Computational vaccinology based development of multi-epitope subunit vaccine for protection against the Norovirus' infections

Irfan Ahmad[1#], Syed Shujait Ali[1#], Ismail Shah[1], Shahzeb Khan[1], Mazhar Khan[5], Saif Ullah[1], Shahid Ali[1], Jafar Khan[1], Mohammad Ali[1], Abbas Khan[2*], Dong-Qing Wei[2*]

[1]Centre for Biotechnology and Microbiology, University of Swat, Swat, Khyber Pakhtunkhwa, Pakistan.

[2]Department of Bioinformatics and Biological statistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, P.R China.

[3]State Key Laboratory of Microbial Metabolism, Shanghai-Islamabad-Belgrade Joint Innovation Center on Antibacterial Resistances, Joint Laboratory of International Cooperation in Metabolic and Developmental Sciences, Ministry of Education and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200030, P.R. China

[4]Peng Cheng Laboratory, Vanke Cloud City Phase I Building 8, Xili Street, Nashan District, Shenzhen, Guangdong, 518055, P.R China.

[5]The CAS Key Laboratory of Innate Immunity and Chronic Diseases, Hefei National Laboratory for Physical Sciences at Microscale, School of Life Sciences, CAS Center for Excellence in Molecular Cell Science, University of Science and Technology of China (USTC), Collaborative Innovation Center of Genetics and Development, Hefei, 230027, Anhui, China.

Irfan Ahmad
irfanmicrouos@gmail.com
Syed Shujait Ali, PhD
shujaitswati@uswat.edu.pk
Shahzeb Khan
khanshahxeb173@gmail.com
Mazhar Khan
mazharsw@mail.ustc.edu.cn
Abbas Khan(corresponding author)
abbaskhan@sjtu.edu.cn
Saif Ullah
Usaif7454@gmail.com
Shahid Ali, PhD
shahidali@uswat.edu.pk
Jafar Khan, PhD
Jafar@uswat.edu.pk
Mohammad Ali, PhD
alimoh@uswat.edu.pk
Dong-Qing Wei, PhD (corresponding author)
dqwei@sjtu.edu.cn
[#]**CONTRIBUTED EQUALLY**

Abstract

Human Norovirus belong to family *Calciviridae*, it was identified in the outbreak of gastroenteritis in Norwalk, due to its seasonal prevalence known as “winter vomiting disease”. Treatment of Norovirus infection is still mysterious because there is no effective antiviral drugs or vaccine developed to protect against the infection, to eradicate the infection an effective vaccine should be developed. In this study capsid protein (A7YK10), small protein (A7YK11) and polyprotein (A7YK09) were utilized. These proteins were subjected to B and T cell epitopes prediction by using reliable immunoinformatics tools. The antigenic and non-allergenic epitopes were selected for subunit vaccine, which can activate cellular and humoral immune responses. Linkers joined these epitopes together. The vaccine structure was modelled and validated by using Errat, ProSA and rampage servers. The modelled vaccine was docked with TLR-7. Stability of the docked complex was evaluated by MD simulation. In order to apply the concept in a wet lab, the reverse translated vaccine sequence was cloned in pET28a (+). The vaccine developed in this study requires experimental validation to ensure its effectiveness against the disease.

Introduction

Norovirus (NVs), also known as "small round-structured virus" and this RNA single-stranded virus is placed within family *Calciviridae.* NVs is responsible for the spread of non-bacterial human gastroenteritis (1). The Human norovirus (HuNV) was first identified in stool specimen, during gastroenteritis outbreak in Norwalk and was coined as Norwalk virus. More than a thousand strains of NVs are isolated which are genetically and serologically different. The infected person has an abdominal cramp, stomach pain, diarrhoea, vomiting and nausea with mild pyrexia (2). The consumption of contaminated food and water is deemed essential for the development and spread of disease (3), (4), and globally 20 % of all diarrheal diseases are caused by HuNV, and nearly 21200 victims succumbed to death annually (5, 6). HuNV interferes with interferon type I & III by influencing MHC-I expression and causing rapid infection. MHC-I plays a key role in providing immunity against viruses. In this process, proteasome-mediated degraded peptides are presented to the CD8+ T-cells for evoking immune reactions (7). The genome of HuNV is 7.5 kb, which consists of three open reading frames (ORF's), ORF1, ORF2, and ORF3. These ORF's (ORF1, ORF2, and ORF3) codes for a nonstructural protein, VP1 major capsid protein, and VP2 minor capsid proteins respectively (8).

In clinical samples, an electron microscope (EM) (9) is used as a diagnostic tool for norovirus identification. ELISA (enzyme-linked immunosorbent assay) and molecular techniques are accessible for the diagnostic purpose of pathogens including norovirus. EM is found in the well-equipped lab and it is used to look for the pathogenic particles in the feces. RT-PCR (10) shows sensitivity for identification, and it also assist in comprehension of these viruses molecular diversity (11-15). Genomic characterization and molecular diversity is assessed by (HMA) hetero duplex mobility assay of various viruses which include; Norovirus (16), measles virus (17), polioviruses (18), hepatitis C virus (19) and polioviruses (18). Cross challenge studies and IEM (immune electron microscopy) studies (20) was previously utilize for NV antigenic diversity assessment, before the development of rNV (recombinant-NV) capsid protein.

Non-bacterial gastroenteritis is still a great challenge, and there is no effective licensed vaccine available for its treatment (7). Researchers are trying their best to launch an effective vaccine against norovirus, however, their investigations are either in clinical trials or in pre-clinical stages (21). These investigations may results in two norovirus vaccines in future, bivalent GI.1/GII.4 intramuscular VLP vaccines (in phase II b clinical trial), and monovalent GI.1 oral pill recombinant adenovirus vaccine (phase I trials) (7). NoV infection is equally pervasive in developing as well as developed countries. Children and elders are severely affected by these infections. Effective vaccines and drug designing would be instrumental in controlling the higher mortality rates caused by these infections (6). Vaccines evoke innate and cellular immune responses to develop antibodies and memory cells, which may provide long-lasting protection from specific serotypes (22). In-silico study based on immunoinformatics approaches was applied to pinpoint effective epitopes or hits as a potential candidate for vaccine or drug designing (23-27). Computational approaches are beneficial to predict the antigen without culturing the pathogenic strain experimentally (28-32).

In this study, computational analysis was performed to predict the epitopes based effective vaccine against the NoV. T-cell, B-cell and HTL epitopes were predicted and analyzed using defined criteria for selecting the potential epitopes for final vaccine construct. Further implementation of molecular docking with TLR receptors, molecular dynamics simulation and codon optimization for expression confirmed the potential of the final vaccine construct. Thus, this study provides a way towards the development of a potential vaccine candidate against the Norovirus.

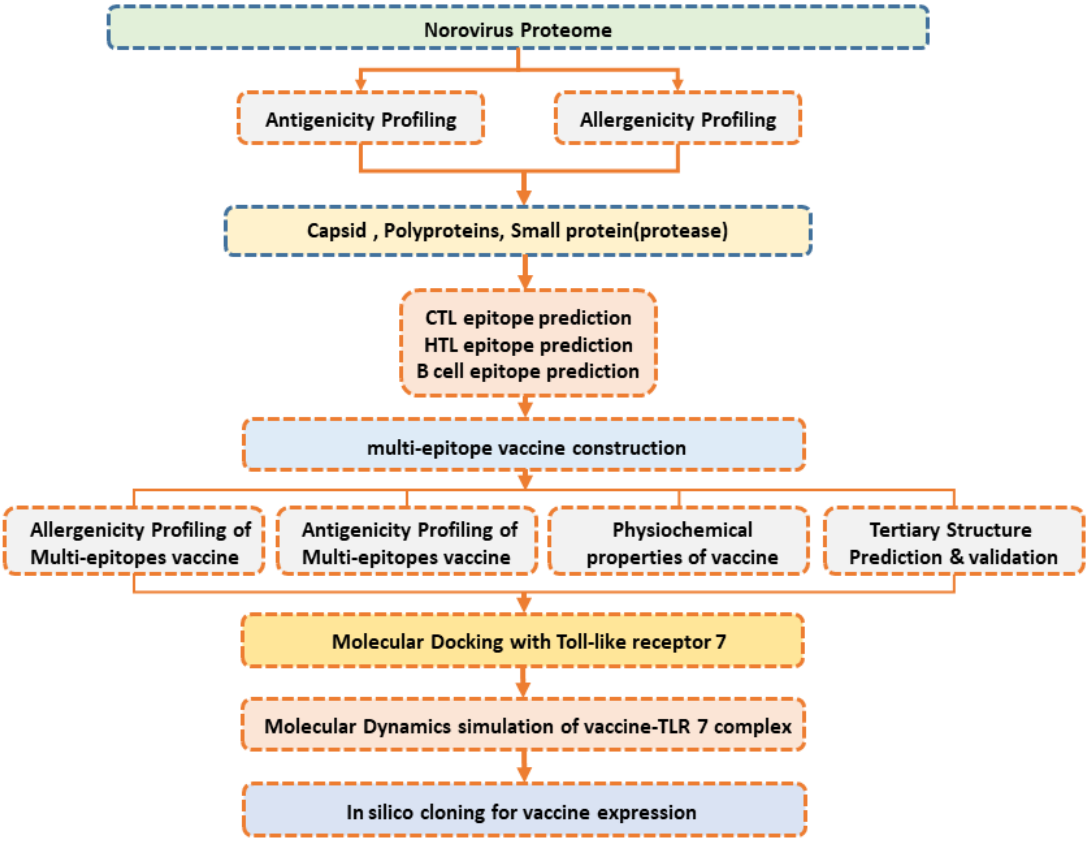


Figure. 1. Schematic representation of the steps involved in epitope-based vaccine designing. A multi-step approach was used to construct the final vaccine candidate. Finally, validation through MD simulation and in silico expression was achieved.

Results

Protein collection

The amino acid sequences of capsid protein (UniProt id: A7YK10), polyprotein (UniProt id: A7YK09) and small protein (A7YK11) of *Norovirus* as well as Beta-defensin 3 (Q5U7J2), an adjuvant was retrieved from UniProtKB in FASTA format. These protein were antigenic in nature based on antigenicity score of 0.48, 0.52, and 0.49 respectively as calculated by VaxiJen server and were selected for the designing of a multi-epitope vaccine by immunoinformatics approach.

CTL (Cytotoxic T Lymphocytes) epitopes prediction

NetCTL1.2 server predicted a total of 51 CTL epitopes of 9-mer in length. In these epitopes, six non-allergenic epitopes **(Table 1)** were selected for vaccine designing based on high binding affinity score. Based on the predicted scores, two epitopes were selected from A7YK09, A7YK10 and A7YK11 each.

Table 1. Selected CTL epitopes. All the epitopes are based on their scores.

Proteins	Sequence ID	Peptide	Affinity	affinity rescale	Cleavage	TAP	Combine score	MHC binding Epitope
A7YK10	469	QSDALLIRY	0.7635	3.2418	0.9510	2.8490	3.5269	Yes
	91	YLAHLSAMY	0.5251	2.2294	0.8881	2.7960	2.5024	Yes
A7YK09	1044	TSSGDFLKY	0.5906	2.5076	0.9717	2.9870	2.8027	Yes
	1513	MQESEFSFY	0.5161	2.1913	0.5334	2.9500	2.4188	Yes
A7YK11	122	AVDWSGTRY	0.5730	2.4330	0.9756	3.1630	2.7375	Yes
	140	FSGGFTPSY	0.3257	1.3828	0.9690	2.6770	1.6620	Yes

HTL (Helper T lymphocytes) prediction

MHC-II prediction module of IEDB was used for Helper T Lymphocytes (HTL) epitopes prediction for HLA-DRB1*01:01, HLA-DRB1*01:02, HLA-DRB1*01:03, HLA-DRB1*01:04, and HLA-DRB1*01:05 Human alleles.. 9 HTL epitopes with the highest binding affinity were selected. The selected epitopes are situated at position 145-159,373-387,117-131 (capsid protein), 1-15, 1631-1645, 816-830, 617-631, 1089-1103 and 193-207 (polyproteins). (Table 2 represent selected HTL epitopes).

Table 2: HTL selected epitopes. The epitopes are based on percentile rank.

Proteins	Allele	Start	End	Peptide	Method	Percentile rank
Capsid protein	HLA-DRB1*01:02	145	159	PHVMCDVRALEPIQL	sturniolo	0.02
	HLA-DRB1*01:03	373	387	KVYASLAAAAPLDLV	NetMHCIIpan	0.20
	HLA-DRB1*01:02	117	131	FTAGKVVVALVPPYF	sturniolo	0.44
Polyprotein	HLA-DRB1*01:02	1	15	MRMATPSSASSVRNT	sturniolo	0.12
	HLA-DRB1*01:01	1631	1645	GEKYYRTVASRVSKE	Consensus (comb.lib./smm/nn)	0.19
	HLA-DRB1*01:03	816	830	ITSILQAAGTAFSIY	NetMHCIIpan	0.25
	HLA-DRB1*01:02	617	631	CRRIDFLVYAESPVV	sturniolo	0.32
	HLA-DRB1*01:03	1089	1103	LALAVRMGSQAAIKI	NetMHCIIpan	0.51
	HLA-DRB1*01:02	193	207	PLDPAELRKCVGMTV	sturniolo	0.59

Interferon production

Online server IFN-epitope was used to identify the HTL epitope with a potential to induce the release of IFN-gamma from T (CD4+) cell. This analysis resulted in three epitopes (out of nine HTL epitopes) with the ability to induce T cells for interferon production **(Table 3)**.

Table 3. Selected HTL epitopes analysis for interferon production.

Serial No	Sequence	Method	Result	Score
1	PHVMCDVRALEPIQL	MERCI	NEGATIVE	1
2	GEKYYRTVASRVSKE	MERCI	NEGATIVE	1
3	ITSILQAAGTAFSIY	MERCI	NEGATIVE	1
4	CRRIDFLVYAESPVV	MERCI	NEGATIVE	1
5	KVYASLAAAAPLDLV	SVM	POSITIVE	0.32853096
6	FTAGKVVVALVPPYF	SVM	NEGATIVE	-0.37555421
7	MRMATPSSASSVRNT	SVM	POSITIVE	0.98957435
8	LALAVRMGSQAAIKI	SVM	POSITIVE	0.087092709
9	PLDPAELRKCVGMTV	SVM	NEGATIVE	-0.024602308

B-cell epitope prediction for norovirus

B-cell epitope prediction was performed on ABCpreds server and six epitopes of 20-mer in length with a score higher than 0.83 were selected for further analysis **(Figure 3A)**. Conformational B-cell epitopes were identified on Discotope 2.0 and out of 359 residues, 50 B-cell residues were predicted. The predicted B-cell Linear and discontinuous epitopes are given in **Table 4** and **5**.

Table 4: Predicted Linear B cell epitopes with their respective scores.

Rank	Sequence	Score	Start position
1	GPGGEKYYRTVASRVSKEGP	**0.91**	222
2	VMCDVRALEPIQLGPGPGHV	**0.89**	127
3	GNSISTGPGPGGGGLMGIIG	**0.87**	334
4	LEPIQLPLLDGPGPGMRMAT	**0.86**	190
5	PGPGPGVMCDVRALEPIQLP	**0.86**	159
6	PGPGITSILQAAGTAFSIYG	**0.84**	241

Table 5. Predicted discontinuous B-cell epitopes residues with their contact number and Discotope score.

S.NO	Residues	Contact number	Discotope score
1	LYS, TYR, PRO, GLN, ASN, GLY, PRO, GLY, GLY, GLU	3, 6, 4, 0, 4, 1, 0, 12, 2, 4	-2.885, -2.029, -3.589 -3.886, -1.879, 1.667 2.652, 3.693, 2.327, 0.024
2	LYS, PRO, TYR, GLY, PRO, GLY, PRO, GLY, HIS, GLY	8, 9, 11,13, 3, 13, 5, 0, 0, 6	-0.356, -2.670, -2.082, 1.008, 2.997, 4.536, 5.611, 5.290, 5.382, 5.370

3	GLU, LYS, TYR, TYR, ARG, THR, ALA, LYS, PRO, GLY	13, 9, 18, 32, 6, 3, 7, 0, 8, 4	5.690,5.180, 4.948, 2.551, 1.623, 0.693,-2.447, -3.572, -2.696, -2.886
4	GLY, ILE, ILE, GLY, ASN,SER,ILE,GLY,PRO,GLY	4, 35, 7, 30, 3, 16, 3, 4, 13, 36	-1.302, 0.705, 2.215, 3.458, 5.151, 4.587, 2.787, 2.029, 0.101, 0.667
5	PRO, GLY, PRO, GLY, PRO, GLY, ILE, ASN, GLY, GLY	11, 28, 7, 0, 7, 12, 6, 0, 16, 15	-0.292, 0.722, 1.126, 1.693, 1.776, 1.315, -0.535, -2.624, 0.504, -2.062

Final multi-epitope vaccine construct

The final multi-epitope vaccine construct was composed of 6 CTL and 9 HTL epitopes selected based on high binding affinity scores. AAY linkers were used to combine CTL epitopes and GPGPG linkers joined HTL epitopes, Whereas EAAAK linker was used for attachment of adjuvant to the N-terminal of vaccine, which amplifies its function.

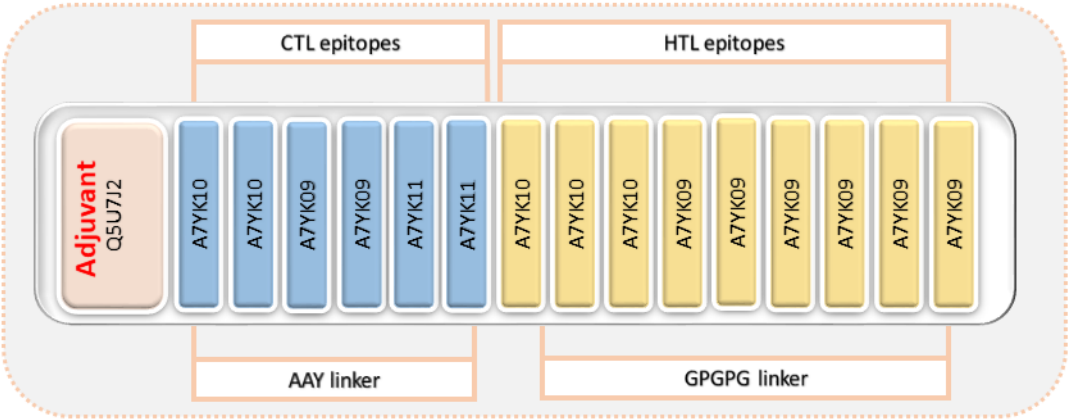


Figure. 2. Final vaccine construct. Six different CTL epitopes while nine HTL epitopes from three different proteins were combined to construct the multi-epitope subunit vaccine using linkers. An adjuvant to the N-terminal has also been added.

Prediction and validation of tertiary structure

3D structure of the vaccine was predicted on Robetta server. Five models were generated, and after evaluation model three (**Figure 3B**) was selected for further analysis. The selected model was validated by RAMPAGE, ProSA-web, and ERRAT (**Figure 4**). Modeled protein Ramachandran plot analysis revealed that most favored regions, additionally allowed regions, generously allowed region and disallowed region contain 81.1%, 6.8%, 1.2% and 0.4% of residues, respectively. ProSA-web and ERRAT were used for the evaluation of potential and quality errors in 3D crude model. Modeled protein overall quality factor was found 69.2% utilizing ERRAT. ProSA-web is utilized for prediction of Z-score prediction, which is found as -4.7.

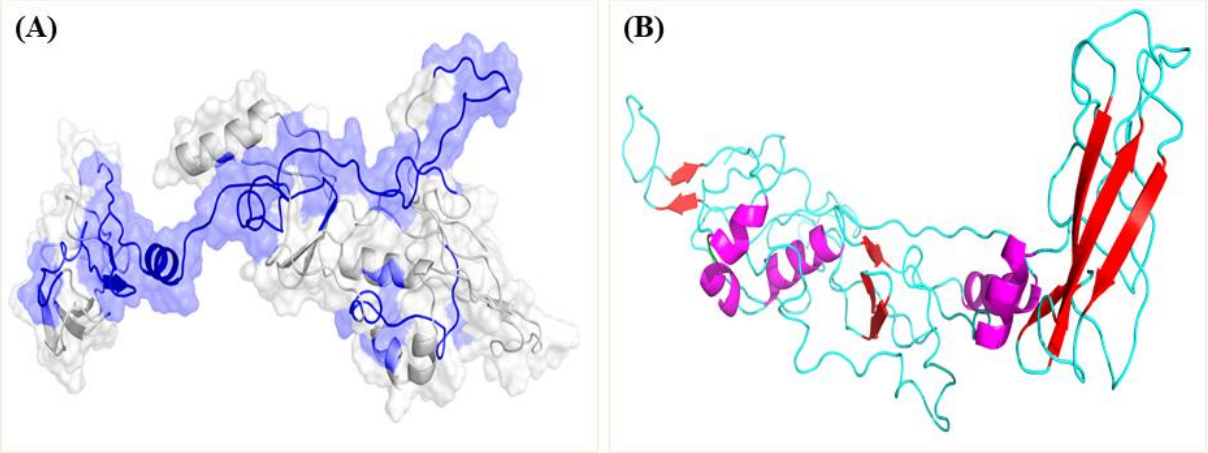


Figure 3: Predicted structure of the vaccine. **(A)** Showing the B-cell epitope (blue) (B) Showing the final structure of multi-epitopes vaccine construct showing loops (cyan), Helix (magenta) and beta-sheets(red).

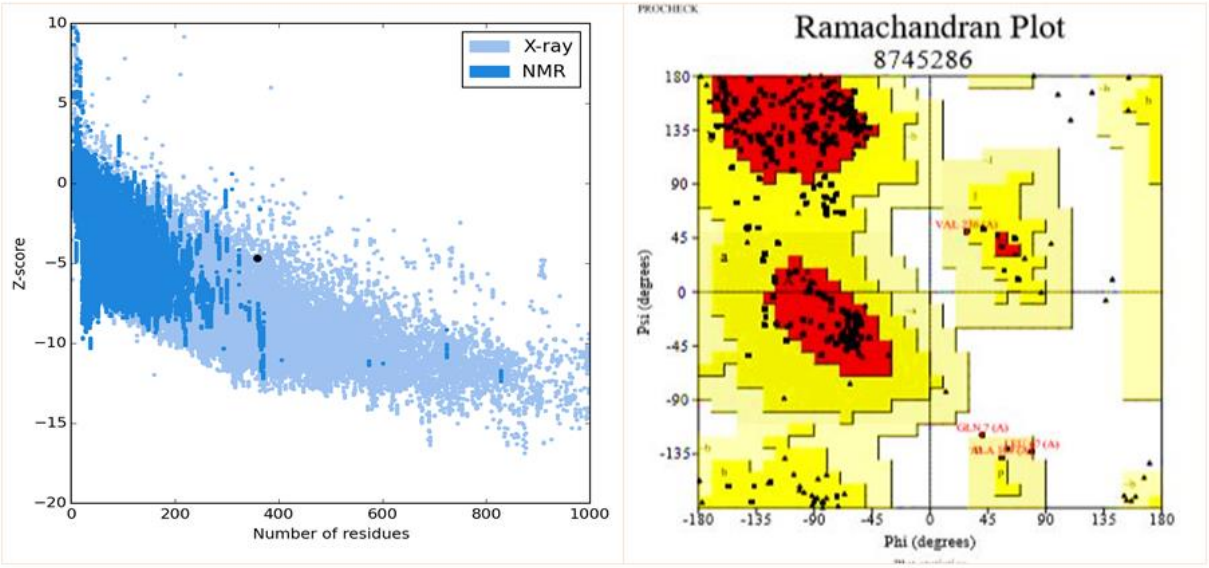


Figure 4: Validation of 3D final vaccine model. **(A)** PROSA showing Z-score (-4.7) for 3D structure validation **(B)** In Ramachandran analysis residues were allocated; most favoured region 81.1%, allowed 6.8%, generously allowed 1.2% and disallowed region 0.4% residues.

Antigenicity, allergenicity and physiochemical parameter prediction of final vaccine construct

The antigenicity of the final vaccine construct was predicted on VaxiJen and ANTIGENpro servers by selecting bacteria model at 0.4% threshold. The antigenicity scores predicted by VaxiJen and ANTIGENpro are 0.8134% and 0.635949%, respectively, which indicates the antigenic nature of the final vaccine. For allergenicity, AlgPred was used and -0.62811 score indicating the non-allergen nature of vaccine using the default threshold of -0.40. The physiochemical parameters including molecular weight (MW) and theoretical isoelectric point value (PI) of vaccine were found 36.56 kDa and 9.17, respectively, as predicted by ProtParam. The PI value (9.17) suggesting the basic nature of the vaccine. Moreover, the half-life in mammalian reticulocytes, yeast, and *E. Coli* was found 30 hours (in vitro); 20 hours (in vivo), and 10 hours (in vivo), respectively. The instability index score of 34.03 refers to the stable nature of the protein. The GRAVY (Grand average of hydropathicity) and the aliphatic index was found 75.88 and -0.040, respectively.

Prediction of secondary structure

Secondary structure for vaccine predicted by PSIPRED program suggests the presence of 20.89% alpha-helix, 16.71% beta-strand, and 62.39 % coil as shown in **Figure 5**.

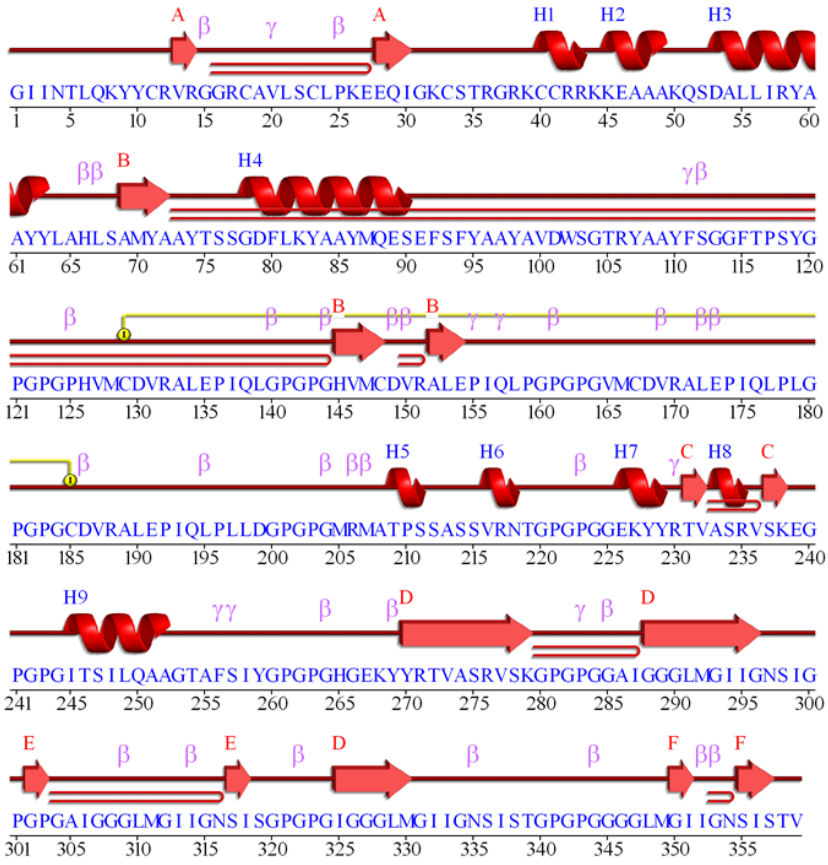


Figure 5: Represent the secondary structure of the final vaccine construct in which alpha helix, beta-strands and coils were identified (20.89%, 16.71% and 62.39%).

Subunit vaccine molecular docking with an immune receptor (TLR-7)

For docking of TLR-7 was docked with the multi-epitopes vaccine, using an online server ZDOCK. Overall, ten complexes were generated and the most suitable vaccine-TLR complex was selected based on correct conformation and binding **(Figure 6)**. The PDBsum server reported 90 interface residues with one salt bridge and eight hydrogen bonds were reported between the vaccine and TLR-7.

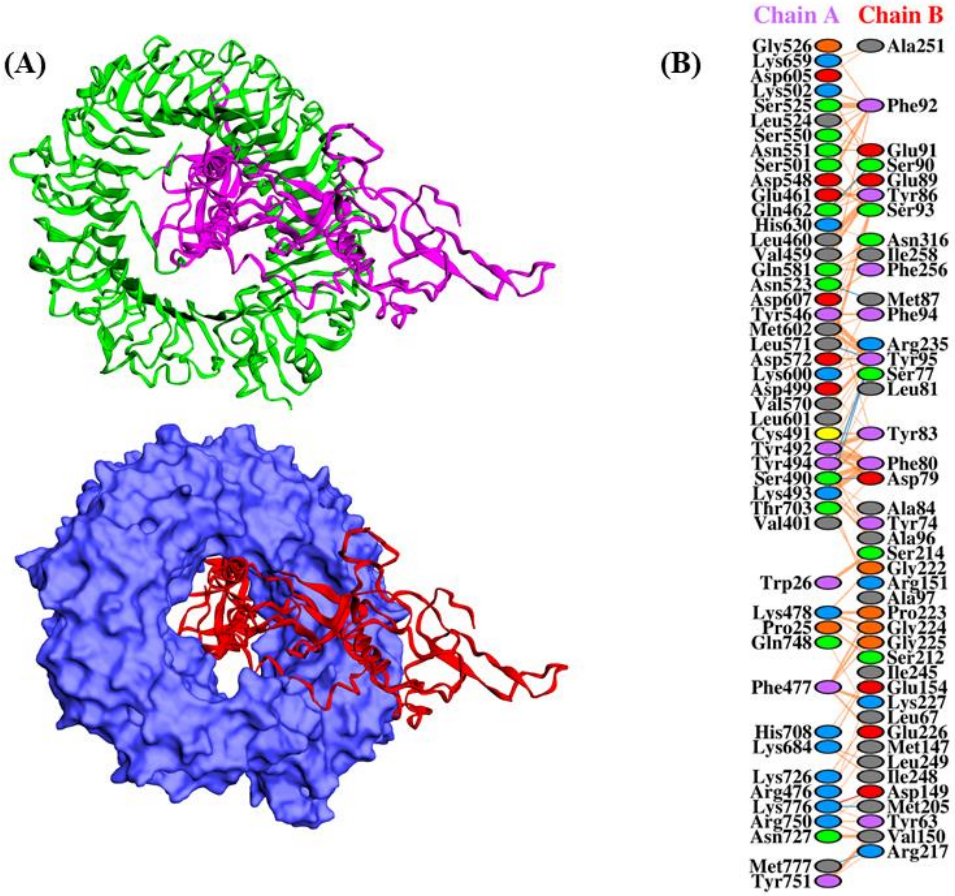


Figure 6: (A) TLR-7 (PDB ID: 5GMF) and vaccine docked complex. Magenta colour represents the vaccine while the green and blue (surface) colour represents receptor. **(B)** Showing the interaction pattern of the TLR-7 and vaccine construct.

MD simulation of immune receptor-vaccine complex

Molecular dynamics (MD) simulation was performed to check the stability and fluctuation of vaccine construct and TLR-7 complex. The computed RMSD and RMSF for the vaccine (protein) and its side chain as well as their graph are shown in **Figure 7**. The RMSD and RMSF of protein and side-chain residues respectively were checked at 30ns time to estimate the stability of the system. Overall fluctuation (RMSD) rate for the simulated system was found 4Å (TLR-7) and RMSF (residual fluctuation) for maximum residues were found in acceptable range while some residues exhibit higher fluctuation.

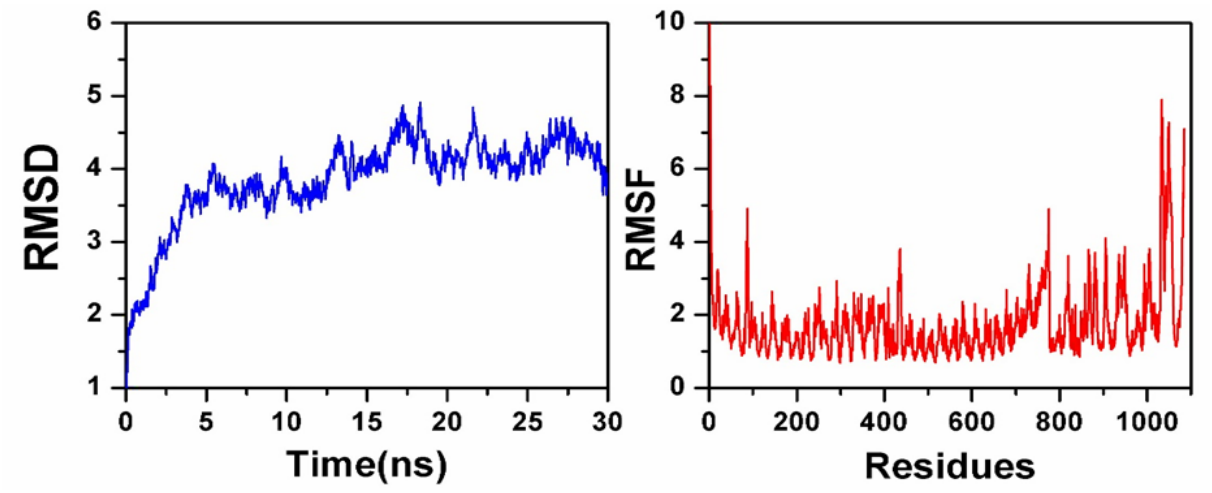


Figure 7: Molecular dynamics simulation of the receptor-vaccine complex. The left graph is showing the RMSD of the complex (X-axis = Time in ns and Y-axis = RMSD) while the right graph is showing the RMSF of the complex (X-axis = Time in residue and Y-axis = RMSF).

Codon optimization and in silico cloning

To assure the maximal expression of the protein, vaccine codon was optimized in *E. coli* (strain k12) using Java Codon Adoption Tool. The optimized length of the codon sequence is 1077 nucleotides. The average GC content was found 56.7% (optimum range 30%-70%) and CAI (codon adaption index) was 0.968, which indicates possibilities of good expression in host *E. coli*. Finally, using restriction enzymes restriction clone was formed and adapted codon sequence was inserted in pET28a (+) plasmid. The designed construct is shown in **Figure 8**.

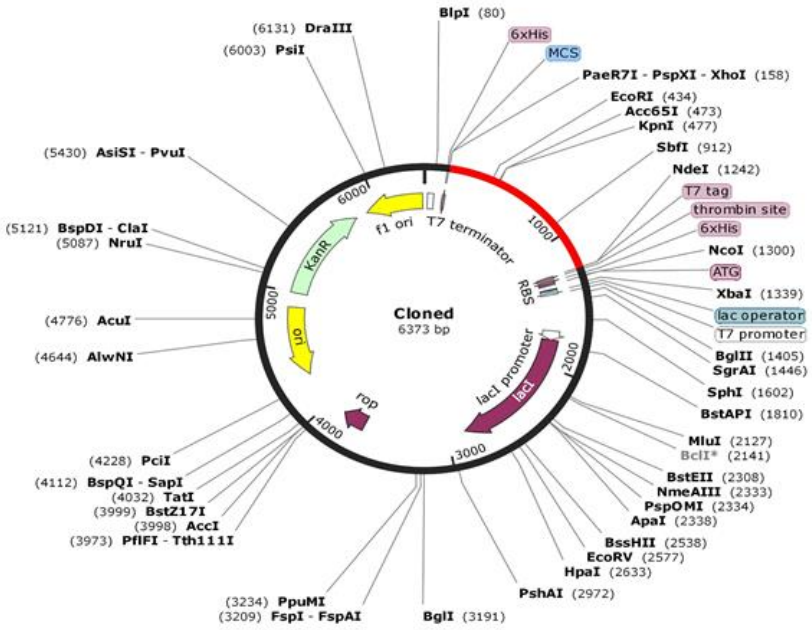


Figure 8: In silico restriction cloning of the final vaccine construct into pET28a (+) expression vector where Red part representing the vaccine insert and black circle showing the vector.

Discussion

Norovirus proteins such as capsid, polyprotein, and protease were found antigenic and are vital for infection and replication within the host. Therefore, they are considered very important for subunit vaccine analysis. Immunization or vaccination is universally recognized method to eradicate or control the infection. The advancement in computational approaches and their applications to biological research ushered a new era of subunit vaccination designing in which the most accurate or exact antigenic part is identified and used as an immunization tool instead of the whole pathogen. The fast accumulation of vast genomic and proteomic data of pathogens including norovirus is helpful in the development of epitope-based effective vaccines to control and eradicate the diseases caused by various pathogens. Computational approaches predicted epitopes (CTL & HTL) based on norovirus proteins and validation scores of these epitopes suggest their use for subunit vaccine construction. MHC (major histocompatibility complex) is of different types, MHC-I carry a peptide of 9-mer to the surface of the cell and act like an impulsive signal for cytotoxic T cell. Which lead to cell destruction by activating immune complementary cascade. MHC-II molecules present peptide of 15-mer to Helper T lymphocytes. Our final subunit vaccine is composed of high-affinity CTL and HTL epitopes to elicit immunity. The allergenicity and antigenicity values of the vaccine were calculated, and these values indicate the non-allergen nature of the vaccine and are capable to provoke immune response due to antigenic nature.

In addition to these epitopes, B cell linear epitopes were also predicted which help in B-cell maturation, in order to produce antibodies. Physiochemical properties of vaccines such as molecular weight, theoretical PI, aliphatic array and thermal stability were calculated. The molecular weight of the vaccine was 36.560kDa, which is a suitable range for subunit vaccine; the theoretical PI score is 9.17, which indicates that vaccine is basic in nature. The aliphatic array suggests that vaccines have aliphatic side chains and instability index endorses the thermally stable nature of the vaccine. For prediction and analysis of the secondary structure of vaccine, PSIPRED V3.3 was used, which indicates the presence of 20.89% alpha-helix, 16.71% beta-strand, and 62.39 % coil. Besides, the 3D structure obtained by homology modeling work comprises of adequate information on the spatial arrangement of such essential protein residues as well as useful guidance in the study of protein normal function, dynamics, and interaction with ligand as well as other proteins. To pinpoint the error in the final 3D structure of vaccine different structural validation tools were used to detect errors. From the main Ramachandran plot, it was found that the overall model is satisfactory because most residues were found in the most-favour region while few were present in the disordered region.

Furthermore, the vaccine was docked with TLR-7 in order to understand the immune response towards vaccine final structure. Energy minimization was conducted to minimize the potential energy of the entire system for the overall conformational stability of the docked vaccine protein-

TLR-7. Energy minimization repairs the structure's unnecessary topology by ditching certain protein atoms and thus forms a more relatively stable structure with adequate stereochemistry is thus formed.

To obtain maximum expression (transcription and translation) of vaccine protein in the host (*E. coli* strain k-12), codon optimization was accomplished by CAI (codon adaptation index). The solubility of overexpressed recombinant protein in the host (*E. coli*) is one of the crucial requirements of many biochemical processes. Our vaccine protein shows a suitable proportion of solubility in the host. The essential goals of many mechanical and biomedical applications are to strengthen the protein. The newly designed vaccine was very effective and showing good immunogenic response in animals, however, when these vaccines were applied to humans the response was not similar due to the complexity of human immunopathology. In this study, modern immunoinformatics approaches were incorporated to develop new thermostable, cheap and effective subunit vaccines. These vaccines are safe and immunogenic to exploit the immune system to provide protection from norovirus infection.

Conclusion

In this study, the main focus was to apply, *in-silico* approaches to design an effective multi-epitope vaccine against norovirus based on three proteins due to their antigenic nature. To design a vaccine, B and T cell epitopes were predicted, which is due to the presentation of pathogen epitope by MHC-I and II. Suitable linkers were used to fuse these epitopes. Vaccine tertiary structure was predicted and validated to ensure the functionality of the vaccine. The physicochemical properties such as antigenicity, allergenicity, and stability were computed. The vaccine was docked with TLR-7 to check the vaccine affinity towards the receptor.

Finally, for codon optimization, the protein was reversely translated which ensure maximum expression of the vaccine in the host (*E. coli*). A wet lab experimental validation is needed to assure the activity of constructed vaccine. This study can help in controlling norovirus infection.

Methodology

Collection of *Norovirus* proteins for vaccine preparation

The three proteins capsid, polyproteins, and small basic proteins were collected from UniProtKB (https://www.uniprot.org) (33). These proteins where found antigenic through VaxiJen server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) and they were considered for vaccine designing (34-36).

Prediction of the (CTL) epitope

NetCTL 1.2 is an online server (http://www.cbs.dtu.dk/services/NetCTL/) in which CTL epitopes were predicted against these proteins (37). Prediction of these CTL epitopes is based upon three essential parameters, including peptide attached to MHC-proteasomal C-terminal degradation activity, TAP (Transporter Associated with Antigen Processing) delivery accuracy and I. Artificial neural network was used to predict the attachment of peptide to MHC-I and

proteasomal C-terminal degradation while TAP deliverance score was predicted by the weight matrix. For CTL epitope identification threshold was set as 0.75.

Prediction of helper T-cell epitopes

IEDB (Immune Epitope Database) is an online server (http://tools.iedb.org/mhcii/), which is employed for prediction of Helper T cells lymphocytes of 15-mer length for five human alleles (HLA-DRB1*01:01, HLA-DRB1*01:02, HLA-DRB1*01:03, HLA-DRB1*01:04, HLA-DRB1*01:05) (38-40). This server speculates epitopes based on receptor affinity, which is figured out from IC50 value (binding score) given to each epitope**.** Epitopes with higher binding affinity mostly have IC50 score <50 nM. The IC50 value determines epitopes affinity, value such is <500 nM represent moderate affinity while <5000 nM instantly show epitopes having low affinity. The affinity of epitopes is inversely related to the value of percentile rank.

Prediction of B cell lymphocytes

ABCpred is an online server (http://crdd.osdd.net/raghava/abcpred/) employed for the prediction of linear B-cell epitopes. The ABCpred predicts B-cell epitopes (linear) with a precision of 75% (0.49 Sensitivity and 0.75 specificity). Conformational B-cell epitopes prediction is carried out by DiscoTope 2.0 an online server (http://www.cbs.dtu.dk/services/DiscoTope/) (41). The process of computation in this method is based on log-odds ratios (ratio among amino acid composition in epitopes and residues of non-epitopes) and surface accessibility. The sensitivity and specificity are found 0.47 and 0.75, respectively, on the default threshold (3.7).

Multi-epitopes vaccine sequence construction

High scoring epitopes including other parameters filtered out the final epitopes from CTL and HTL. These selected epitopes were used to construct the final multi-epitopes subunit vaccine. Linker such as AAY and GPGPG interconnect these epitope sand also enhanced effective separation and presentation of epitopes. These linkers have two important roles; first, they assist binding of HLA-II epitopes and also immune processing, and second, they restrain epitopes numbers by specifying cleavage point (42-45). Further, to boost the immune responses an adjuvant (TLR-7 agonist Beta-defensin 3 (Q5U7J2)) was affixed to the vaccine construct N-terminus by EAAAK linker (46).

Subunit vaccine and interferon

Online server IFNepitope (http://crdd.osdd.net/raghava/ifnepitope/predict.php) assist in designing and predicting the amino acid sequence from proteins, which have the potential to release IFN-gamma from CD4+ T cells. This server helps in the designing of effective and better subunit vaccine by identifying peptide which binds to MHC II and releases IFN-gamma (47)

Prediction of vaccine allergenicity

For allergenicity prediction of multi-epitope subunit vaccine AlgPred (http://crdd.osdd.net/raghava/algpred/) web tool was used, in which algorithm like (IgEpitope+SVMc+MAST+ARPs BLAST) was applied (48). This hybrid prediction veracity is

about 85% at 0.4 thresholds. The server employed six distant routes to act. These approaches abet allergic protein prediction with perfection.

Vaccine antigenicity prediction

Online server ANTIGENpro (http://scratch.proteomics.ics.uci.edu/) was used for antigenicity prediction (49). Server to predict protein antigenicity used two approaches; primary protein sequence is presented multiple time, and five different machine-learning algorithms were applied to produce the result based on protein microarray data analysis.

Physiochemical parameters and identification of domain

ProtParam (http://web.expasy.org/protparam/), an online server was used for the prediction of many physiochemical parameters such as theoretical PI, instability index, amino acid composition, in vitro and in vivo half-life, molecular weight, aliphatic index and also the (GRAVY) which is a grand average of hydropathicity (50).

Prediction of secondary structure

The primary amino-acid sequence was used to predict the secondary structure of Protein by PDBsum. Sequences that shows homology to vaccine protein were selected for structure prediction. Position position-specific iterated (PSI-BLAST) identified homologous residues.

Prediction of the tertiary structure

For prediction and analysis of protein tertiary structure, an online server Robetta (http://robetta.bakerlab.org) was utilized, which is an automated tool(51). Protein sequences were submitted in FASTA format to predict 3D structure. Models were generated after parsing the structure into respective domains. This model is based either on comparative modelling or de novo structure; for comparative modelling homologs sequence was identified by BLAST, 3D-Jury or FFAS03, and PSI-BLAST, which is then used as a templet. If homologs were not found, de novo structure was generated using Rosetta fragment insertion method.

Validation of tertiary structure

To validate the predicted model is an essential step, ProSA-web an online server (https://prosa.services.came.sbg.ac.at/prosa.php) was used for validation of tertiary structure. For input structure, overall quality scoring was computed. The predicted protein will have an error when the scoring characteristic was found arbitrary then native proteins. For statistical analysis of non-bonded interaction ERRAT server was used (https://servicesn.mbi.ucla.edu/ERRAT/) (52), and Ramachandran plot analysis was performed on RAMPAGE server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) (53). This server utilized PROCHECK principle to validate the structure of protein through Ramachandran plot and separate plots for Proline and Glycine residues.

Vaccine and TLR-7 Docking

ZDOCK server (http://zdock.umassmed.edu/) was used for the docking of the final vaccine with TLR-7 (54). The ZDOCK webserver produces quick and accurate complexes. Refinement and post-processing of the complexes were subjected to Firedock server(55).

MD (Molecular dynamic simulation) of Receptor-Vaccine complex

For selected complexes, amber 14 (56) was used to conduct MD simulation. The system was neutralized and solvated with TIP3P water box. Two stages of energy minimization, followed by gentle heating and equilibration, were performed (57). After equilibration, 30ns simulation was conducted. For post-simulation trajectories analysis (RMSD and RMSF), 2.0 ps time scale was used and for trajectory, sampling using CPPTRAJ and PTRAJ(58)implemented in AMBER 14.

Codon optimization and in silico cloning of vaccine

For effective expression of the vaccine in a host (*E. coli* strain K12) was performed. Reverse translation and codon optimization were performed using JCAT (Java Codon Adaption tool). Norovirus genome expression is distinct from the expression of vector genome, codon optimization ensures maximum expression of vaccine with in the vector. In order to get the desired result, three additional options were selected, such as prokaryote ribosome binding site, restriction enzymes cleavage and rho-independent transcription termination. JCat output includes a codon adaptation index (CAI) and to ensure the high-level protein expression percentage GC content is used (25). In order to clone the desired gene in *E. coli* pET-28a (+) two restriction sites NdeI and XhoI were added to the C and N terminal of the sequence, respectively. Finally, the adapted sequence having the restriction site was inserted to pET-28a (+) to maximize the vaccine expression.

Acknowledgements:

Dong-Qing Wei is supported by the grants from the Key Research Area Grant 2016YFA0501703 of the Ministry of Science and Technology of China, the National Natural Science Foundation of China (Contract no. 61832019, 61503244), the Science and Technology Commission of Shanghai Municipality (Grant: 19430750600), the Natural Science Foundation of Henan Province (162300410060) and Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2017ZD14). The computations were partially performed at the Pengcheng Lab and the Center for High-Performance Computing, Shanghai Jiao Tong University.

529

ReferencesPrimary Sources

Secondary Sources

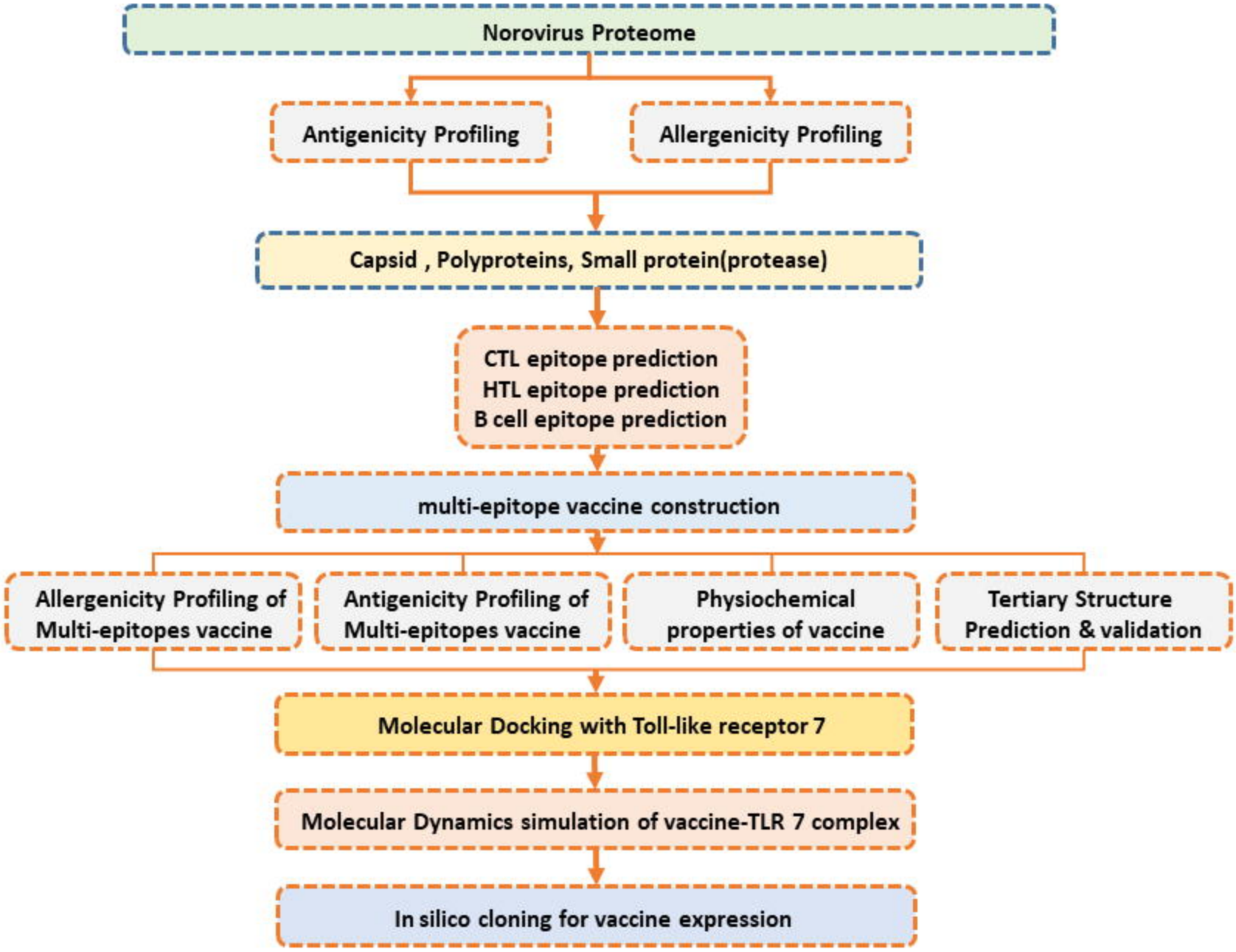
Uncategorized References

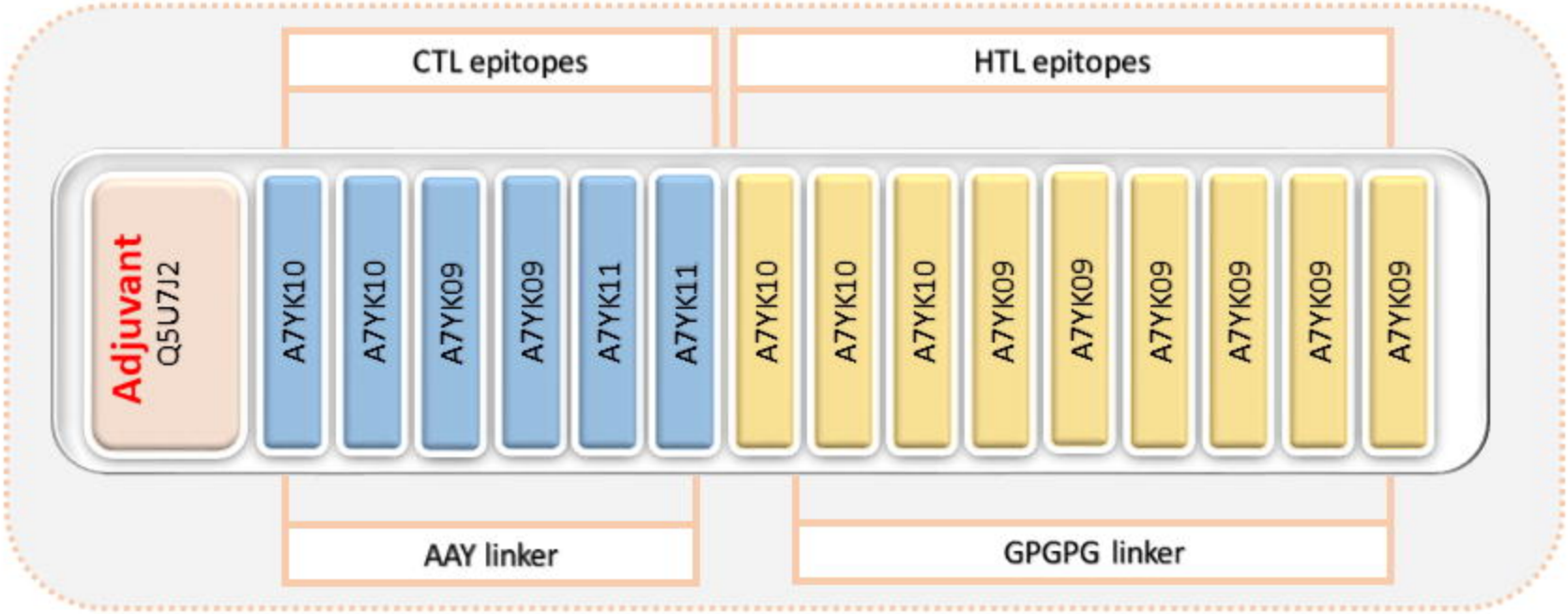
1. Lew, J. F., Kapikian, A. Z., Valdesuso, J. and Green, K. Y. (1994) title. The Journal of infectious diseases 170, 535-542.
2. Robilotti, E., Deresinski, S. and Pinsky, B. A. (2015) title. Clinical microbiology reviews 28, 134-164.
3. White, M. B., Rajagopalan, S. and Yoshikawa, T. T. (2016) title. Clinics in geriatric medicine 32, 509-522.
4. Mead, P. S., Slutsker, L., Dietz, V., McCaig, L. F., Bresee, J. S., Shapiro, C., Griffin, P. M. and Tauxe, R. V. (1999) title. Emerging infectious diseases 5, 607-625.
5. Ahmed, S. M., Hall, A. J., Robinson, A. E., Verhoef, L., Premkumar, P., Parashar, U. D., Koopmans, M. and Lopman, B. A. (2014) title. The Lancet. Infectious diseases 14, 725-730.
6. Pires, S. M., Fischer-Walker, C. L., Lanata, C. F., Devleesschauwer, B., Hall, A. J., Kirk, M. D., Duarte, A. S., Black, R. E. and Angulo, F. J. (2015) title. PloS one 10, e0142927.
7. Mattison, C. P., Cardemil, C. V. and Hall, A. J. (2018) title. Expert review of vaccines 17, 773-784.
8. McFadden, N., Bailey, D., Carrara, G., Benson, A., Chaudhry, Y., Shortland, A., Heeney, J., Yarovinsky, F., Simmonds, P., Macdonald, A. and Goodfellow, I. (2011) title. PLoS pathogens 7, e1002413.
9. Caul, E. and Appleton, H. (1982) title. Journal of medical virology 9, 257-265.
10. Ando, T., Monroe, S. S., Gentsch, J. R., Jin, Q., Lewis, D. C. and Glass, R. I. (1995) title. Journal of Clinical Microbiology 33, 64-71.
11. Green, J., Norcott, J., Lewis, D., Arnold, C. and Brown, D. (1993) title. Journal of Clinical Microbiology 31, 3007-3012.
12. Norcott, J., Green, J., Lewis, D., Estes, M., Barlow, K. and Brown, D. (1994) title. Journal of medical virology 44, 280-286.
13. Vinje, J. and Koopmans, M. P. (1996) title. Journal of Infectious Diseases 174, 610-615.
14. Fankhauser, R. L., Noel, J. S., Monroe, S. S., Ando, T. and Glass, R. I. (1998) title. The Journal of infectious diseases 178, 1571-1578.
15. Maguire, A. J., Green, J., Brown, D. W., Desselberger, U. and Gray, J. J. (1999) title. Journal of Clinical Microbiology 37, 81-89.
16. Mattick, K. L., Green, J., Punia, P., Belda, F. J., Gallimore, C. I. and Brown, D. W. (2000) title. Journal of virological methods 87, 161-169.
17. Kreis, S. and Whistler, T. (1997) title. Virus research 47, 197-203.
18. Chezzi, C. and Schoub, B. D. (1996) title. Journal of virological methods 62, 93-102.
19. Weiner, A., Houghton, M., Polito, A., Massaro, A., Bonino, F., Brunetto, M., Danielle, F., Calabrese, G., Guaschino, R. and Dinello, R. (1998) title.
20. Kapikian, A. Z., Wyatt, R. G., Dolin, R., Thornhill, T. S., Kalica, A. R. and Chanock, R. M. (1972) title. Journal of virology 10, 1075-1081.
21. Cortes-Penfield, N. W., Trautner, B. W. and Jump, R. L. P. (2017) title. Infect Dis Clin North Am 31, 673-688.
22. Pulendran, B. and Ahmed, R. (2011) title. Nature immunology 12, 509-517.
23. Dash, R., Das, R., Junaid, M., Akash, M. F. C., Islam, A. and Hosen, S. Z. (2017) title. Adv Appl Bioinform Chem 10, 11-28.
24. Adhikari, U. K., Tayebi, M. and Rahman, M. M. (2018) title. Journal of Immunology Research 2018, 22.

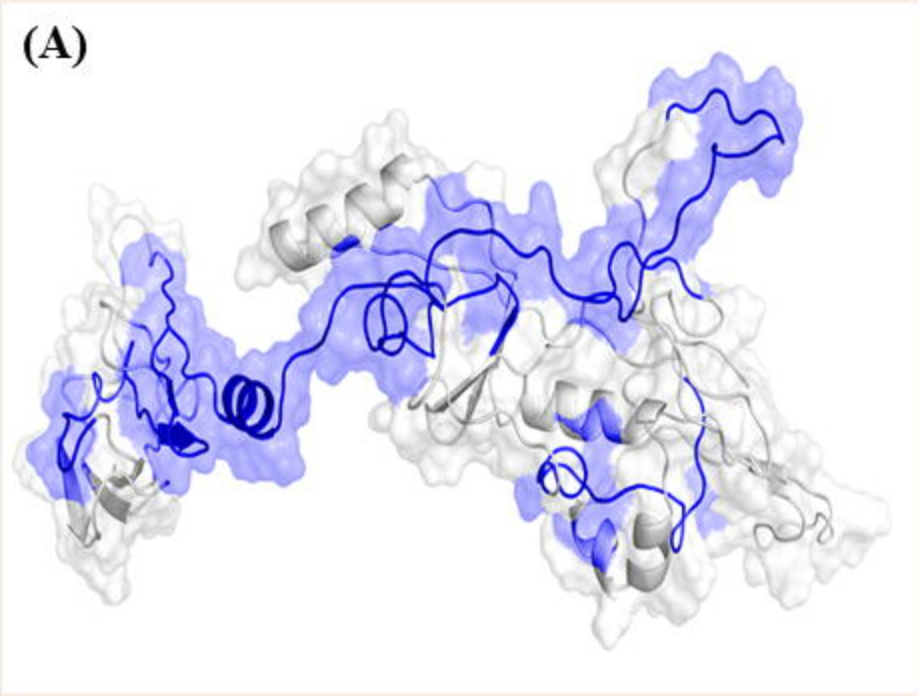
574 25. Khan, S., Khan, A., Rehman, A. U., Ahmad, I., Ullah, S., Khan, A. A., Ali, S. S., Afridi, S. G. and Wei, D.
575 Q. (2019) title. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary
576 genetics in infectious diseases 73, 390-400.
577 26. Kaliamurthi, S., Selvaraj, G., Junaid, M., Khan, A., Gu, K. and Wei, D.-Q. (2018) title. Current
578 pharmaceutical design 24, 3791-3817.
579 27. Khan, A., Junaid, M., Kaushik, A. C., Ali, A., Ali, S. S., Mehmood, A. and Wei, D.-Q. (2018) title. PloS
580 one 13.
581 28. Rappuoli, R. (2000) title. Current Opinion in Microbiology 3, 445-450.
582 29. Sette, A. and Fikes, J. (2003) title. Current opinion in immunology 15, 461-470.
583 30. Agrawal, P. and Raghava, G. P. S. (2018) title. Frontiers in microbiology 9, 2551.
584 31. Khan, M., Khan, S., Ali, A., Akbar, H., Sayaf, A. M., Khan, A. and Wei, D.-Q. (2019) title. Scientific
585 reports 9, 1-13.
586 32. Ali, A., Khan, A., Kaushik, A. C., Wang, Y., Ali, S. S., Junaid, M., Saleem, S., Cho, W. C., Mao, X. and
587 Wei, D.-Q. (2019) title. Scientific reports 9, 1-12.
588 33. Consortium, T. U. (2018) title. Nucleic Acids Research 47, D506-D515.
589 34. Doytchinova, I. A. and Flower, D. R. (2007) title. BMC bioinformatics 8, 4.
590 35. Doytchinova, I. A. and Flower, D. R. (2007) title. Vaccine 25, 856-866.
591 36. Doytchinova, I. A. and Flower, D. R. (2008) title. Open Vaccine J 1, 4.
592 37. Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O. and Nielsen, M. (2007) title. BMC
593 Bioinformatics 8, 424.
594 38. Wang, P., Sidney, J., Kim, Y., Sette, A., Lund, O., Nielsen, M. and Peters, B. (2010) title. BMC
595 Bioinformatics 11, 568.
596 39. Wang, P., Sidney, J., Dow, C., Mothe, B., Sette, A. and Peters, B. (2008) title. PLoS computational
597 biology 4, e1000048.
598 40. Nielsen, M. and Lund, O. (2009) title. BMC Bioinformatics 10, 296.
599 41. Kringelum, J. V., Lundegaard, C., Lund, O. and Nielsen, M. (2012) title. PLoS computational biology 8,
600 e1002829.
601 42. Saadi, M., Karkhah, A. and Nouri, H. R. (2017) title. Infection, Genetics and Evolution 51, 227-234.
602 43. Livingston, B., Crimi, C., Newman, M., Higashimoto, Y., Appella, E., Sidney, J. and Sette, A. (2002)
603 title. The Journal of Immunology 168, 5499-5506.
604 44. Dorosti, H., Eslami, M., Negahdaripour, M., Ghoshoon, M. B., Gholami, A., Heidari, R., Dehshahri, A.,
605 Erfani, N., Nezafat, N. and Ghasemi, Y. (2019) title. Journal of Biomolecular Structure and Dynamics 37,
606 3524-3535.
607 45. Bergmann, C. C., Yao, Q., Ho, C. K. and Buckwold, S. L. (1996) title. The Journal of Immunology 157,
608 3242-3249.
609 46. Arai, R., Ueda, H., Kitayama, A., Kamiya, N. and Nagamune, T. (2001) title. Protein engineering 14,
610 529-532.
611 47. Dhanda, S. K., Vir, P. and Raghava, G. P. S. (2013) title. Biology Direct 8, 30.
612 48. Saha, S. and Raghava, G. P. S. (2006) title. Nucleic Acids Research 34, W202-W209.
613 49. Cheng, J., Randall, A. Z., Sweredoski, M. J. and Baldi, P. (2005) title. Nucleic Acids Research 33, W72-
614 W76.
615 50. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. e., Wilkins, M. R., Appel, R. D. and Bairoch, A.
616 (2005) Protein Identification and Analysis Tools on the ExPASy Server, in The Proteomics Protocols
617 Handbook, (J. M. Walker ed, Humana Press, Totowa, NJ: pp. 571-607.
618 51. Kim, D. E., Chivian, D. and Baker, D. (2004) title. Nucleic acids research 32, W526-W531.
619 52. Colovos, C. and Yeates, T. O. (1993) title. Protein science 2, 1511-1519.
620 53. Colovos, C. and Yeates, T. O. (1993) title. Protein science : a publication of the Protein Society 2,
621 1511-1519.

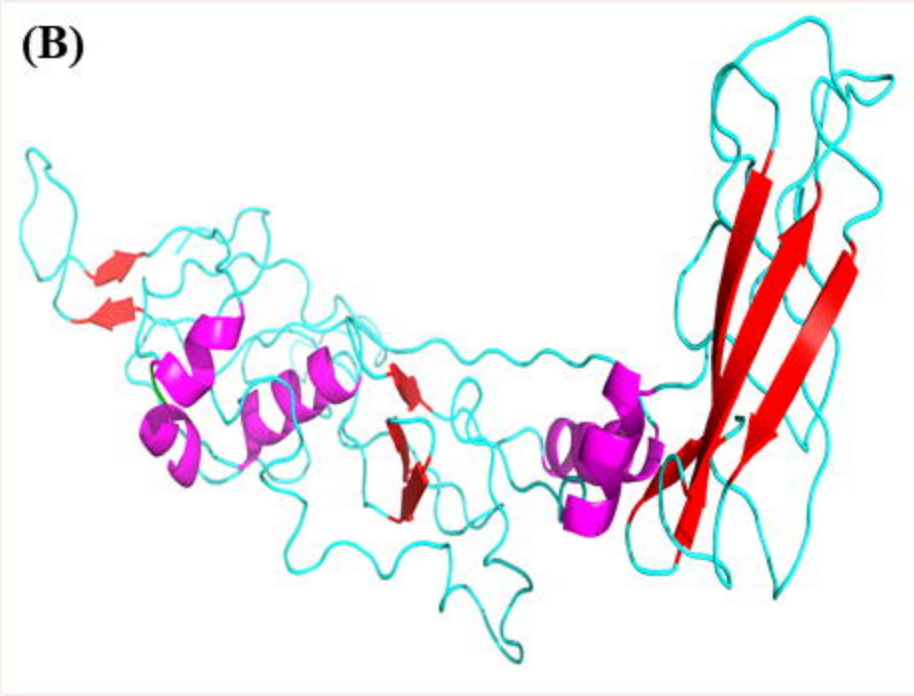
622 54. Enosi Tuipulotu, D., Netzler, N. E., Lun, J. H., Mackenzie, J. M. and White, P. A. (2018) title. 62.
623 55. Pierce, B. G., Wiehe, K., Hwang, H., Kim, B.-H., Vreven, T. and Weng, Z. (2014) title. Bioinformatics
624 30, 1771-1773.
625 56. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., DeBolt, S., Ferguson, D.,
626 Seibel, G. and Kollman, P. (1995) title. Computer Physics Communications 91, 1-41.
627 57. Salomon-Ferrer, R., Case, D. A. and Walker, R. C. (2013) title. Wiley Interdisciplinary Reviews:
628 Computational Molecular Science 3, 198-210.
629 58. Roe, D. R. and Cheatham, T. E. (2013) title. Journal of Chemical Theory and Computation 9, 3084-
630 3095.

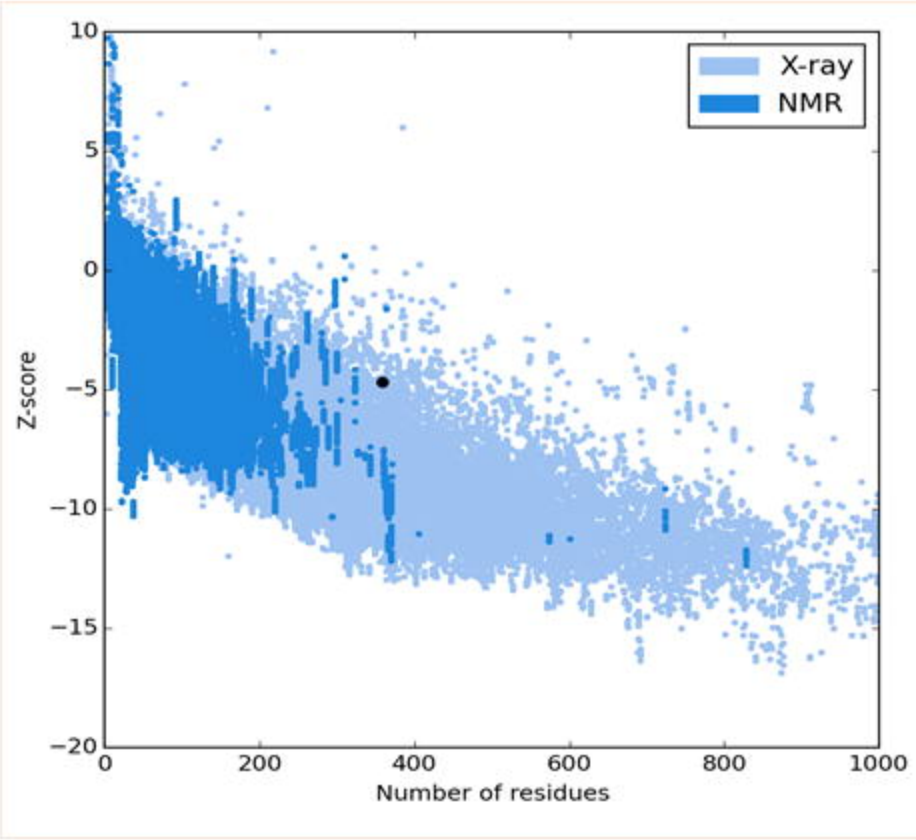
631

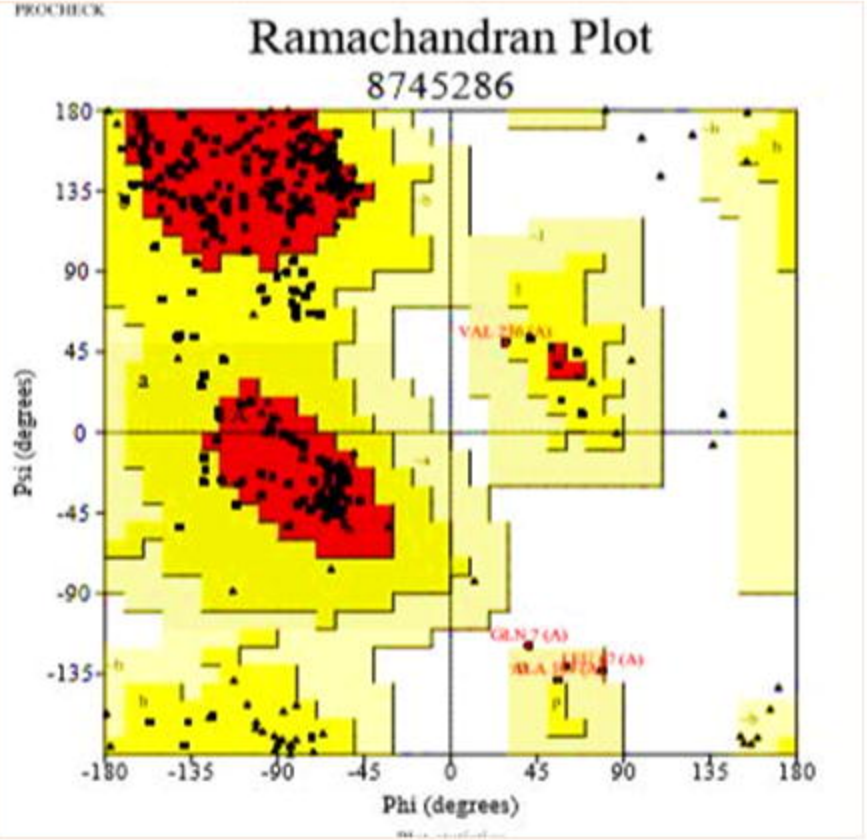

Norovirus Proteome
Antigenicity Profiling
Allergenicity Profiling
Capsid , Polyproteins, Small protein(protease)
CTL epitope prediction
HTL epitope prediction
B cell epitope prediction
multi-epitope vaccine construction
Allergenicity Profiling of Multi-epitopes vaccine
Antigenicity Profiling of Multi-epitopes vaccine
Physiochemical properties of vaccine
Tertiary Structure Prediction & validation
Molecular Docking with Toll-like receptor 7
Molecular Dynamics simulation of vaccine-TLR 7 complex
In silico cloning for vaccine expression

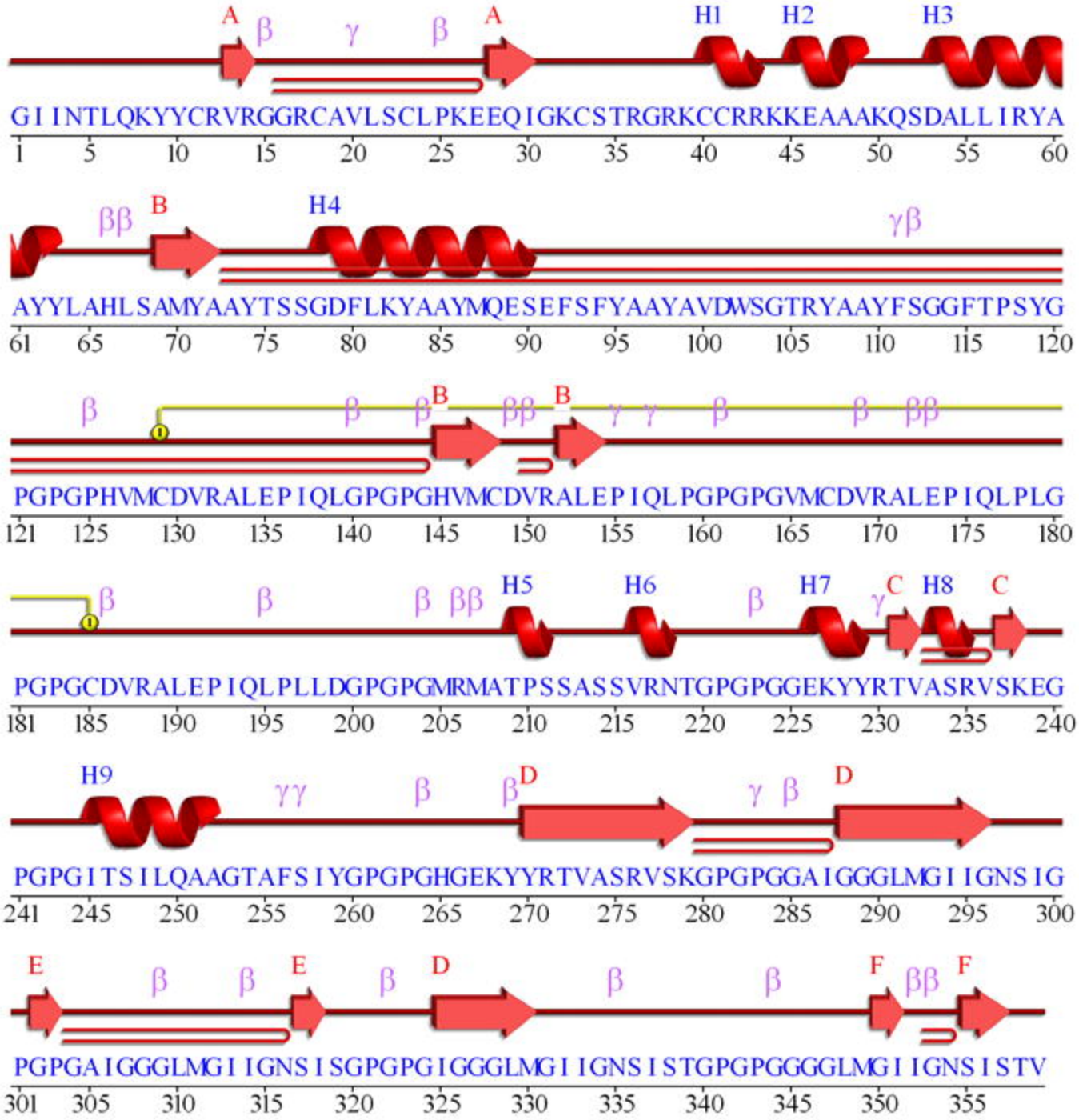

CTL epitopes
HTL epitopes
Adjuvant
Q5U7J2
A7YK10
A7YK10
A7YK09
A7YK09
A7YK11
A7YK11
A7YK10
A7YK10
A7YK10
A7YK09
A7YK09
A7YK09
A7YK09
A7YK09
AAY linker
GPGPG linker

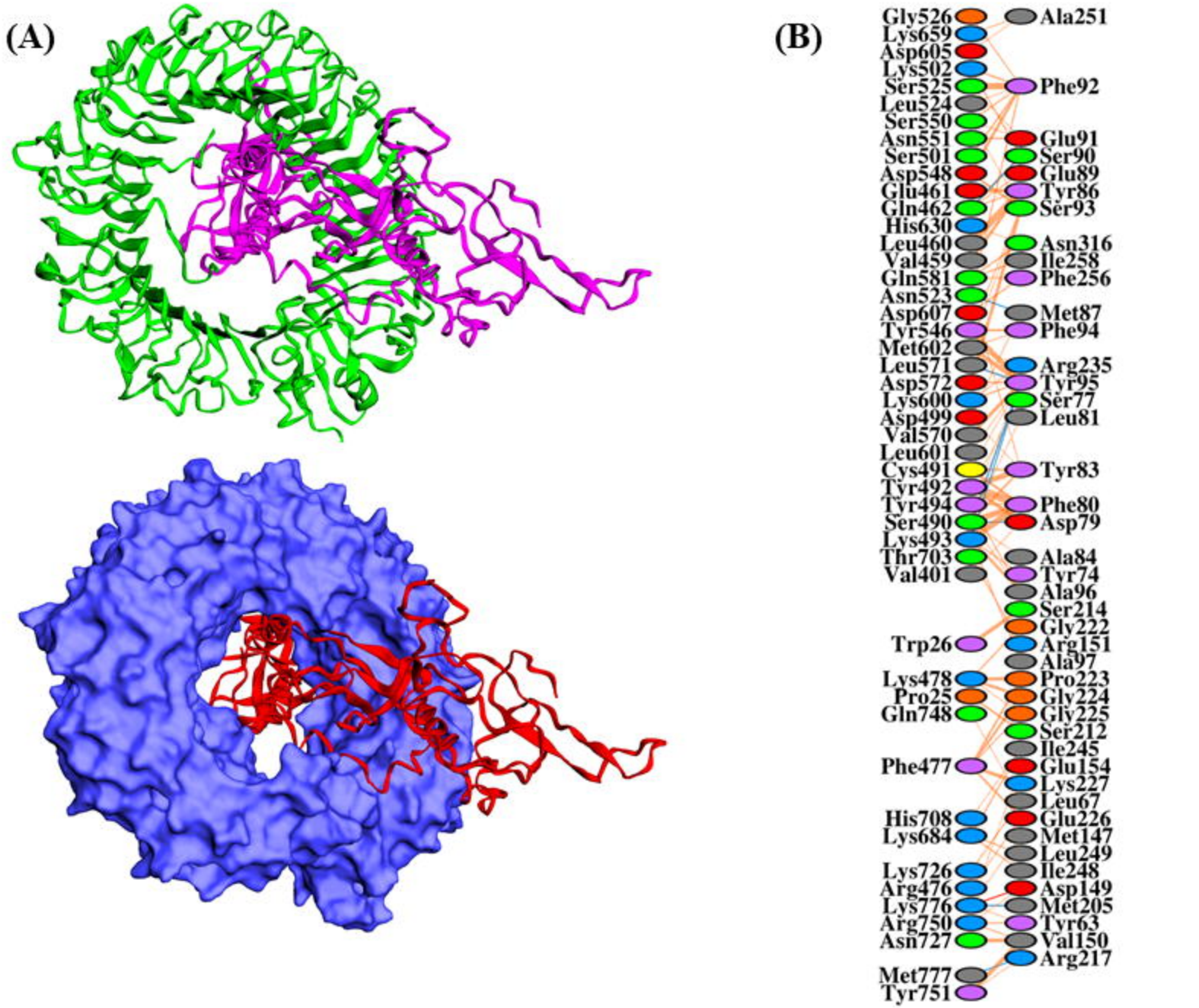

(A)

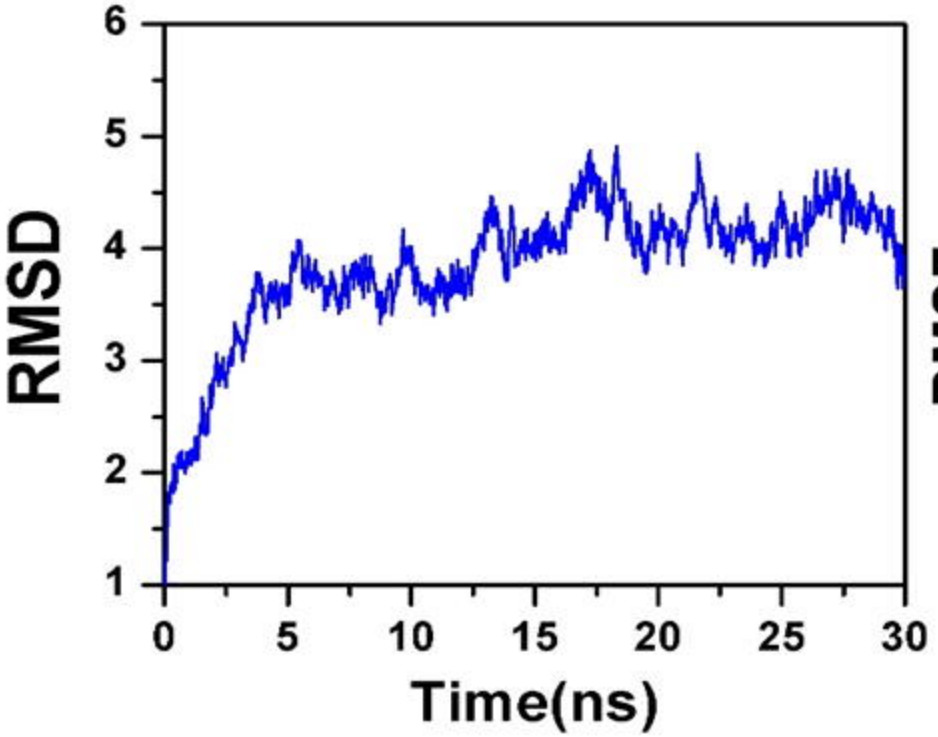

(B)

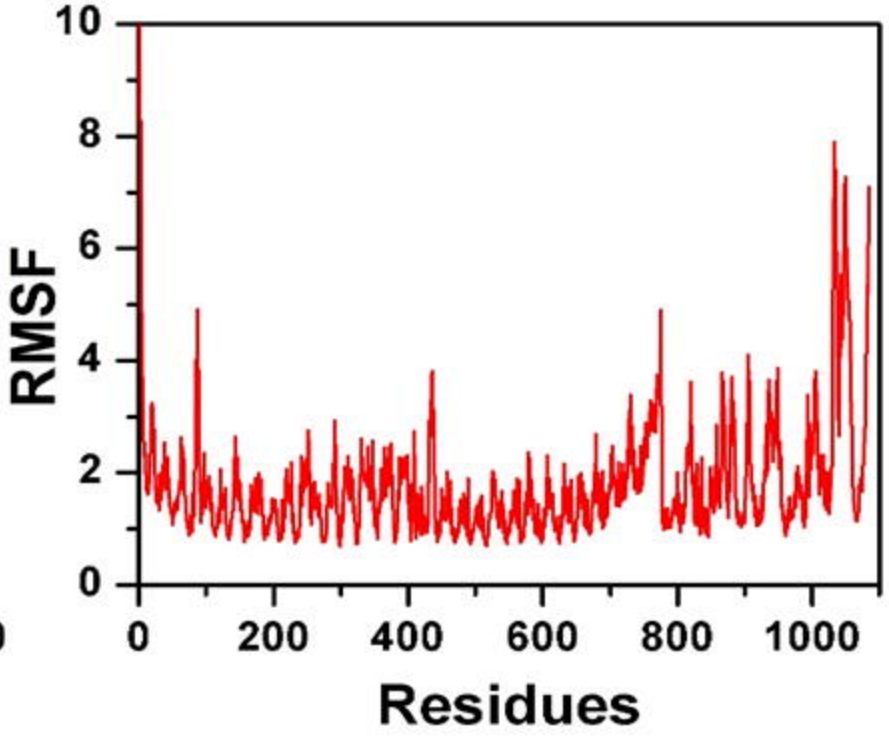

X-ray
NMR
Z-score
Number of residues
10
5
0
−5
−10
−15
−20
0
200
400
600
800
1000

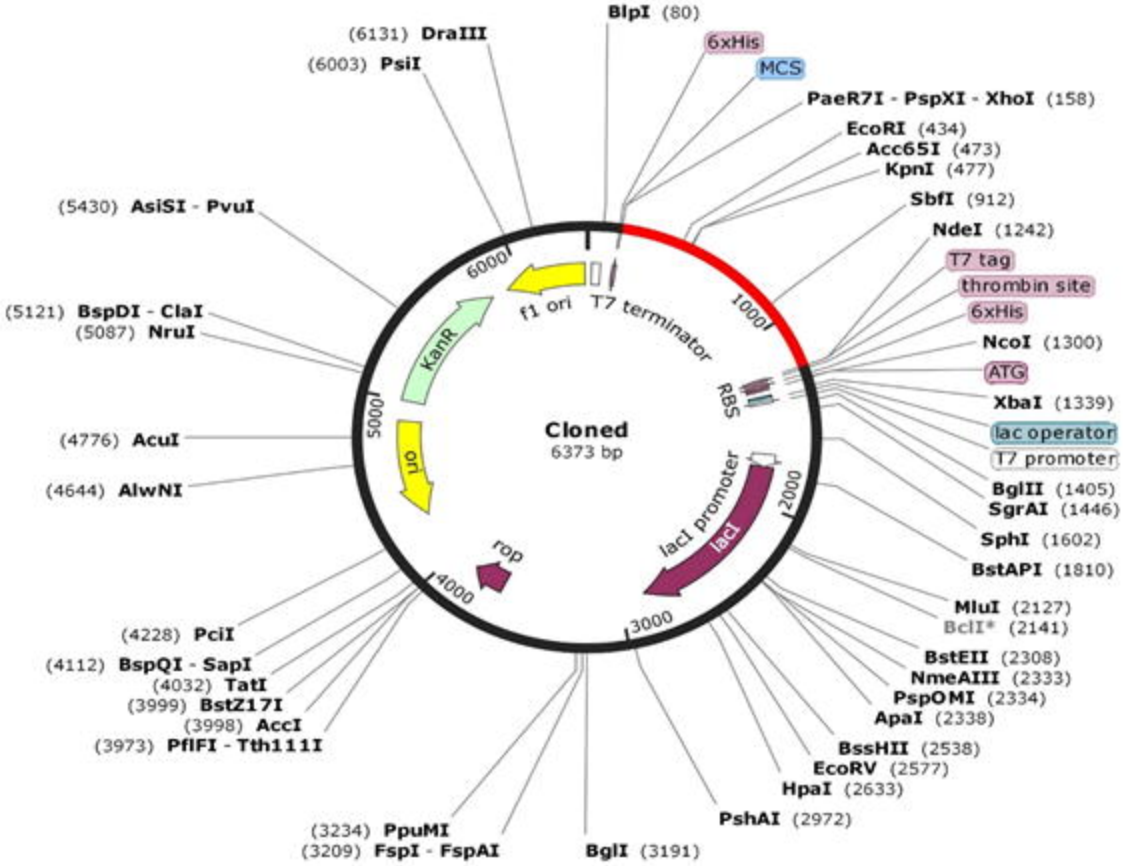

PROCHECK
Ramachandran Plot
8745286
Psi (degrees)
Phi (degrees)
180
135
90
45
0
-45
-90
-135
-180


A β γ β A H1 H2 H3
GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRKKEAAAKQSDALLIRYA
1 5 10 15 20 25 30 35 40 45 50 55 60
ββ B H4 γβ
AYYLAHLSAMYAAYTSSGDFLKYAAYMQESEFSFYAAYAVDWSGTRYAAYFSGGFTPSYG
61 65 70 75 80 85 90 95 100 105 110 115 120
β β β B ββ B γ γ β β ββ
PGPGPHVMCDVRALEPIQLGPGPGHVMCDVRALEPIQLPGPGPGVMCDVRALEPIQLPLG
121 125 130 135 140 145 150 155 160 165 170 175 180
β β β ββ H5 H6 β H7 C H8 C
PGPGCDVRALEPIQLPLLDGPGPGMRMATPSSASSVRNTGPGPGGEKYYRTVASRVSKEG
181 185 190 195 200 205 210 215 220 225 230 235 240
H9 γγ β β D γ β D
PGPGITSILQAAGTAFSIYGPGPGHGEKYYRTVASRVSKGPGPGGAIGGGLMGIIGNSIG
241 245 250 255 260 265 270 275 280 285 290 295 300
E β β E β D β β F ββ F
PGPGAIGGGLMGIIGNSISGPGPGIGGGLMGIIGNSISTGPGPGGGGLMGIIGNSISTV
301 305 310 315 320 325 330 335 340 345 350 355


(A)
(B)
Chain A
Chain B
Gly526
Lys659
Asp605
Lys502
Ser525
Leu524
Ser550
Asn551
Ser501
Asp548
Glu461
Gln462
His630
Leu460
Val459
Gln581
Asn523
Asp607
Tyr546
Met602
Leu571
Asp572
Lys600
Asp499
Val570
Leu601
Cys491
Tyr492
Tyr494
Ser490
Lys493
Thr703
Val401
Trp26
Lys478
Pro25
Gln748
Phe477
His708
Lys684
Lys726
Arg476
Lys776
Arg750
Asn727
Met777
Tyr751
Ala251
Phe92
Glu91
Ser90
Glu89
Tyr86
Ser93
Asn316
Ile258
Phe256
Met87
Phe94
Arg235
Tyr95
Ser77
Leu81
Tyr83
Phe80
Asp79
Ala84
Tyr74
Ala96
Ser214
Gly222
Arg151
Ala97
Pro223
Gly224
Gly225
Ser212
Ile245
Glu154
Lys227
Leu67
Glu226
Met147
Leu249
Ile248
Asp149
Met205
Tyr63
Val150
Arg217


RMSD
Time(ns)
1
2
3
4
5
6
0
5
10
15
20
25
30


RMSF
Residues
0
2
4
6
8
10
0
200
400
600
800
1000


Cloned
6373 bp
BlpI (80)
6xHis
MCS
PaeR7I - PspXI - XhoI (158)
EcoRI (434)
Acc65I (473)
KpnI (477)
SbfI (912)
NdeI (1242)
T7 tag
thrombin site
6xHis
NcoI (1300)
ATG
XbaI (1339)
lac operator
T7 promoter
BglII (1405)
SgrAI (1446)
SphI (1602)
BstAPI (1810)
MluI (2127)
BclI* (2141)
BstEII (2308)
NmeAIII (2333)
PspOMI (2334)
ApaI (2338)
BssHII (2538)
EcoRV (2577)
HpaI (2633)
PshAI (2972)
BglI (3191)
(3234) PpuMI
(3209) FspI - FspAI
(4228) PciI
(4112) BspQI - SapI
(4032) TatI
(3999) BstZ17I
(3998) AccI
(3973) PflFI - Tth111I
(4644) AlwNI
(4776) AcuI
(5087) NruI
(5121) BspDI - ClaI
(5430) AsiSI - PvuI
(6003) PsiI
(6131) DraIII
f1 ori
T7 terminator
KanR
ori
rop
RBS
lacI promoter
lacI
1000
2000
3000
4000
5000
6000