In-Silico Design of Envelope based Multi-epitope Vaccine Candidate Against Kyasanur Forest Disease Virus

Sathishkumar Arumugam (sathishkumar.a@manipal.edu)
Manipal Academy of Higher Education

Research Article

Keywords: KFDV, Epitopes, Vaccine, Envelop protein, Docking, Immuno-informatics

Posted Date: February 5th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-155690/v1

In-Silico design of envelope based multi-epitope vaccine candidate against Kyasanur Forest Disease Virus.

Sathishkumar Arumugam*

Manipal Institute of Virology, Manipal Academy of Higher Education, Manipal-576104, Karnataka, India.

(E.mail: sathishkumar.a@manipal.edu)

ABSTRACT

Kyasanur Forest Disease Virus (KFDV) causing common tick-borne hemorrhagic fever in south India, there is no approved anti-viral or efficacious vaccine against this disease. Recent KFDV spread into new geographic locations gives urgent call for development of new vaccine and drugs. In this study, we adapted in-silico approach to design multi-epitope subunit vaccine for KFDV. Conserved regions of KFDV envelope protein sequences reported during 1962 to 2016 were identified. Eight different immuno-informatics tools were employed to predict the linear B-cell and T-cell epitopes, high scored and/or multi-immunogenic epitopes were linked together and obtained two vaccine candidates (VC1 and VC2). Obtained vaccine candidates were found to be non-allergic and had good antigenic properties, also gives the cross-protection against to Alkhurma Hemorrhagic Fever virus (AHFV). The 3D structures of vaccine candidates were built and validated. Docking of vaccine candidates with toll-like receptor-8 (TLR-8) was performed by Hex 8.0 and Cluspro, highest binding energy observed between VC2 and TLR8. JCAT sever confirmed cloning efficiency of both vaccine constructs and in-silico cloning into pET30a (+) vector by SnapGene suggests successful translation of vaccine constructs. In this study, multi-epitope vaccine candidates were designed and validated, it paves the way for up-coming vaccine and diagnostic kit development.

Keywords: KFDV, Epitopes, Vaccine, Envelop protein, Docking, Immuno-informatics.

INTRODUCTION

Kyasanur Forest Disease (KFD) is a southern Indian endemic zoonotic disease caused by KFDV that belongs to the family *Flaviviridae*. National institute of Allergy and Infectious Disease has recognized it as category C priority virus. KFDV was first identified in 1957 from isolates of the sick and dying monkeys of black-faced langur (*Presbytis entellus*) and red-faced bonnet (*Macaca radiate*) species in Kyasanur Forest, state of Karnataka, India[1,2]. Humans are the dead-end host of KFDV life cycle, initial stage KFDV circulates among monkeys, rodents then it transfers into humans by bites of infected ticks, primarily through *Haemaphysalis* sp[3].

The KFDV is an enveloped virus with icosahedral nucleocapsid, spherical in shape and about 40-65 nm in size. It contains 10,774 bases of positive single stranded RNA, encodes

3146 amino acid. Post-translational cleavage of polyprotein into three structural proteins (Capsid, Membrane and Envelope) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5)[4,5].

During 1957 to 2017, totally 9594 KFDV cases were reported in India[6]. Recently, KFDV outbreak was reported in a new geographic location Sindhudurg district of Maharashtra[7]. In future, it may spread to other parts of India. But still there is no cure for KFDV, currently formalin inactivated vaccine being used as primary strategy for controlling KFD[8], it gives partial effect against to KFDV infections and provides short term immunity to those have received vaccination[9]. It indicates low efficiency of current vaccine, there is need for development of new vaccine candidates against KFDV.

There was a study reported the KFDV envelope protein structural variation and different in B-cell epitope antigenicity[10]. There is no much studies on KFDV vaccine development, so we tried to construct the epitope based subunit vaccine which elite T-cell and B-cell immune response upon KFDV infection. T-cells and B-cells drive the process of adaptive immunity and develop immunological memory by recognizing portions of antigens called epitopes. These epitopes contains short amino acid sequence that can induce more direct and specific immune response, which overcome the disadvantages of live attenuated vaccines[11]. In-silico based epitope prediction methods have reduced the time consumption and money spent on false epitope candidates and this tools has been employed to develop multi-epitope vaccine against several diseases such as Dengue, SARS-CoV-2, *Mycobacterium tuberculosis*, *Steaphylococcus aureus* etc[12–17]. Hence, the present investigation has adapted this approach for rational design of epitope-based vaccine against KFDV. Here, first time we have identified T-cell and B-cell epitopes from highly conversed regions of KFDV envelop protein. And we have successfully designed the multi-immunogenic, non-allergic and high cloning efficient multi-epitope vaccine candidates by utilizing immune-informatics and bioinformatics tools.

RESULTS

Conserved region of KFDV envelope protein.

The multiple sequence alignment analysis (Supplementary Table S1) reveals that except few amino acids, the entire KFDV envelope protein sequences remain conserve. Phylogenetic tree was constructed from the alignment file, shows separate cluster of 2012 KFDV envelope protein, rest of the sequence were almost the same (Figure 1a). Based on the conserve score (Supplementary Table S2) conservation graph has been drawn, value 1.0 on the Y-axis represents more conserved (Figure 1b). Amino acids, which got >0.90 conserve score was

considered as a conserved region and taken into further study, conserved amino acid sequence and their position were given in table 1.

Conserved regions of KFDV envelope protein containing T-cell and B-cell epitopes.

Here we found that the conserved region contains epitopes for T-cell and B-cell targets. Totally 35 epitopes were obtained IEDB-MHC-I tools (Supplementary Table S3), whereas the MAPPP tool produced 300 epitopes from 43 human HLA alleles (Supplementary Table S4). IEDB-MHC-I top 10 percentile epitopes were also predicted in MAPPP tool, they are listed in table 2. Combined predictor of proteasomal cleavage, TAP transport, and MHC process has produced 34 epitopes (Supplementary Table S5), and top 10 ranked epitopes are listed in table 3.

There are 20 MHC class II epitopes (Supplementary Table S6) were identified by IEDB-MHC-II tools, whereas MHC2Pred tool recognized 504 epitopes (Supplementary Table S7) on the conserved region of the envelope protein. Based on IEDB=MHC-II percentile > 0.10 (FILT-1) and MHC2Pred epitopes which scored >0.60 as well having overlap sequence with FILT-1, 19 epitopes were selected as MHC II targets (table 4). B-cell liner epitopes prediction tools ABCpred, BCpred and SVMTriP were identified 15, 13 and 4 epitopes respectively (table 5).

Selected epitopes display cross-protection against Alkhumra Hemorrhagic fever Virus.

Uniportkb_human BLAST analysis showed similarity with four human proteins such as shish-7, Isoform 2 of Ribonuclease T2, Ribonuclease T2 and Extra-cellular ribonuclease with E-value of 2e-1, 7.7e-1, 2.1e0 and 2.1e0 respectively (Supplementary Table S8). These similarity hits can be ignored, as E-value is ≥ 0.1, because as rule of thumb an E-value should be <10-4 to assure the homology. Whereas UniProtKB BLAST revealed that selected epitopes were sharing the highest similarity with Alkhumra Hemorrhagic Fever Virus (AHFV) polyprotein, E-value was 1.7e-43 (Supplementary Table S9). So, the proposed vaccine candidates also confer immunity against AHFV.

Multi-epitope vaccine candidate sequences their allergenicity and antigenicity.

Totally, two multiple-epitope vaccine candidates were designed from selected high ranked, multi-immunogenic and over lapping epitope sequences. Vaccine construct 1 (VC1) composed with an adjuvant protein β defensin, 6 MHC- I epitope and 11 B-cell epitopes. Vaccine constructs 2 (VC2) composed of VC1 and 4 MHC-II epitopes. Complete information of vaccine construct has given in table 6. Vaccine candidates VC1 and VC2 were found to be non-allergic in behavior, VC1 got a higher antigenic score (0.6667) and identified as a better

antigen, while VC2 antigenic score was 0.5835.

Physiochemical and solubility properties of vaccine construct.

The molecular weight (MW) of constructed vaccine candidates VC1 and VC2 were found to 34.6 kDa and 39.4 kDa respectively. Theoretical pI value 9.38(VC1) and 9.25 (VC2) were observed and proteins are expected to alkaline in nature. Estimated half-life for both VC1 and VC2 candidates were 30 hours in mammalian reticulocytes in vitro. Instability index were found to be 32.86 and 33.61 for VC1 and VC2 respectively and confirmed as stable proteins. Negative GRAVY values of VC1 (-0.455) and VC2 (-0.511) indicate they are hydrophilic in nature and they could interact with water. PROSO II predicted VC1 as soluble protein with score of 0.667, whereas VC2 scored 0.489 and it identified as insoluble protein, obtained physiochemical characteristics strengthening the vaccine candidate's potency.

3D-strucure of designed vaccine candidates.

Phyre2 server has built three-dimensional structure of multi-epitope vaccine candidates. The cryo-em structure of tick borne encephalitis virus complex was used as template for VC1 and VC2 tertiary structure, and both the model got 100% score in terms of model confidence. Initial refinement of vaccine candidates VC1 and VC2 on ModRefiner have produced refine structure with TM-score of 0.9507 and 0.9622 respectively. Further refinement of VC1 and VC2 were done with GalaxyRefine server, based on quality score the best refined structure were chosen and named as VC1R and VC2R respectively. Those were used as ligand molecules in docking study (Figure 2a &b), quality scores of ligands were given in table 7.

The Ramachandran plot analysis of modeled vaccine candidates revealed that 97.4% of VC1R protein residues were in favored region and 2.6% of residues in allowed region (Figure 2c). Similarly, 95.9% of VC2R protein residues were in favored region and 4.1% of the residues in allowed region (Figure 2d). None of the amino acids residues were fallen in outlier region, which indicates the good quality of protein structure. ProSA-web analysis of yielded Z-score of -2.74 and -2.66 for vaccine candidates of VC1R and VC2R respectively. (Figure 2e &f).

Identification of binding energy between multi-epitope vaccine candidates and TLR8.

Docking analysis by Hex and Cluspro tools has revealed that designed vaccine candidates have strong binding capacity with TLR8. Hex docking of VC1 and VC2 with TLR8 showed binding energy value of -473.71 kcal/mol and -491.35 kcal/mol respectively (Figure3a &b). Cluspro protein-protein docking result page displayed 10 best-docked confirmations, the top ranked model from balanced coefficient-based calculation was considered as best model.

TLR8 with VC1 exhibited center energy values of -903.9 kcal/mol and -1234.4 kcal/mol as lowest energy (Figure 3c). Both center and lowest energy value of VC2 and TLR8 complex was found to -1256.1 kcal/mol(Figure 3d). These results clearly indicate that designed vaccine candidates have binding capacity with TLR8.

Codon Optimization and in silico cloning of KFD- VC1 and KFD-VC2.

Java Codon Adaptation Tool was used to check codon optimism of vaccine candidate's in *E.coli* (strain k12) expression system. It revealed that VC1 and VC2 multi-epitope vaccine construct composed of 927 and 1050 nucleotides respectively. The Codon Adaptation Index (CAI) was observed, 1.0 for VC1 and VC2 was 0.98. The GC content of VC1 found to be 51.13% and VC2 was 52.19%. The obtained values indicate that both vaccine candidates were having cloning efficiency. SnapGene tool was used for in-silico insertion of adapted codon sequence of vaccine construct into pET30a (+), between XhoI and NedI restriction sites and clones were obtained successfully (Figure 4).

Discussion

This genome era enabled with bulk inflow of genomics and proteomics information of almost all clinically important organisms. It facilitates us to choosing the targets for drug design, identifying new strains or drug resistance, diagnosis kit development and personalized medicine etc. Similarity, computational approaches simplified vaccine development process, vaccine informatics is a fast-growing field where in-silico vaccine design ventures are possible. Previous report on KFDV immune response has suggested to target both T cell and B cells mediated immune response for successful KFDV vaccine development[18]. Here, we have applied few immune informatic tools to identify the epitope which elicit both cellular and humoral immune responses.

In our study, KFDV envelope protein sequence was chosen as target for epitope prediction. The envelope protein of flaviviruses is major protective antigen and consists of three domains EI, EII and EIII. Especially EIII domain contains specific and sub-complex specific neutralizing epitopes, moreover it is easy to express by recombinant techniques[19]. Another challenge is genetic diversity of virus that resulted in immune escape, always-recombinant vaccines design prefers to target conserved antigenic regions of virus, it might accelerates better immune response[20]. The KFDV vaccinated individuals were reported for disease occurrence, it may because of variations in circulating virus[21]. Therefore, our study has picked the conserved region of KFDV envelope protein sequence, to overcome the genetic diversity of viral strains.

So far, there is no experimental study on identification of T-cell and B-cell epitopes. Here we have identified both of them by using immuno-informatics tools. Epitopes or viral antigens arise the specific immune responses in the body, which induce adaptive immune responses of T-cells mediated cellular immunity and B-cells/antibody mediated humoral immunity[22,23]. All three MHC-I epitope prediction tools identified KTILTLGDY and KTAEHLPKAW. Also, MHC II epitope KAWQVHRDWFEDLSL overlap with MHC class I epitope KAWQVHRDW, it may serve as target for both CD8+ and CD4+ cells. B-cell epitopes KRPPTDSGHD, TVVLELDKTAEH, DDIHQENPAKTR, AANESHSNRKTASF, TGSKPCRIPVRAVA were identified by any of two-prediction tools out of three.

The high ranked, multi-immunogenic and overlapped epitopes were linked together and multi-epitope construct was obtained. Typical purified proteins are not inherently immunogenic, it requires adjuvant to accelerate the innate immune system and enhance vaccine potency[24]. We have added β-defensin at N terminal end of the multi-epitope vaccine construct. β-defensin is known to induce lymphokines production which promotes T-cell mediated cellular immunity and antigen-specific Ig production[25].

Totally, we made two multi-epitopes vaccines construct against KFDV, 3D structure, and refinement of the vaccine candidates were achieved. The Ramachandran plots of built structures showed that none of the resided were in outlier region, >95% of the amino acids fallen in favored region. It clearly indicates that modeled structures of are in good quality and overall vaccine constructs are satisfactory.

TLRs plays vital role in innate immunity, especially they detect virus and leading to innate immune activation followed by activation of the adaptive immune response. Several reports confirm that TLR8 act as a host sensor for single stranded RNA (ssRNA) virus. ssRNA detected in compartments of TLR protein and subsequently activates the innate immune system[26–28]. Hence, Docking used TLR8 to see the interaction with designed vaccine molecules, there is no previous studies in this aspect. Our docking studies suggest that designed vaccine construct have the binding capacity with TLR8, both HEX and Cluspro docking tools have identified VC2 as potential binder with TLR8. Very importantly constructed vaccine should express in *E. coli* K12 strain to obtain more quantity of our multi-epitope recombinant vaccine protein. *E.coli* is most preferable choice for production of recombinant proteins[29]. In order to succeed the high-level production, codon optimization of epitope sequences were carried out. Codon adaptability index was >0.99% and GC content was > 50% for both vaccine candidates, it confirmed the favorable content of vaccine constructs for high-level protein expression in *E.coli* host.

Conclusion:

The present study has made an attempt to design the multi-epitope vaccine against KFDV by using immune-informatics tools. To our knowledge this is first report for identifying T-cell and B-cell targeting epitopes of KFDV. The designed chimeric vaccine peptide could elicit immune response, but still it need to be tested on in-vitro and in-vivo models. Interestingly, our constructed vaccine has cross-protection effect against AHFV. This study has given foresight for development of new prophylaxis for KFDV control in India, and gives the directions in selecting epitopes for KFDV vaccine development.

MATERIAL AND METHODS

Retrieval of amino acid sequence and conservation analysis.

Totally 13 KFDV envelope protein sequences reported during 1962 to 2016, were retrieved from NCBI database (http://www.ncbi.nlm.nih.gov).

Multiple sequence alignment (MSA) of retrieved sequences were performed in CLCworkbench (https://www.qiagenbioinformatics.com/products/clc-main-workbench/), alignment file (.aln) was generated from all retrieved envelop protein sequences, neighbor-joining method was used to the phylogenetic tree and conservation score of each amino acids were obtained.

Prediction of T-cell and B-cell epitopes.

Eight different immunoinformatic tools were used to predict T-cell (MHC class I and MHC class II) and B-cell epitope regions on KFDV envelope protein. Three different tools (IEDB-MHC -I, IEDB- combined and MAPPP) were used to obtain MHC class I epitope, proteasomal cleavage score, Transporter of Antigenic Peptide (TAP) and MHC scores. MHC class II epitope was prediction by IEDB- MHC-II and MHC2Pred tools. B-cell epitopes were identified by ABCpred, BCpred, and SVMTriP tools, the detailed information on tools algorithm, URL site and threshold values were in Table.8.

Self-antigen and cross-protection analysis.

Initially, all top ranked T-cell and B-cell epitope sequences were checked for self-antigen and cross-protection analysis. Uniport-BLAST tool(https://www.uniprot.org/blast/) engaged for this propose. Self–antigen or similarity of human proteins with predicted epitopes were identified by Uniport- BLAST search against uniportkb_human databases. And UniProtKB database BLAST search was used to identify the cross protection of selected epitopes with other pathogenic organisms.

Construction of multi-epitope vaccine candidate.

Based on high score, multi-immunogenic and overlapping sequence of T-cell and B-cell epitopes were conjugated to construct the multi-epitope vaccine candidates. Construct of vaccine has started with adjuvant β defensin (ACK99045.1) peptide sequence, in order to augment the immunogenicity of the vaccine candidate. Adjuvant was connected to MHC-I epitopes with EAAAK linker, every individual epitopes of MHC-I were linked by GGGS and all MHC-II epitopes assembled with GPGPG linker. Whereas B-cell epitopes were combined by KK and linkers.

Allergenicity and antigenicity of the vaccine candidate.

Algpred (https://webs.iiitd.edu.in/raghava/algpred/submission.html) used to predict the allergenicity of vaccine candidate based on the similarity of known epitope. The vaccine candidate sequence was uploaded into the server, the IgE epitope and PID-BLAST search on allergen representative peptide algorithms were chosen for allergen prediction[30]. Antigenicity capacity of the vaccine candidate were evaluate by VaxiJen 2.0 server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) [31–33].

Physiochemical properties and solubility prediction

Physiochemical parameters of vaccine candidate scuh as molecular weight, theoretical pI, instability index, aliphatic index, estimated half-life and grand average of hydropathicity (GRAVY) were identified by ProtParam (https://web.expasy.org/protparam/) tool[34]. PROSO II server (http://mbiljj45.bio.med.uni-muenchen.de:8888/prosoII/prosoII.seam) was used to evaluate the solubility of vaccine candidates [35].

3D Structure prediction, refinement and validation.

Phyre2 server (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index)[36] used to build the 3D-structure of the designed vaccine candidates. The server is build the 3D model

based on homologues with known protein structure, and produce accurate model of about 70% of the domains in a typical genome. The obtained 3D structures of designed multi-epitope vaccine candidates were subjected into two-step refinement. Initial refinement was performed in ModRefiner (https://zhanglab.ccmb.med.umich.edu/ModRefiner/), an atomic-level protein structure refinement tool. This program makes protein into full-atomic relaxation, where initial model or reference model does not restrain the refined model. This improves both global and local structures with more accurate side chain positions, better hydrogen-bonding networks, and fewer atomic overlaps. Then second refinement was done with GalaxyRefine server (http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE)[37,38], it performs repeated structure perturbation to side-chains and for secondary structure elements and loops are also applied followed by overall structural relaxation by molecular dynamics simulation. The obtained 3D-structures of vaccine candidates were validated in RAMPAGE server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php)[39] and ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php)[40,41]. ProSA recognize the errors in experimental and theoretical model of protein structure and z-core will tell about the quality of the protein structure.

Molecular docking of designed vaccine candidate with TLR8.

TLR mediating innate immune activation upon viral infections, especially TLR8 and it is closely related protein TLR7 were reported for recognition of viral ssRNA and/or antiviral responses (Heil; jurk). Hence, TLR considered as promising target for adjuvants and antiviral compounds. This part of investigation has aimed to see the interaction between TLR8 (PDB ID: 3W3G) and designed multiepitope vaccine candidate by two different docking tools. Hex 8.0.0 (http://hex.loria.fr) docking tool is First Fourier Transform (FTT) based analysis software[42]. In this method, rigid docking is undertaken taking into consideration different orientation through 6D analysis. The docking was performed with default parameters such as correlation types-shape, FFT mode-3D, grid dimension 0.6, receptor range-180, ligand range-180, twist range-360, distance range-40, translation step-0.8 and score threshold <0.0. The ClusPro server (https://cluspro.org) is a widely used tool for protein-protein docking[43,44].

In silico cloning optimization of KFD multi-epitope vaccine candidates.

The Java Codon Adaptation Tool (http://www.prodoric.de/JCat) was used for reverse translation and codon optimization of designed vaccine candidates [45]. Codon optimization was executed in order to express the KFD multi epitope vaccine construct in Escherichia coli (strain

K12). The output files were checked for Codon Adaptation Index (CAI) (>0.8-1.0), and GC content (30—70%), to ensure transcription and translational efficiency of designed vaccine construct. Resulted optimized codon sequence of vaccine candidates were introduced with XhoI and NdeI restriction sties at N-terminal and C-terminal ends respectively. SnapGene (www.snapgene.com) was utilized to insert the vaccine sequence in expression vector pET-30a(+), between XhoI and NdeI cloning site, final clones of vaccine candidates VC1 and VC2 were obtained.

References:

1. T H WORK, H. T. A new virus disease in India. Summary of preliminary report on investigations of the Virus Research Center on an epidemic disease affecting forest villagers and wild monkeys of Shimoga District, Mysore. *Indian Journal of Medical Sciences* **11**, 341–342 (1957).
2. Sreenivasanh, M. A., Bhat, R. & Rajagopalan, P. K. The epizootics of kyasanur forest disease in wild monkeys during 1964 to 1973. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **80**, 810–814 (1986).
3. Naren babu, N. *et al.* Spatial distribution of Haemaphysalis species ticks and human Kyasanur Forest Disease cases along the Western Ghats of India, 2017–2018. *Experimental and Applied Acarology* **77**, 435–447 (2019).
4. Dodd, K. A. *et al.* Ancient ancestry of KFDV and AHFV revealed by complete genome analyses of viruses isolated from ticks and Mammalian hosts. *PLoS Neglected Tropical Diseases* **5**, 1–7 (2011).
5. Gritsun, D. J., Jones, I. M., Gould, E. A. & Gritsun, T. S. Molecular Archaeology of Flaviviridae Untranslated Regions : Duplicated RNA Structures in the Replication Enhancer of Flaviviruses and Pestiviruses Emerged via Convergent Evolution. **9**, 1–11 (2014).
6. Chakraborty, S., Andrade, F. C. D., Ghosh, S., Uelmen, J. & Ruiz, M. O. Historical Expansion of Kyasanur Forest Disease in India From 1957 to 2017: A Retrospective Analysis. *GeoHealth* **3**, 44–55 (2019).
7. Gurav, Y. K. *et al.* Kyasanur Forest Disease Prevalence in Western Ghats Proven and Confirmed by Recent Outbreak in Maharashtra, India, 2016. *Vector-Borne and Zoonotic Diseases* **18**, 164–172 (2018).
8. Mehendale, S. *et al.* Kyasanur Forest Disease Outbreak and Vaccination Strategy , Shimoga District , India , 2013 – 2014. **21**, 2013–2014 (2019).
9. Kasabi, G. S., Murhekar, M. V, Sandhya, V. K. & Raghunandan, R. Coverage and

Effectiveness of Kyasanur Forest Disease (KFD) Vaccine in Karnataka , South India , 2005 – 10. **7**, 13–16 (2013).

10. Shil, P., Yadav, P. D., Patil, A. A., Balasubramanian, R. & Mourya, D. T. Bioinformatics characterization of envelope glycoprotein from Kyasanur Forest disease virus. 195–201 (2018). doi:10.4103/ijmr.IJMR
11. Palatnik-de-Sousa, C. B., Soares, I. da S. & Rosa, D. S. Editorial: Epitope discovery and synthetic vaccine design. *Frontiers in Immunology* **9**, 9–11 (2018).
12. Ali, M., Pandey, R. K., Khatoon, N., Narula, A. & Mishra, A. Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Scientific Reports* 1–13 (2017). doi:10.1038/s41598-017-09199-w
13. Can, H., Köseoğlu, A. E., Alak, S. E., Güvendi, M. & Döşkaya, M. In silico discovery of antigenic proteins and epitopes of SARS - CoV - 2 for the development of a vaccine or a diagnostic approach for COVID - 19. *Scientific Reports* 1–16 (2020). doi:10.1038/s41598-020-79645-9
14. Kar, T. *et al.* OPEN A candidate multi - epitope vaccine against SARS - CoV - 2. *Scientific Reports* 1–24 (2020). doi:10.1038/s41598-020-67749-1
15. Behmard, E., Soleymani, B., Najafi, A. & Barzegari, E. Immunoinformatic design of a COVID - 19 subunit vaccine using entire structural immunogenic epitopes of SARS - CoV - 2. *Scientific Reports* 1–12 (2020). doi:10.1038/s41598-020-77547-4
16. Bibi, S., Ullah, I., Zhu, B., Adnan, M. & Liaqat, R. In silico analysis of epitope - based vaccine candidate against tuberculosis using reverse vaccinology. *Scientific Reports* 1–16 (2021). doi:10.1038/s41598-020-80899-6
17. Klimka, A. *et al.* Epitope-speci fi c immunity against Staphylococcus aureus coproporphyrinogen III oxidase. *npj Vaccines* doi:10.1038/s41541-020-00268-2
18. Devadiga, S., McElroy, A. K., Prabhu, S. G. & Arunkumar, G. Dynamics of human B and T cell adaptive immune responses to Kyasanur Forest disease virus infection. *Scientific Reports* **10**, 1–9 (2020).
19. Liu, G. *et al.* Immunogenicity and Efficacy of Flagellin-Envelope Fusion Dengue. **22**, 516–525 (2015).
20. Bauer, A. *et al.* Preferential Targeting of Conserved Gag Regions after Vaccination with a Heterologous DNA Prime-Modified Vaccinia Virus Ankara Boost HIV-1 Vaccine Regimen. *Journal of virology* **91**, (2017).

21. Rajaiah, P. Kyasanur Forest Disease in India: innovative options for intervention. *Human Vaccines and Immunotherapeutics* **15**, 2243–2248 (2019).
22. Sette, A. *et al.* Definition of epitopes and antigens recognized by vaccinia specific immune responses: their conservation in variola virus sequences, and use as a model system to study complex pathogens. *Vaccine* **27 Suppl 6**, G21–G26 (2009).
23. Welsh, R. M. & Fujinami, R. S. PERSPECTIVES immunity and vaccine design. **5**, 555–563 (2007).
24. De Gregorio, E., Caproni, E. & Ulmer, J. B. Vaccine adjuvants: Mode of action. *Frontiers in Immunology* **4**, 1–6 (2013).
25. Kenji Tani *et al.* Defensins act as potent adjuvant taht promote cellular and humoral immune response in mice to a lymphhoma idiotype and carrier antigents. *International Immunology* **12**, 691–700 (2000).
26. Triantafilou, K. *et al.* TLR8 and TLR7 are involved in the host ' s immune response to human parechovirus 1. 2416–2423 (2005). doi:10.1002/eji.200526149
27. Carty, M. & Bowie, A. G. Recent insights into the role of Toll-like receptors in viral infection. *Clinical and Experimental Immunology* **161**, 397–406 (2010).
28. Martinez, J., Huang, X. & Yang, Y. Toll-like receptor 8-mediated activation of murine plasmacytoid dendritic cells by vaccinia viral DNA. (2010). doi:10.1073/pnas.0913291107
29. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli : advances and challenges. **5**, 1–17 (2014).
30. Saha, S. & Raghava, G. P. S. AlgPred: Prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Research* **34**, 202–209 (2006).
31. Doytchinova, I. A. & Flower, D. R. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* **8**, 1–7 (2007).
32. Doytchinova, I. A. & Flower, D. R. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. *Vaccine* **25**, 856–866 (2007).
33. Doytchinova, I. & Flower DR. Bioinformatic approach for identifying parasite and fungal candidate subunit vaccines. *Open Vaccine J* **1**, 4 (2008).
34. Wilkins, M. R. *et al.* Protein identification and analysis tools in the ExPASy server. *Methods in molecular biology (Clifton, N.J.)* **112**, 531–552 (1999).
35. Smialowski, P., Doose, G., Torkler, P., Kaufmann, S. & Frishman, D. PROSO II – a new method for protein solubility prediction. **279**, 2192–2200 (2012).

36. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling , prediction and analysis. *Nature Protocols* **10**, 845–858 (2015).
37. Shin, W.-H., Lee, G. R., Heo, L., Lee, H. & Seok, C. Prediction of Protein Structure and Interaction by GALAXY Protein Modeling Programs. *Bio Design* **2**, 1–11 (2014).
38. Ko, J., Park, H., Heo, L. & Seok, C. GalaxyWEB server for protein structure prediction and refinement. **40**, 294–297 (2012).
39. Carugo, O. & Djinović-Carugo, K. A proteomic Ramachandran plot (PRplot). *Amino acids* **44**, 781–790 (2013).
40. Wiederstein, M. & Sippl, M. J. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research* **35**, 407–410 (2007).
41. Sippl, M. J. Recognition of errors in three-dimensional structures of proteins. *Proteins: Structure, Function, and Bioinformatics* **17**, 355–362 (1993).
42. Macindoe, G., Mavridis, L., Venkatraman, V., Devignes, M. D. & Ritchie, D. W. HexServer: An FFT-based protein docking server powered by graphics processors. *Nucleic Acids Research* **38**, 445–449 (2010).
43. Vajda, S. *et al.* New additions to the ClusPro server motivated by CAPRI. *Proteins: Structure, Function and Bioinformatics* **85**, 435–444 (2017).
44. Kozakov, D. *et al.* HHS Public Access. **12**, 255–278 (2018).
45. Grote, A. *et al.* JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Research* **33**, 526–531 (2005).
46. Kim, Y., Sidney, J., Pinilla, C., Sette, A. & Peters, B. Derivation of an amino acid similarity matrix for peptide:MHC binding and its application as a Bayesian prior. *BMC Bioinformatics* **10**, 1–11 (2009).
47. Nielsen, M. *et al.* Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Science* **12**, 1007–1017 (2003).
48. Lundegaard, C., Lund, O. & Nielsen, M. Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers. *Bioinformatics* **24**, 1397–1398 (2008).
49. Peters, B. & Sette, A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics* **6**, 1–9 (2005).
50. Sidney, J. *et al.* Quantitative peptide binding motifs for 19 human and mouse MHC

class i molecules derived using positional scanning combinatorial peptide libraries. *Immunome Research* **4**, 1–14 (2008).

51. Tenzer, S. *et al.* Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cellular and Molecular Life Sciences* **62**, 1025–1037 (2005).
52. Peters, B., Bulik, S., Tampe, R., van Endert, P. M. & Holzhütter, H.-G. Identifying MHC Class I Epitopes by Predicting the TAP Transport Efficiency of Epitope Precursors. *The Journal of Immunology* **171**, 1741–1749 (2003).
53. Hakenberg J, Nussbaum AK, Schild H, Rammensee HG, Kuttler C, Holzhütter HG, Kloetzel PM, Kaufmann SH, M. H. MAPPP: MHC class I antigenic peptide processing prediction. *Applied Bioinformatics* **2**, 155–158 (2003).
54. Lin, H. H., Ray, S., Tongchusak, S., Reinherz, E. L. & Brusic, V. Evaluation of MHC class I peptide binding prediction servers: Applications for vaccine research. *BMC Immunology* **9**, 1–13 (2008).
55. Wang, P. *et al.* Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics* **11**, 568 (2010).
56. Wang, P. *et al.* A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS Computational Biology* **4**, (2008).
57. Lata, S., Bhasin, M. & Raghava, G. P. S. Application of Machine Learning Techniques in Predicting MHC Binders BT - Immunoinformatics: Predicting Immunogenicity In Silico. in (ed. Flower, D. R.) 201–215 (Humana Press, 2007). doi:10.1007/978-1-60327-118-9_14
58. Saha, S. & Raghava, G. P. S. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* **65**, 40–48 (2006).
59. Chen, J., Liu, H., Yang, J. & Chou, K. C. Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. *Amino Acids* **33**, 423–428 (2007).
60. El-Manzalawy, Y., Dobbs, D. & Honavar, V. Predicting linear B-cell epitopes using string kernels. *Journal of Molecular Recognition* **21**, 243–255 (2008).
61. El-Manzalawy, Y., Dobbs, D. & Honavar, V. Predicting flexible length linear B-cell epitopes. *Computational systems bioinformatics / Life Sciences Society. Computational Systems Bioinformatics Conference* **7**, 121–132 (2008).
62. Yao, B., Zhang, L., Liang, S. & Zhang, C. SVMTriP: A Method to Predict Antigenic Epitopes Using Support Vector Machine to Integrate Tri-Peptide Similarity and Propensity. *PLoS ONE* **7**, 5–9 (2012).

Figure 1: Multiple sequence Alignment of KFDV envelop protein A) Phylogenetic tree of KFDV envelope protein. B) Conservation graph of KFDV envelop protein sequence (value1.0 is highest conserved).

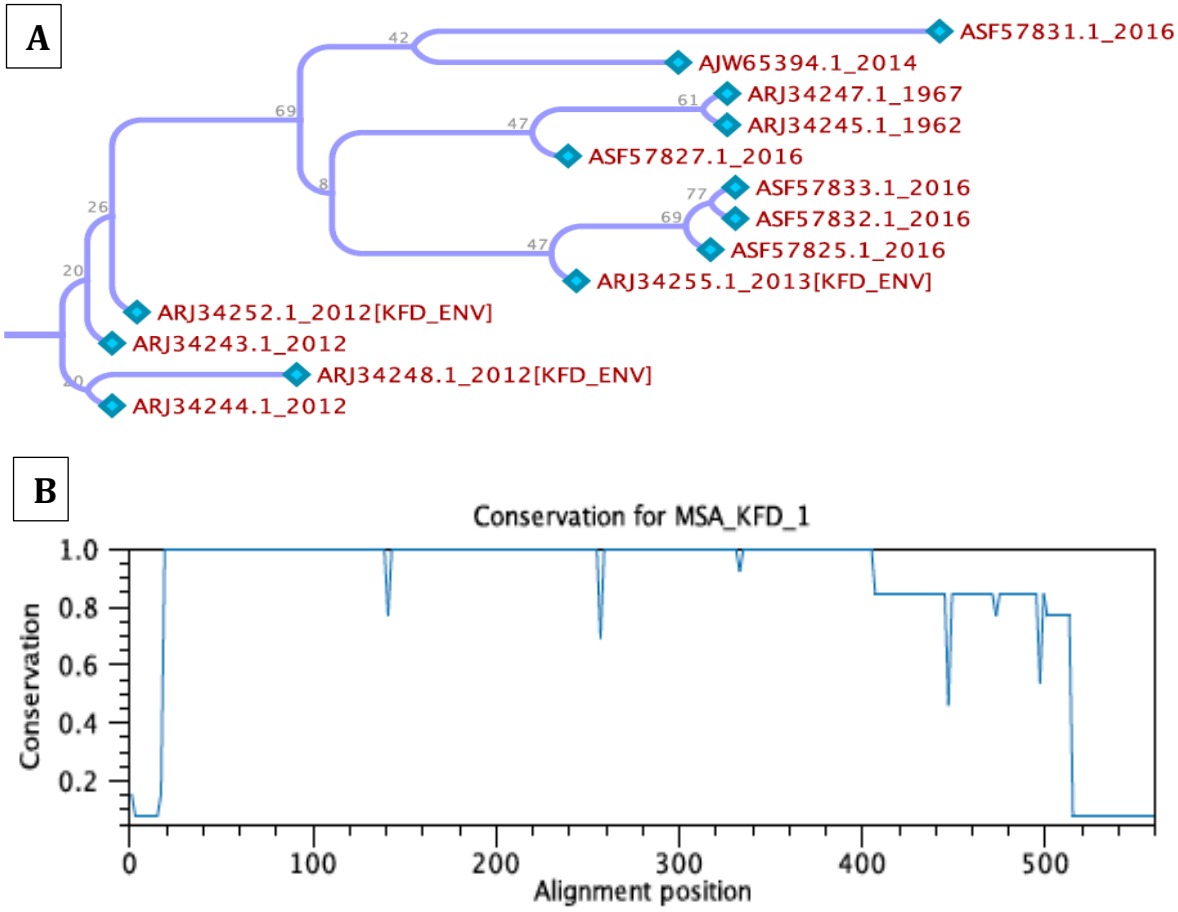

Figure2: 3D-modeling, refinement and validation of multi epitope vaccine constructs. a) Refined 3D structure of VC1 construct b) Refined 3-D structure of VC2 construct c) Ramachandran plot analysis of refined VC1 structure d) Ramachandran plot analysis of refined VC2 structure e) ProSA-web shows Z-score of -2.74 for VC1 and f) -2.66 for VC2.

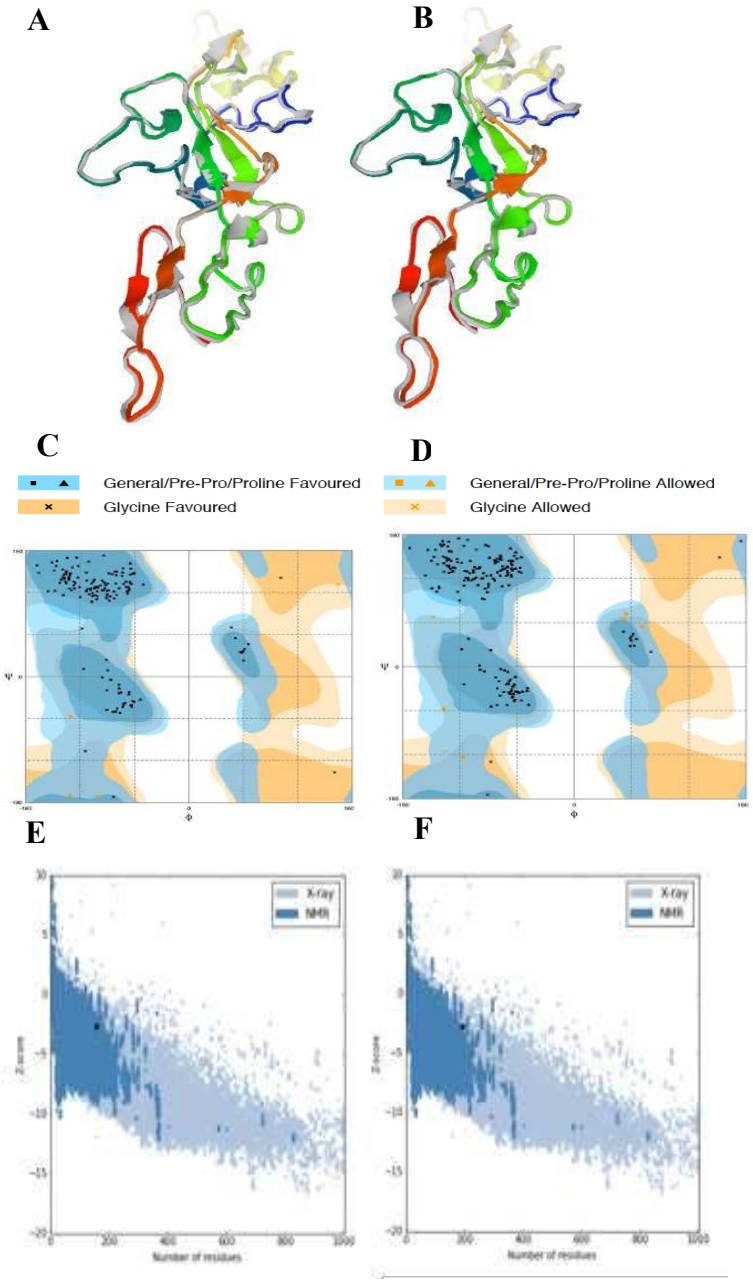

Figure 3: Identification of binding energy by Hex and Cluspro docking tools A) Hex docked complex of VC1 and TLR8 B) VC2 and TLR8 docking confirmation C) Cluspro best model of VC1 and TLR8 complex D) VC2 and TLR8 complex (Ligand-Yellow color; Receptor-brown color).

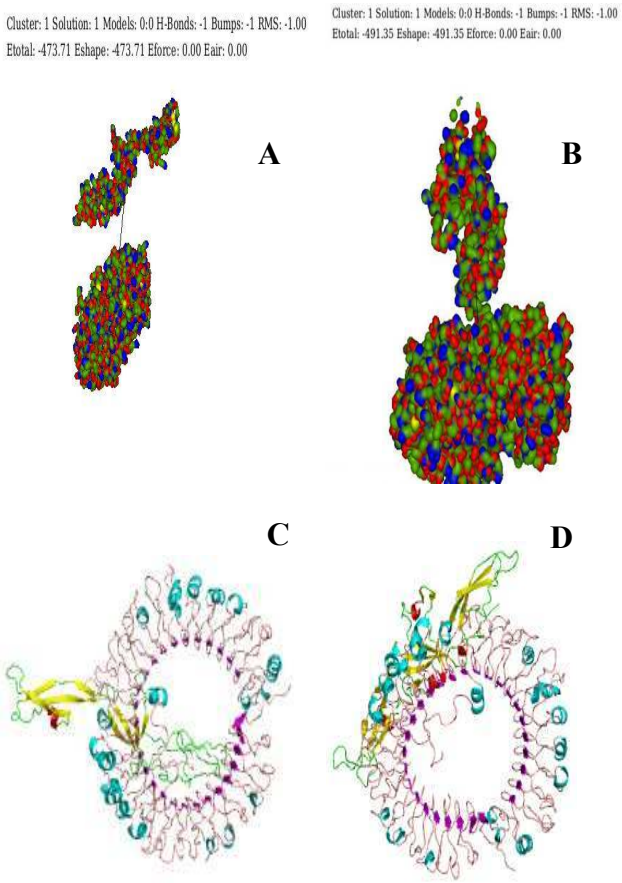

Figure4: In silico restriction cloning of multiepitope subunit vaccine sequence into pET30a(+) expression vector, red color part represent vaccine sequence and black circle represent vector sequence. Restriction sites are heighted in yellow color and translation sequence parts were mention in green color arrows.

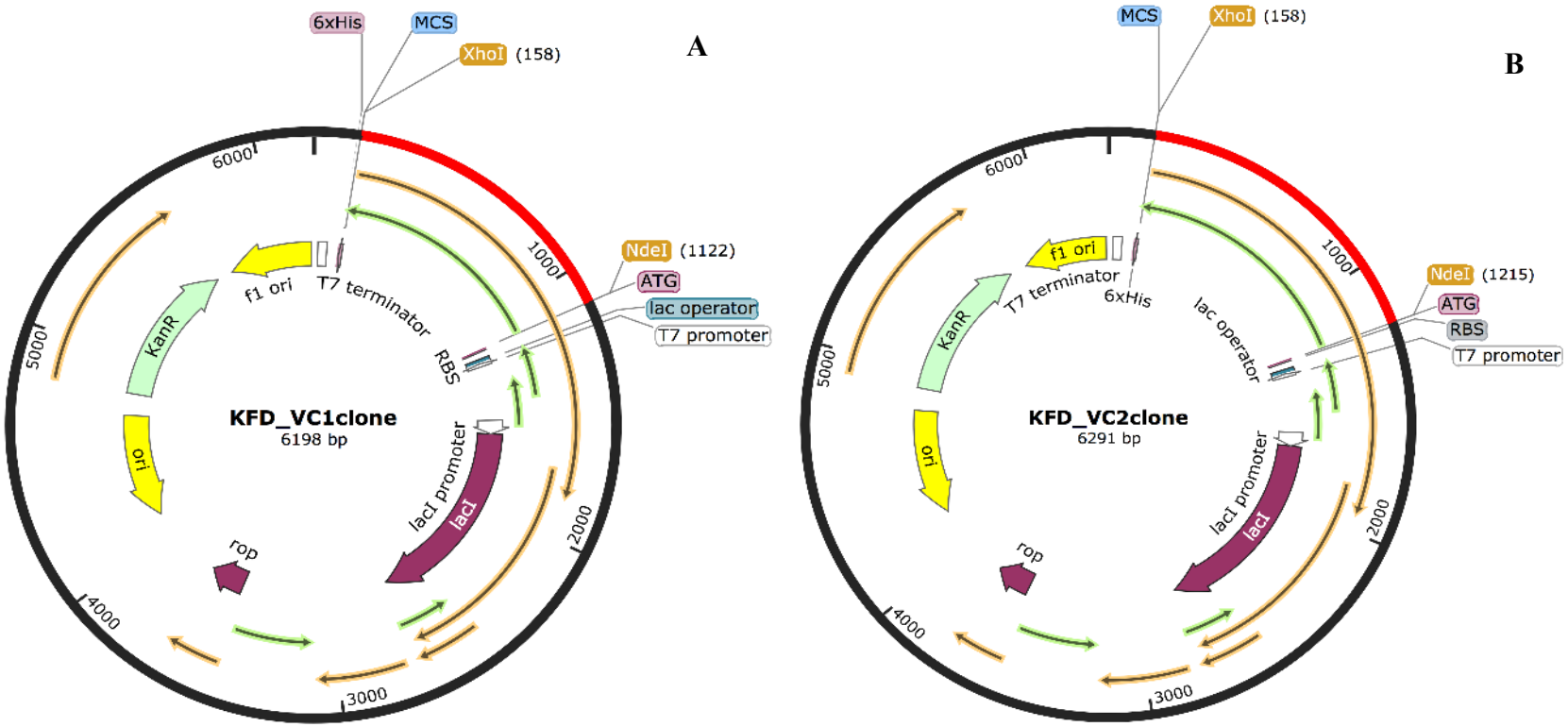

Table 1: Conserved amino acid sequences of KFDV envelop protein.

Name	Conserved amino acid sequence	Position	Length
CS1	TRVSLVLELGGCVTLTAEGKPSVDVWLDDIHQENPAKTREYC LHAKLANSKVAARCPAMGPATLPEEHQASTVCRRDQSDRGW GNHCGLFGKGSIVACAKFSCETKKKATGYVYDVNKITYV	19 to 140	121
CS2	KVEPHTGDYLAANESHSNRKTASFTTQSEKTILTLGDYGDISL TCRVTSGVDPAQTVVLELDKTAEHLPKAWQVHRDWFEDLSL PWRHEGAQEWNHADRLVEFGEPHAVKMDIFN	142 to 256	114
CS3	GDQTGILLKSLAGVPVANIEGSKYHLQSGHVTCDVGLEKLKM KGMTYTVCEGSKFAWKRPPTDSGHDTVVMEVTYTGSKPCRI PVRAVAHGEPNVNVASLITPNPSMETTGGGFVELQLPPGDNII YVGELSHQWFQKGSTIGRVLEKT	258 to 406	148

Table 2: List of top 10 MHC1 binding epitopes (overlapping sequence found in both IEDB-MHC I and MAPPP tool were bolded)

S.No	Allele	Peptide Length	Sequence	Proteasome Score	TAP Score	MHC Score	Processing Score	Total Score
1	HLA-B*35:01	10	YVYDVNKITY	1.48	1.39	-1.5	2.87	1.37
2	HLA-A*30:02	9	**KTILTLGDY**	1.24	1.26	-1.23	2.51	1.27
3	HLA-B*40:01	9	REYCLHAKL	1.54	0.55	-0.93	2.09	1.16
4	HLA-B*40:01	9	LELGGCVTL	1.76	0.39	-1.04	2.15	1.11
5	HLA-A*30:02	9	KVEPHTGDY	0.96	1.26	-1.37	2.22	0.86
6	HLA-B*35:01	9	LPPGDNIIY	1.45	1.11	-1.77	2.57	0.8
7	HLA-A*26:01	9	ETKKKATGY	1.2	1.17	-1.59	2.37	0.78
8	HLA-B*58:01	10	**KTAEHLPKAW**	1.5	0.37	-1.12	1.87	0.75
9	HLA-B*15:01	10	YVYDVNKITY	1.48	1.39	-2.18	2.87	0.69
10	HLA-A*23:01	10	IYVGELSHQW	1.31	0.55	-1.18	1.86	0.68

Table 3: List of top 10 MHC1 binding epitopes based on combined score of proteasome, cleave, TAP and MHC processing score. (overlapping epitope sequence found in IEDB-MHC I, IEDB-MHC-I-combined and MAPPP tool were bolded).

S.No	Allele	Length	Peptide	IEDB-MHC-I Percentile rank
1	HLA-B*57:01	9	KAWQ**VHRDW**	0.06
2	HLA-A*30:02	9	**KTILTLGDY**	0.06
3	HLA-B*57:01	10	**KTAEHLPKAW**	0.09
4	HLA-A*11:01	9	**STIGRVLEK**	0.11
5	HLA-A*33:01	10	**DYGDISLTCR**	0.15
6	HLA-B*58:01	10	**KTAEHLPKAW**	0.17
7	HLA-B*07:02	10	**KPCRIPVRAV**	0.18
8	HLA-A*11:01	9	**ASFTTQSEK**	0.2
9	HLA-B*58:01	9	**KAWQVHR**DW	0.2
10	HLA-A*11:01	9	**VTCDVGLE**K	0.2

Table 4: List of MHC2 binding epitopes (overlapping epitope sequence found in IEDB-MHC-II and MHC2Pred tools were bolded).

S.No	Allele	Method	Peptide	IEDB-percentile rank/ MHC2PRED Peptide score
1	HLA-DQA1*01:01/ DQB1*05:01	ANN/IEDB	**KAWQVHRDWFEDLSL**	0.02
2	HLA-DRB1*0802	SVM/MHC2PRED	**KAWQVHRDW**	0.672
3	HLA-DR3	SVM/MHC2PRED	P**KAWQVHRD**	0.819
4	HLA-DR15	SVM/MHC2PRED	LP**KAWQVHR**	0.839
5	HLA-DQ8	SVM/MHC2PRED	**RDWFEDLSL**	1.797
6	HLA-DQA1*01:01/ DQB1*05:01	ANN/IEDB	**AWQVHRDWFEDLSLP**	0.02
7	HLA-DRB5*0101	SVM/MHC2PRED	**AWQVHRDWF**	0.836
8	HLA-DQB1*0301	SVM/MHC2PRED	PK**AWQVHRD**	0.897
9	HLA-DQ4	SVM/MHC2PRED	**DWFEDLSLP**	0.847
10	HLA-DQA1*01:01/ DQB1*05:01	ANN/IEDB	**WQVHRDWFEDLSLPW**	0.02
11	HLA-DRB5*0101	SVM/MHC2PRED	**AWQVHRDWF**	0.836
12	HLA-DQB1*0301	SVM/MHC2PRED	**WFEDLSLPW**	0.751
13	HLA-DR3	SVM/MHC2PRED	**FEDLSLPWR**	1.168
14	HLA-DQA1*01:01/ DQB1*05:01	ANN/IEDB	**QVHRDWF**EDLSLPWR	0.04
15	HLA-DRB5*0101	SVM/MHC2PRED	AW**QVHRDWF**	0.836
16	HLA-DQA1*01:01/ DQB1*05:01	ANN/IEDB	VHRDWF**EDLSLPWRH**	0.05
17	HLA-DQ6	SVM/MHC2PRED	**EDLSLPWRH**	0.639
18	HLA-DRB3*01:01	ANN/IEDB	KPSVDV**WLDDIHQEN**	0.1
19	HLA-DQB1*03	SVM/MHC2PRED	**WLDDIHQEN**	0.917

Table 5: List of Bcell epitopes predicated by ABCpred, BCpred and SVMTriP tools (overlapping sequence of B-cell epitopes sequences were bolded).

S.No	ABCpred		BCpred		SVMTriP	
	Epitope	**Score**	**Epitope**	**Score**	**Epitope**	**Score**
1	AQ**TVVLELDKTAEH**	0.91	AARCPAMGPATLPE	0.999	KTREYCLHAKLANS	1
2	ASTVCRRDQSDRGW	0.87	L**DDIHQENPAKTRE**	0.944	TDSGHDTVVMEVTY	1
3	DKTAEHLPKAWQVH	0.87	RDQSDRGWGNHCGL	0.895	WFQKGSTIGRVLEK	0.886
4	**DDIHQENPAKTRE**Y	0.85	KKKATGYVYDVNKI	0.877	**TVVLELDKTAEH**LP	0.834
5	**NVNVASLITPNP**SM	0.83	WRHEGAQEWNHADR	0.902		
6	VGLEKLKMKGMTYT	0.83	LTLGDYGDI**SLTCR**	0.845		
7	VACAKFSCETKKKA	0.82	**AANESHSNRKTASF**	0.843		
8	**SLTCR**VTSGVDPAQ	0.82	EFGEPHAVKMDIFN	0.652		
9	GVPVANIEGSKYHL	0.82	GGGFVEL**QLPPGDN**	0.98		
10	FGKGSIVACAKFSC	0.81	KFAW**KRPPTDSGHD**	0.98		
11	**KRPPTDSGHD**TVVM	0.81	**TGSKPCRIPVRAVA**	0.974		
12	**QLPPGDN**IIYVGEL	0.81	SHQWFQKGSTIGRV	0.761		
13	**AANESHSNRKTASF**	0.8	EP**NVNVASLITPNP**	0.723		
14	**SNRKTASF**TTQSEK	0.78				
15	**TGSKPCRIPVRAVA**	0.78				

Table 6: Multi-epitope vaccine candidate sequence

Vaccine Construct	Composition Of Construct	Complete sequence of designed construct	Allegicity (AlgPred)	VaxiJen score Status
VC1	β defensin adjuvant, high scored MHC-1 epitopes and B-cell epitopes	MSYLRNSTSLVRVPKAFLKPFRVCCFVIAG HGGIINTLQKYYCRVRGGRCAVLSCLPKEE QIGKCSTRGRKCCRRKK**EAAAK**LELGGCV TL**GGGS**REYCLHAKL**GGGS**YVYDVNKITY **GGGS**KVEPHTGDY**GGGS**KTILTLGDY**GG GS**DYGDISLTCR**GGGS**KTAEHLPKAW**KK**L DDIHQENPAKTREYCLHAKLANS**KK**AARC PAMGPATLPE**KK**ASTVCRRDQSDRGWGN HCGL**KK**VACAKFSCETKKKATGYVYDVN KI**KK**LTLGDYGDISLTCRVTSGVDPAQAQT VVLELDKTAEHLPKAWQVH	Non-allergen	0.6667 Portable antigen
VC2	β defensin adjuvant, high scored MHC-I epitopes, B-Cell epitopes, MHC-II epitopes	MSYLRNSTSLVRVPKAFLKPFRVCCFVIAG HGGIINTLQKYYCRVRGGRCAVLSCLPKEE QIGKCSTRGRKCCRRKK**EAAAK**LELGGCV TL**GGGS**REYCLHAKL**GGGS**YVYDVNKITY **GGGS**KVEPHTGDY**GGGS**KTILTLGDY**GG GS**DYGDISLTCR**GGGS**KTAEHLPKAW**KK**L DDIHQENPAKTREYCLHAKLANS**KK**AARC PAMGPATLPE**KK**ASTVCRRDQSDRGWGN HCGL**KK**VACAKFSCETKKKATGYVYDVN KI**KK**LTLGDYGDISLTCRVTSGVDPAQAQT VVLELDKTAEHLPKAWQVH**GPGPG**LPKA WQVHRDWFEDLSLPWRHKPSVDVWLDDI HQEN	Non-allergen	0.5835 Portable antigen

Table 7: GalaxyRefine server quality scores of vaccine candidates.

Vaccine model	GDT-HA ,	RMSD	MolProbity	Clash score	Poor Rotamers	Rama score
VC1_model3 (VC1R)	0.9143	0.488	1.567	8.4	0.0	97.4
VC2_model2 (VC2R)	0.9477	0.437	1.961	13.5	0.0	95.4

Table 8: List of T-cell and B-cell prediction tools.

Epitope Target	**Tool**	**Algorithm**	**Threshold Value**	**URLsite**
T-cell MHC class I	IEDB-MHC-I	Artificial Network	Percentile rank >0.5	http://tools.iedb.org/mhci/ [46–50]
	IEDB-combined predictor	Artificial Network	Total score >0.5	http://tools.iedb.org/processing/ [51,52]
	MAPPP	SYFPEITHI	>1.0	http://www.mpiib-berlin.mpg.de/MAPPP/binding.html [53,54]
T-cell MHC class II	IEDB-MHC-II	Artificial Network	Percentile rank >0.5	http://tools.iedb.org/mhcii/[55,56]
	MHC2Pred	Support Vector Machine (SVM)	> 1.5	http://crdd.osdd.net/raghava/mhc2pred/[57]
B-Cell	ABCpred	Artificial Network	>0.5	https://webs.iiitd.edu.in/raghava/abcpred/ABC_submission.html [58]
	BCpred	Artificial Network	Specificity 75%	http://ailab.ist.psu.edu/bcpred/predict.html [59–61]
	SVMTriP	Support Vector Machine (SVM)	Epitope length 14	http://sysbio.unl.edu/SVMTriP/prediction.php[62]

Figures

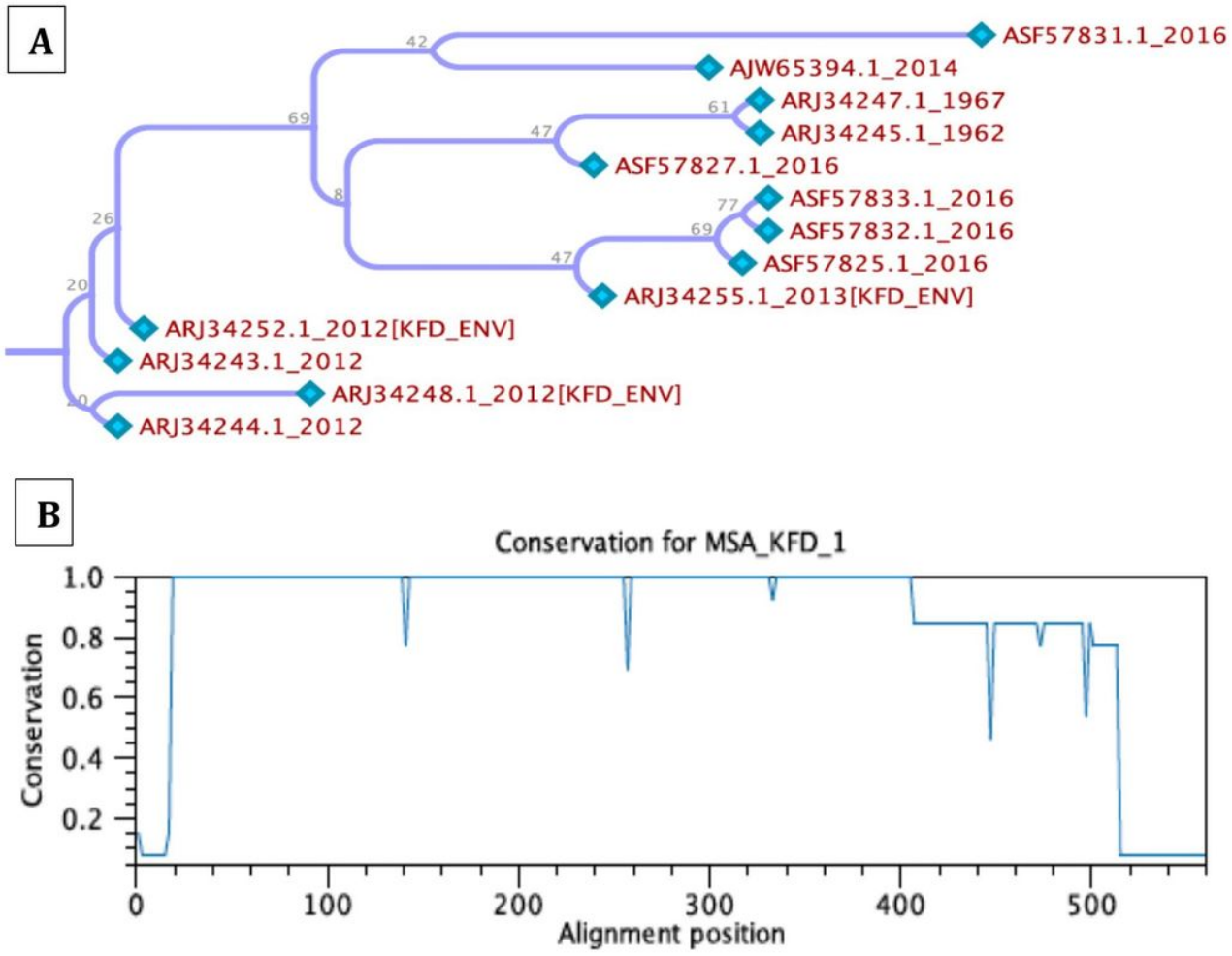

Figure 1

Multiple sequence Alignment of KFDV envelop protein A) Phylogenetic tree of KFDV envelope protein. B) Conservation graph of KFDV envelop protein sequence (value1.0 is highest conserved).

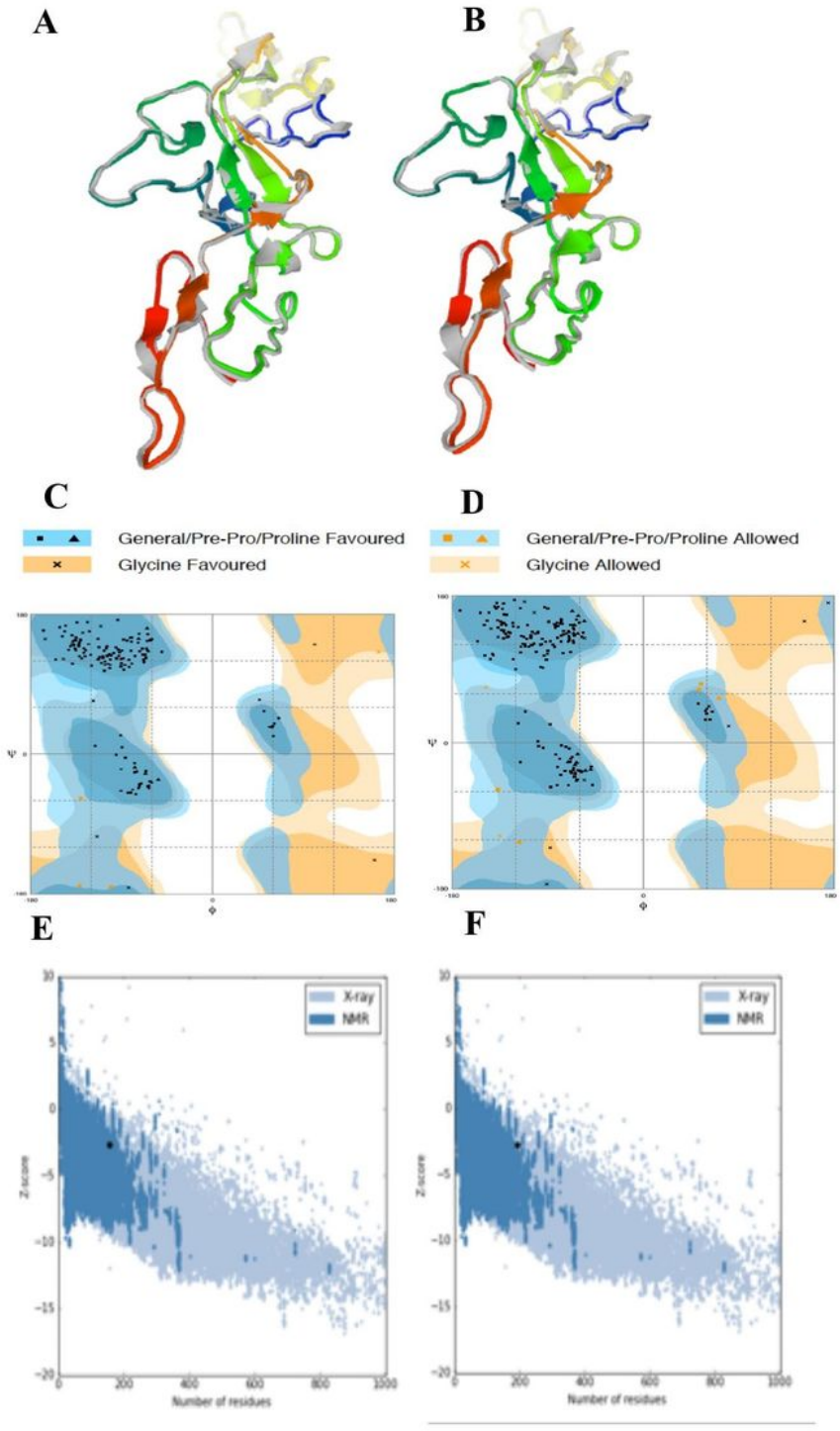

Figure 2

3D-modeling, refinement and validation of multi epitope vaccine constructs. a) Refined 3D structure of VC1 construct b) Refined 3-D structure of VC2 construct c) Ramachandran plot analysis of refined VC1 structure d) Ramachandran plot analysis of refined VC2 structure e) ProSA-web shows Z-score of -2.74 for VC1 and f) -2.66 for VC2.

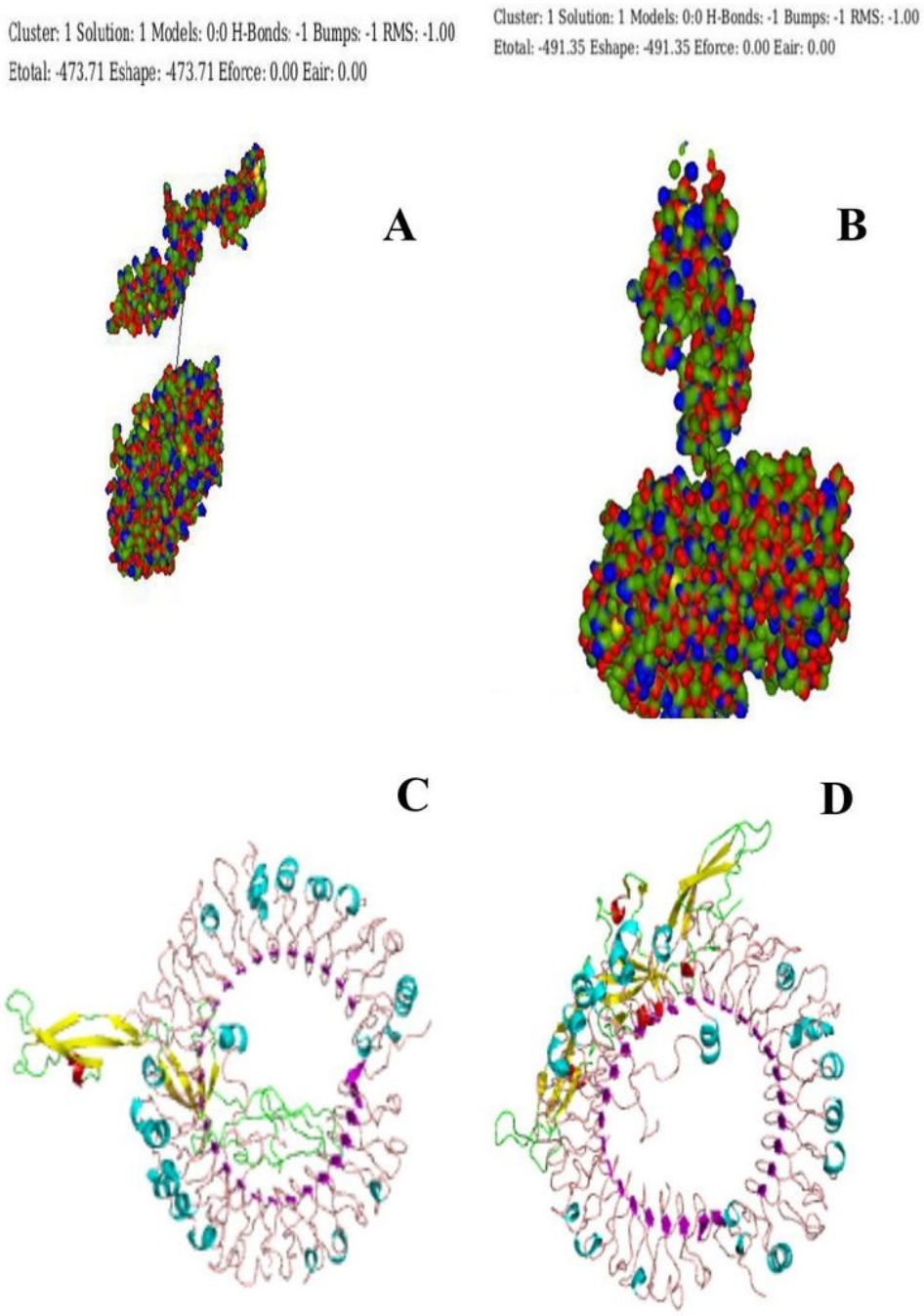

Figure 3

Identification of binding energy by Hex and Cluspro docking tools A) Hex docked complex of VC1 and TLR8 B) VC2 and TLR8 docking confirmation C) Cluspro best model of VC1 and TLR8 complex D) VC2 and TLR8 complex (Ligand-Yellow color; Receptorbrown color).

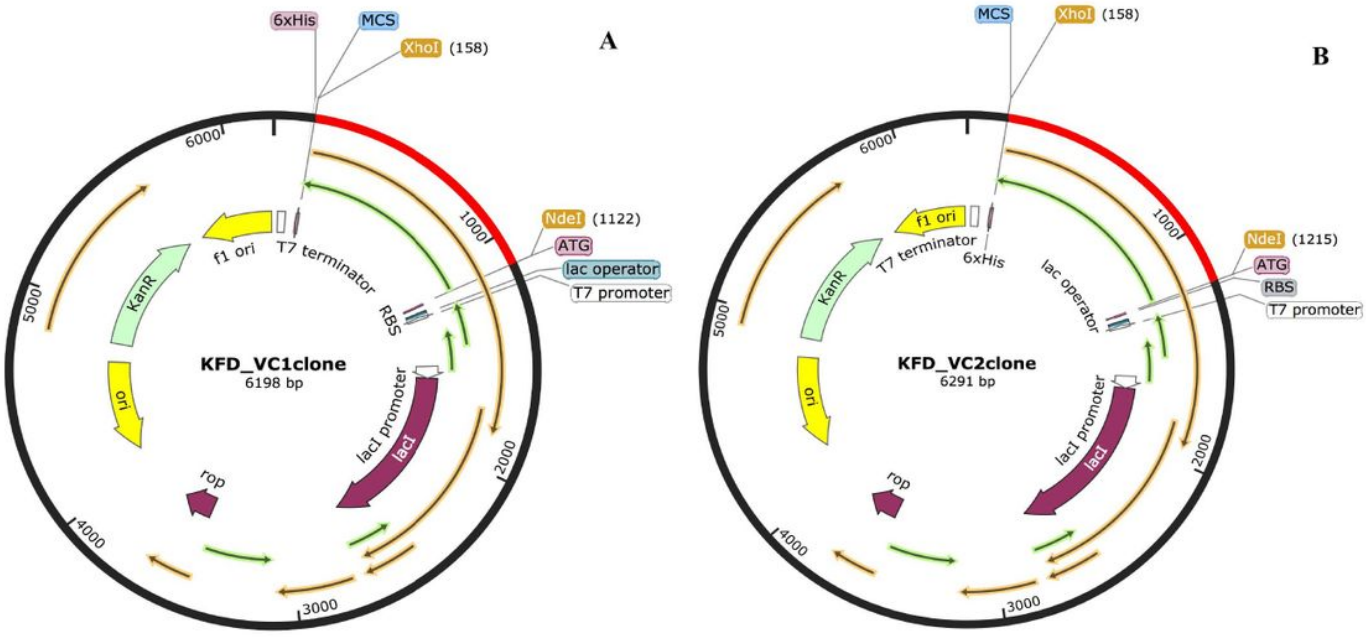

Figure 4

In silico restriction cloning of multiepitope subunit vaccine sequence into pET30a(+) expression vector, red color part represent vaccine sequence and black circle represent vector sequence. Restriction sites are heighted in yellow color and translation sequence parts were mention in green color arrows.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryinformation.pdf