



Designing novel epitope-based polyvalent vaccines against herpes simplex virus-1 and 2 exploiting the immunoinformatics approach

Bishajit Sarkar, Md. Asad Ullah, Yusha Araf, Sowmen Das, Md. Hasanur Rahman & Abu Tayab Moin

To cite this article: Bishajit Sarkar, Md. Asad Ullah, Yusha Araf, Sowmen Das, Md. Hasanur Rahman & Abu Tayab Moin (2021) Designing novel epitope-based polyvalent vaccines against herpes simplex virus-1 and 2 exploiting the immunoinformatics approach, *Journal of Biomolecular Structure and Dynamics*, 39:17, 6585-6605, DOI: [10.1080/07391102.2020.1803969](https://doi.org/10.1080/07391102.2020.1803969)

To link to this article: <https://doi.org/10.1080/07391102.2020.1803969>



[View supplementary material](#)



Published online: 07 Aug 2020.



[Submit your article to this journal](#)



Article views: 147



[View related articles](#)



[View Crossmark data](#)



[Citing articles: 1](#) [View citing articles](#)



Designing novel epitope-based polyvalent vaccines against herpes simplex virus-1 and 2 exploiting the immunoinformatics approach

Bishajit Sarkar^a , Md. Asad Ullah^a , Yusha Araf^b , Sowmen Das^c, Md. Hasanur Rahman^d and Abu Tayab Moin^e

^aDepartment of Biotechnology and Genetic Engineering, Faculty of Biological Sciences, Jahangirnagar University, Dhaka, Bangladesh;

^bDepartment of Genetic Engineering and Biotechnology, School of Life Sciences, Shahjalal University of Science and Technology, Sylhet, Bangladesh;

^cDepartment of Computer Science and Engineering, School of Physical Sciences, Shahjalal University of Science and Technology, Sylhet, Bangladesh;

^dDepartment of Biotechnology and Genetic Engineering, Faculty of Life Sciences, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh;

^eDepartment of Genetic Engineering and Biotechnology, Faculty of Biological Sciences, University of Chittagong, Chattogram, Bangladesh

Communicated by Ramaswamy H. Sarma

ABSTRACT

Herpes Simplex Virus (HSV) is a highly infectious virus that is responsible for various types of orofacial and genital infections. Two types of HSV exist i.e. HSV-1 and HSV-2, that are infecting millions of people around the world. However, no satisfactory treatment or counter-measure has yet been discovered to fight against the HSV infections. In this study, three possible polyvalent subunit vaccines against multiple strains of HSV-1 and HSV-2, targeting the envelope glycoproteins- E, B, and D, were designed using the tools of reverse vaccinology and immunoinformatics. The highly antigenic, non-allergenic, non-toxic, non-homolog (to the human proteome), and 100% conserved epitopes across the selected strains and species (eight epitopes from each of the CTL, HTL, and BCL epitope groups), were selected for vaccine construction. These designed vaccines are expected to be effective against the selected viral types simultaneously (as a polyvalent vaccine), without producing any unwanted adverse reaction within the body. Finally, from the three vaccine constructs, one best vaccine was determined by molecular docking analysis and thereafter, the MD simulation and immune simulation studies of the best vaccine construct also yielded satisfactory results, pointing towards quite good stability of the complex. Finally, *in silico* cloning was performed for analyzing the effective mass production strategy of the best vaccine construct. However, wet lab-based study should be conducted on the suggested vaccines for validating their potentiality, safety, and efficacy.

ARTICLE HISTORY

Received 15 July 2020

Accepted 22 July 2020

KEYWORDS

Herpes simplex virus-1; herpes simplex virus-2; vaccine; immunity; immunoinformatics; envelope glycoprotein

1. Introduction

Herpes Simplex Virus (HSV) is regarded as one of the most common sexually transmitted viruses around the world (Straface et al., 2012). It is a member of the *Herpesviridae* family. The HSV-1 and HSV-2 are the two HSV types that are known to infect and cause diseases in humans (Ryan & Ray, 2004; Whitley & Roizman, 2001). The HSV-1 and 2 viruses consist of a capsid which contains a large, double-stranded DNA genome. The capsid is wrapped in a lipid bilayer structure, known as the envelope (Mettenleiter et al., 2006). The HSV-1 virus is transmitted from individual to individual primarily by oral-oral contact and causes a disease which is known as cold sores. HSV-1 infection is wide-spread throughout the world and in many cases, the infection may remain latent or undetected throughout the whole life of an infected person (Looker et al., 2020). On the other hand, the HSV-2 mainly causes the genital herpes which is transmitted during sexual intercourse (Wald et al., 2001). Although HSV

infection is not fatal in most cases, however, complications can rise if the infection remains untreated for many years. HSV infections leading to other lethal diseases like the Alzheimer's disease, liver failure, meningitis, and herpes simplex encephalitis have also been reported over the years (<http://www.antimicrobe.org/e7.asp>. Retrieved May 5, 2020; Itzhaki et al., 1997; Riediger et al., 2009). According to the WHO, as of 2020, about 3.7 billion people under the age 50 and about 491 million people from the age 15 to 49, have the HSV-1 and HSV-2 infections worldwide. Although in most cases, the HSV infections remain asymptomatic, however, the transmission from an asymptomatic individual can still be occurred. For this reason, HSV is considered as one of the most contagious viruses in the world (<https://www.who.int/news-room/fact-sheets/detail/herpes-simplex-virus>. Retrieved May 5, 2020).

Since HSV establishes a lifelong infection in humans, therefore, this virus can't be completely eradicated from the human body. Several antiviral therapies have been

developed in recent years. Acyclovir, famciclovir, and valacyclovir are the most commonly used antiviral drugs to combat HSV infections. But they can only reduce the severity and frequency of the symptoms. These treatments don't have the capability to fully cure the infections (<https://www.who.int/news-room/fact-sheets/detail/herpes-simplex-virus>). Retrieved May 5, 2020; Koelle & Corey, 2008). Moreover, the extensive use of these drugs sometimes leads to the development of resistance of HSV towards these drugs, which makes the treatment of HSV even more difficult (Balfour, 1983; Tremel et al., 2020). To prevent the HSV infections, scientists are working on to develop potential vaccines against the virus. Several pharmaceutical companies are also involved in developing and testing HSV vaccines, including the Sanofi Pasteur and GlaxoSmithKline (GSK) (WHO, 2014). To date, many vaccine candidates are developed and although some of them showed potential hopes in the initial stages of development, however, in the later stages of clinical trials, all of them had failed to achieve the market access for a variety of safety concerns (Johnston et al., 2016). Therefore, in this study, the reverse vaccinology and immunoinformatics approaches were exploited to develop effective polyvalent vaccines against both HSV types i.e. HSV-1 and HSV-2.

Reverse vaccinology and immunoinformatics are the processes of vaccine development where the novel antigens of a virus or pathogen are identified by analyzing its genomic information. In contrast to the conventional methods of vaccine development, today's cutting-edge technology and the availability of information about the genome of almost every pathogen have made it possible to develop novel peptide-based "subunit vaccines" comprised of only the antigenic protein segments of a target pathogen. The computation-based methods of vaccine designing have also acquired great acceptance among the scientific community around the world in recent years due to their inexpensive and cost-effective nature. In these methods, different tools of bioinformatics and *in silico* biology are used for vaccine development by dissecting the genome and studying the genetic makeup of a target pathogen (Chong & Khan, 2019; Kumar et al., 2015; Rappuoli, 2000). In our study, epitope based polyvalent vaccines were designed which might confer immunity towards the both types of HSV, HSV-1 and 2, targeting the envelope glycoprotein (EG)- E, B, and D of these viruses. Two different strains of HSV-1 i.e. strain-17 and strain-F and two different strains of HSV-2 i.e. strain-HG52 and strain-333 were used for the vaccine construction in our experiment. These four strains are the most widely studied and commonly found strains of the HSV. Using the HSV-1, strain-17 as the model, the vaccines were constructed by the T-cell and B-cell epitopes which were found to be 100% conserved across all the four selected strains. The HSV-1 EG-E, B, and D were used for the epitope prediction and then the epitopes with 100% conservancy across the four selected strains were finally considered for the vaccine construction. Therefore, these vaccines are expected to be effective against all these four selected HSV strains, targeting their EG-E, B, and D. These three glycoproteins of HSV mediate the fusion of the virus with its target cell and thus aid in the

viral entry (Farnsworth et al., 2007; Han et al., 2012; Johnson et al., 2001; Whitbeck et al., 1997). Therefore, these three proteins were used as potential targets in this study for designing the vaccines because inhibiting these viral proteins should prevent the viral entry and thus interfere with the viral life cycle. Figure 1 illustrates a flowchart depicting the step-by-step procedures adapted in the designing of the polyvalent vaccines against the HSV viruses.

2. Materials and methods

2.1. Strain identification and retrieval of protein sequences

The HSV-1 (strain- 17 and strain- F) and HSV-2 (strain- HG52 and strain- 333) strains were identified and selected by reviewing the literatures of the National Center for Biotechnology Information or NCBI (<https://www.ncbi.nlm.nih.gov/>) database. Then, the EG- E, B, and D proteins of the selected HSV strains were retrieved from the UniProt (<https://www.uniprot.org/>) database.

2.2. Antigenicity and physicochemical property analysis of the proteins

The antigenicity of the protein sequences was predicted by the online server, VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>), keeping the prediction accuracy parameter threshold at 0.4. The accuracy, sensitivity, and specificity of a prediction depends on the accuracy prediction threshold and the 0.4 threshold improves the prediction accuracy of the server. The server uses auto cross covariance (ACC) transformation method for predicting the antigenicity of query proteins or peptides and provides results with 70% to 89% accuracy. Therefore, this server is the widely used and accepted server for *in silico* determination of the antigenicity of query proteins (Doytchinova & Flower, 2007). Thereafter, various physicochemical properties of the selected antigenic protein sequences were analyzed by the ExPASy's online tool ProtParam (<https://web.expasy.org/protparam/>). This server computes the physicochemical properties of a query protein where no additional information is required (Gasteiger et al., 2005).

2.3. T-cell and B-cell epitope prediction

An effective multi-epitope subunit vaccine must comprise of cytotoxic T-cell, helper T-cell, and B-cell epitopes so that during the immune response, the vaccine would be able to stimulate these immune cells (Zhang, 2018). The epitopes of the selected protein sequences were predicted using the online epitope prediction server, Immune Epitope Database or IEDB (<https://www.iedb.org/>) and during the predictions, all the parameters in the server were kept at their default values (Vita et al., 2018). The MHC class-I restricted CD8+ cytotoxic T-lymphocyte (CTL) epitopes of the selected sequences were obtained using the recommended NetMHCpan EL 4.0 prediction method (<http://tools.iedb.org/>

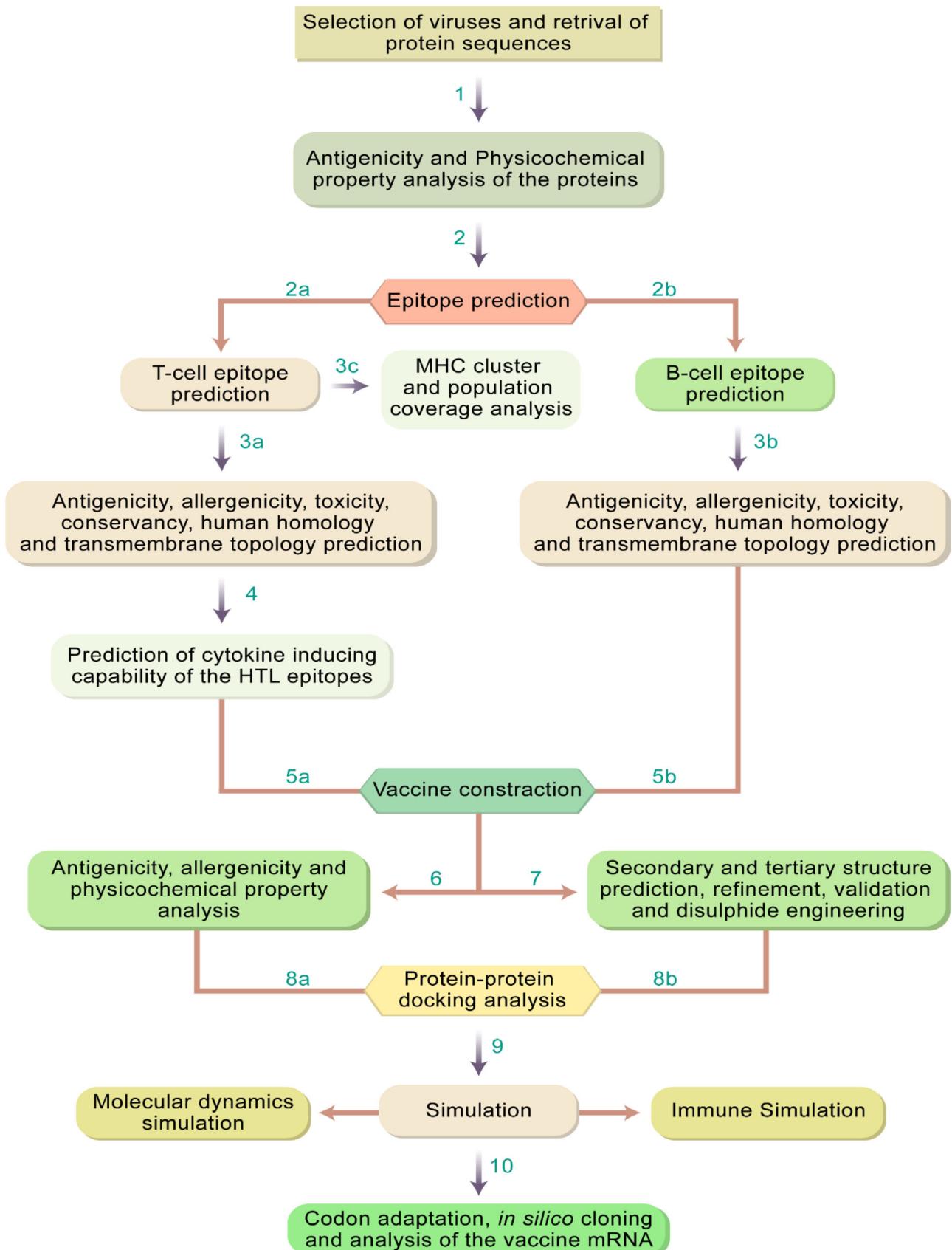


Figure 1. The step-by-step procedures adapted in the vaccine designing experiment.

mhci/) for some common HLA alleles, i.e. HLA A*01:01, HLA A*02:01, HLA A*02:06, HLA A*03:01, HLA A*11:01, and HLA A*29:02 and the length of the epitopes were kept at 9 (9-

mer epitopes). The MHC class-II restricted CD4+ helper T-lymphocyte (HTL) epitopes (15-mer epitopes) were also obtained for some common HLA alleles i.e. DRB1*03:01,

DRB1*04:01, DRB1*15:01, DRB5*01:01, DRB4*01:01, and DRB3*01:01, using the IEDB recommended 2.22 prediction method (<http://tools.iedb.org/mhcii/>). The top ten MHC class-I and MHC class-II epitopes which were found to be common for all the mentioned HLA alleles, were taken for further analyses. On the other hand, B-cell lymphocyte epitopes (BCL) were selected based on their lengths (the epitope that had lengths of more than ten amino acids) and obtained using the BepiPred linear epitope prediction method, keeping all the parameters default.

2.4. Antigenicity, allergenicity, toxicity, and transmembrane topology prediction

The antigenicity of the selected epitopes was predicted using the VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.htm>) server again, keeping the prediction accuracy threshold at 0.4. After that, the allergenicity of the selected epitopes was determined using two online tools, AllerTOP v2.0 (<https://www.ddg-pharmfac.net/AllerTOP/>) and AllergenFP v1.0 (<http://ddg-pharmfac.net/AllergenFP/>). However, the results predicted by AllerTOP v2.0 were given priority since the server has better accuracy of 88.7% than AllergenFP server (87.9%) (Dimitrov et al., 2013, 2014). The toxicity prediction of the selected epitopes was carried out using ToxinPred server (<http://crdd.osdd.net/raghava/tox-inpred/>), keeping all the parameters default. The support vector machine (SVM) method was used for the toxicity prediction in the server and all the parameters were kept default. The SVM is a widely accepted machine learning technique for toxicity prediction since it can differentiate the toxic and non-toxic epitopes quite efficiently (Gupta et al., 2013). Finally, the transmembrane topology of the selected epitopes was determined using the transmembrane topology of protein helices determinant, TMHMM v2.0 server (<http://www.cbs.dtu.dk/services/TMHMM/>). This server generates the transmembrane topology prediction based on the Hidden Markov Model (HMM) with quite good accuracy (Krogh et al., 2001).

2.5. Cytokine inducing capability determination of the epitopes

The helper T-cells (HTLs) produce different types of cytokines like IFN-gamma, IL-4, and IL-10, which later function in activation of other types of immune cells i.e. cytotoxic T-cells, B-cells, macrophages etc. (Luckheeram et al., 2012). Therefore, the ability of the HTL epitopes to induce these cytokines was determined in this experiment. The IFN-gamma, IL-4, and IL-10 inducing capability of the HTL epitopes were predicted by the IFNepitope (<http://crdd.osdd.net/raghava/ifnepitope/>), IL4pred (<http://crdd.osdd.net/raghava/il4pred/>), and IL10pred (<http://crdd.osdd.net/raghava/IL-10pred/>) servers, respectively. All these servers are easy, widely used, and user-friendly tools for predicting the cytokine inducing ability of query peptides or proteins (Dhanda, Gupta, et al., 2013; Dhanda, Vir, et al., 2013; Nagpal et al., 2017). During the INF-gamma inducing capability prediction, Hybrid (Motif + SVM)

prediction method was used and both the IL4pred and IL10pred predictions were conducted based on the SVM prediction method where the default threshold parameters and values were used.

2.6. Conservancy and human homology determination of the epitopes

The conservancy analysis of the selected epitopes were performed using the 'epitope conservancy analysis' module of IEDB server (<https://www.iedb.org/conservancy/>) (Bui et al., 2007). The epitopes that were found to be 100% conserved (along with some other criteria) among the selected strains, were considered for vaccine construction because this will ensure the broad spectrum activity of the polyvalent vaccine over the selected strains. The homology of the epitopes to the human proteome was determined to find out any epitope that is homolog to the human proteome. Only the non-homolog epitopes were taken because if any epitope is homolog to the human proteome, then that epitope might not have the capability to induce potential immune response. The protein BLAST module (blastP) of BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) tool was used in the human homology determination, where *Homo sapiens* (taxid: 9606) was used for comparison keeping all other parameters default (Altschul et al., 1990). An e-value cut-off value of 0.05 was set in the experiment and the epitopes that had no hits below the e-value inclusion threshold, were selected as non-homologous peptides (Mehla & Ramana, 2016).

Based on the antigenicity, allergenicity, toxicity, conservancy, and human homology analysis, the most promising epitopes were selected for the vaccine construction. The epitopes that showed high antigenicity, non-allergenicity, non-toxicity, high conservancy (100% conservancy), and non-homology to the human proteome, were considered as the most promising epitopes or the best selected epitopes. But the HTL that at least one cytokine inducing capability along with all these mentioned criteria, were selected as the best and most promising HTL epitopes.

2.7. Population coverage analysis

The distribution of specific HLA alleles among different ethnicities and population around the world is an important prerequisite for designing a multi-epitope vaccine. The expression of different HLA alleles may vary from one population to another. The IEDB population coverage tool (<http://tools.iedb.org/population/>) was used to determine the population coverage of the best-selected epitopes across multiple HLA alleles in different regions around the world (Adhikari & Rahman, 2017; Bui et al., 2007).

2.8. Cluster analysis of the MHC alleles

Cluster analysis of the MHC alleles helps to identify the alleles of the MHC class-I and class-II molecules that have similar binding specificities. The cluster analysis of the MHC alleles were carried out by the online tool MHCcluster 2.0

(<http://www.cbs.dtu.dk/services/MHCcluster/>) (Thomsen et al., 2013). During the analysis, the number of peptides to be included was kept 50,000, the number of bootstrap calculations were kept 100 and all the HLA super-type representatives (MHC class-I) and HLA-DR representatives (MHC class-II) were selected. The server generates results in the form of MHC specificity tree and MHC specificity heat-map.

2.9. Vaccine construction

Three possible polyvalent vaccines were constructed to work against the HSV strains. The predicted CTL, HTL, and BCL epitopes were conjugated together for constructing the vaccines. All the vaccines were generated maintaining the sequence: adjuvant, PADRE sequence, CTL epitopes, HTL epitopes and BCL epitopes. Three different adjuvant sequences were used for constructing three different vaccines: human beta defensin-3, L7/L12 ribosomal protein, and HABA protein (*M. tuberculosis*, accession number: AGV15514.1). Beta-defensin-3 adjuvant induces the activation of the toll like receptors (TLRs): 1, 2, and 4. And both L7/L12 ribosomal protein and HABA protein activate the TLR-4. During the vaccine construction, various linkers were used i.e. EAAAK, GGGGS, GPGPG, and KK linkers (Solanki & Tiwari, 2018; Ullah et al., 2020). The EAAAK linker provides effective separation of the domains of a bifunctional fusion protein (Arai et al., 2001). Again, the flexible linker, GGGGS protects a protein from degradation by protease enzymes (Wen et al., 2013). Furthermore, the GPGPG linker has the ability to facilitate the immune processing and presentation and it also prevents the generation of junctional epitopes (Saadi et al., 2017). And the bi-lysine (KK) linker preserves the independent immunological activities of the epitopes of a vaccine (Gu et al., 2017). For these reasons, these linkers were used to construct the vaccines in this study. Again, many studies have proved that the PADRE sequence improves the CTL response of the vaccines that contain it (Sarkar, Ullah, & Araf, 2020). Total three vaccines were constructed in the experiment. Figure 2 depicts a generalized structure of the three vaccine constructs.

2.10. Antigenicity, allergenicity, and physicochemical property analyses

The antigenicity of the constructed vaccines was determined by the online server VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.htm>), where again the accuracy threshold was kept at 0.4. Thereafter, AlgPred (<http://crdd.osdd.net/raghava/algpred/>) and AllerTop v2.0 (<https://www.ddg-pharmfac.net/AllerTOP/>) servers were used for the prediction of the allergenicity of the vaccine constructs. The AlgPred server predicts the possible allergens based on similarity of known epitope of any of the known region of a protein (Saha & Raghava, 2006). The MEME/MAST motif prediction approach was used in the allergenicity prediction of the vaccines by AlgPred server. Moreover, various physicochemical properties of the vaccines were examined by the online server, ProtParam (<https://web.expasy.org/protparam/>).

Along with the physicochemical property analysis, the solubility of vaccine constructs upon over-expression in *E. coli* host was also predicted by the SOLpro module of the SCRATCH protein predictor (<http://scratch.proteomics.ics.uci.edu/>) and later further clarified by the Protein-Sol server (<https://protein-sol.manchester.ac.uk/>). Both these servers can predict the solubility of a query protein with quite good accuracy. The SolPro generates its predictions based on the SVM method whereas the Protein-Sol utilizes a fast sequence based method for determining the results (Hebditch et al., 2017; Magnan et al., 2009). During the solubility analysis, all the parameters of the servers were kept at their default values.

2.11. Secondary and tertiary structure prediction of the vaccine constructs

The secondary structures of the vaccine constructs were generated using online tool PRISPRED (<http://bioinf.cs.ucl.ac.uk/prispred/>), keeping all the parameters default. PRISPRED is a simple secondary structure generator which can be used to predict the transmembrane topology, transmembrane helix, fold, and domain recognition etc. along with the secondary structure prediction (Buchan & Jones, 2019; Jones, 1999). Thereafter, the tertiary or 3D structures of the vaccines were generated using online tool RaptorX (<http://raptorgx.uchicago.edu/>) server. The server predicts the tertiary structure of query proteins by template based method and generates p-values for the predicted structures. The p-values represent the model quality of the predicted protein structures and the lowest p-value always represents the best quality model (Sarkar, Ullah, Johora, et al., 2020; Wang et al., 2016).

2.12. 3D structure refinement and validation

When protein 3D structures are predicted by computational methods, they may lack their true, native structures. Therefore, the 3D structure refinement was performed to convert the low resolution predicted models to models with higher resolution that closely resemble the native protein structures. The generated 3D structures of the vaccines were refined by GalaxyRefine module of the GalaxyWEB server (<http://galaxy.seoklab.org/>), which uses CASP10 tested refinement method and dynamics simulation to provide better-refined structures (Ko et al., 2012; Nugent et al., 2014). After structure refinement, the vaccine constructs were validated by analyzing the Ramachandran plots, generated by PROCHECK (<https://servicesn.mbi.ucla.edu/PROCHECK/>) server (Laskowski et al., 2006; Morris et al., 1992). Then another online tool, ProSA-web (<https://prosa.services.came.sbg.ac.at/prosa.php>) was also used for protein validation. The server works on various statistical methods to generate z-score which is used to express the quality of a query protein structure. A z-score within the range of the z-scores of all the experimentally determined protein chains in the PDB database represents better quality of a query protein (Wiederstein & Sippl, 2007).

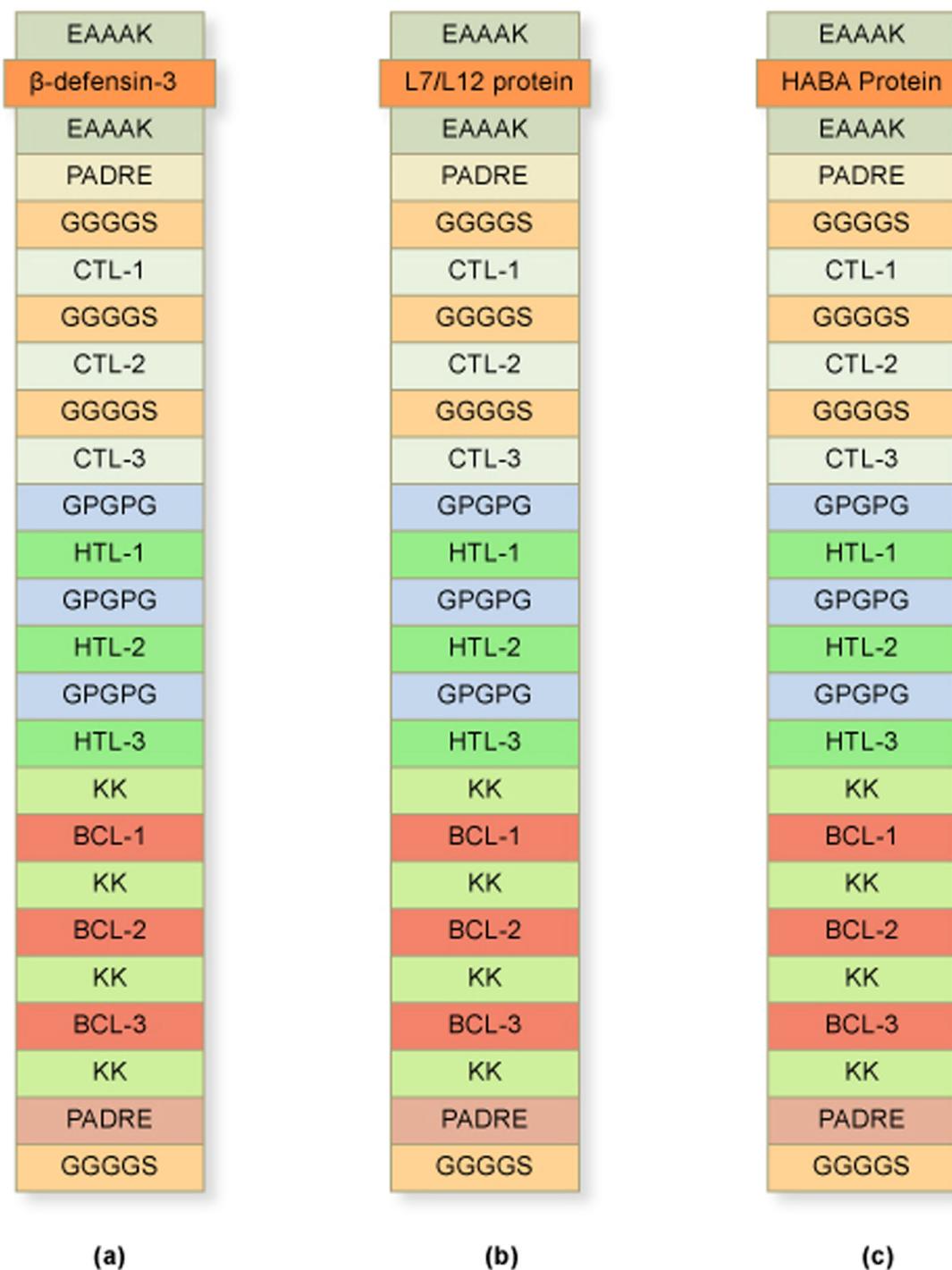


Figure 2. A schematic representation of the three possible vaccine constructs with linkers (EAAAK, GGGGS, GPGPG, KK), PADRE sequence, adjuvants (human beta-defensin-3, L7/L12 protein, and HABA protein) and epitopes (CTL, HTL, BCL) in sequential and appropriate manner. (a) is the first vaccine constructed using the beta-defensin-3 adjuvant, (b) is the second vaccine constructed using L7/L12 adjuvant protein, and (c) is the third vaccine constructed using the HABA protein as an adjuvant. CTL; cytotoxic T lymphocytic epitope, HTL; helper T lymphocytic epitope, BCL; B cell lymphocytic epitope. The three vaccine constructs differ from each other only in their adjuvant sequences.

2.13. Vaccine protein disulfide engineering

The vaccine protein disulfide engineering was carried out by the online tool, Disulfide by Design 2 v12.2 (<http://cptweb.cpt.wayne.edu/DbD2/>). The server predicts the possible sites within a protein structure which have the greater possibility of undergoing disulfide bond formation. When engineering the disulfide bonds, the intra-chain, inter-chain, and C_{β} for

glycine residue were selected. The χ_3 angle was kept -87° or $+97^{\circ} \pm 5$ and $C_{\alpha}\text{-}C_{\beta}\text{-}S_{\gamma}$ angle was kept $114.6^{\circ} \pm 10$ and the amino acid pairs with less than 2.2 kcal/mol bond energy were selected for mutation into cysteine residues to form the disulfide bonds among themselves. The bond energy 2.2 kcal/mol was used as a threshold in this study because 90% of native disulfide bonds are generally found to have

energy value less than 2.2 Kcal/mol (Craig & Dombkowski, 2013).

2.14. Protein-protein docking

In the protein-protein docking analysis, the three constructed vaccines were analyzed by docking against various toll like receptors (TLRs). One best vaccine was selected based on its performances in the docking experiment. The TLRs are some of the key players that mediate immune responses of the immune cells i.e. B-cells, T-cells, dendritic cells, macrophages etc. In this experiment, the vaccines constructs were docked against TLR-1 (PDB ID: 6NIH), TLR-2 (PDB ID: 3A7C), TLR-3 (PDB ID: 2A0Z), TLR-4 (PDB ID: 4G8A), TLR-8 (PDB ID: 3W3M), and TLR9 (PDB ID: 4QDH). The vaccine that showed the best performance in the docking study, was considered as the best vaccine construct. The protein-protein docking was carried out using various online docking tools to improve the prediction accuracy. First, the docking was carried out by ClusPro 2.0 (<https://cluspro.bu.edu/login.php>). The server ranks the clusters of docked complexes based on their center and lowest energy scores (Kozakov et al., 2013, 2017; Vajda et al., 2017). The ClusPro server calculates the energy score based on the following equation:

$$E = 0.40E_{\text{Rep}} + (-0.40E_{\text{Att}}) + 600E_{\text{Elec}} + 1.00ED_{\text{ARS}} \quad (\text{Craig \& Dombkowski, 2013; Wiederstein \& Sippl, 2007})$$

Thereafter, the docking was again performed by PatchDock (<https://bioinfo3d.cs.tau.ac.il/PatchDock/php.php>) server keeping all the parameters at their default values (Schneidman-Duhovny et al., 2005). The tool works on specific algorithms that act by analyzing the Connolly dot surface representations and root mean square deviation (RMSD) clustering scores of the candidate compounds. After docking by the PatchDock server, the results were refined and re-scored by the FireDock server (<http://bioinfo3d.cs.tau.ac.il/FireDock/php.php>) (Andrusier et al., 2007). The FireDock server ranks the docked complexes based on their global energy scores and the lower the energy score, the better the result. Finally, the docking was conducted using the HawkDock server (<http://cadd.zju.edu.cn/hawkdock/>) for the third time. The Molecular Mechanics/Generalized Born Surface Area (MM-GBSA) study was also carried out using the HawkDock server, where all the parameters were kept default (Feng et al., 2017; Hou et al., 2011; Weng et al., 2019). From the docking experiment, one best vaccine was selected for further analyses. After successful docking, the TLR-3 docked with the best vaccine construct, was taken for visualization and performing the molecular dynamics (MD) simulation study. The docked complex was picked randomly from all the docked complex and was visualized by the Discovery Studio Visualizer (Biovia, 2017).

2.15. Conformational B-cell epitope prediction

The humoral immunity of the body is dependent on the B-cells that produce antibodies when they encounter an antigen. Therefore, the vaccines should have effective conformational B-cell epitopes to provide much stronger immunity.

The conformational B-cell epitopes of the best-predicted vaccine protein from the 2.14 sub-section were determined by IEDB ElliPro tool (<http://tools.iedb.org/ellipro/>), using the default parameters (Ponomarenko et al., 2008).

2.16. Molecular dynamics simulation

MD simulation was performed to observe the change of state and effects of the environment on the protein in a biological environment. For this purpose, GROMACS (GROningen MAchine for Chemical Simulations) (Abraham et al., 2015) a Linux based command-line program had been used. MD simulation undergoes many stages where the topology is generated and the protein complex's energies are minimized before performing the final run. The protein 'pdb' file had first been cleaned to remove any unwanted environment substrates. Then the pdb2gmx was run using OPLS-AA (Optimized Potential for Liquid Simulation-All Atom) force field (Kaminski et al., 2001) to generate the topology. The protein was positioned in the center of a 2 nm sized cube resulting in it being 1 nm from each edge and its periodic image being generated 2 nm apart. The volume of this cubic box was found to be 2389.27 nm³. The box had been solvated with water molecules which were spatially placed with a force constant of 1000 kJ mol⁻¹ nm⁻². The system had then been neutralized using the 'gmx ionize' command. Next energy minimization had been performed to stabilize the protein complex. Supplementary Figure S1 depicts the potential energy of the structure had quickly reduced below the order of 10⁶. From this, we inferred that the system was stable enough to conduct further simulations. To stabilize the temperature, NVT (Number Volume Temperature) equilibration had been performed for 100 ps. Following that, the NPT (Number Pressure Temperature) equilibration had been performed for 100 ps and thereafter, the pressure and density had been calculated. This resulting stabilized structure was later subjected to the MD simulation for 50 ns. The RMSD (Root-Mean-Square Deviation) of backbone of the energy minimized structure had been predicted and radius of gyration (Rg) was also calculated. All plots and simulation graphs had been analyzed using the XmGrace (QtGraceTcpServer and Client-code, 2014–2017) and QtGrace (Turner, 2005) tool.

2.17. Immune simulation

To characterize the immunogenicity and immune response profile of the best selected vaccine, the immune simulation study was performed by C-IMMSIM server (<http://150.146.2.1/C-IMMSIM/index.php>). The server generates predictions of the real-life-like immune interactions using the machine learning methods and position-specific scoring matrix (PSSM) techniques (Rapin et al., 2010). While conducting the experiment, all the parameters were kept default, except the time steps which were set at 1, 84, and 170 (time step 1 is injection at time = 0) and the number of simulation steps was set at 1050. So, three injections would be given at four weeks apart because the recommended interval between

two doses of most of the available commercial vaccines is four weeks (Castiglione et al., 2012). The Simpson's Diversity index, D was calculated from the figures.

2.18. Codon adaptation and *in silico* cloning

The codon adaptation and *in silico* cloning were carried out only for the best selected vaccine protein. For conducting these experiments, the vaccine protein was reverse translated to the possible DNA sequence. Later, the DNA sequence was adapted according to the desired organism, so that the cellular machinery of that specific organism could use the codons of the adapted DNA sequence efficiently and provide better production of the desired product. Codon adaptation is a necessary step of *in silico* cloning since the same amino acid can be encoded by different codons in different organisms, a phenomenon which is known as codon bias. Moreover, the cellular mechanisms of an organism may be completely different from another organism and a codon for a specific amino acid may not work in another organism (Carbone et al., 2003; Sharp & Li, 1987). Therefore, codon adaptation step is performed which predicts the suitable codon for a specific amino acid in a specific organism. The codon adaptation of the selected vaccine protein was carried out by the Java Codon Adaptation Tool or JCat server (<http://www.jcat.de/>) (Grote et al., 2005). Eukaryotic *E. coli* strain K12 was selected and rho-independent transcription terminators, prokaryotic ribosome binding sites and Eael and StyI cleavage sites of restriction enzymes, were avoided. Then the optimized DNA sequence was taken and Eae1 and StyI restriction sites were conjugated at the N-terminal and C-terminal sites, respectively. Finally, the SnapGene restriction cloning software was used for inserting the newly adapted DNA sequence between the Eae1 and StyI restriction sites of the pETite vector (Lucigen, USA) (Solanki & Tiwari, 2018). The pETite vector contains two tags within its DNA sequence i.e. the SUMO tag (small ubiquitin-like modifier) and 6X-His tag. These tags facilitates the solubilization and effective affinity purification of the recombinant protein (Biswal et al., 2015).

2.19. Analysis of the mRNA secondary structure

After the *in silico* cloning experiment, the mRNA secondary structure of the best vaccine protein was determined by two servers i.e. Mfold (<http://unafold.rna.albany.edu/?q=mfold>) and RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>). These two servers predict the minimum free energy (ΔG Kcal/mol) for each of the generated mRNA secondary structures and the lower minimum free energy represents to more stably folded mRNA structure (Gruber et al., 2008; Mathews et al., 1999, 2007; Zuker, 2003). To predict the mRNA secondary structure of the best-selected vaccine, the optimized DNA sequence from JCat server was taken and the converted to RNA sequence by the DNA \leftrightarrow RNA \rightarrow Protein tool (<http://biomodel.uah.es/en/lab/cybertory/analysis/trans.htm>). After that, the RNA sequence was collected

Table 1. List of the proteins used in this study with their UniProt accession numbers. NA; data not available.

No	Virus Strain	Name of the protein	UniProt accession number
01	HSV-1, strain-17	Envelope glycoprotein E	P04488
02		Envelope glycoprotein B	P10211
03		Envelope glycoprotein D	Q69091
04	HSV-1, strain-F	Envelope glycoprotein E	Q703F0
05		Envelope glycoprotein B	P06436
06		Envelope glycoprotein D	Q05059
07	HSV-2, strain-HG52	Envelope glycoprotein E	P89475
08		Envelope glycoprotein B	P08666
09		Envelope glycoprotein D	Q69467
10	HSV-2, strain-333	Envelope glycoprotein E	NA
11		Envelope glycoprotein B	P06763
12		Envelope glycoprotein D	P03172

from the tool and used in the Mfold and RNAfold servers for prediction keeping all the parameters default.

3. Results and discussions

3.1. Identification, selection, and retrieval of viral protein sequences

The four strains of HSV and their EG proteins were selected by reviewing literatures from the NCBI database and then the protein sequences were retrieved from the UniProt database. Table 1 lists the UniProt accession IDs of the selected protein sequences.

3.2. Antigenicity and physicochemical property analysis of the proteins

In the physicochemical property analysis, all the selected proteins were found to be antigenic, so all of them should stimulate potential immune response within the human body. The theoretical pI describes the pH at which a protein does not have any net charge. Only the EG-E proteins from HSV-1 strain-17, strain-F and HSV-2 strain-HG52 were found to possess acidic pI (pH less than 7.0). So, these proteins would require less than 7.0 pH to be neutralized which is quite achievable. The aliphatic index of a protein refers to the relative volume of the amino acids in its side chains occupied by the aliphatic amino acids i.e. alanine, valine, etc. (Ikai, 1980). Again, the GRAVY value determines the hydrophilic and hydrophobic characteristics of a compound. The negative GRAVY value represents hydrophilic characteristic, whereas the positive GRAVY value corresponds to hydrophobic characteristic of a compound (Chang & Yang, 2013; Kyte & Doolittle, 1982). With the negative GRAVY value, all the proteins were found to be hydrophilic in nature. So, the proteins would be easily soluble in the water. All of the query proteins were also found to have quite high and similar half-life of 30 h in mammalian cell culture system which is quite satisfactory and they also had high aliphatic index (over 70.00). The aliphatic index represents the protein's thermal stability and the higher aliphatic index of a protein corresponds to its more thermostable state (Panda & Chandra, 2012). So, with quite high aliphatic indexes, all the proteins were declared to be thermostable. The EG-B protein from

Table 2. List of the epitopes finally selected for vaccine construction (selection criteria: antigenicity, non-allergenicity, non-toxicity, 100% conservancy, and non-homology to the human proteome for all the epitopes and at least one cytokine production capability for the HTL epitopes only).

Epitope class	Envelope glycoprotein E	Envelope glycoprotein B	Envelope glycoprotein D
MHC class-I epitopes	YTLSVGDIK ITISTAAQY AVKSRSAGK	KVTDVMVRK YAYSHQLSR ASANASVER	KMADPNRFR SIQDAATPY
MHC class-II epitopes	GAVMGAALLLSALGL AVMGAAALLSALGGS	IRYMALVSAMERTEH	ILFVVIVGLHGVRSK GAVILFVVIVGLHGVS VILFVVIVGLHGVR AVILFVVIVGLHGVR FVVIVGLHGVRSKYA
B-cell epitopes	PECLSPADAPCAAST AYAPPAPSATGGL VWEQPLPQRGADLAEPTHPHVGAPPHTHG YADWSSDSEGERDQVWLAPPERPDSPSTNGSG	TTKARATAPTRN RTEHKAKKKG REQSRKPPNPTPPPGASANAS	IREDDQPSSHQ

the HSV-1 strain-17 had the lowest GRAVY value of -0.403 . All these results of the physicochemical analysis were found to be quite sound and satisfactory. The results of the physicochemical property analysis are listed in [Supplementary Table S1](#).

3.3. T-cell and B-cell epitope prediction

After determining the physicochemical properties, the T-cell and B-cell epitopes were predicted for vaccine construction. The cytotoxic T-cells function in the recognition of antigens, whereas the helper T-cells aid in activating the B-cell, macrophages, and even the cytotoxic T-cells (Chaudhri et al., 2009; Zhu & Paul, 2008). Furthermore, the B-cells mediate the humoral immune response by producing antibodies. However, the humoral immune response is not so robust like the cell mediated immune response and may get weaker overtime (Bacchetta et al., 2005; Cooper & Nemerow, 1984). On the other hand, the cell mediated immune response can provide much broader and life-long immunity by secreting antiviral cytokines and specifically identifying and destroying the infected cells (Cano & Lopera, 2013; Garcia et al., 1999). To determine the epitopes, the HSV-1, strain-17 was selected as the model and the epitopes were selected for HSV-1, strain-17. Then the epitopes that were found to be highly antigenic, non-allergenic, non-toxic, and 100% conserved across other strains were finally selected for vaccine construction. This 100% conservancy ensures the broad-spectrum activity of the vaccines over all the four selected strains of HSV-1 and HSV-2. Among the hundreds of epitopes generated by the server, the top T-cell epitopes were selected for further analysis. And the B-cell epitopes with lengths of more than ten amino acids, were selected after predicting by the BepiPred linear prediction method of the IEDB server. [Supplementary Tables S2](#) and [S3](#) list the potential T-cell epitopes of EG-E, [Supplementary Tables S4](#) and [S5](#) list the potential T-cell epitopes of EG-B, [Supplementary Tables S6](#) and [S7](#) list the potential T-cell epitopes of EG-D, and [Supplementary Table S8](#) lists the potential B-cell epitopes with their respective topologies.

3.4. Antigenicity, allergenicity, toxicity, conservancy, and human homology analyses

In the antigenicity, allergenicity, toxicity, conservancy, and human-homology analyses, the epitopes that were found to

be highly antigenic (so that the epitopes would stimulate high antigenic response), non-allergenic (so that the epitopes would not cause any unwanted allergic reaction to occur within the body), and non-toxic along with the 100% conservancy and non-homology to the human proteome (so that the epitopes won't be regarded as self-antigens) were considered as the "best selected epitopes" or "most promising epitopes" and selected for further analyses. Most of the epitopes were found to be non-homolog to the human proteome. So, all of them should be recognized as foreign particles or antigens within the human body. Again, the HTL epitopes that were at least one cytokine inducer along with all these criteria, were considered as the best HTL epitopes. [Table 2](#) lists the best selected epitopes used in the vaccine construction.

3.5. Cytokine inducing ability prediction

Various types of cytokines i.e. the IFN-gamma, IL-10, and IL-4 are required for the proper activation and proper functioning of many immune cells (Luckheeram et al., 2012), therefore, the cytokine production ability of the HTL epitopes was also determined. The IFN-gamma, IL-4, and IL-10 inducing capacity prediction of the HTL epitopes showed that most of the HTL epitopes had at least one of these cytokine producing capability and the HTL epitopes which were selected for vaccine construction, were also found to have the ability to induce at least one cytokine ([Supplementary Table S3](#), [Supplementary Table S5](#), and [Supplementary Table S7](#)). This cytokine production or inducing ability would impact greatly on the immunogenic activities of the vaccines.

3.6. Population coverage analysis

The population coverage analysis showed that the MHC class-I and class-II epitopes covered 90.48% and 88.70%, respectively of the world population and the combined MHC class-I and class-II covered 85.12% of the world population. North America had the highest percentage of population coverage of the MHC class-I epitopes (81.22%). North America also had the highest percentage of population coverage of the MHC class-I and class-II epitopes in combination (88.61%). However, India had the highest population coverage of the MHC class-II epitopes of 81.73% ([Supplementary Figure S3](#)). From the population coverage

Table 3. List of the vaccines constructed against the four selected strains. The bolded letters represent the linker sequences.

Name of the vaccines	Vaccine constructs
Herpes Simplex Virus vaccine-1 (HV-1)	EAAAKGIINTLQKYYCRVRRGRCAVLSCLPKEEIQIGKCSTRGRKCCRRKKEAAAKAKFVAAWTLKAAAGGGGS YTLSVGDIKG ^{GGGGSITISTAAQYGGGGSAVSRAS} GKVTDMVRKGGGSYAYSHQLSRGGGGSA SANASVERGGGGSKMADPNRFRGGGGSSIQDAATPYGPGPGGAVMGAALLLSALGLGP ^{PGAVMGAALLL} S ALGLSGP ^{GP} GI ^{RY} MALVSAMERTEHGP ^{GP} ILFVVIVGLHGVR ^{SK} GP ^{PG} GA ^V ILFVVIVGLHG ^V GP ^{GP} GV ^I FVVIVG LHGVRSGP ^{GP} GA ^V ILFVVIVGLHGVR ^{GP} GFVVIVGLHGVR ^{SKY} AKKPECLSPAD ^{CA} STKKAYAPPAPSATGGL KKVVEQPLPQRGADLAEP ^T H ^P HVGAPP ^H PTHGKKYADWSSDSEGERDQVPWLAPPERPDSPSTNGSKK TTKARATAPTRNKK RTEHKAKKGKKREQS ^R KPPNPTPPP ^G ASANASKKIREDDQPSHQKKAKFVAAWTLKAAAGGGGS
Herpes Simplex Virus vaccine-2 (HV-2)	EAAAKMAKLSTDELLDAFKEMTLL^ESDFV^KFEETF^EV^AAA^PVA^AAGA^PAA^VEEAAEQSEFDVILEAAGDKKIGVIKV VREIVSGLGLKEAKDVL ^D GAP ^K PLLEKVAKEADEAKAKLEAAGATVT ^K EEAAAKAKFVAAWTLKAAAGGGGS YTLSVGDIKG ^{GGGGSITISTAAQYGGGGSAVSRAS} GKVTDMVRKGGGSYAYSHQLSRGGGGSASANASVER GGGGSKMADPNRFRGGGGSSIQDAATPYGPGPGGAVMGAALLLSALGLGP ^{PG} GA ^V MGAA ^L LSALGLSGP ^{GP} GI ^{RY} MA LVSAMERTEHGP ^{GP} ILFVVIVGLHGVR ^{SK} GP ^{PG} GA ^V ILFVVIVGLHG ^V GP ^{GP} GV ^I FVVIVGLHGVRSGP ^{GP} GA ^V ILFVVIVGLH GRV ^{GP} GP ^{GP} GA ^V ILFVVIVGLHGVR ^{SKY} AKKPECLSPAD ^{CA} STKKAYAPPAPSATGGLKKVVEQPLPQRGADLAEP ^T H ^P HVGAPP ^H PTHGKKYADWSSDSEGERDQVPWLAPPERPDSPSTNGSKKTTKARATAPTRNKKRTEHKAKKGKK REQS ^R KPPNPTPPP ^G ASANASKKIREDDQPSHQKKAKFVAAWTLKAAAGGGGS
Herpes Simplex Virus vaccine-3 (HV-3)	EAAAKMAENPNIDDL^PAP^LLA^GAD^LALATVN^DLI^NLR^EET^ATR^RTR^VEERR^RRL^TKFQEDL^PEQFIELR DKFTTEEL ^R KA ^E GYLEAT ^N RYNELVERGEAE ^L QRLRSQTAFED ^A SARAEGYV ^D QAVEL ^T QEALGTVASQTR AVGERAAKLV ^G IELEAAAKFVAAWTLKAAAGGGGS ^T Y ^L TSVGDIKG ^{GGGGSITISTAAQYGGGGSAVSRAS} GKGGGGS KVTDVMVRKGGGGSYAYSHQLSRGGGGSASANASVERGGGGSKMADPNRFRGGGGSSIQDAATPYGPGPG GAVMGAALLLSALGLGP ^{PG} GA ^V MGAA ^L LSALGLSGP ^{GP} GI ^{RY} MA ^V LSAMERTEHGP ^{GP} ILFVVIVGLHGVR ^{SK} GP ^{PG} GAVILFVVIVGLHG ^V GP ^{GP} GA ^V ILFVVIVGLHGVR ^{SK} GP ^{PG} GA ^V ILFVVIVGLHG ^V GP ^{GP} GV ^I VGLHGVR ^{SKY} AKKPE CLSPAD ^{CA} STKKAYAPPAPSATGGLKKVVEQPLPQRGADLAEP ^T H ^P HVGAPP ^H PTHGKKYADWSSDSEGERD QVPWLAPPERPDSPSTNGSKKTTKARATAPTRNKKRTEHKAKKGKKREQS ^R KPPNPTPPP ^G ASANASKKIRE DDQPSSHQKKAKFVAAWTLKAAAGGGGS

analysis, it is clear that, the epitopes and their alleles are quite common among the population from different ethnicities and countries across the globe.

3.7. Cluster analysis of the MHC alleles

The cluster analysis of the possible MHC class-I and MHC class-II alleles that may interact with the predicted epitopes of the three selected proteins, were performed by the online tool MHCcluster 2.0. The cluster analysis showed that all the alleles used in this study were related to each other as well as some other alleles, as predicted by the MHCcluster 2.0 tool in phylogenetic manner. *Supplementary Figure S4* illustrates the outcome of the experiment where the red zone indicates strong interaction and the yellow zone corresponds to weaker interaction.

3.8. Vaccine construction

After successful docking, three vaccines were constructed, that could be used effectively to fight against the selected viral strains. For vaccine constructions, three different adjuvants were used to construct three different vaccines: human beta defensin-3, L7/L12 ribosomal protein, and HABA protein, were used as adjuvants. Adjuvants are known to enhance the antigenicity, immunogenicity, stability, and longevity of the constructed vaccines (Lee & Nguyen, 2015; Meza et al., 2017). The PADRE sequence was also used for constructing the vaccine which was proved to enhance the activity of the vaccines. The three different vaccine constructs differed from each other only in their adjuvant sequences. During vaccine construction, EAAAK, GGGGS, GPGPG, and KK linkers were used at their appropriate positions, as depicted in the *Figure 2*. The newly constructed vaccines were designated as: HV-1, HV-2, and HV-3 (*Table 3*).

3.9. Antigenicity, allergenicity, and physicochemical property analyses of the vaccine constructs

All the vaccines were found to be antigenic as well as non-allergenic in the antigenicity and allergenicity analyses. Therefore, they might provoke high immune responses without causing any unwanted allergic reaction within the body. Thereafter, the physicochemical property analysis of the three vaccine constructs was performed. In the physicochemical property analysis, all the vaccines were found to have quite similar theoretical pI which implies that all the three vaccines could be neutralized at the same pH. Since all the vaccines were found to possess quite high aliphatic indexes, so all of them were considered to be thermostable (Chaudhri et al., 2009). So, the vaccines should be stable at the normal body temperature. Moreover, the negative GRAVY value of the three vaccine proteins revealed that all of them might be hydrophilic. The hydrophilic characteristic of the proteins should aid in easy purification and formulation of the vaccines (Chang & Yang, 2013; Kyte & Doolittle, 1982). And since all of the vaccine constructs were found to have half-life of more than 10 h in *E. coli*, therefore, their half-life might not cause any problem during the mass production and purification of the vaccines in the *E. coli* cell culture system. Solubility is a major factor for the post-production studies of vaccines because higher solubility of a protein represents better purification during the downstream processing (Magnan et al., 2009). All the vaccine proteins were found to be soluble upon over-expression in *E. coli* by the both servers (SolPro and Protein-Sol). And the instability indexes (less than 40) of the vaccine proteins indicated that they might be quite stable in the biological environment because a compound with instability index less than 40 is considered to be stable (Guruprasad et al., 1990). Considering all these aspects of the physicochemical property analysis, it can be

Table 4. Antigenicity, allergenicity, and physicochemical property analyses of the three vaccines constructs. AN; antigenicity, AG; allergenicity, AI; aliphatic index, II; instability index, GRAVY; grand average of hydropathicity.

Name of the vaccine	AN	AG	pl	Number of positively charged amino acids	Number of negatively charged amino acids	Ext. coefficient (in M ⁻¹ cm ⁻¹)	Est. half-life (in mammalian and <i>E. coli</i> cell culture system)	AI	II	GRAVY	Solubility
HV-1	Antigenic	Non-allergenic	10.24	72	28	38890	1 h, >10 h	71.01	Stable (37.45)	-0.269	Soluble (SolPro: 0.721; ProteinSol: 0.682)
HV-2	Antigenic	Non-allergenic	9.72	75	51	35535	1 h, >10 h	78.29	Stable (33.23)	-0.141	Soluble (SolPro: 0.630; ProteinSol: 0.686)
HV-3	Antigenic	Non-allergenic	9.77	80	57	40005	1 h, >10 h	76.29	Stable (38.56)	-0.295	Soluble (SolPro: 0.648; ProteinSol: 0.674)

considered that the three predicted vaccines might be suitable as potential vaccine candidates (Table 4).

3.10. Secondary and tertiary structure prediction of the vaccine constructs

From the secondary structure analysis, it was predicted that, the HV-1 had the highest percentage of the amino acids (63.4%) in the coil formation as well as the highest percentage of amino acids (14.6%) in the beta-strand formation. However, HV-3 had the highest percentage of 39.8% of its amino acids in the alpha-helix formation (Supplementary Figure S5 and Supplementary Table S9). So, it can be concluded that the difference of the adjuvant sequences had caused some significant changes in the secondary structures of the three vaccines.

The 3D structures of the vaccine constructs were predicted by the online server RaptorX. All the three vaccines had 3 domains and HV-2 had the lowest p-value of 8.91e-05. So, HV-2 vaccine were predicted to have relatively good quality 3D structure than the other two vaccines. The homology modeling of the three vaccine constructs were carried out using 1KJ6A (for HV-1), 1DD3A (for HV-2), and 4TQLA (for HV-3) as templates from the Protein Data Bank. The results of the 3D structure analysis are listed in Supplementary Table S10 and illustrated in Figure 3.

3.11. Protein 3D structure refinement and validation

The protein structures generated by the RaptorX server were refined using Galaxy-web server, which were then analyzed by Ramachandran plots generated by the PROCHECK server and the z-scores generated by the ProSA-web server. The Ramachandran plot analysis showed that HV-1 vaccine had 63.9% of the amino acids in the most favored region, 30.6% of the amino acids in the additional allowed regions, 3.7% of the amino acids in the generously allowed regions, and 1.8% of the amino acids in the disallowed regions. The HV-2 vaccine had 70.4% of the amino acids in the most favored regions, 26.2% of the amino acids in the additional allowed regions, 2.9% of the amino acids in the generously allowed regions, and 0.5% of the amino acids in the disallowed regions. The HV-3 vaccine had 74.7% of the amino acids in the most favored regions, 23.7% of the amino acids in the additional allowed regions, 1.1% of the amino acids in the generously allowed regions, and 0.5% of the amino acids in

the disallowed regions. Moreover, HV-1, HV-2, and HV-3 had the z-scores of -7.59, -5.91, and -5.12, which represented that all of them had scores well within the range of experimentally proven X-ray crystal structures of proteins from the Protein Data Bank (Supplementary Figure S6). In the tertiary structure refinement and validation study, all the three vaccine proteins were found to possess quite good quality protein structures

3.12. Vaccine protein disulfide engineering

In protein disulfide engineering, the DbD2 server identifies the pairs of amino acids that have the capability to form disulfide bonds based on the given selection criteria. In this experiment, we selected only those amino acid pairs that had bond energy value was less than 2.2 kcal/mol. The HV-1 generated 4 pairs of amino acids with bond energy less than 2.2 kcal/mol: 19 Arg and 148 Gly, 21 Gly and 47 Arg, 177 Val and 162 Gly, 233 Pro and 251 Ser. HV-2 generated 6 pairs with the selected bond energy: 60 Glu and 67 Glu, 68 Phe and 118 Ala, 174 Ser and 217 Ser, 275 Gly and 319 Lys, 314 Ser and 325 Pro, 360 Gly and 369 Met. And HV-3 generated only 3 pairs of the amino acids: 191 Val and 248 Gly, 253 Ser and 290 Ser, 255 Asn and 282 Trp. The selected amino acid pairs formed the mutant version of the original vaccines (with disulfide bonds) in the DbD2 server (Supplementary Figure S7). Since HV-2 was predicted to have 6 possible pairs of amino acid residues (the highest among the three vaccines) with the capability to form potential disulfide bonds, therefore, it can be declared that HV-2 might be most stable among the three vaccine constructs, but the other two vaccine proteins were also predicted to be quite stable.

3.13. Protein-protein docking study

The protein-protein docking of the three vaccine constructs with several TLRs was performed using three online tools because the protein-protein docking was used to predict the best vaccine construct among the three designed vaccines. The docking step is one of the necessary steps in the vaccine designing experiment because it determines the best vaccine construct and also the possibility of the interactions of the constructed vaccines with different TLRs, which might occur during an original immune response. From analyzing the protein-protein docking, it can be declared that HV-1 was the best vaccine construct. It generated the best and lowest

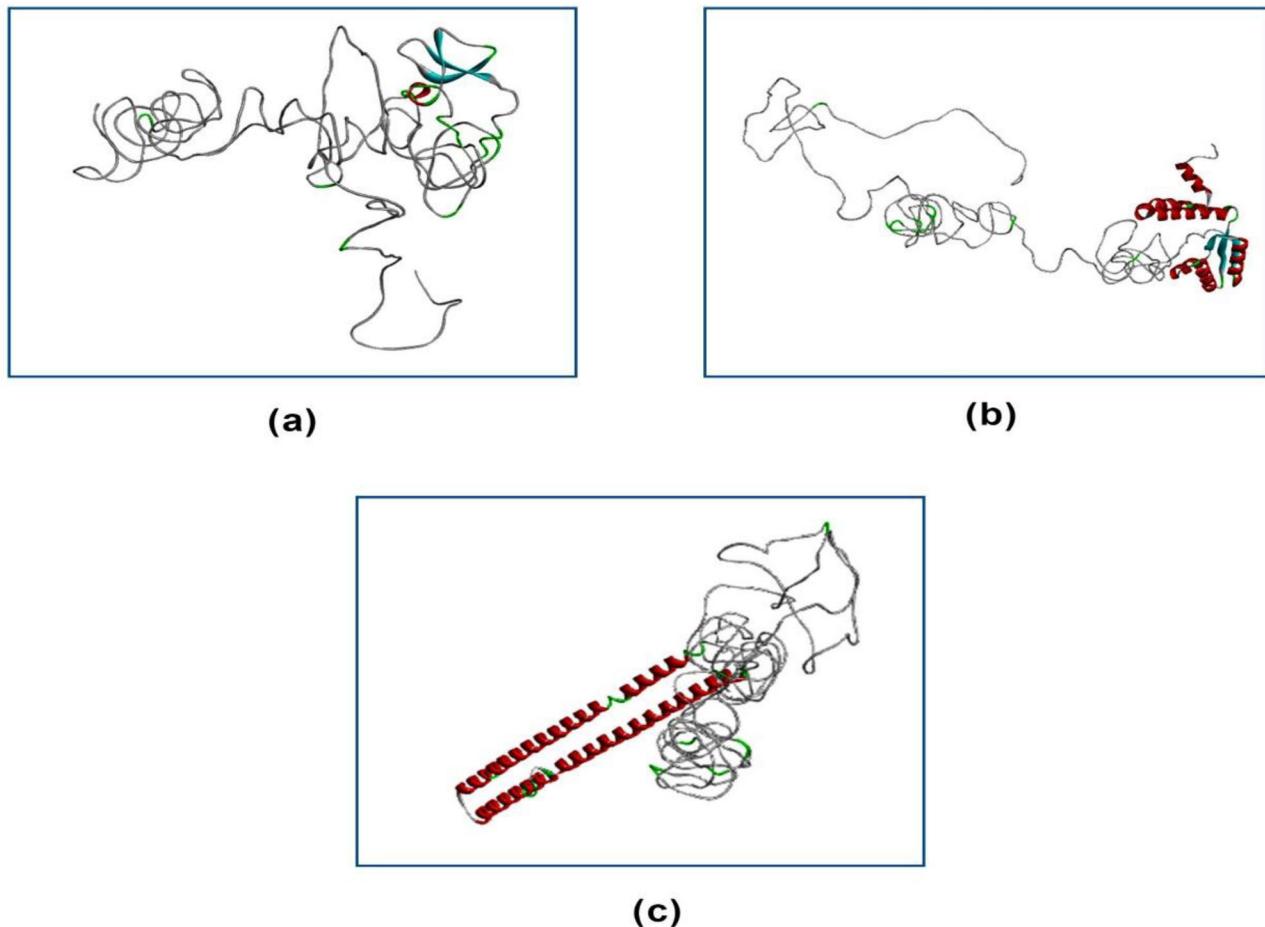


Figure 3. The tertiary structures of the three HSV vaccines. Here, (a) is the HV-1, (b) is the HV-2, and (c) is the HV-3.

Table 5. Results of the docking study of all the three vaccine constructs.

Name of the vaccines	Name of the Targets	PDB IDs of the targets	Binding affinity, ΔG (kcal mol $^{-1}$)	Global energy	HawkDock score (the lowest score)	MM-GBSA (binding free energy, in kcal mol $^{-1}$)
HV-1	TLR-1	6NIH	-17.8	-25.95	-5736.14	-64.19
	TLR-2	3A7C	-29.5	-18.12	-6147.60	-99.15
	TLR-3	2A0Z	-19.3	-8.71	-6450.66	-136.25
	TLR-4	4G8A	-19.2	-2.71	-5975.92	-64.87
	TLR-8	3W3M	-1.8	-32.33	-5086.44	-54.85
	TLR-9	4QDH	-10.6	-13.32	-5850.40	-75.96
	TLR-1	6NIH	-17.0	-7.43	-3248.84	-38.51
	TLR-2	3A7C	-19.8	-16.52	-3766.88	-32.33
	TLR-3	2A0Z	-11.4	-1.28	-3804.33	-16.05
HV-2	TLR-4	4G8A	-19.0	-7.28	-3795.29	-40.44
	TLR-8	3W3M	-14.3	-5.68	-3884.87	-33.34
	TLR-9	4QDH	-1.5	-7.39	-3538.55	-50.96
	TLR-1	6NIH	-16.9	-19.01	-3022.91	-23.72
	TLR-2	3A7C	-19.8	-5.56	-3146.50	-10.41
	TLR-3	2A0Z	-19.3	-17.60	-2880.47	-12.88
	TLR-4	4G8A	-22.6	-11.16	2946.17	-6.85
	TLR-8	3W3M	-21.2	-18.40	-2989.85	-13.90
	TLR-9	4QDH	-7.4	-6.38	-2601.94	-19.75
Remarks	-	-	Best vaccine construct: HV-1 (considering the docking scores with the targets except TLR-4 and TLR-8)	Best vaccine construct: HV-1 (considering the docking scores with the targets except TLR-4)	Best vaccine construct: HV-1 (considering the docking scores with all the targets)	Best vaccine construct: HV-1 (considering the docking scores with all the targets)

scores in the MM-GBSA study and HawkDock study. HV-1 also generated best results in most of the aspects when the docking was carried out by the PatchDock and ClusPro 2.0 servers. And most importantly, HV-1 showed very good and satisfactory results when docked with the TLRs by all

the servers. Since HV-1 showed superior results in the protein-protein docking study, it was considered as the best vaccine construct (Table 5). The interaction between the TLR-3 and the best selected vaccine HV-1 is depicted in Figure 4.

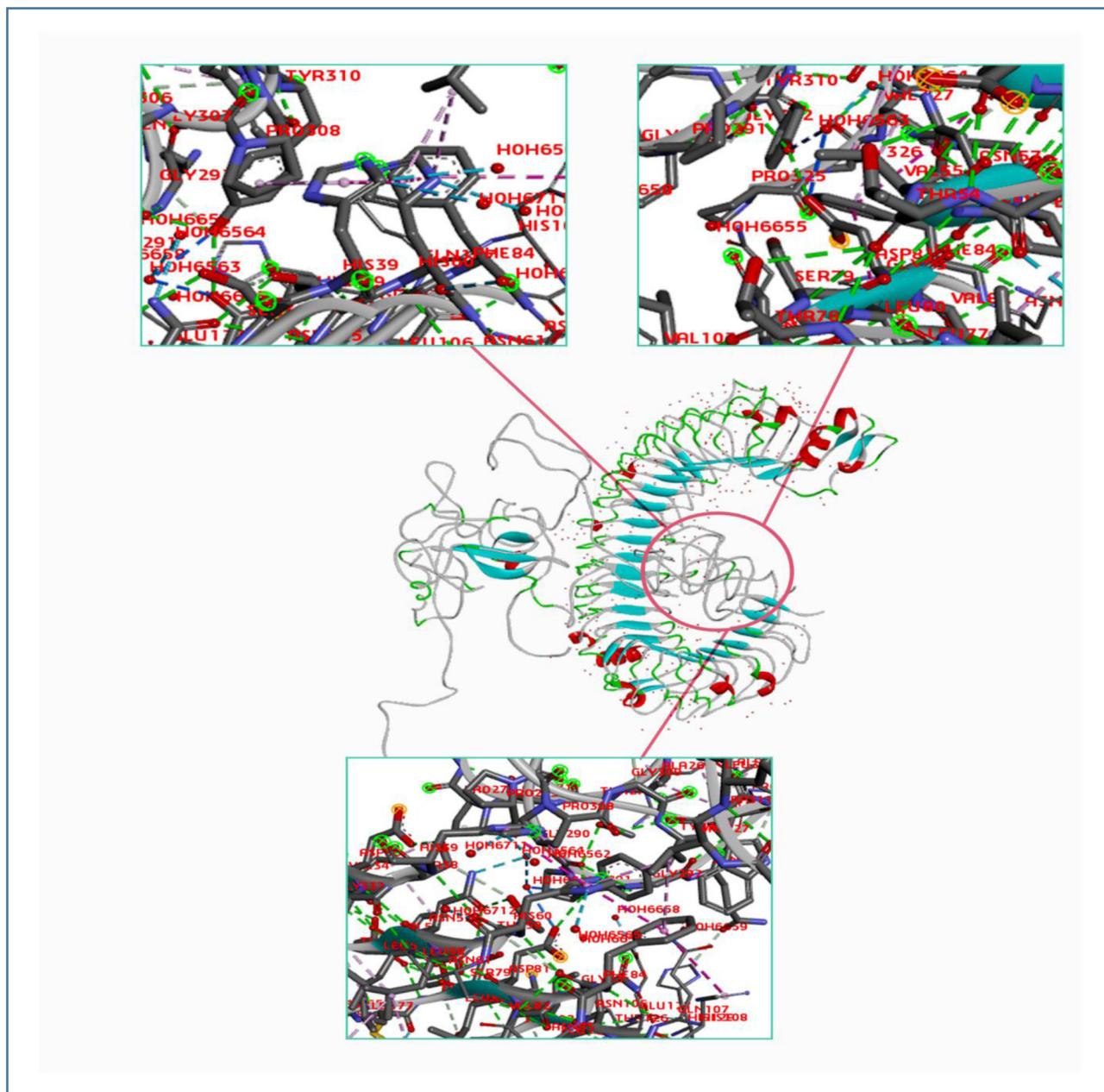


Figure 4. The interaction between TLR-3 (receptor protein on right) and the HV-1 vaccine (ligand protein on left). Here the interacted amino acids are: His 60 (receptor) - Val 327 (ligand), Phe 84 (receptor) - Val 327 (ligand), His 60 (receptor) - Pro 307 (ligand), His 39 (receptor) - Pro 308 (ligand), Phe 84 (receptor) – Pro 308 (ligand), Val 34 (receptor) - Pro 272 (ligand), Asp 81 (receptor) - Tyr 310 (ligand), His 129 (receptor) - Gln 293 (ligand).

3.14. Conformational B-cell epitope prediction

After the docking study, the conformational B-cell epitopes of the HV-1 vaccine construct were predicted. The prediction showed four potential regions of the vaccine with quite satisfactory scores from 0.555 to 0.823 and covered a total of 274 amino acids (Supplementary Table S11 and Supplementary Figure S8).

3.15. Molecular dynamics simulation

MD simulation was conducted on the HV-1-TLR3 docked complex. The mass of one chain was found to be 76278.606 amu as well as a charge of -7.0 e was also observed. The other chain had a mass of 52655.029 amu and a charge of 44.00 e. A total of 157763 water molecules had been added to the system

after solvation from which 37 were replaced by CL ions during ionization to neutralize the system's charge.

Energy minimization was completed in 2028 steps when the steepest descent had converged and the force had reached <1000 KJ/mol. The average potential energy had been calculated to be $-4.03091e + 06$ KJ/mol. From the temperature equilibration plot of Supplementary Figure S9(a) we can see that the target minimization value of 300 K remained stable over the remainder of the equilibration and fluctuates only by ± 1 K. The pressure value had been calculated after the NPT equilibration showed fluctuations around 0 bar with a range of ± 150 bar. The average pressure had been found to be -4.34269 bar. Similarly, density had also been calculated for 100 ps, where the average density was measured to be 1011.56 kgm^3 . Both graphs are shown in Supplementary

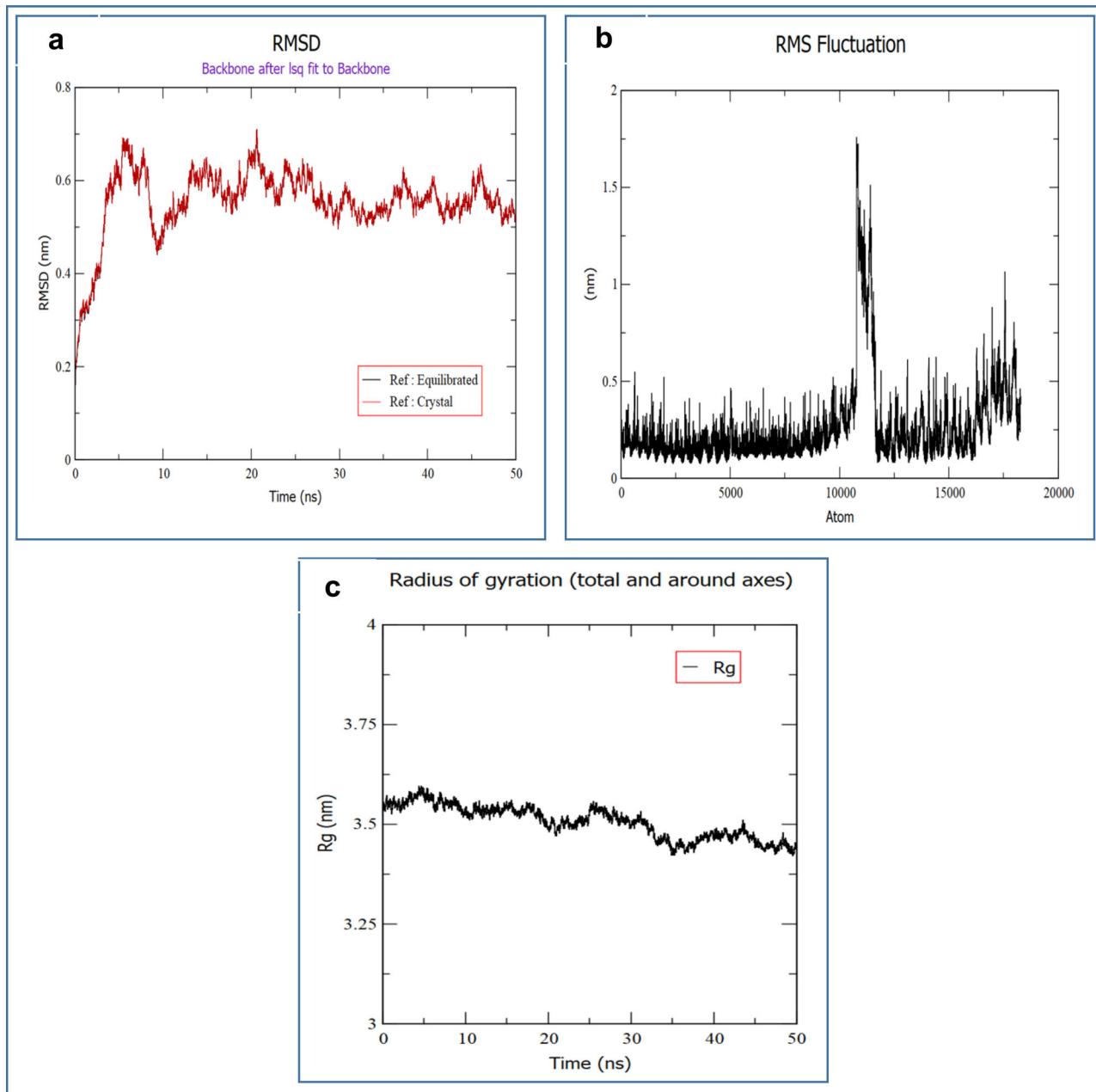


Figure 5. (a) RMSD plot of backbone. RMSD graph shows the structure had maintained a stable 0.5 nm deviation with minimum fluctuations. (b) RMS Fluctuations of all the atoms about its average position. The peaks and dips in the graph denote the flexibility of the corresponding region in the molecular structure. (c) Radius of gyration of the HV-1-TLR-3 complex.

Figure S9(b, c) with a 10 ps running average plotted. The density values remained mostly stable over time, indicating that the system was well equilibrated.

Trajectory analysis was carried out after completion of the 50 ns simulation. After performing trajectory conversion to account for periodicity in the system RMSD had been calculated. The plot of RMSD backbone revealed that RMSD levels had gone up to ~0.5 nm and maintained during the course of the simulation Figure 5(a). The black line refers to the RMSD relative to the structure present in the minimized, equilibrated system. The red line is the RMSD relative to the crystal structure. Both the plots are almost identical with little to no difference. These plots indicated that the structure remained quite stable during the simulation. The RMSF and the radius of gyration were also calculated and they showed

quite sound and satisfactory results with good stability of the protein complex Figure 5(b, c). A low average potential energy of $-4.03091e+06$ KJ/mol, an average temperature fluctuation of only ± 1 K from the target 300 K, and an average RMSD of ~ 0.5 nm (after simulating for 50 ns) were observed in the MD simulation of the protein complex which pointed towards the fact that the energy minimized HV-1-TLR-3 docked structure was quite stable throughout the experiment.

3.16. Immune simulation

The immune simulation of the best selected vaccine, HV-1 was performed by the C-ImmSimm server which predicts the

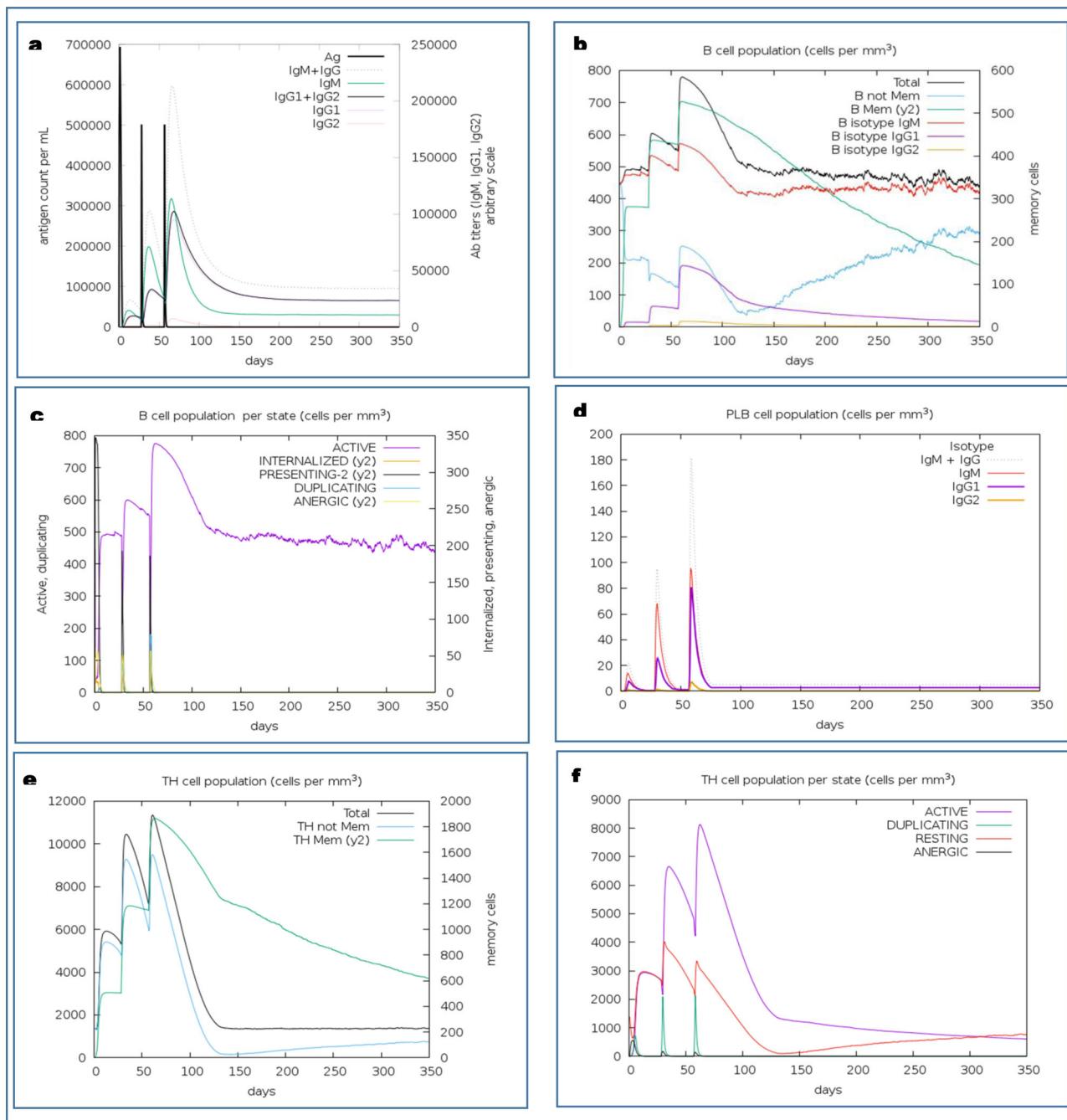
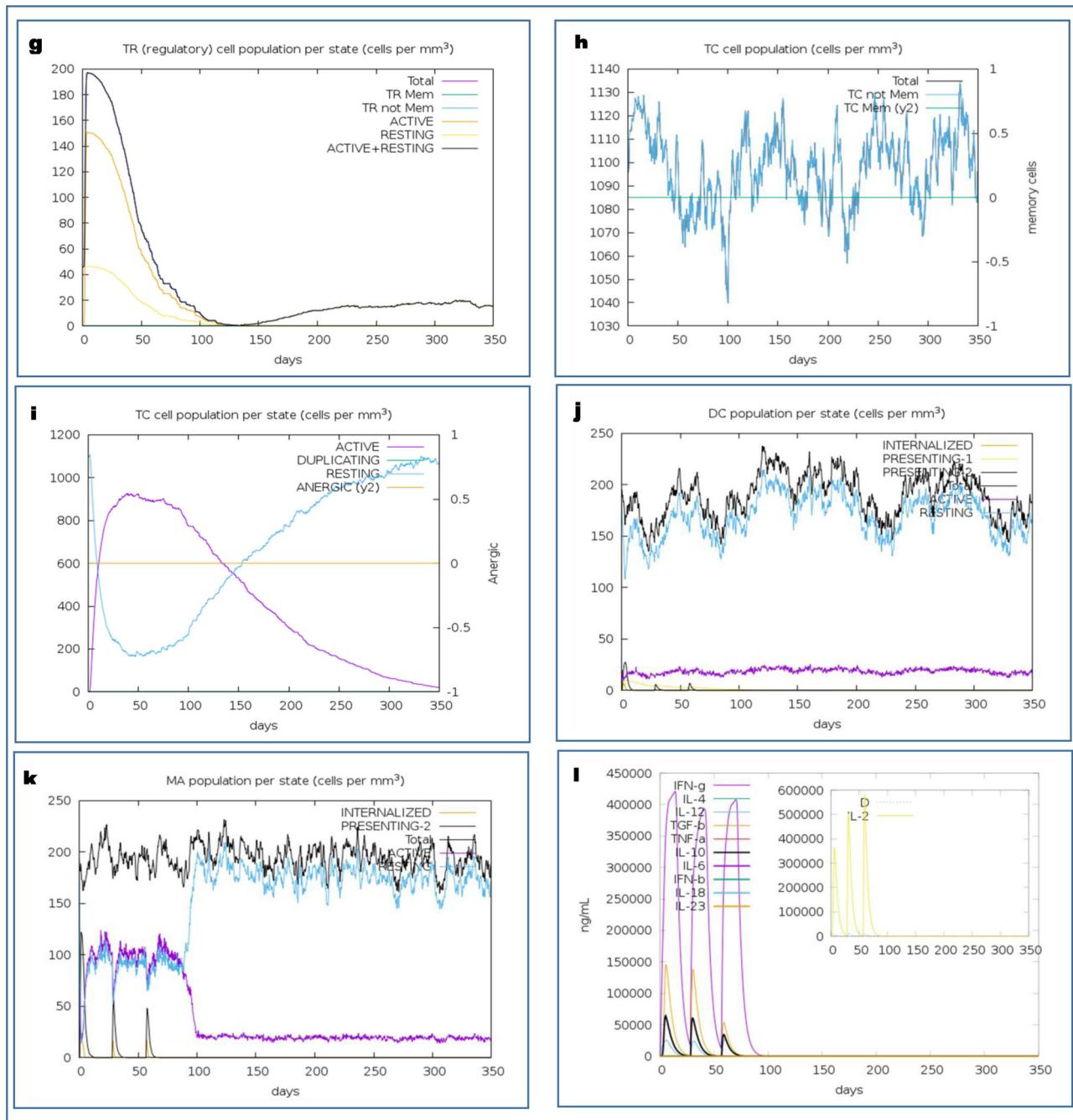


Figure 6. C-IMMSIMM representation of the immune simulation of the best predicted vaccine, HV-1. **a.** The immunoglobulin and immunocomplex response to the HV-1 vaccine inoculations (black colored lines) and specific subclasses are induced by colored lines, **b.** Increase in the B-cell population over the course of the three injections, **c.** Augmentation in the B-cell population per state over the course of vaccination, **d.** Elevation in the plasma B-cell population over the course of the injections, **e.** Increase in the helper T-cell population over the course of the three injections, **f.** Elevation of the helper T-cell population per state over the course of the vaccination, **g.** Rise in the regulatory T lymphocyte over the course of the three injections, **h.** Elevation in the cytotoxic T lymphocyte population over the course of the injections, **i.** Increase in the active cytotoxic T lymphocyte population per state over the course of the three injections, **j.** Rise in the active dendritic cell population per state over the course of the three injections, **k.** Increase in the macrophage population per state over the course of the injections, **l.** Augmentation in the concentrations of different types of cytokines over the course of the three injections.

generation of adaptive immunity as well as the immune interactions of epitopes with their specific targets (Luckheeram et al., 2012). The immune simulation study predicted that the HV-1 vaccine might produce an immune response which is consistent with the typical and natural immune system. The study revealed that after each of the three injections, the primary immune response against the vaccine was predicted to stimulate significantly as indicated

by the gradual increase in the levels of different immunoglobulins (Figure 6(a)). Along with the primary immune response, the secondary immune responses were also found to be augmented. The gradual increase in the concentrations of active B-cell (Figure 6(b, c)), plasma B-cell (Figure 6(d)), helper T-cell (Figure 6(e, f)), regulatory T-cell, and cytotoxic T-cell (Figure 6(g-i)) were also predicted, which pointed towards a very strong secondary immune response, very

**Figure 6.** Continued.

good immune memory generation and the increased clearance of the antigen. The stimulation of helper T-cells indicated good adaptive immunity provided by the vaccine (Almofti et al., 2018; Carvalho et al., 2002). Furthermore, the augmentation in the concentration of dendritic cells and macrophages indicated a very good antigen presentation by these antigen presenting cells (APCs) (Figure 6(j, k)). The vaccine was also predicted to produce good amount of different types of cytokines like the IFN-gamma, IL-23, IL-10, and IFN-beta which are some of the most significant cytokines for generating immune response against viruses (Hoque et al., 2019; Kambayashi & Laufer, 2014; Shey et al., 2019) (Figure 6(l)). The negligible Simpson index (D) suggested a diverse immune response of the vaccine HV-1 (Kaminski et al., 2001).

Overall, the immune simulation study had revealed that with the predicted ability of generating high amount of immunoglobulins, cytokines, APCs, active B-cells, and T-cells, the polyvalent vaccine HV-1 might be able to provoke good immunogenic responses after administration.

3.17. Codon adaptation, *in silico* cloning, and analysis of the vaccine mRNA structure

Finally the codon adaptation and *in silico* cloning were performed to construct a recombinant plasmid which could be used for the mass production of the HV-1 vaccine in the *E. coli* strain K12. *E. coli* cell culture system is the recommended

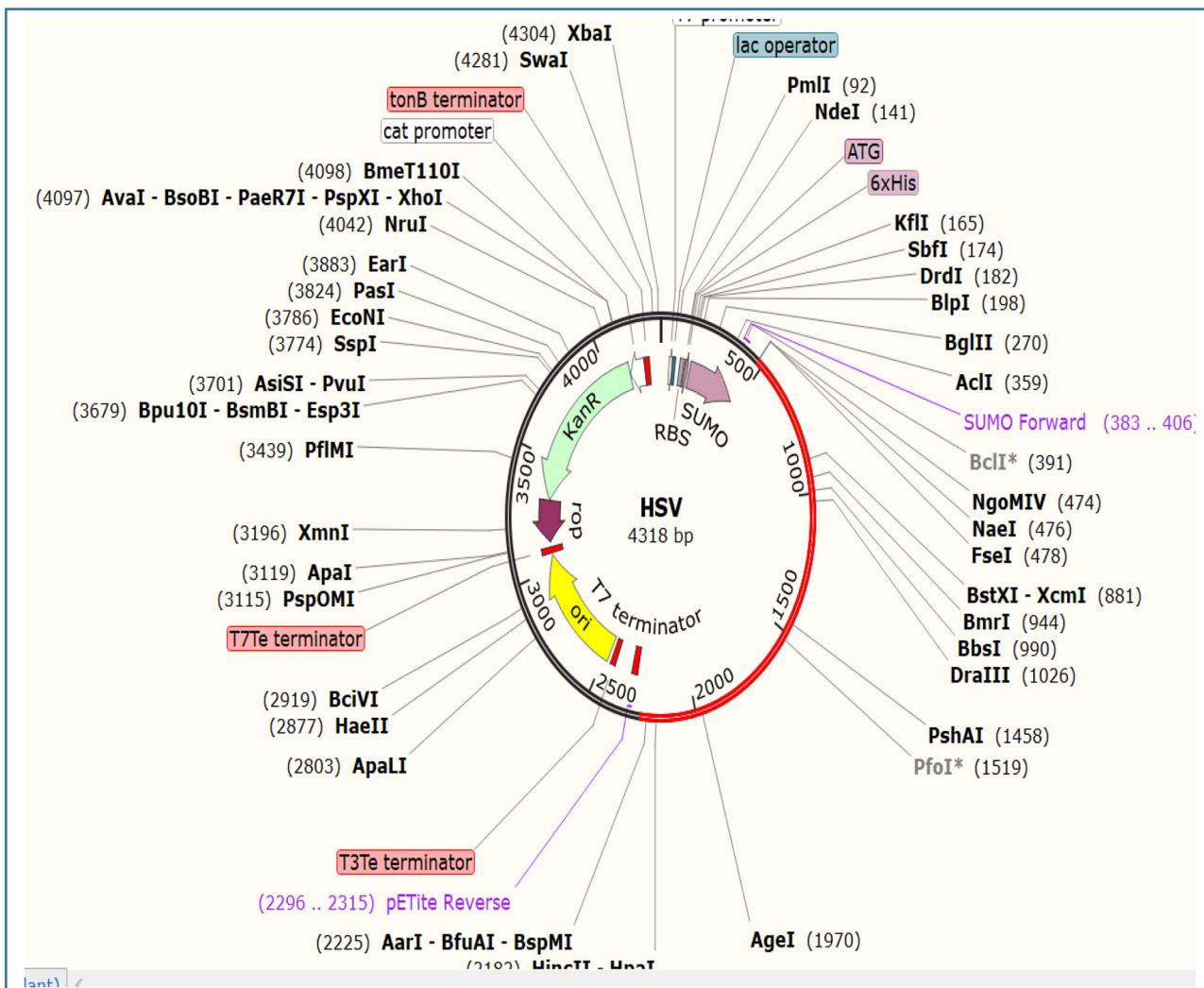


Figure 7. In silico restriction cloning of the HV-1 vaccine sequence in the pETite plasmid between the EaeI and StyI restriction enzyme sites. The red colored marked site contains the DNA insert of the vaccine.

system for the production of recombinant proteins (Grote et al., 2005). For *in silico* cloning and plasmid construction, the protein sequence of the best selected vaccine, HV-1 was adapted by the JCcat server. The codon adaptation index (CAI) value of 0.98 of HV-1 indicated that the DNA sequences contained higher proportion of the codons that are most likely to be used in the cellular machinery of the target organism, *E. coli* strain K12 (codon bias) (Supplementary Figure S10). For this reason, the production of the HV-1 vaccine should be carried out efficiently in the *E. coli* strain K12 (Mathews et al., 1999; Zuker, 2003). The GC content of the improved sequence was found to be 54.49%.

In the codon adaptation study, quite sound results were achieved with the CAI value of 0.98 and the GC content of 54.49% because any CAI value over 0.80 and the GC content from 30% to 70% are considered as good scores (Khatoon et al., 2017; Morla et al., 2016; Shey et al., 2019). After codon adaptation, the predicted DNA sequence of HV-1 was inserted into the pETite vector plasmid between the *Eae*l and *Sty*l restriction sites. Since the plasmid contains sequences of the SUMO tag and 6X His tag, so the vaccine protein was expected to be purified easily during downstream

processing. The newly constructed cloned vector plasmid was designated as "HSV" (Figure 7).

The secondary structure of the HV-1 mRNA was predicted by Mfold and RNAfold servers. The Mfold server generated minimum free energy score of -556.50 kcal/mol , which was in agreement with the prediction of RNAfold server, which predicted the minimal free energy of -533.90 kcal/mol . Since lower minimal free energy corresponds to better mRNA stability, therefore it can be declared that the predicted HV-1 vaccine might be quite stable upon expression *in vivo* (Hamasaki-Katagiri et al., 2017). Figure 8 illustrates the predicted secondary structure of the V3 vaccine.

Overall, this study suggests the HV-1 as the possible preventative measure to combat the HSV infections around the world. However, further *in vivo* and *in vitro* researches are necessary to strengthen the outcomes of this study.

4. Conclusion

With very high infection rate, Herpes Simplex Virus (HSV) is one of the most infectious and common sexually

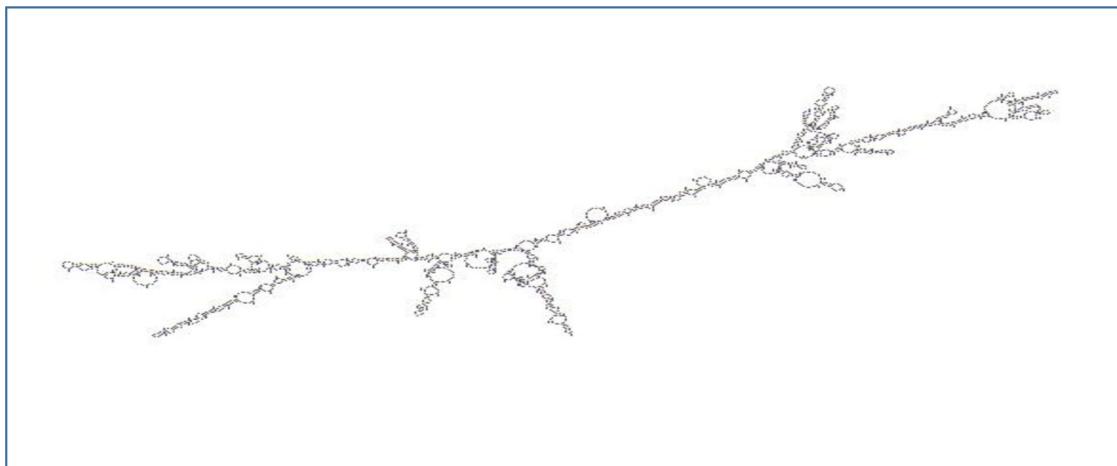


Figure 8. The mRNA secondary structure of the best vaccine HV-1 predicted by the RNAfold server.

transmitted diseases in the world. However, only the HSV-2 is transmitted sexually. The other type of HSV, HSV-1, transmits through oral-oral contact. Although researches are going on, no vaccine is yet available with satisfactory results, to control the spread of HSV infection. In this study, possible epitope-based polyvalent vaccines were designed against the four strains of HSV, HSV-1, strain-17; HSV-1, strain-F; HSV-2, strain-HG52 and HSV-2, strain-333, using the tools of bioinformatics, immunoinformatics, and vaccinomics. Since the vaccines contained multiple conserved T-cell and B-cell epitopes, therefore, they might be quite effective against all the four selected viral strains. The highly antigenic, non-allergenic, and non-toxic as well as 100% conserved and non-homolog epitopes were considered for final vaccine construction, as a result, the vaccines would be able to generate high immunogenic response and at the same time they should not cause any harmful reaction within the body. Results of different experiments performed in the study indicated that these polyvalent vaccines should be quite safe, effective, and responsive to use. However, since all these predictions were done based only on the computational methods, more wet lab-based research is needed to finally confirm the outcomes of this study. With high cost requirements and multiple limitations for developing the live, attenuated or inactivated vaccine preparation for such contagious agents, these peptide-based vaccine candidates might be relatively cheap and effective options to reach the entire world to combat the HSV infections.

Acknowledgements

Authors are thankful to Swift Integrity Computational Lab, Dhaka, Bangladesh, a virtual platform of young researchers, for providing the support to conduct this research.

Disclosure statement

The authors declare that they have no conflict of interest among them.

ORCID

Bishajit Sarkar <http://orcid.org/0000-0001-8439-6994>

Md. Asad Ullah <http://orcid.org/0000-0002-6615-6433>

Yusha Araf <http://orcid.org/0000-0002-0144-5875>

References

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>
- Adhikari, U. K., & Rahman, M. M. (2017). Overlapping CD8+ and CD4+ T-cell epitopes identification for the progression of epitope-based peptide vaccine from nucleocapsid and glycoprotein of emerging Rift Valley fever virus using immunoinformatics approach. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 56, 75–91. <https://doi.org/10.1016/j.meegid.2017.10.022>
- Almofti, Y. A., Abd-Elrahman, K. A., Gassmallah, S. A. E., & Salih, M. A. (2018). Multi epitopes vaccine prediction against Severe Acute Respiratory Syndrome (SARS) Coronavirus using immunoinformatics approaches. *American Journal of Microbiological Research*, 6(3), 94–114. <https://doi.org/10.12691/ajmr-6-3-5>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Andrusier, N., Nussinov, R., & Wolfson, H. J. (2007). FireDock: Fast interaction refinement in molecular docking. *Proteins*, 69(1), 139–159. <https://doi.org/10.1002/prot.21495>
- 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 Engineering*, 14(8), 529–532. <https://doi.org/10.1093/protein/14.8.529>
- Bacchetta, R., Gregori, S., & Roncarolo, M. G. (2005). CD4+ regulatory T cells: Mechanisms of induction and effector function. *Autoimmunity Reviews*, 4(8), 491–496. <https://doi.org/10.1016/j.autrev.2005.04.005>
- Balfour, H. H. (1983). Resistance of herpes simplex to acyclovir. *Annals of Internal Medicine*, 98(3), 404–406. <https://doi.org/10.7326/0003-4819-98-3-404>
- Biovia, D. S. (2017). Discovery studio modeling environment.
- Biswal, J. K., Bisht, P., Mohapatra, J. K., Ranjan, R., Sanyal, A., & Pattnaik, B. (2015). Application of a recombinant capsid polyprotein (P1) expressed in a prokaryotic system to detect antibodies against foot-and-mouth disease virus serotype O. *Journal of Virological Methods*, 215–216, 45–51. <https://doi.org/10.1016/j.jviromet.2015.02.008>

- Buchan, D. W., & Jones, D. T. (2019). The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Research*, 47(W1), W402–W407. <https://doi.org/10.1093/nar/gkz297>
- 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(1), 361. <https://doi.org/10.1186/1471-2105-8-361>
- Cano, R. L. E., & Lopera, H. D. E. (2013). Introduction to T and B lymphocytes. In *Autoimmunity: From bench to bedside*. El Rosario University Press.
- Carbone, A., Zinovyev, A., & Képes, F. (2003). Codon adaptation index as a measure of dominating codon bias. *Bioinformatics (Oxford, England)*, 19(16), 2005–2015. <https://doi.org/10.1093/bioinformatics/btg272>
- Carvalho, L. H., Sano, G. I., Hafalla, J. C., Morrot, A., De Lafaille, M. A. C., & Zavala, F. (2002). IL-4-secreting CD4+ T cells are crucial to the development of CD8+ T-cell responses against malaria liver stages. *Nature Medicine*, 8(2), 166–170. <https://doi.org/10.1038/nm0202-166>
- Castiglione, F., Mantile, F., De Berardinis, P., & Prisco, A. (2012). How the interval between prime and boost injection affects the immune response in a computational model of the immune system. *Computational and Mathematical Methods in Medicine*, 2012, 1–9. <https://doi.org/10.1155/2012/842329>
- Chang, K. Y., & Yang, J. R. (2013). Analysis and prediction of highly effective antiviral peptides based on random forests. *PloS One*, 8(8), e70166. <https://doi.org/10.1371/journal.pone.0070166>
- Chaudhri, G., Quah, B. J., Wang, Y., Tan, A. H., Zhou, J., Karupiah, G., & Parish, C. R. (2009). T cell receptor sharing by cytotoxic T lymphocytes facilitates efficient virus control. *Proceedings of the National Academy of Sciences of the United States of America*, 106(35), 14984–14989. <https://doi.org/10.1073/pnas.0906554106>
- Chong, L. C., & Khan, A. M. (2019). Vaccine target discovery. In *Encyclopedia of bioinformatics and computational biology*. Elsevier. <https://doi.org/10.1016/b978-0-12-809633-8.20100-3>
- Cooper, N. R., & Nemerow, G. R. (1984). The role of antibody and complement in the control of viral infections. *Journal of Investigative Dermatology*, 83(1), S121–S127. <https://doi.org/10.1038/jid.1984.33>
- Craig, D. B., & Dombkowski, A. A. (2013). Disulfide by Design 2.0: A web-based tool for disulfide engineering in proteins. *BMC Bioinformatics*, 14(1), 346. <https://doi.org/10.1186/1471-2105-14-346>
- Dhanda, S. K., Gupta, S., Vir, P., & Raghava, G. P. (2013). Prediction of IL4 inducing peptides. *Clinical and Developmental Immunology*, 2013, 1–9. <https://doi.org/10.1155/2013/263952>
- Dhanda, S. K., Vir, P., & Raghava, G. P. (2013). Designing of interferon-gamma inducing MHC class-II binders. *Biology Direct*, 8(1), 30. <https://doi.org/10.1186/1745-6150-8-30>
- Dimitrov, I., Flower, D. R., & Doytchinova, I. (2013). AllerTOP-a server for in silico prediction of allergens. *BMC Bioinformatics*, 14(Suppl 6), S4. <https://doi.org/10.1186/1471-2105-14-S6-S4>
- Dimitrov, I., Naneva, L., Doytchinova, I., & Bangov, I. (2014). AllergenFP: Allergenicity prediction by descriptor fingerprints. *Bioinformatics (Oxford, England)*, 30(6), 846–851. <https://doi.org/10.1093/bioinformatics/btt619>
- Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, 8(1), 4. <https://doi.org/10.1186/1471-2105-8-4>
- Farnsworth, A., Wisner, T. W., Webb, M., Roller, R., Cohen, G., Eisenberg, R., & Johnson, D. C. (2007). Herpes simplex virus glycoproteins gB and gH function in fusion between the virion envelope and the outer nuclear membrane. *Proceedings of the National Academy of Sciences of the United States of America*, 104(24), 10187–10192. <https://doi.org/10.1073/pnas.0703790104>
- Feng, T., Chen, F., Kang, Y., Sun, H., Liu, H., Li, D., Zhu, F., & Hou, T. (2017). HawkRank: A new scoring function for protein-protein docking based on weighted energy terms. *Journal of Cheminformatics*, 9(1), 66. <https://doi.org/10.1186/s13321-017-0254-7>
- Garcia, K. C., Teyton, L., & Wilson, I. A. (1999). Structural basis of T cell recognition. *Annual Review of Immunology*, 17(1), 369–397. <https://doi.org/10.1146/annurev.immunol.17.1.369>
- Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook* (pp. 571–607). Humana press. <https://doi.org/10.1385/1-59259-890-0:571>
- Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D. C., & Jahn, D. (2005). JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Research*, 33(Web Server issue), W526–W531. <https://doi.org/10.1093/nar/gki376>
- Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., & Hofacker, I. L. (2008). The Vienna RNA website. *Nucleic Acids Research*, 36(Web Server issue), W70–W74. <https://doi.org/10.1093/nar/gkn188>
- Gu, Y., Sun, X., Li, B., Huang, J., Zhan, B., & Zhu, X. (2017). Vaccination with a paramyosin-based multi-epitope vaccine elicits significant protective immunity against *Trichinella spiralis* infection in mice. *Frontiers in Microbiology*, 8, 1475. <https://doi.org/10.3389/fmicb.2017.01475>
- Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., & Raghava, G. P. S. (2013). In silico approach for predicting toxicity of peptides and proteins. *PLoS One*, 8(9), e73957. <https://doi.org/10.1371/journal.pone.0073957>
- Guruprasad, K., Reddy, B. B., & Pandit, M. W. (1990). Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Engineering*, 4(2), 155–161. <https://doi.org/10.1093/protein/4.2.155>
- Hamasaki-Katagiri, N., Lin, B. C., Simon, J., Hunt, R. C., Schiller, T., Russek-Cohen, E., Komar, A. A., Bar, H., & Kimchi-Sarfaty, C. (2017). The importance of mRNA structure in determining the pathogenicity of synonymous and non-synonymous mutations in haemophilia. *Haemophilia: The Official Journal of the World Federation of Hemophilia*, 23(1), e8–e17. <https://doi.org/10.1111/hae.13107>
- Han, J., Chadha, P., Starkey, J. L., & Wills, J. W. (2012). Function of glycoprotein E of herpes simplex virus requires coordinated assembly of three tegument proteins on its cytoplasmic tail. *Proceedings of the National Academy of Sciences of the United States of America*, 109(48), 19798–19803. <https://doi.org/10.1073/pnas.1212900109>
- Hebditch, M., Carballo-Amador, M. A., Charonis, S., Curtis, R., & Warwicker, J. (2017). Protein-Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics (Oxford, England)*, 33(19), 3098–3100. <https://doi.org/10.1093/bioinformatics/btx345>
- Hoque, M. N., Istiaq, A., Clement, R. A., Sultana, M., Crandall, K. A., Siddiki, A. Z., & Hossain, M. A. (2019). Metagenomic deep sequencing reveals association of microbiome signature with functional biases in bovine mastitis. *Scientific Reports*, 9(1), 1–14. <https://doi.org/10.1038/s41598-019-49468-4>
- Hou, T., Wang, J., Li, Y., & Wang, W. (2011). Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *Journal of Chemical Information and Modeling*, 51(1), 69–82. <https://doi.org/10.1021/ci100275a>
- Ikai, A. (1980). Thermostability and aliphatic index of globular proteins. *The Journal of Biochemistry*, 88(6), 1895–1898. <https://doi.org/10.1093/oxfordjournals.jbchem.133168>
- Itzhaki, R. F., Lin, W. R., Shang, D., Wilcock, G. K., Faragher, B., & Jamieson, G. A. (1997). Herpes simplex virus type 1 in brain and risk of Alzheimer's disease. *The Lancet*, 349(9047), 241–244. [https://doi.org/10.1016/S0140-6736\(96\)10149-5](https://doi.org/10.1016/S0140-6736(96)10149-5)
- Johnson, D. C., Webb, M., Wisner, T. W., & Brunetti, C. (2001). Herpes simplex virus gE/gI sorts nascent virions to epithelial cell junctions, promoting virus spread. *Journal of Virology*, 75(2), 821–833. <https://doi.org/10.1128/JVI.75.2.821-833.2001>
- Johnston, C., Gottlieb, S. L., & Wald, A. (2016). Status of vaccine research and development of vaccines for herpes simplex virus. *Vaccine*, 34(26), 2948–2952. <https://doi.org/10.1016/j.vaccine.2015.12.076>
- Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*, 292(2), 195–202. <https://doi.org/10.1006/jmbi.1999.3091>
- Kambayashi, T., & Laufer, T. M. (2014). Atypical MHC class II-expressing antigen-presenting cells: Can anything replace a dendritic cell? *Nature Reviews. Immunology*, 14(11), 719–730. <https://doi.org/10.1038/nri3754>

- Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., & Jorgensen, W. L. (2001). Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B*, 105(28), 6474–6487. <https://doi.org/10.1021/jp003919d>
- Khatoon, N., Pandey, R. K., & Prajapati, V. K. (2017). Exploring Leishmania secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. *Scientific Reports*, 7(1), 1–12. <https://doi.org/10.1038/s41598-017-08842-w>
- Ko, J., Park, H., Heo, L., & Seok, C. (2012). GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Research*, 40(Web Server issue), W294–W297. <https://doi.org/10.1093/nar/gks493>
- Koelle, D. M., & Corey, L. (2008). Herpes simplex: Insights on pathogenesis and possible vaccines. *Annual Review of Medicine*, 59, 381–395. <https://doi.org/10.1146/annurev.med.59.061606.095540>
- Kozakov, D., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., Hall, D. R., & Vajda, S. (2013). How good is automated protein docking? *Proteins*, 81(12), 2159–2166. <https://doi.org/10.1002/prot.24403>
- Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., & Vajda, S. (2017). The ClusPro web server for protein–protein docking. *Nature Protocols*, 12(2), 255–278. <https://doi.org/10.1038/nprot.2016.169>
- Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology*, 305(3), 567–580. <https://doi.org/10.1006/jmbi.2000.4315>
- Kumar, H., Frischknecht, F., Mair, G. R., & Gomes, J. (2015). In silico identification of genetically attenuated vaccine candidate genes for Plasmodium liver stage. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 36, 72–81. PMID: 26348884, <https://doi.org/10.1016/j.meegid.2015.09.002>
- Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydrophobic character of a protein. *Journal of Molecular Biology*, 157(1), 105–132. [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0)
- Laskowski, R. A., MacArthur, M. W., & Thornton, J. M. (2006). PROCHECK: Validation of protein-structure coordinates. In *International tables for crystallography*. Wiley. <https://doi.org/10.1107/97809553602060000882>
- Lee, S., & Nguyen, M. T. (2015). Recent advances of vaccine adjuvants for infectious diseases. *Immune Network*, 15(2), 51–57. <https://doi.org/10.4110/in.2015.15.2.51>
- Looker, K. J., Welton, N. J., Sabin, K. M., Dalal, S., Vickerman, P., Turner, K. M., Boily, M. C., & Gottlieb, S. L. (2020). Global and regional estimates of the contribution of herpes simplex virus type 2 infection to HIV incidence: A population attributable fraction analysis using published epidemiological data. *The Lancet. Infectious Diseases*, 20(2), 240–249. [https://doi.org/10.1016/S1473-3099\(19\)30470-0](https://doi.org/10.1016/S1473-3099(19)30470-0)
- Luckheeram, R. V., Zhou, R., Verma, A. D., & Xia, B. (2012). 2012. CD4+ T cells: Differentiation and functions. *Clinical and Developmental Immunology*, 2012, 1–12. <https://doi.org/10.1155/2012/925135>
- Magnan, C. N., Randall, A., & Baldi, P. (2009). SOLpro: Accurate sequence-based prediction of protein solubility. *Bioinformatics (Oxford, England)*, 25(17), 2200–2207. <https://doi.org/10.1093/bioinformatics/btp386>
- Mathews, D. H., Sabina, J., Zuker, M., & Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology*, 288(5), 911–940. <https://doi.org/10.1006/jmbi.1999.2700>
- Mathews, D. H., Turner, D. H., & Zuker, M. (2007). RNA secondary structure prediction. *Current Protocols in Nucleic Acid Chemistry*, 28(1), 11–12. <https://doi.org/10.1002/0471142700.nc1102s28>
- Mehla, K., & Ramana, J. (2016). Identification of epitope-based peptide vaccine candidates against enterotoxigenic *Escherichia coli*: A comparative genomics and immunoinformatics approach. *Molecular Biosystems*, 12(3), 890–901. <https://doi.org/10.1039/c5mb00745c>
- Mettenleiter, T. C., Klupp, B. G., & Granzow, H. (2006). Herpesvirus assembly: A tale of two membranes. *Current Opinion in Microbiology*, 9(4), 423–429. <https://doi.org/10.1016/j.mib.2006.06.013>
- Meza, B., Ascencio, F., Sierra-Beltrán, A. P., Torres, J., & Angulo, C. (2017). A novel design of a multi-antigenic, multistage and multi-epitope vaccine against *Helicobacter pylori*: An in silico approach. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 49, 309–317. <https://doi.org/10.1016/j.meegid.2017.02.007>
- Morla, S., Makhija, A., & Kumar, S. (2016). Synonymous codon usage pattern in glycoprotein gene of rabies virus. *Gene*, 584(1), 1–6. <https://doi.org/10.1016/j.gene.2016.02.047>
- Morris, A. L., MacArthur, M. W., Hutchinson, E. G., & Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. *Proteins*, 12(4), 345–364. <https://doi.org/10.1002/prot.340120407>
- Nagpal, G., Usmani, S. S., Dhanda, S. K., Kaur, H., Singh, S., Sharma, M., & Raghava, G. P. (2017). Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Scientific Reports*, 7, 42851. <https://doi.org/10.1038/srep42851>
- Nugent, T., Cozzetto, D., & Jones, D. T. (2014). Evaluation of predictions in the CASP10 model refinement category. *Proteins: Structure, Function, and Bioinformatics*, 82, 98–111. <https://doi.org/10.1002/prot.24377>
- Panda, S., & Chandra, G. (2012). Physicochemical characterization and functional analysis of some snake venom toxin proteins and related non-toxin proteins of other chordates. *Bioinformation*, 8(18), 891–896. <https://doi.org/10.6026/97320630008891>
- Ponomarenko, J., Bui, H. H., Li, W., Fusseder, N., Bourne, P. E., Sette, A., & Peters, B. (2008). ElliPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*, 9(1), 514. <https://doi.org/10.1186/1471-2105-9-514>
- QtGraceTcpServer and Client-code. (2014–2017). Vadim Engelson and Nimalendiran Kailasanathan. Wolfram MathCore AB.
- Rapin, N., Lund, O., Bernaschi, M., & Castiglione, F. (2010). Computational immunology meets bioinformatics: The use of prediction tools for molecular binding in the simulation of the immune system. *PLoS One*, 5(4), e9862. <https://doi.org/10.1371/journal.pone.0009862>
- Rappuoli, R. (2000). Reverse vaccinology. *Current Opinion in Microbiology*, 3(5), 445–450. [https://doi.org/10.1016/S1369-5274\(00\)00119-3](https://doi.org/10.1016/S1369-5274(00)00119-3)
- Riediger, C., Sauer, P., Matevossian, E., Müller, M. W., Büchler, P., & Friess, H. (2009). Herpes simplex virus sepsis and acute liver failure. *Clinical Transplantation*, 23, 37–41. <https://doi.org/10.1111/j.1399-0012.2009.01108.x>
- Ryan, K. J., & Ray, C. G. (2004). *Medical microbiology* (Vol. 4, p. 370). McGraw Hill.
- Saadi, M., Karkhah, A., & Nouri, H. R. (2017). Development of a multi-epitope peptide vaccine inducing robust T cell responses against brucellosis using immunoinformatics based approaches. *Infection, Genetics and Evolution*, 51, 227–234. <https://doi.org/10.1016/j.meegid.2017.04.009>
- Saha, S., & Raghava, G. P. (2006). AlgPred: Prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Research*, 34(Web Server issue), W202–W209. <https://doi.org/10.1093/nar/gkl343>
- Sarkar, B., Ullah, M. A., & Araf, Y. (2020). A systematic and reverse vaccinology approach to design novel subunit vaccines against dengue virus type-1 and human Papillomavirus-16. *Informatics in Medicine Unlocked*, 19, 100343. <https://doi.org/10.1016/j.imu.2020.100343>
- Sarkar, B., Ullah, M. A., Johora, F. T., Taniya, M. A., & Araf, Y. (2020). Immunoinformatics-guided designing of epitope-based subunit vaccine against the SARS Coronavirus-2 (SARS-CoV-2). *Immunobiology*, 225(3), 151955. <https://doi.org/10.1016/j.imbio.2020>
- Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005). PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Research*, 33(Web Server issue), 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 Research*, 15(3), 1281–1295. <https://doi.org/10.1093/nar/15.3.1281>
- Shey, R. A., Ghogomu, S. M., Esoh, K. K., Nebangwa, N. D., Shintouo, C. M., Nongley, N. F., Asa, B. F., Ngale, F. N., Vanhamme, L., & Souopgui, J. (2019). In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Scientific Reports*, 9(1), 1–18. <https://doi.org/10.1038/s41598-019-40833-x>
- Solanki, V., & Tiwari, V. (2018). Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric



- vaccine against *Acinetobacter baumannii*. *Scientific Reports*, 8(1), 1–19. <https://doi.org/10.1038/s41598-018-26689-7>
- Straface, G., Selmin, A., Zanardo, V., De Santis, M., Ercoli, A., & Scambia, G. (2012). Herpes simplex virus infection in pregnancy. *Infectious Diseases in Obstetrics and Gynecology*, 2012, 385697. <https://doi.org/10.1155/2012/385697>
- Thomsen, M., Lundsgaard, C., Buus, S., Lund, O., & Nielsen, M. (2013). MHCcluster, a method for functional clustering of MHC molecules. *Immunogenetics*, 65(9), 655–665. <https://doi.org/10.1007/s00251-013-0714-9>
- Treml, J., Gazdová, M., Šmejkal, K., Šudomová, M., Kubatka, P., & Hassan, S. T. (2020). Natural products-derived chemicals: Breaking barriers to novel anti-HSV drug development. *Viruses*, 12(2), 154. <https://doi.org/10.3390/v12020154>
- Turner, P. J. (2005). XMGRACE (Version 5.1.19). Center For coastal and land-margin research, Oregon graduate institute of science and technology.
- Ullah, A., Sarkar, B., & Islam, S. S. (2020). Exploiting the reverse vaccinology approach to design novel subunit vaccine against ebola virus. *Immunobiology*, 225(3), 151949. <https://doi.org/10.1016/j.imbio.2020.151949>
- Vajda, S., Yueh, C., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., Hall, D. R., & Kozakov, D. (2017). New additions to the ClusPro server motivated by CAPRI. *Proteins*, 85(3), 435–444. <https://doi.org/10.1002/prot.25219>
- Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., Wheeler, D. K., Sette, A., & Peters, B. (2018). The immune epitope database (IEDB): 2018 update. *Nucleic Acids Research*, 47(D1), D339–D343. <https://doi.org/10.1093/nar/gky1006>
- Wald, A., Langenberg, A. G., Link, K., Izu, A. E., Ashley, R., Warren, T., Tyring, S., Douglas Jr, J. M., & Corey, L. (2001). Effect of condoms on reducing the transmission of herpes simplex virus type 2 from men to women. *JAMA*, 285(24), 3100–3106. <https://doi.org/10.1001/jama.285.24.3100>
- Wang, S., Li, W., Liu, S., & Xu, J. (2016). RaptorX-Property: A web server for protein structure property prediction. *Nucleic Acids Research*, 44(W1), W430–W435. <https://doi.org/10.1093/nar/gkw306>
- Wen, D., Foley, S. F., Hronowski, X. L., Gu, S., & Meier, W. (2013). Discovery and investigation of O-xylosylation in engineered proteins containing a (GGGGS) n linker. *Analytical Chemistry*, 85(9), 4805–4812. <https://doi.org/10.1021/ac400596g>
- Weng, G., Wang, E., Wang, Z., Liu, H., Zhu, F., Li, D., & Hou, T. (2019). HawkDock: A web server to predict and analyze the protein–protein complex based on computational docking and MM/GBSA. *Nucleic Acids Research*, 47(W1), W322–W330. <https://doi.org/10.1093/nar/gkz397>
- Whitbeck, J. C., Peng, C., Lou, H., Xu, R., Willis, S. H., Ponce de Leon, M., Peng, T., Nicola, A. V., Montgomery, R. I., Warner, M. S., Soulka, A. M., Spruce, L. A., Moore, W. T., Lambris, J. D., Spear, P. G., Cohen, G. H., & Eisenberg, R. J. (1997). Glycoprotein D of herpes simplex virus (HSV) binds directly to HVEM, a member of the tumor necrosis factor receptor superfamily and a mediator of HSV entry. *Journal of Virology*, 71(8), 6083–6093. <https://doi.org/10.1128/JVI.71.8.6083-6093.1997>
- Whitley, R. J., & Roizman, B. (2001). Herpes simplex virus infections. *The Lancet*, 357(9267), 1513–1518. [https://doi.org/10.1016/S0140-6736\(00\)04638-9](https://doi.org/10.1016/S0140-6736(00)04638-9)
- Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, 35(Web Server issue), W407–W410. <https://doi.org/10.1093/nar/gkm290>
- World Health Organization (WHO). (2014). Status of vaccine research and development of vaccines for herpes simplex virus prepared for WHO PD-VAC. Pub. Sep. 24.
- Zhang, L. (2018). Multi-epitope vaccines: A promising strategy against tumors and viral infections. *Cellular & Molecular Immunology*, 15(2), 182–184. <https://doi.org/10.1038/cmi.2017.92>
- Zhu, J., & Paul, W. E. (2008). CD4 T cells: Fates, functions, and faults. *Blood: The Journal of the American Society of Hematology*, 112(5), 1557–1569. <https://doi.org/10.1182/blood-2008-05-078154>
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13), 3406–3415. <https://doi.org/10.1093/nar/gkg595>