the sudden emergence of the 2009 influenza A/H1N1 pandemic strain (Hamilton, 2009) and the recent appearance of the highly pathogenic avian A/H7N9 strain in 2013 in China (Zhang et al., 2014), underscores the unpredictable nature of the virus and the difficulty in comprehending the emergence of new strains. Hitherto, FDA-approved, licensed seasonal vaccines include trivalent/quadrivalent Inactivated Influenza Vaccines (IIVs), Recombinant Influenza Vaccine (RIV) and Live Attenuated Influenza Vaccine (LAIV), which target the three most representative virus types in circulation

* Corresponding author at: Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai, Maharashtra 400076, India.
E-mail address: kirankondabagil@iitb.ac.in (K. Kondabagil).

¹ Contributed equally.

(two subtypes of influenza A viruses and one influenza B virus) in the succeeding year (Grohskopf et al., 2018). But despite the commercial availability of vaccines, the weakening vaccine-specific antibodies and the antigenic drift of influenza viruses over time necessitates annual revaccination (Houser and Subbarao, 2015). Hence, the currently available vaccines have proven to be inefficient in providing complete protection against pandemic and seasonal strains of influenza across all age groups. Therefore, there is an urgent need to design and develop a “universal vaccine” that induces extensive cross-protection and long-lasting immunity against all subtypes of influenza viruses, including novel and existing strains.

Influenza viruses is classified into four types (A, B, C & D) based on the highly conserved internal proteins; matrix protein 1 (M1), matrix protein 2 (M2) and nucleoprotein (NP) (Freed and Martin, 2013; Hause et al., 2014). IAVs are further sub-divided into subtypes based on the antigenicity of their hemagglutinin (HA) and neuraminidase (NA) surface glycoproteins. The HA and NA surface proteins of IAV are the major targets of immune responses elicited by vaccination, especially humoral responses (Krammer and Palese, 2015). The ectodomain of matrix protein 2 (M2e) of IAV is another prominent target for an influenza A vaccine (Deng et al., 2015). However, the long-drawn production cycle of vaccines and the resultant unavailability during a pandemic fails the very purpose of the licensed vaccines. This necessitates alternate approaches to vaccine development.

Unlike the traditional live-attenuated or inactivated vaccines, subunit vaccines lack an infectious particle and are thus considered to be safer than the former. The increasing commercial use of subunit vaccines to counter various diseases has also paved the way for its use in the design of a novel influenza vaccine. Recombivax Hb, a recombinant subunit vaccine composed of hepatitis B virus surface antigen, is administered for prevention of infection caused by all known subtypes of hepatitis B virus (Tan and Jiang, 2016).

Lately, several approaches have been developed to overcome these challenges and improve the immunogenicity and efficacy of influenza vaccines. A recent IAV vaccine strategy, computationally optimized broadly reactive antigen (COBRA), has been employed to develop COBRA HA VLP that triggered wider antibody responses against multiple clades of H5N1 in nonhuman primates, mice, and ferrets (Giles et al., 2012; Giles and Ross, 2011). Further, COBRA-based HA vaccines have been developed that elicits a broadly reactive response in vaccinated mice challenged with pandemic H1N1 influenza virus (Carter et al., 2016). A new vaccine, Multimeric-001 comprising conserved, linear epitopes has been developed, that activate both cellular and humoral immune responses against several influenza A and B strains in elderly participants (Atsmon et al., 2014).

Our approach to developing a universal flu vaccine is to combine the diverse antigenic epitopes with the most conserved epitopes of the IAV, which may have the capability to elicit broadly protective immune response. To achieve this, we selected three proteins of IAV, namely, HA, M1, and NP, as candidates for the development of the universal subunit vaccine. HA is the surface glycoprotein of IAV and a major antigenic determinant with the highest variability in its sequence. M1 and NP are multi-functional structural proteins with comparatively conserved sequences (Zheng et al., 2014). Mass Spectroscopy analysis of purified virion has shown that with 2200, 530, and 294 copies, respectively, the M1, NP, and HA are the most abundant proteins in the virion (Hutchinson et al., 2014). Additionally, structural proteins NP and M1 are targeted for developing strong cross-protective T-cell responses as virus-specific CD8⁺ cytotoxic T-lymphocytes (CTL). Here, we report an efficient *in silico* design strategy for the construction of a recombinant subunit vaccine. For this purpose, we analyzed sequences of 50 strains of IAV including H1N1, H2N2, H3N2, H5N1, and H7N9 (Supplemental Table 1), isolated from Asia (Indian subcontinent, China, South East Asia), Africa, Australia, North America, and South America (Fig. 1A) between 1918 and 2016. Suitable CTLs, Helper-T-lymphocytes (HTL) and B-cell lymphocytes (BCL) epitopes were selected from all the

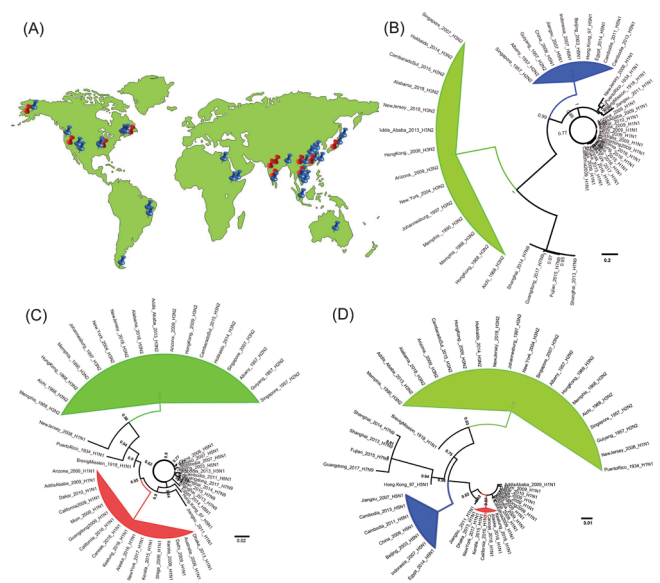


Fig. 1. Selection of strains of Influenza A virus. (A) The world map depicting the origin of 50 strains from different continents selected for this study. Red pin refers to pandemic strains. Blue pins refer to seasonal strains. (B–D) Maximum likelihood phylogeny based on amino acid sequences of HA (B), NP (C) and M1 (D) of the IAV strains obtained through flu database. The clustering of similar strains is shown in the coloured clade. Scale for each phylogeny is represented below the tree. Bootstrap values below 0.1 are not shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

selected subtypes and a vaccine construct of 79.5 kDa was designed with appropriate linkers and two adjuvants. CTLs induced after infection with seasonal H1N1 and H3N2 influenza viruses are known to cross-react with IAVs of the H5N1 subtype (Kreijtz et al., 2008; Lee et al., 2008). Hence, such a vaccine construct can potentially elicit a cell-mediated and humoral immune response against different subtypes of IAV, both currently circulating strains and newly evolved ones.

2. Methods

2.1. Selection of strains of IAV for efficient vaccine construct

The 50 strains selected for this study included representative strains of four pandemic outbreaks of IAV (1918, 1957, 1968, and 2009) as well as seasonally circulating strains from different parts of the world in the last hundred years (Supplemental Table 1). These strains belong to five subtypes of IAV, namely, H1N1, H2N2, H3N2, H5N1, and H7N9. Of them, H1N1, H2N2 and H3N2 strains selected in the study have known to cause widespread disease in humans, whereas, the other two subtypes' strains are highly pathogenic for poultry and have crossed the species barrier to infect humans at different times (Freed and Martin, 2013). Amino acid sequences of HA, NP, and M1 proteins from these fifty strains of IAV were obtained from the NCBI flu database (<https://www.ncbi.nlm.nih.gov/genomes/FLU/Database>).

Only full-length protein sequences were taken for epitope prediction except for the HA sequence of Brevig Mission 1918 H1N1 (Accession ID: AAD17218.1) and Guiyang 1957 H2N2 (Accession ID: ACD85231.1) as their complete protein sequence is not available. IAV strains of various subtypes from Europe (A/Bretagne/7608/2009 (H1N1); A/Germany/2018/H3N2; A/Netherlands/56/1963 (H2N2)) were also analyzed, but have not been included in the study as their predicted epitopes were similar to the epitopes predicted for the strains selected from Asia and North America.

2.2. Phylogenetic analysis of NP, M1 and HA sequences used in the study

For reconstructing the phylogenies, amino acid sequences were aligned using clustal Ω and the trees were generated with default parameters in FastTree v.2., using a maximum likelihood method based on Jones–Taylor–Thornton (JTT) substitution model and 1000 bootstrap replicates (Price et al., 2009).

2.3. Identification and prediction of CTL binding epitopes

NetCTL 1.2 server (<http://www.cbs.dtu.dk/services/NetCTL/>) was used to predict CTL epitopes in HA, M1, and NP proteins (Larsen et al., 2007). For each of the three proteins, amino acid sequences from all 50 strains were submitted collectively in the FASTA format. Taking the default parameters, we selected four MHC class I supertypes (A1, A2, A3, and B7) individually for each type of protein sequence. It has been reported that this group of MHC class-I supertypes covers approximately 83.0–88.5% of the population in different ethnic groups (Lund et al., 2004).

2.4. Conservancy analysis of CTL epitopes

Immune epitope database (IEDB) epitope conservancy analysis tool (<http://tools.iedb.org/conservancy/>) was used to identify the conserved CTL epitopes in HA, M1, and NP protein sequences of all 50 strains (Bui et al., 2007). In the case of M1 and NP protein, the top five epitopes predicted for one strain were randomly selected and screened against the complete set of fifty sequences of the respective protein. In the case of HA, the top five epitopes per MHC class I supertype were selected separately for each subtype group (H1N1, H2N2, H3N2, H5N1, and H7N9) and then screened for their conservancy against HA sequences of fifty strains. The threshold for linear sequence identity was 100%.

2.5. Identification and prediction of helper T-lymphocytes (HTL) binding epitopes

IEDB MHC-II binding predictions tool (<http://tools.iedb.org/mhcii/>) was used for identifying HTL binding epitopes in HA, M1, and NP protein. “IEDB recommended” prediction method was used along with a reference allele set of 27 MHC Class II HLA alleles. The alleles included in 27 reference allele set are as follows: HLA-DRB1*01:01, HLA-DRB1*03:01, HLA-DRB1*04:01, HLA-DRB1*04:05, HLA-DRB1*07:01, HLA-DRB1*08:02, HLA-DRB1*09:01, HLA-DRB1*11:01, HLA-DRB1*12:01, HLA-DRB1*13:02, HLA-DRB1*15:01, HLA-DRB3*01:01, HLA-DRB3*02:02, HLA-DRB4*01:01, HLA-DRB5*01:01, HLA-DQA1*05:01/DQB1*02:01, HLA-DQA1*05:01/DQB1*03:01, HLA-DQA1*03:01/DQB1*03:02, HLA-DQA1*04:01/DQB1*04:02, HLA-DQA1*01:01/DQB1*05:01, HLA-DQA1*01:02/DQB1*06:02, HLA-DPA1*02:01/DPB1*01:01, HLA-DPA1*01:03/DPB1*02:01, HLA-DPA1*01/DPB1*04:01, HLA-DPA1*03:01/DPB1*04:02, HLA-DPA1*02:01/DPB1*05:01, HLA-DPA1*02:01/DPB1*14:01. This reference allele set is reported to cover 99% of the population (Wang et al., 2008). In the case of M1 and NP, sequences from fifty strains were submitted for prediction of MHC-II binding peptides against 27 reference allele sets through IEDB recommended method. For HA protein, amino acid sequences of five different subtype groups were submitted individually with the same parameters as mentioned above. The primary selection of HTL epitopes was performed based on the mean percentile rank. Lower percentile translates to a higher binding affinity of the predicted peptide to the respective HLA (MHC-class II) alleles. The top five epitopes were selected manually for M1, NP and five subtype groups of HA protein based on their lowest percentile rank irrespective of the HLA-allele (MHC-class II) binding to these epitopes and overlapping epitopes were avoided.

2.6. Conservancy analysis of selected HTL binding epitopes and allele coverage identification

IEDB epitope conservancy analysis tool was used for further sorting of the top 5 HTL epitopes. Screening of M1 and NP HTL epitopes for their linear sequence conservancy with a threshold of 100% sequence identity was done against sequences from all the fifty strains, whereas HA HTL epitopes were screened against sequences of their respective subtype groups. In this study, we have defined ‘allele coverage’ as the number of different HLA-alleles (MHC class II) to which a particular epitope binds below a percentile rank of 10, irrespective of its linear sequence conservancy. Allele coverage was manually identified for only those HTL epitopes which exhibited 100% conservancy in a maximum number of respective sequences.

2.7. Population coverage analysis of CTLs and HTLs

The population coverage analysis of CTLs and HTLs was executed using the IEDB population coverage server (<http://tools.iedb.org/population/>) (Bui et al., 2006). For this exercise, the entire global population was studied. For 9 HTLs, we considered all MHC class II alleles that bind to each epitope, up to the percentile rank of 10. Analogous to the prediction of epitopes by NetCTL 1.2 server against MHC class I super-types, in the case of CTLs, additional information was required about the binding of MHC class I alleles to these epitopes. Using IEDB MHC-I binding prediction tool (<http://tools.iedb.org/mhci/>) (Lundegaard et al., 2008), we predicted the MHC class I alleles binding to each CTL epitope against a reference allele set. The reference allele set given on IEDB MHC-I binding prediction tool comprises of allele variants which bind to epitopes with the size of 9 or 10 amino acids (Weiskopf et al., 2013). However, we had examined 27 alleles that only bind to epitopes with a size of 9 amino acids. The reference set of 27 alleles follows: HLA-A*01:01, HLA-A*02:01, HLA-A*02:03, HLA-A*02:06, HLA-A*03:01, HLA-A*11:01, HLA-A*23:01, HLA-A*24:02, HLA-A*26:01, HLA-A*30:01, HLA-A*30:02, HLA-A*31:01, HLA-A*32:01, HLA-A*33:01, HLA-A*68:01, HLA-A*68:02, HLA-B*07:02, HLA-B*08:01, HLA-B*15:01, HLA-B*35:01, HLA-B*40:01, HLA-B*44:02, HLA-B*44:03, HLA-B*51:01, HLA-B*53:01, HLA-B*57:01, HLA-B*58:01. CTL epitopes binding to one or more MHC class I alleles within percentile rank 1 were selected and their respective combinations were used in the population coverage analysis.

2.8. Prediction of linear B-cell epitopes and conservancy analysis

B-Cell epitope prediction server (BCPREDS <http://ailab.ist.psu.edu/bcpred/predict.html>) was used to identify B-cell binding epitopes in HA, M1, and NP proteins (El-Manzalawy et al., 2008). The protein sequence from 50 strains of IAV was submitted individually (total 150 sequences) for B-cell epitope prediction. The final selection of B-cell epitopes was done based on conservancy analysis. For each type of protein, epitope from one strain was randomly selected as a reference and screened for its linear sequence conservancy at a threshold of 100% linear sequence identity, against 50 sequences of the respective protein.

2.9. Physicochemical characterization of predicted construct

The vaccine construct was analyzed for several physicochemical parameters (composition, molecular weight, theoretical isoelectric point (pI), number of charged residues, extinction coefficient, half-life, instability index, and aliphatic index) using the ProtParam tool (<http://web.expasy.org/protparam/>) (Gasteiger et al., 2005).

2.10. Allergenicity and antigenicity prediction

Allergenicity and antigenicity were predicted by the AllerTop v.2.0 (<https://www.ddg-pharmfac.net/AllerTOP/>) (Dimitrov et al., 2013)

and ANTIGENpro (<http://scratch.proteomics.ics.uci.edu/>) tools, respectively.

2.11. Secondary and tertiary structure prediction and validation

For secondary and tertiary structure prediction RaptorX (<http://raptorx.uchicago.edu/about/>) (Källberg et al., 2012) webserver was used. Upon providing an input sequence, the server predicts the final 3D structure without close homologs in the Protein Data Bank (PDB). Moreover, it assigns confidence scores to describe the quality of the predicted model. The structure was further refined using ProSA web-server (<https://prosa.services.came.sbg.ac.at/prosa.php>) (Wiederstein and Sippl, 2007) and validated by generating Ramachandran Plot using WinCoot software (Emsley et al., 2010).

2.12. Disulfide engineering

The final vaccine construct was mutated to incorporate disulfide bonds *in silico* using “Disulfide by Design” server v2.0 (<http://cptweb.cpt.wayne.edu/DbD2/>). The server displays amino acid pairs that can be mutated to cysteines so that a disulfide bond is formed. The pairs are selected based on the B-score which is a measure of dynamic mobility of an atom (Craig and Dombkowski, 2013). The best pair was selected and replaced with cysteine in the original construct and a PDB file of the same was generated using RaptorX.

2.13. Molecular docking with TLR3, TLR7, TLR8, MHC-I and MHC-II receptors

Molecular docking was performed by the PatchDock server (Schneidman-Duhovny et al., 2005) (<http://bioinfo3d.cs.tau.ac.il/PatchDock/>). The results from the PatchDock were further refined using the FireDock server (<http://bioinfo3d.cs.tau.ac.il/FireDock/php.php>). The FireDock selects and displays the top ten solutions as the final docking results (Mashiach et al., 2008). The crystal structure 1ZIW.pdb of TLR3, 3W3G.pdb of TLR8, 2X4O.pdb of MHC-I, and 1T5W.pdb of MHC-II were used for molecular docking with the vaccine construct. The missing residues of 1ZIW.pdb were modeled using Modeller (Webb and Sali, 2017) through the Chimera interface (Pettersen et al., 2004). Three-dimensional models of the TLR7 receptor protein (Uniprot ID: Q9YK1) were build using the Swiss-MODEL server (Waterhouse et al., 2018). The least energy docked structures of vaccine construct with receptor proteins were further used as a starting structure for molecular dynamics simulation.

2.14. Molecular dynamics simulation

The structural stability of the vaccine construct with receptor proteins was checked by the molecular dynamics (MD) simulation using the Gromacs2018.1 (Van Der Spoel et al., 2005). The amber ff99SBildn force field was used for the MD simulation. The docked structure of vaccine and receptor proteins such as the vaccine-TLR3, vaccine-TLR7, vaccine-TLR8, vaccine-MHC-I, and vaccine-MHC-II complexes were solvated with TIP3P waters in a cubic box of 10 Å length and neutralized by adding the required number of counter ions using ‘xleap’ module of AmberTools18 (Case et al., 2018). The amber generated topology and coordinates files were then converted to the Gromacs friendly topology using the ‘ParmEd tool’. The energy minimization was performed using the steepest descent followed by the conjugate gradient method over all the complexes. Next, the 500 ps MD simulations using the NVT ensembles and 500 ps steps of MD simulation using NPT ensembles with position restraint were performed to equilibrate the systems. Finally, a 20 ns production simulation was performed for all the vaccine-receptor complexes. The cut-off distance of 1.0 nm with a Fourier spacing of 0.16 nm and an interpolation order of 4 was used for the simulation. The particle mesh Ewald (PME) method was used to calculate the long-range

electrostatic interactions. The trajectories were incremented at every 2-fs time step, similar to a previous study (Kumbhar and Bhandare, 2020). The simulated trajectories were analyzed using Chimera and Visual molecular dynamics (VMD) (Humphrey et al., 1996).

2.15. Codon optimization and in-silico cloning

For expressing the predicted protein construct in an expression system, first, the nucleotide sequence was generated using Sequence Manipulation Suite: Reverse Translate (http://www.bioinformatics.org/sms2/rev_trans.html). It generates two types of output nucleotide sequences, one using most common codons and the other using consensus codons. Optimization of codons is necessary to achieve maximum protein expression as the expression system, *E.coli* K12 differs from the human host. Java codon adaptation tool (<http://www.jcat.de/>) was employed to obtain the Codon adaptation index (CAI) and GC content, which also checks for the desired restriction sites to be absent within the construct. Finally, *E.coli* pET-28a (+) vector was used to clone the constructed gene using restriction sites flanking the final construct sequence.

3. Results

3.1. Evolutionary analysis of different strains of human IAV across numerous subtypes

IAV is subjected to host immune selection pressure and undergoes rapid evolution, especially when the virus crosses the species barrier (Joseph et al., 2017). The phylogenetic analyses using NP, M1, and HA indicate that the 2009 pandemic H1N1 HA, NP, and M1 (Fig. 1B, C, and D, respectively) have diverged from seasonal or H5 influenza viruses. The clustering of seasonal or H5 influenza forming a separate clade suggests their evolutionary relationship whereas they showed collective divergence from the 2009 pandemic H1N1 strain. The divergence of HA protein is higher as compared to NP and M1 suggesting rapid evolution in HA protein. M1 is an extremely slow evolving protein as compared to the other two (Fig. 1D). Thus, phylogenies suggest that M1 and NP are highly conserved proteins of IAV. Vaccines aimed at evolutionarily conserved influenza virus proteins have been reported to protect against multiple IAV subtypes in animal models (Epstein et al., 2005) and, pending further successful clinical development, remain promising alternatives for the HA-based, strain-specific vaccines currently in use. Therefore, to build an efficient universal flu vaccine, we selected a rapidly evolving (HA), a moderately evolving (NP) and slow evolving (M1) proteins of IAV.

A subunit vaccine employing recombinant DNA techniques is an effective yet affordable method for disease control. Therefore, in the present study, we have designed a subunit vaccine comprising of conserved antigenic peptides (CTL, HTL, and BCL epitopes) of three proteins of IAV that may elicit innate, cellular and humoral immune responses.

3.2. Selection of CTL epitopes

CTL epitopes are peptides that bind to MHC class I proteins and activate cytotoxic T- lymphocytes. NetCTL 1.2 server predicts CTL epitopes of 9-mer length based on three parameters, namely, C-terminal proteosomal cleavage, transport efficiency of transporter associated with antigen processing (TAP), and binding affinity to MHC class-I protein. Top five CTL epitopes of M1 and NP protein against each of the class I MHC supertypes (A1, A2, A3, and B7) were selected for conservancy analysis. In the case of HA protein, CTL epitopes were selected separately for each subtype group. One epitope for each MHC class I supertype having 100% conservancy (100% linear sequence identity) in a maximum number of strains was selected. A total of 18 CTL epitopes of HA protein and 4 each of M1 and NP protein were confirmed

for the final construct (Supplemental Table 2).

3.3. Selection of HTL epitopes

HTL epitopes are MHC class II protein binding peptides that activate helper T-lymphocytes. IEDB MHC-II binding prediction tool predicts MHC class II binding peptides based on the type of alleles selected and the combination of different prediction methods. Top five epitopes of M1, NP, and five subtype HA proteins having the lowest combined percentile rank (best score) were selected irrespective of the HLA (MHC class II) allele they bind to. A total of 25 HTL epitopes of HA and 5 HTL epitopes each of M1 and NP protein were screened for their 100% linear sequence conservancy in all the fifty strains. Further, the stringency of the selection of HTL epitopes for each category of protein was achieved based on their conservancy and HLA (MHC class II) allele coverage. A total of 9 HTL epitopes (5 for HA and 2 each for M1 and NP) with 100% conservancy in the maximum number of strains and maximum allele coverage were selected for the subunit vaccine (Supplemental Table 3).

3.4. Selection of linear B cell epitopes

Linear B Cell Lymphocytes (BCL) epitopes were predicted using BCPREDS server and the top two epitopes of each category (M1, NP and five subtypes of HA) were screened for their conservancy. Due to the length of the epitopes (20 amino acids), 100% sequence conservancy was not found in all fifty strains. Hence, 5 linear BCL epitopes for HA protein and 1 linear BCL epitope each for M1 and NP protein were selected based on the highest conservancy in a maximum number of strains (Supplemental Table 4).

3.5. Construction of subunit vaccine, their physicochemical properties and population coverage

Our subunit vaccine construct for IAV contains 26 CTL epitopes, 9 HTL epitopes, and 7 linear BCL epitopes. Two peptide adjuvants, human β -defensin-3 (45 amino acids) and LL37 (37 amino acids, also known as hCAP18) are also included to boost immunogenicity of the construct. Human β -defensin-3 is a cationic antimicrobial peptide (Bartels et al., 2000) which has been reported to specifically interact with Toll-like receptors1/2 (TLR1/TLR2) heterodimer and activate several downstream pathways involved in both innate and adaptive immune response (Funderburg et al., 2007). LL37 is also a cationic anti-microbial peptide and is the only member of the cathelicidin family found in humans. Unlike human β -defensin-3, LL37 is able to modulate the activity of multiple TLRs (TLR-3, 4, 5, 7, 8, 9, and 2/1 heterodimer) (Lai et al., 2011). The “combination adjuvant” (Levast et al., 2014) of human β -defensin-3 and LL37 is a novel formulation which has not only generated higher antigenicity without any allergenicity but is also

expected to act synergistically to produce the Th-1 based adaptive immune response. Th-1 immune response plays an important role in the elimination of intracellular pathogens like viruses by phagocyte activation and antibody-dependent cellular cytotoxicity (Netea et al., 2005).

Both the adjuvants and different epitopes are linked by a variety of suitable two to five amino acid residue linkers that are known to enhance the structural and physicochemical stability of the construct. Without linkers, junctional epitopes (new epitopes formed by tandem positioning of two or more naturally existing epitopes) might be formed that could compromise the functionality and stability of the construct (Livingston et al., 2002; Nezafat et al., 2014). EAAAK has been used between human β -defensin-3 and LL-37 as well as between LL-37 and rest of the construct (Arai et al., 2001). Similarly, CTL epitopes are separated by AAY linker, HTL epitopes by GPGPG linker, and BCL epitopes by KK linker (Fig. 2). Different linker epitope combinations not only avoid formation of junctional epitopes but also facilitate the processing of antigens for further presentation by acting as the cleavage site for their respective proteolytic pathways (Arai et al., 2001; Hajjighahramani et al., 2017; Livingston et al., 2002; Nezafat et al., 2014; Yano et al., 2005). The design of the final construct is shown in Fig. 2. The subunit vaccine construct with all the linkers and epitopes consists of 736 amino acids (79.5 kDa) with a theoretical pI of 9.55. The instability index of the vaccine construct was predicted to be 35.05, suggesting that the protein is stable. Additionally, the half-life of the subunit vaccine protein was predicted to be 30 h in mammalian reticulocytes, 10 h in *E. coli* and 20 h in yeast. The aliphatic index and the grand average of the hydropathicity (GRAVY) were found to be 84.95 and -0.034 , suggesting that the protein is thermodynamically stable and hydrophilic in nature, respectively. As predicted by ANTIGENpro, our designed vaccine construct has the probable antigenicity score of 0.569, suggesting the antigenic nature of the construct. Additionally, AllerTOP output suggested the non-allergenic nature of the subunit vaccine (Supplemental Table 5). Due to the more stringent selection of MHC class I alleles (within 1 percentile rank, as suggested by IEDB analysis tool) binding to CTL epitopes, only 18 out of our set 26 CTL epitopes could be used for the population coverage calculation. The combined world population coverage of 18 CTLs and all 9 HTLs was obtained to be 99%.

3.6. Secondary and tertiary structure prediction

RaptorX prediction showed that the construct is composed of 38% alpha-helix, 14% beta-sheets, and 46% coiled region. The RaptorX-predicted three-dimensional structure of the construct is shown in Fig. 3A. RaptorX used four different templates to model the final tertiary structures, namely, PDB id: 1kj6A (Human β -Defensin 3), 4av3A (membrane pyrophosphatase of *Thermotoga Maritima*), 5ykca (H5 hemagglutinin from chicken/Taiwan/0502/2012), and 4ysha (glycine oxidase from *Geobacillus kaustophilus*). The p-value of the modelled

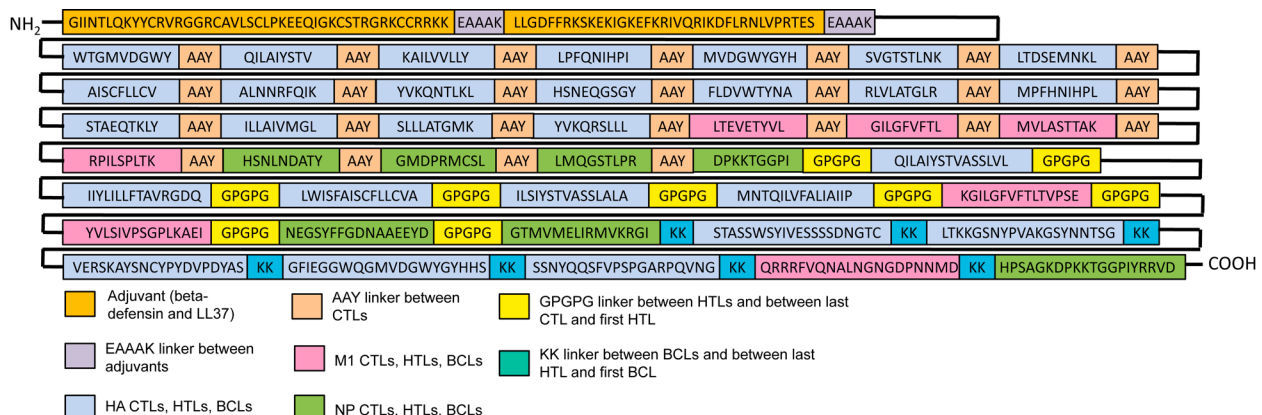


Fig. 2. Schematic of vaccine construct consists of two adjuvants, CTL and HTL epitopes from HA, NP and M1 fused to each other with the appropriate linkers.

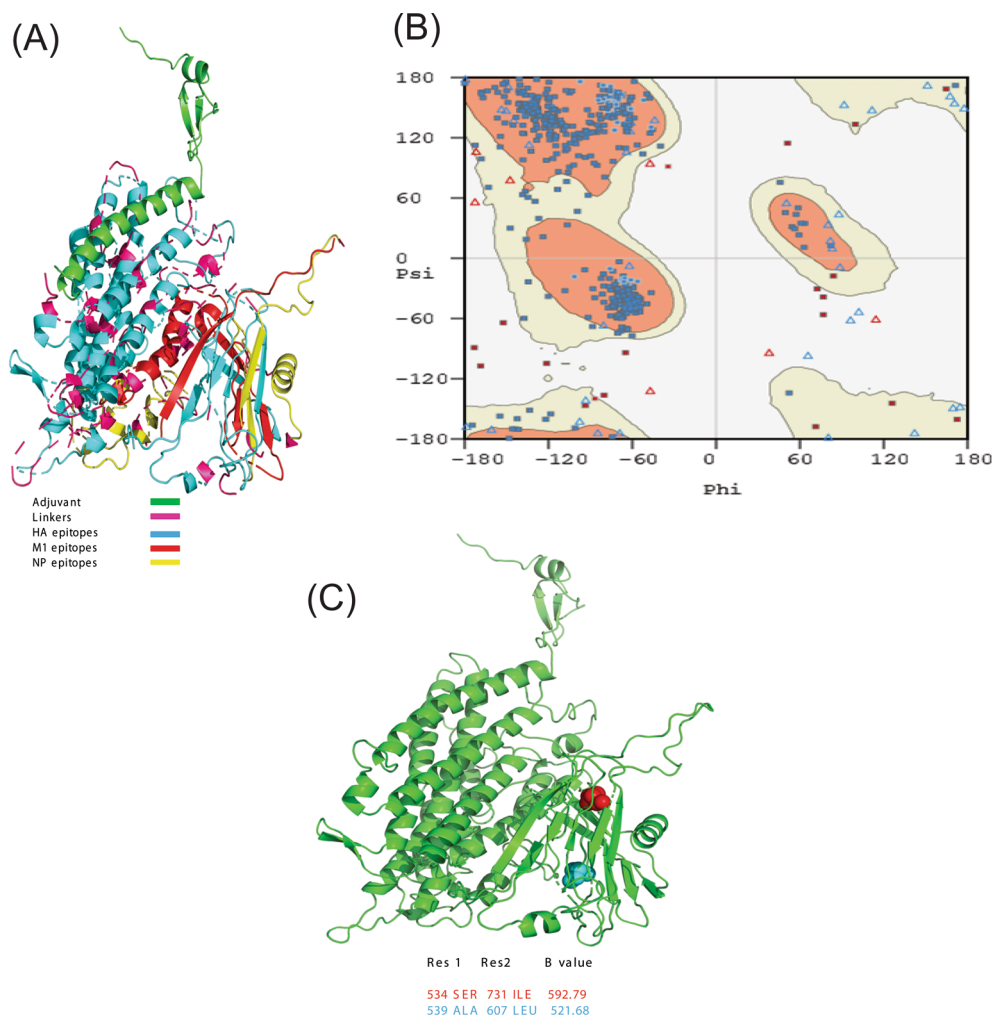


Fig. 3. Tertiary model preparation and validation. (A) Tertiary structure of the construct showing all epitopes of HA (cyan), NP (yellow) and M1 (red) with linkers (purple) and adjuvant (green). (B) Ramachandran plot of the vaccine model generated by WinCoot software, where phi (x-axis) and psi (y-axis) represent the torsion angles between N-C α and C α -C respectively in a polypeptide chain, shows that 92.05% residues are located in preferred regions, 3.84% residues fall in allowed regions and 4.11% are outliers. (C) Disulfides were introduced in vaccine construct using Disulfide by design server v.2 by mutating two pairs of amino acids to cysteine shown in red and cyan. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

structure was 1.19×10^{-4} indicating high quality. To validate the model, the Ramachandran plot was generated using WinCoot software (Fig. 3B). It showed that 92% (672) of the residues are located in the most preferred regions, 3.84% (28) residues are in the allowed regions and 4.11% (30) residues are found to be in the outlier region. Further, ProSA webserver was used to validate each domain of the predicted model and the Z-scores for domain 1, 2, 3, and 4 were -3.33 , -3.63 , -2.78 and -3.43 , respectively; all of which fall within the Z-score ranges of the native structures of the known proteins (determined by X-Ray and NMR), thus validating the model.

3.7. Disulfide engineering

To enhance the stability of the construct, disulfide bonds were introduced using Disulfide by design server v.2. The server predicted a total of 100 pairs of amino acid residues that could be mutated to cysteine to introduce a disulfide bridge. Two pairs, Ser 534-Ile 731 and Ala 539-Leu 607 with the highest B-factor, located in the epitopic region were mutated to Cys as shown in Fig. 3C.

3.8. Docking of vaccine model with TLR3, TLR7, TLR8, MHC-I and MHC-II

We studied the binding mode of the vaccine construct with the receptor proteins such as TLR3, TLR7, TLR8, MHC-I, and MHC-II, to examine the possibility of the immune response through these receptors. The PatchDock server was used to perform the *in silico* docking of

vaccine construct with TLR3, TLR7, TLR8, MHC-I, and MHC-II receptors. The best 10 outputs from PatchDock were selected and fed to FireDock for refinement of docked models. The final output generated by FireDock, with the lowest global energy with a value of -12.89 for solution number 3 of vaccine-TLR3 (Fig. 4A), -6.08 for solution number 4 of vaccine-TLR7 (Fig. 4B), -2.62 for solution number 7 for vaccine-TLR8 (Fig. 4C), -7.83 for solution number 1 of vaccine-MHC-I (Fig. 4D), and -4.52 of solution number 1 of vaccine-MHC-II (Fig. 4E) was considered over other models. The negative value of global energy reflects the possibility of a favourable interaction between the receptor and vaccine construct. These selected models of vaccine construct with receptor TLR3, TLR7, TLR8, MHC-I, and MHC-II proteins were further used as a starting structure for the molecular dynamics simulations, to investigate the vaccine-receptor protein structural stability.

3.9. Molecular dynamics simulations

To investigate the stability of vaccine construct with receptor proteins such as TLR3, TLR7, TLR8, MHC-I, and MHC-II proteins, molecular dynamics simulations were performed for 20 ns using Gromacs2018.1 (Van Der Spoel et al., 2005). For simulation, vaccine-TLR3 (Fig. 4A), vaccine-TLR7 (Fig. 4B), vaccine-TLR8 (Fig. 4C), vaccine-MHC-I (Fig. 4D), and vaccine-MHC-II (Fig. 4E) were used as the starting structures. The stability of the vaccine-receptor protein complexes was checked by plotting the backbone C α root mean square deviations (RMSD) and root mean square fluctuations (RMSF) for the period of 20 ns as shown in Fig. 5A and B. Overall, the RMSD of the MD simulated

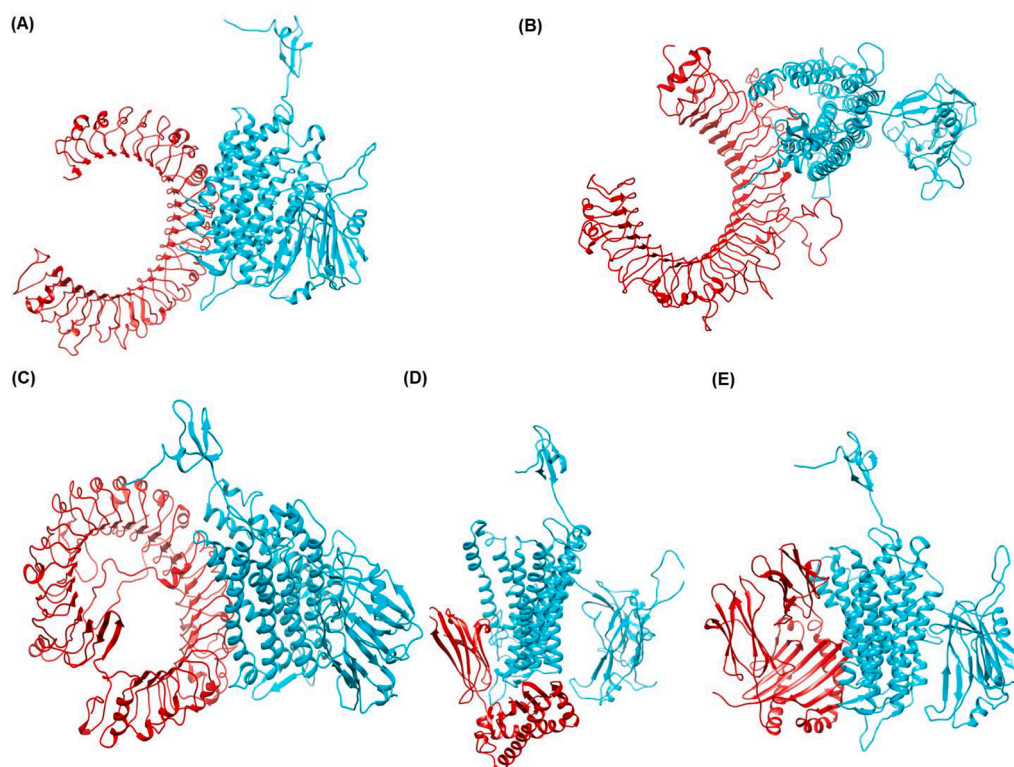


Fig. 4. Interaction of the proposed vaccine construct with TLR3, TLR7, TLR8, MHC-I and MHC-II. (A–E) *In silico* docking between TLR3, TLR7, TLR8, MHC-I, MHC-II and subunit vaccine construct was predicted using patchdock and FireDock tools, where the receptor is shown in red and the vaccine construct is shown in cyan colors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

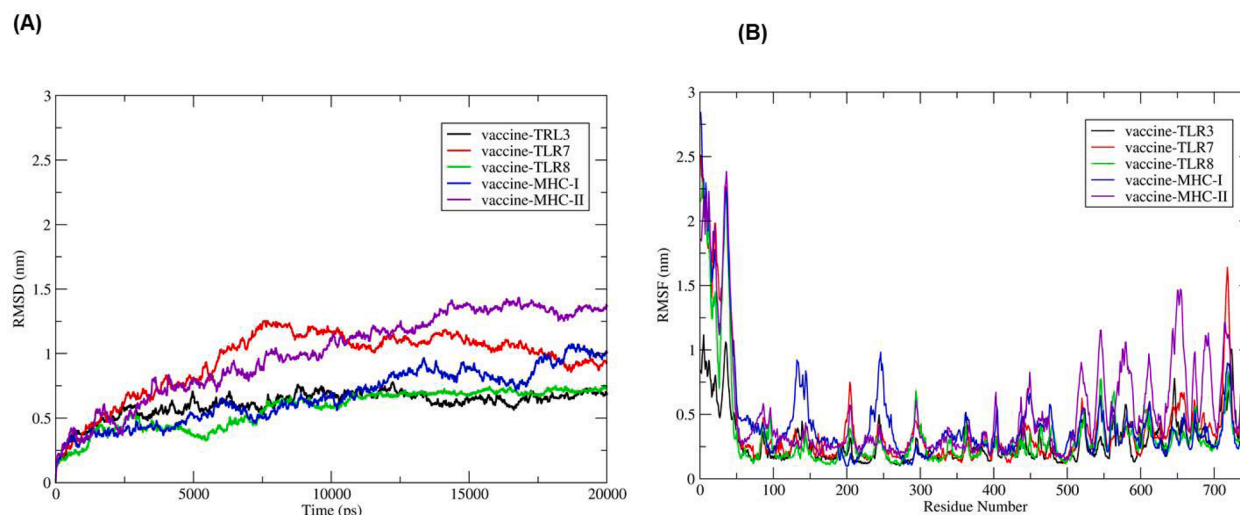


Fig. 5. Molecular dynamics simulation of vaccine constructs with receptor TLR3, TLR7, TLR8, MHC-I, and MHC-II proteins. (A) RMSD plot of the backbone C_{α} of vaccine constructs and TLR3 (black), TLR7 (red), TLR8 (green), MHC-I (blue), and MHC-II (violet) complex at 20 ns time steps. (B) The RMSF plot of the docked vaccine constructs in complex with receptors. The highly flexible regions showed higher RMSF values while the constrained regions showed lower RMSF values during the simulation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

complexes was found to be below 1.5 nm. The fluctuations in the RMSD plot are very low thereby reveals the stability of the vaccine construct and receptor complex. Furthermore, the RMSF plot for the C_{α} backbone atom of the vaccine construct was found to be below 1.5 nm (Fig. 5B). The RMS fluctuation shows the stability of the vaccine and receptor protein complexes. The higher fluctuations in the RMSF plot (Fig. 5B) indicate the flexibility in the regions of the vaccine construct in complex with receptor proteins. The highly flexible regions showed higher RMSF

values, while the constrained regions showed lower RMSF values (Fig. 5B). Overall, the molecular dynamics simulation results reveal the stability of all the vaccine construct and receptor protein complexes during the time of 20 ns.

3.10. Codon optimization and in silico cloning

The protein sequence of vaccine construct was converted into

nucleotide sequence using Reverse Translate which returned a DNA sequence with 2208 bases. The codons were optimized for expression in *E. coli* K12 using the Java codon adaptation tool. The Codon Adaptation Index (CAI) was found to be 0.557, a higher value (close to 1) indicates a greater level of heterologous gene expression in *E. coli* (Sharp and Li, 1987). Three restriction enzyme cleavage sites (*EcoRI*, *BamHI*, and *XhoI*) were found to be absent within the nucleotide sequence that would aid in the cloning of the gene into the desired pET vector (Fig. 6).

4. Discussion

With the threat of another IAV pandemic looming large, it is important to develop innovative ways of combating IAV. Conventional LAIVs containing a pathogenic component may cause reversion to virulent strain, while IIVs often elicit a weak immune response. In contrast, subunit vaccines provide long-term protection and can be administered to people with compromised immunity. These vaccines pose minuscule risks from adverse reactions, as only antigenic epitopes are used in their construction (Vartak and Sucheck, 2016). Besides, it is difficult to manufacture strain-specific vaccines at the onset of a pandemic in a short time. Thus, the subunit vaccine approach based on immunoformatics tools would help to extend vaccine production and distribution widely in undeveloped and underdeveloped nations, where the vaccines are in high demand. Furthermore, a poor immune response generated by a subunit vaccine can be overcome by the addition of an adjuvant, which may contribute to expanding the breadth of antibody recognition, by stimulating a suite of TLRs on antigen-presenting cells (Vogel, 2000).

Studies have shown that the influenza virus is recognized by different TLRs, such as TLR7/8 and TLR3, which bind to ssRNA and dsRNA, respectively (Pulendran and Maddur, 2015). A group, using both *in vitro* (lung epithelial cells) and *in vivo* (mouse model) approaches, recently demonstrated that TLR3 plays a major role in the synthesis of inflammatory cytokines and chemokines IL-6, G-CSF, RANTES during influenza infection (Guillot et al., 2005; Le Goffic et al., 2006). Importantly, TLR-7 triggers the induction of humoral B-cell and long-term memory responses to IAV infection and vaccination (Jeisy-Scott et al., 2012). Likewise, another study proved that synthetic TLR7/8 ligands elicit heightened innate and adaptive immune response when used alone, or in combination with synthetic TLR-4 ligand (Miller et al., 2018). Thus, to improve the vaccine efficacy of our construct, we added two antimicrobial peptides, namely, human β -defensin-3 (Bartels et al., 2000) and

LL-37, commonly used as adjuvants in vaccines. LL37-ssRNA complex has been reported to modulate TLRs7-9 and enhance downstream signalling to further activate cytokines. It is also known to act synergistically with TLR1/TLR2 heterodimer, TLR3, and TLR 5 agonists to facilitate IL-6 and IL-8 production (Lai et al., 2011). Of specific interest to us is the synergy of LL37 with TLR1/TLR2 because human β -defensin-3 functioning is primarily channelled through interaction with this heterodimer and downstream adapter molecules Myd88 and IL-1 receptor associated kinase-1 (IRAK-1). Human β -defensin-3 activates co-stimulatory molecules expression at monocytes (antigen-presenting cells) by this pathway (Funderburg et al., 2007). Furthermore, human β -defensin-3 has also been reported to induce interferon gamma (IFN- γ) secretion, natural killer (NK) cells activation as well as myeloid dendritic cells (mDC) dependent cytolytic activities (Judge et al., 2015).

We propose a novel 'combination adjuvant' to direct the immune response towards Th-1 type bias since the role of Th-1 type immune response in controlling intracellular pathogens is already known (Berger, 2000; Jankovic et al., 2001). TLR3, 4, 7, and 9 are considered to be crucial for TLR-derived Th1 responses which is mostly mediated by adapter molecule Myd88 (Netea et al., 2005). Additionally, IFN- γ is the most important Th-1 cytokine which as described above, is known to be secreted by human β -defensin-3 in a TLR1/TLR2 dependent manner (Judge et al., 2015; Netea et al., 2005).

Including two adjuvants significantly increased the predicted antigenicity value in our studies and did not show any allergenicity. However, peptide adjuvants or peptide combination adjuvants are not very common with viral vaccines. Furthermore, specific cytokine profile which might be generated by LL-37 and human β -defensin-3 combination and subsequent synergy of TLR1/TLR2 with TLR3, 7, and 9 is yet unexplored. Hence, further studies are required to comment on the impact of these two adjuvants in combination on immunogenicity as well as Th-1 response.

To design an efficacious subunit vaccine, it is important to identify the conserved CTL and HTL epitopes of the IAV proteins, an approach which has been considerably streamlined using computational tools and reverse vaccinology. As shown in Fig. 1B, NP is highly conserved across influenza subtypes and therefore, is a potential antigen for the universal influenza vaccine. After infection, NP is the main antigen recognized by influenza-specific CD8⁺ T lymphocytes, which have the potential to kill virus-infected cells (Yewdell et al., 1985). Infection with an IAV is known to induce heterosubtypic immunity, i.e., protective immunity to the infection with another subtype of IAV. In various animal models, including mice and non-human primates, it was demonstrated that virus-specific T lymphocytes, in particular, CD8⁺ T cells contribute to heterosubtypic immunity (Hillaire et al., 2011). This suggests that one or more CTL epitopes of all three proteins can bind to a different type of MHC class-I alleles present in people of different ethnic groups and elicit a significant immune response.

CTL epitopes have been selected against four supertypes of MHC class I (A1, A2, A3, and B7) for all three proteins, HA, M1, and NP. This allowed coverage of more than 83% of people of different ethnic groups. All HTL epitopes are significantly conserved and can potentially bind to different HLA MHC class II alleles with slightly different binding affinities.

Linear BCL epitopes have been incorporated in our vaccine construct. Since the conservancy in the case of HA BCL epitopes is not high in each subtype of IAV, therefore cross-reactive antibodies against HA of various subtypes will not be generated by BCL epitopes, although M1 and NP BCL epitopes are comparatively more conserved than HA BCL epitopes. The BCL epitopes were included in the vaccine for the generation of memory B-cells which can elicit strain-specific antibody-mediated immunity (Wang et al., 2016). However, the presence of numerous CTL, HTL, and BCL epitopes in the vaccine design may activate both innate and cell-mediated immunity.

The physicochemical characteristics of the vaccine construct are sufficiently within the threshold level, making it a suitable vaccine

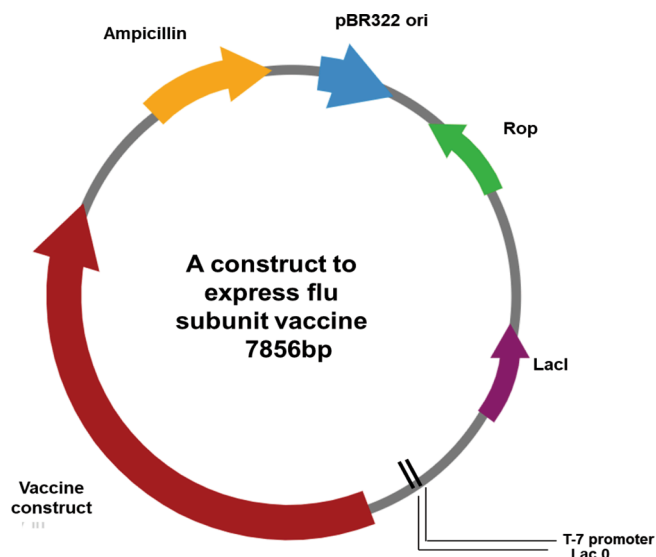


Fig. 6. *In silico* cloning of the codon-optimized construct into an *E. coli* expression vector.

candidate. The instability index is well under the threshold limit of 40, suggesting that the protein is highly stable. The pI value of 9.55 indicates that the protein is basic in nature. A negative GRAVY value (−0.034) makes the construct hydrophilic in nature. The disulfide bonds were introduced in the final tertiary structure for making the construct thermodynamically more stable. The docking and molecular dynamics simulation study was performed to examine the vaccine-receptors (TLR3, TLR7, TLR8, MHC-I and MHC-II) binding affinity and stability. The RMSD and RMSF plot show strong stability of the complexes.

Taken together, we suggest a conceptual framework to achieve the practical application of universal vaccine to combat IAV infections. However, this study needs experimental validation. The experimental work may include the synthesis of the designed subunit vaccine followed by the *in vitro* and *in vivo* analyses to determine the immunogenicity and delivery of the same for inducing a protective immune response. Formulation, safety, and protocols for immunizing with such vaccine construct are also important aspects to be considered.

CCRediT authorship contribution statement

Shipra Sharma: Conceptualization, Methodology, Investigation, Writing - original draft, Visualization. **Vibha Kumari:** Methodology, Investigation, Writing - original draft, Visualisation, Validation. **Bajrang Kumbhar:** Methodology, Investigation, Formal analysis. **Amrita Mukherjee:** Investigation, Validation. **Ruchika Pandey:** Investigation, Validation. **Kiran Kondabagil:** Conceptualization, Methodology, Formal analysis, Visualization, Writing - original draft, Supervision, Project administration, Funding acquisition.

Acknowledgements

SS is supported by DST-Young Scientist Grant (YSS/2015/002059). VK is supported by IITB-Monash academy fellowship, and AM is supported by Council of Scientific and Industrial Research (CSIR) Junior Research Fellowship.

Conflict of interest

The authors declare that they have no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.imbio.2021.152053>.

References

- Arai, R., Ueda, H., Kitayama, A., Kamiya, N., Nagamune, T., 2001. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng. Des. Sel.* 14, 529–532. <https://doi.org/10.1093/protein/14.8.529>.
- Atsmon, J., Caraco, Y., Ziv-Sefer, S., Shaikovich, D., Abramov, E., Volokhov, I., Bruzil, S., Haima, K.Y., Gottlieb, T., Ben-Yedidia, T., 2014. Priming by a novel universal influenza vaccine (Multimeric-001)—A gateway for improving immune response in the elderly population. *Vaccine* 32, 5816–5823. <https://doi.org/10.1016/j.vaccine.2014.08.031>.
- Bartels, J., Christophers, E., Schrö, J.-M., 2000. Isolation and characterization of human N_L-defensin-3, a novel human inducible peptide antibiotic. *Jü rgen Harder* 276, 5707–5713. <https://doi.org/10.1074/jbc.M008557200>.
- Berger, A., 2000. Th1 and Th2 responses: what are they? *BMJ* 321, 424. <https://doi.org/10.1136/bmj.321.7258.424>.
- Bright, R.A., Shay, D.K., Shu, B., Cox, N.J., Klimov, A.I., 2006. Adamantane resistance among Influenza A viruses isolated early during the 2005–2006 Influenza Season in the United States. *JAMA* 295, 891–894. <https://doi.org/10.1001/jama.295.8.joc60020>.
- Bui, H.-H., Sidney, J., Dinh, K., Southwood, S., Newman, M.J., Sette, A., 2006. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics* 7, 153. <https://doi.org/10.1186/1471-2105-7-153>.
- Bui, H.-H., Sidney, J., Li, W., Fusseder, N., Sette, A., 2007. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics* 8, 361. <https://doi.org/10.1186/1471-2105-8-361>.
- Carter, D.M., Darby, C.A., Lefoley, B.C., Crevar, C.J., Alefantis, T., Oomen, R., Anderson, S.F., Strugnell, T., Cortés-García, G., Vogel, T.U., Parrington, M., Kleanthous, H.,

- Ross, T.M., 2016. Design and characterization of a computationally optimized broadly reactive hemagglutinin vaccine for H1N1 influenza viruses. *J. Virol.* 90, 4720 LP – 4734. <https://doi.org/10.1128/JVI.03152-15>.
- Case, D.A., Ben-Shalom, I.Y., Brozell, S.R., Cerutti, D.S., Cheatham III, T.E., Cruzeiro, V. W.D., Darden, T.A., Duke, R.E., Ghoreishi, D., Gilson, M.K., Gohlke, H., Goetz, A.W., Greene, D., Harris, R., Homeyer, N., Huang, Y., Izadi, S., Kovalenko, A., Kurtzman, T., Lee, T.S., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Mermelstein, D.J., Merz, K.M., Miao, Y., Monard, G., Nguyen, C., Nguyen, H., Omelyan, I., Onufriev, A., Pan, F., Qi, R., Roe, D.R., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shen, J., Simmerling, C.L., Smith, J., Ferrer, R.S., Swails, J., Walker, R. C., Wang, J., Wei, H., Wolf, R.M., Wu, X., Xiao, L., York, D.M., Kollman, P.A., 2018. *Amber 2018*. University of California, San Francisco.
- Craig, D.B., Dombkowski, A.A., 2013. Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinformatics* 14, 346. <https://doi.org/10.1186/1471-2105-14-346>.
- Deng, L., Cho, K.J., Fiers, W., Saelens, X., 2015. M2e-based universal Influenza A vaccines. *Vaccines* 3, 105–136. <https://doi.org/10.3390/vaccines3010105>.
- Dimitrov, I., Flower, D.R., Doytchinova, I., 2013. AllerTOP—a server for in silico prediction of allergens. *BMC Bioinformatics* 14 Suppl 6, S4–S4. <https://doi.org/10.1186/1471-2105-14-S6-S4>.
- El-Manzalawy, Y., Dobbs, D., Honavar, V., 2008. Predicting linear B-cell epitopes using string kernels. *J. Mol. Recognit.* 21, 243–255. <https://doi.org/10.1002/jmr.893>.
- Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* 66, 486–501. <https://doi.org/10.1107/S0907444910007493>.
- Epstein, S.L., Kong, W., Misplon, J.A., Lo, C.-Y., Tumpey, T.M., Xu, L., Nabel, G.J., 2005. Protection against multiple influenza A subtypes by vaccination with highly conserved nucleoprotein. *Vaccine* 23, 5404–5410. <https://doi.org/10.1016/j.vaccine.2005.04.047>.
- Freed, E., Martin, M., 2013. Fields virology, *Fields Virology*. <https://doi.org/10.1016/j.imbio.2021.152053>.
- Funderburg, N., Lederman, M.M., Feng, Z., Drage, M.G., Jadowsky, J., Harding, C.V., Weinberg, A., Sieg, S.F., 2007. Human -defensin-3 activates professional antigen-presenting cells via Toll-like receptors 1 and 2. *Proc. Natl. Acad. Sci. U. S. A.* 104, 18631–18635. <https://doi.org/10.1073/pnas.0702130104>.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein Analysis Tools on the ExPASy Server. *Proteomics Protoc. Handb. Protein Identif. Anal. Tools ExPASy Serv.* 571–607. <https://doi.org/10.1385/1592598900>.
- Giles, B.M., Crevar, C.J., Carter, D.M., Bissell, S.J., Schultz-Cherry, S., Wiley, C.A., Ross, T.M., 2012. A computationally optimized hemagglutinin virus-like particle vaccine elicits broadly reactive antibodies that protect nonhuman primates from H5N1 infection. *J. Infect. Dis.* 205, 1562–1570. <https://doi.org/10.1093/infdis/jis232>.
- Giles, B.M., Ross, T.M., 2011. A computationally optimized broadly reactive antigen (COBRA) based H5N1 VLP vaccine elicits broadly reactive antibodies in mice and ferrets. *Vaccine* 29, 3043–3054. <https://doi.org/10.1016/j.vaccine.2011.01.100>.
- Grohskopf, L.A., Sokolow, L.Z., Broder, K.R., Walter, E.B., Fry, A.M., Jernigan, D.B., 2018. Prevention and control of seasonal influenza with vaccines: recommendations of the advisory committee on immunization practices—United States, 2018–19 influenza season. *MMWR. Recomm. Rep. Morb. Mortal. Wkly. Report. Recomm. Reports* 67, 1–20. <https://doi.org/10.15585/mmwr.mm6703a1>.
- Guillot, L., Le Goffic, R., Bloch, S., Escricou, N., Akira, S., Chignard, M., Si-Tahar, M., 2005. Involvement of Toll-like receptor 3 in the immune response of lung epithelial cells to double-stranded RNA and influenza A virus. *J. Biol. Chem.* 280, 5571–5580. <https://doi.org/10.1074/jbc.M410592200>.
- Hajjigahramani, N., Nezafat, N., Eslami, M., Negahdaripour, M., Rahmatabadi, S.S., Ghasemi, Y., 2017. Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against *Staphylococcus aureus*. *Infect. Genet. Evol.* 48, 83–94. <https://doi.org/10.1016/j.meegid.2016.12.010>.
- Hamilton, K., 2009. The emergence of the pandemic A/H1N1 2009 virus and its characteristics. *Bull. Mem. Acad. R. Med. Belg.* 164, 260–263.
- Hause, B., Collin, E., Liu, R., Huang, B., Sheng, Z., Lu, W., 2014. Characterization of a novel influenza virus strain in cattle and swine: proposal for a new genus in the Orthomyxoviridae family. *MBio* 5, 1–10. <https://doi.org/10.1128/mBio.00031-14>.
- Hillaire, M.L.B., Osterhaus, A.D.M.E., Rimmelzwaan, G.F., 2011. Induction of virus-specific cytotoxic T lymphocytes as a basis for the development of broadly protective influenza vaccines. *J. Biomed. Biotechnol.* 2011, 939860. <https://doi.org/10.1155/2011/939860>.
- Houser, K., Subbarao, K., 2015. Influenza vaccines: challenges and solutions. *Cell Host Microbe* 17, 295–300. <https://doi.org/10.1016/j.chom.2015.02.012>.
- Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- Hurt, A.C., Holien, J.K., Parker, M., Kelso, A., Barr, I.G., 2009. Zanamivir-resistant influenza viruses with a novel neuraminidase mutation. *J. Virol.* 83, 10366–10373. <https://doi.org/10.1128/JVI.01200-09>.
- Hutchinson, E.C., Charles, P.D., Hester, S.S., Thomas, B., Trudgian, D., Martinez-Alonso, M., Fodor, E., 2014. Conserved and host-specific features of influenza virion architecture. *Nat. Commun.* 5, 4816. <https://doi.org/10.1038/ncomms5816>.
- Jankovic, D., Liu, Z., Gause, W.C., 2001. Th1- and Th2-cell commitment during infectious disease: asymmetry in divergent pathways. *Trends Immunol.* 22, 450–457. [https://doi.org/10.1016/s1471-4906\(01\)01975-5](https://doi.org/10.1016/s1471-4906(01)01975-5).
- Jeisy-Scott, V., Kim, J.H., Davis, W.G., Cao, W., Katz, J.M., Sambhara, S., 2012. TLR7 Recognition is dispensable for Influenza virus A infection but important for the induction of hemagglutinin-specific antibodies in response to the 2009 pandemic

- split vaccine in mice. *J. Virol.* 86, 10988 LP–10998. <https://doi.org/10.1128/JVI.01064-12>.
- Joseph, U., Su, Y.C.F., Vijaykrishna, D., Smith, G.J.D., 2017. The ecology and adaptive evolution of influenza A interspecies transmission. *Influenza Other Respi. Viruses* 11, 74–84. <https://doi.org/10.1111/irv.12412>.
- Judge, C.J., Reyes-Aviles, E., Conry, S.J., Sieg, S.S., Feng, Z., Weinberg, A., Anthony, D. D., 2015. HBD-3 induces NK cell activation, IFN- γ secretion and mDC dependent cytolytic function. *Cell. Immunol.* 297, 61–68. <https://doi.org/10.1016/j.cellimm.2015.06.004>.
- Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., Xu, J., 2012. Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* 7, 1511–1522. <https://doi.org/10.1038/nprot.2012.085>.
- Krammer, F., Palese, P., 2015. Advances in the development of influenza virus vaccines. *Nat. Rev. Drug Discov.* 14, 167.
- Kreijtz, J.H.C.M., de Mutsert, G., van Baalen, C.A., Fouchier, R.A.M., Osterhaus, A.D.M. E., Rimmelzwaan, G.F., 2008. Cross-recognition of avian H5N1 influenza virus by human cytotoxic T-lymphocyte populations directed to human influenza A virus. *J. Virol.* 82, 5161 LP – 5166. <https://doi.org/10.1128/JVI.02694-07>.
- Kumbhar, B.V., Bhandare, V.V., 2020. Exploring the interaction of Peloruside-A with drug resistant α II and α III tubulin isotypes in human ovarian carcinoma using a molecular modeling approach. *J. Biomol. Struct. Dyn.* <https://doi.org/10.1080/07391102.2020.1745689>.
- Lai, Y., Adhikarakunnathu, S., Bhardwaj, K., Ranjith-Kumar, C.T., Wen, Y., Jordan, J.L., Wu, L.H., Dragnea, B., San Mateo, L., Kao, C.C., 2011. LL37 and cationic peptides enhance TLR3 signaling by viral double-stranded RNAs. *PLoS One* 6, e26632–e26632. <https://doi.org/10.1371/journal.pone.0026632>.
- Larsen, M.V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., Nielsen, M., 2007. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics* 8, 1–12. <https://doi.org/10.1186/1471-2105-8-424>.
- Le Goffic, R., Balloy, V., Lagranderie, M., Alexopoulou, L., Escirou, N., Flavell, R., Chignard, M., Si-Tahar, M., 2006. Detrimental contribution of the Toll-like receptor (TLR)3 to influenza A virus-induced acute pneumonia. *PLoS Pathog.* 2, e53–e53. <https://doi.org/10.1371/journal.ppat.0020053>.
- Lee, L.Y.-H., Ha, D.L.A., Simmons, C., de Jong, M.D., Chau, N.V.V., Schumacher, R., Peng, Y.C., McMichael, A.J., Farrar, J.J., Smith, G.L., Townsend, A.R.M., Askonas, B. A., Rowland-Jones, S., Dong, T., 2008. Memory T cells established by seasonal human influenza A infection cross-react with avian influenza A (H5N1) in healthy individuals. *J. Clin. Invest.* 118, 3478–3490. <https://doi.org/10.1172/JCI32460>.
- Livingston, B., Crimi, C., Newman, M., Higashimoto, Y., Appella, E., Sidney, J., Sette, A., 2002. A rational strategy to design multi-epitope immunogens based on multiple Th lymphocyte epitopes. *J. Immunol.* 168, 5499 LP – 5506. <https://doi.org/10.4049/jimmunol.168.11.5499>.
- Lund, O., Nielsen, M., Kesmir, C., Petersen, A.G., Lundegaard, C., Wornig, P., Sylvester-Hvid, C., Lamberth, K., Røder, G., Justesen, S., Buus, S., Brunak, S., 2004. Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics* 55, 797–810. <https://doi.org/10.1007/s00251-004-0647-4>.
- Lundegaard, C., Lamberth, K., Harndahl, M., Buus, S., Lund, O., Nielsen, M., 2008. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. *Nucleic Acids Res.* 36, W509–W512. <https://doi.org/10.1093/nar/gkn202>.
- Mashiach, E., Schneidman-Duhovny, D., Andrusier, N., Nussinov, R., Wolfson, H.J., 2008. FireDock: a web server for fast interaction refinement in molecular docking. *Nucleic Acids Res.* 36, W229–W232. <https://doi.org/10.1093/nar/gkn186>.
- Miller, S.M., Cybulski, V., Walsh, L., Livesay, M., Bess, L., Burkhart, D.J., Bazin-Lee, H., Evans, J.T., 2018. Investigation of novel TLR7/8 ligands in combination with TLR4 ligands as adjuvants to drive cell mediated anti-influenza immunity. *J. Immunol.* 200, 125.16 LP-125.16.
- Moscona, A., 2005. Oseltamivir resistance — disabling our influenza defenses. *N. Engl. J. Med.* 353, 2633–2636. <https://doi.org/10.1056/nejmp058291>.
- Netea, M.G., Van der Meer, J.W.M., Sutmoller, R.P., Adema, G.J., Kullberg, B.-J., 2005. From the Th1/Th2 paradigm towards a toll-like receptor/T-helper bias. *Antimicrob. Agents Chemother.* 49, 3991 LP – 3996. <https://doi.org/10.1128/AAC.49.10.3991-3996.2005>.
- Nezafat, N., Ghasemi, Y., Javadi, G., Khoshnoud, M.J., Omidinia, E., 2014. A novel multi-epitope peptide vaccine against cancer: an in silico approach. *J. Theor. Biol.* 349, 121–134. <https://doi.org/10.1016/j.jtbi.2014.01.018>.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* <https://doi.org/10.1002/jcc.20084>.
- Price, M.N., Dehal, P.S., Arkin, A.P., 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26, 1641–1650. <https://doi.org/10.1093/molbev/msp077>.
- Pulendran, B., Maddur, M.S., 2015. Innate immune sensing and response to influenza. *Curr. Top. Microbiol. Immunol.* 386, 23–71. https://doi.org/10.1007/82_2014_405.
- Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., Wolfson, H.J., 2005. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* 33, W363–W367. <https://doi.org/10.1093/nar/gki481>.
- Sharp, P.M., Li, W.H., 1987. The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15, 1281–1295.
- Tan, M., Jiang, X., 2016. Recent advancements in combination subunit vaccine development. *Hum. Vaccin. Immunother.* 13, 180–185. <https://doi.org/10.1080/21645515.2016.1229719>.
- Taubenberger, J.K., Morens, D.M., 2010. Influenza: the once and future pandemic. *Public Health Rep.* 125, 16–26. <https://doi.org/10.1177/003335491012508305>.
- Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., Berendsen, H.J.C., 2005. GROMACS: fast, flexible, and free. *J. Comput. Chem.* <https://doi.org/10.1002/jcc.20291>.
- Vartak, A., Suchack, S.J., 2016. Recent advances in subunit vaccine carriers. *Vaccines* 4, 12. <https://doi.org/10.3390/vaccines4020012>.
- Vogel, F.R., 2000. Improving vaccine performance with adjuvants. *Clin. Infect. Dis.* 30 (Suppl 3), S266–S270. <https://doi.org/10.1086/313883>.
- Wang, P., Sidney, J., Dow, C., Mothé, B., Sette, A., Peters, B., 2008. A Systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLOS Comput. Biol.* 4, e1000048.
- Wang, X., Sun, Q., Ye, Z., Hua, Y., Shao, N., Du, Y., Zhang, Q., Wan, C., 2016. Computational approach for predicting the conserved B-cell epitopes of hemagglutinin H7 subtype influenza virus. *Exp. Ther. Med.* 12, 2439–2446. <https://doi.org/10.3892/etm.2016.3636>.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gky427>.
- Webb, B., Sali, A., 2017. Protein structure modeling with MODELLER, in: *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-4939-7231-9_4.
- Weiskopf, D., Angelo, M.A., de Azeredo, E.L., Sidney, J., Greenbaum, J.A., Fernando, A. N., Broadwater, A., Kolla, R.V., De Silva, A.D., de Silva, A.M., Mattia, K.A., Doranz, B.J., Grey, H.M., Shrestha, S., Peters, B., Sette, A., 2013. Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. *Proc. Natl. Acad. Sci. U. S. A.* 110, E2046–E2053. <https://doi.org/10.1073/pnas.1305227110>.
- Wiederstein, M., Sippl, M.J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35, W407–W410. <https://doi.org/10.1093/nar/gkm290>.
- Yano, A., Onozuka, A., Asahi-Ozaki, Y., Imai, S., Hanada, N., Miwa, Y., Nisizawa, T., 2005. An ingenious design for peptide vaccines. *Vaccine* 23, 2322–2326. <https://doi.org/10.1016/j.vaccine.2005.01.031>.
- Yewdell, J.W., Bennink, J.R., Smith, G.L., Moss, B., 1985. Influenza A virus nucleoprotein is a major target antigen for cross-reactive anti-influenza A virus cytotoxic T lymphocytes. *Proc. Natl. Acad. Sci. U. S. A.* 82, 1785–1789.
- Zhang, J., Jin, Z., Sun, G.-Q., Sun, X.-D., Wang, Y.-M., Huang, B., 2014. Determination of original infection source of H7N9 avian influenza by dynamical model. *Sci. Rep.* 4, 4846. <https://doi.org/10.1038/srep04846>.
- Zheng, M., Luo, J., Chen, Z., 2014. Development of universal influenza vaccines based on influenza virus M and NP genes. *Infection* 42, 251–262. <https://doi.org/10.1007/s15010-013-0546-4>.

SUPPLEMENTARY MATERIAL

Table 1: Accession numbers of HA, NP, and M1 proteins from the 50 strains of influenza A virus strains considered for this study.

Sr. No.	Influenza A Virus strain	HA	NP	M1
1	Addis Ababa_2009_H1N1	ADM14975.2	ADM14976.1	ADM14978.1
2	Addis Ababa 2013_H3N2	ALM06006.1	ALM04798.1	ALM04512.1
3	Aichi (Japan)_1968_H3N2	AFM71857.1	AFM71861.1	AFM71858.1
4	Alabama_2018_H3N2	AVP06274.1	AVP06273.1	AVP06271.1
5	Alaska_2016_H1N1	AOZ86117.1	AOZ86118.1	AOZ86120.1
6	Albany_1957_H2N2	ABO38307.1	ABO38311.1	ABP49460.1
7	Arizona(US)_2009_H1N1	ACQ76347.1	ACQ76341.1	ACQ76345.1
8	Arizona(US)_2009_H3N2	ACT67779.1	AEK28972.1	ACT67828.1
9	Australia_2009_H1N1	ADD21864.1	ADD21868.1	ADD21865.1
10	Beijing_2003_H5N1	ABQ58979.1	ABQ58980.1	ABQ58982.1
11	Brazil(cambara)_2015_H3N2	ARI68499.1	ARI68500.1	ARI68501.1
12	Brazil(canoas)_2016_H1N1	ARI68560.1	ARI68619.1	ARI68568.1
13	Brevig Mission(USA)_1918_H1N1	AAD17218.1	AAV48837.1	AAN06597.1
14	California_2009_H1N1	ACP41953.1	AMV49024.1	AGI55496.1
15	California_2016_H1N1	ANM87248.1	ANE27542.1	ANE27543.1
16	Cambodia_2011_H5N1	AEN68938.1	AFF59184.1	AFF59185.1
17	Cambodia_2013_H5N1	AGM16225.1	AGM16226.1	AGM16228.1
18	China_2006_H5N1	ABR10842.1	ABI16505.1	ABI16508.1
19	Delhi_2009_NIV_H1N1	AEM63469.1	AEM62827.1	AEM62829.1
20	Dhaka_2013_(H1N1)	APF29591.1	AVR58470.1	AVR58471.1
21	Egypt_2014_H5N1	AIO11739.1	AJM70747.1	AJM70737.1
22	Fujian(China)_2015_H7N9	APR73185.1	APR73186.1	APR73189.1
23	Guangdong (China)_2017_H7N9	AQY18946.1	AQY18947.1	AQY18949.1
24	Guangdong (China)_2009_H1N1	AFN87931.1	ADL14538.1	ADL14533.1
25	Guiyang_1957_H2N2	ACD85231.1	ACD85235.1	ACD85232.1
26	Hong Kong_1968_H3N2	ACU79871.1	ACU79875.1	APO39313.1
27	Hong Kong_1997_H5N1	AAC40508.1	CAC19704.1	AAC32090.1
28	Hong Kong_2009_H3N2	AEK84972.1	AFM72909.1	AFM72906.1
29	Indonesia_2007_H5N1	ABM90467.1	ABM90435.1	ABM90470.1
30	Japan(Hokkaido)_2014_H3N2	BAQ58289.1	BAQ58300.1	BAQ58293.1
31	Jiangsu(China)_2007_H5N1	ACB87573.1	ACB87577.1	ACB87574.1
32	Jiangsu(China)_2011_H1N1	AGN69338.1	AGN69339.1	AGN69341.1
33	Johannesburg_1997_H3N2	AFG72713.1	AFG72717.1	AFG72714.1
34	Keelung(Taiwan)_2016_H1N1	AQN74611.1	AQN75088.1	AQN74326.1
35	Kerala_2009_H1N1	AJE62443.1	AMU04343.1	AJE62171.1
36	Kerala_2015_H1N1	ARG42800.1	AMU04346.1	AMU04356.1
37	Memphis(USA)_1968_H3N2	ABB54514.1	ABB54518.1	ABB54515.1
38	Memphis(USA)_1990_H3N2	ABA43336.1	ABA43340.1	ABA43337.1
39	Mum_NIV_2009_H1N1	AEM63479.1	AEM62836.1	AEM62838.1
40	New Jersey_2008_H1N1	ACD85759.1	ACU79985.1	ACU31267.1
41	New Jersey_2018_H3N2	AVP07366.1	AWB13549.1	AVP07362.1

42	New York_2004_H3N2	AAX76623.1	YP_308843.1	AAX76624.1
43	New York_2017_H1N1	ARB16853.1	ARB16854.1	ARB16856.1
44	Puerto Rico_8_1934_H1N1	CAA24272.1	ACV49549.1	ACV49546.1
45	Senegal(Africa)_2010_H1N1	ADX96258.1	ADX96259.1	ADX96261.1
46	Shanghai_2013_H7N9	AGL44438.1	AGL44439.1	AGL44441.1
47	Shanghai_2014_H7N9	AHK10800.1	AHK10804.1	AHK10801.1
48	Shiga(Japan)_2009_H1N1	ACS34968.1	ACS36631.1	ACS36626.1
49	Singapore_1957_H2N2	BAF48641.1	AAA52251.1	CAA30892.1
50	Singapore_2007_H3N2	AJB43494.1	AJB43500.1	AJB43498.1

Sr. No.	Protein-MHC I Supertype	Epitope Sequence	Conservancy
1	H1N1 HA A1	WTGMVDGWY	90.47 %
2	H1N1 HA A2	QILAIYSTV	90.47 %
3	H1N1 HA A3	KAILVLLY	80.5 %
4	H1N1 HA B7	LPFQNIHPI	76.1%
5	H2N2 HA A1	MVDGWYGYH	100%

Table 2: Linear sequence conservancy of CTL epitopes in HA, M1 and NP proteins of 50 strains

6	H2N2 HA A3	SVGTSTLNK	100%
7	H3N2 HA A1	LTDSEMKNL	100%
8	H3N2 HA A2	AISCFLLCV	100%
9	H3N2 HA A3	ALNNRFQIK	100%
10	H3N2 HA B7	YVKQNTLKL	78.5%
11	H5N1 HA A1	HSNEQGSY	87.5%
12	H5N1 HA A2	FLDVWYNA	100%
13	H5N1 HA A3	RLVLATGLR	75%
14	H5N1 HA B7	MPFHNIHPL	100%
15	H7N9 HA A1	STAEQTKLY	100%
16	H7N9 HA A2	ILLAIVMGL	100%
17	H7N9 HA A3	SLLLATGMK	100%
18	H7N9 HA B7	YVKQRSLLL	100%
19	M1 A1	LTEVETYVL	100%
20	M1 A2	GILGFVFTL	96%
21	M1 A3	MVLASTTAK	100%
22	M1 B7	RPILSPLTK	100%
23	NP A1	HSNLNDATY	80%
24	NP A2	GMDPRMCSL	100%
25	NP A3	LMQGSTLPR	100%
26	NP B7	DPKKTGGPI	98%

Table 3: Linear sequence conservancy of HTL epitopes in HA, M1, and NP proteins of 50 strains and their allele coverage

Sr. No.	Protein	Epitope sequence	Conservancy	Allele Coverage
1	H1N1 HA	QILAIYSTVASSLVL	90 %	100%
2	H2N2 HA	IIYLILLFTAVRGDQ	100 %	100%
3	H3N2 HA	LWISFAISCFLLCVA	64.2 %	40%
4	H5N1 HA	ILSIYSTVASSLALA	100 %	63%
5	H7N9 HA	MNTQILVFALIAIIP	100 %	74%
6	M1	KGILGFVFTLTVPSE	94%	37%
7	M1	YVLSIVPSGPLKAEI	38 %	37%
8	NP	NEGSYFFGDNAEEYD	96 %	26%
9	NP	GTMVMELIRMVKRGI	14 %	30%

Table 4: Linear sequence conservancy of linear BCL epitopes in HA, M1, and NP proteins of 50 strains.

Table 5: Physiochemical properties of the vaccine construct.

Sr.No.	Physiochemical Properties	Value	Inference
1	Number of amino acid	736	
2	Molecular weight	79.5 kDa	
3	Theoretical pI	9.55	Alkaline
4	Half-Life (Mammal, Yeast, E.coli)	30h, 20h, 10h	Highly Stable
5	Instability Index ^a	35.05	Stable
6	GRAVY	-0.034	Hydrophilic
7	Antigenicity	0.569	Immunogenic
8	Allergenicity	Non-allergen	No over-reaction by host immune response
9	Aliphatic Index ^b	84.95	Thermodynamically Stable

Sr. No.	Protein	Epitope sequence	Conservancy
1	H1N1 HA	STASSWSYIVESSSSDNGTC	48%
2	H2N2 HA	LTKKGSNYPVAKGSYNNTSG	67%
3	H3N2 HA	VERSKAYSNCYPYDVDPDYAS	57%
4	H5N1 HA	GFIEGGWQGMVDGWYGYHHS	100%
5	H7N9 HA	SSNYQQSFVPSPGARPVNG	100%
6	M1	QRRRFVQNALNGNDPNNMD	84%
7	NP	HPSAGKDPKKTGGPIYRRVD	30%

a-Threshold for instability index of a protein is 40. *b*-Aliphatic index is defined as the fraction of relative volume of protein occupied by aliphatic side chains. A high index (84.95) for our subunit vaccine protein indicates that it is thermodynamically stable. *Aliphatic* index is defined as the fraction of relative volume of protein occupied by aliphatic side chains. A high index (84.95) for our subunit vaccine protein indicates that it is thermodynamically stable.