Drug and vaccine design against Novel Coronavirus (2019-nCoV) spike protein through Computational approach

Suresh Kumar*

Department of Diagnostic and Allied Health Science, Faculty of Health & Life Sciences, Management & Science University, 40100 Shah Alam, Selangor, Malaysia

*Correspondence: sureshkumar@msu.edu.my

Abstract:

The recent outbreak of the new virus in Wuhan city, China from the sea food market has led to the identification of a new strain called the corona virus and named as novel corona virus (2019-nCoV) belonging to Coronaviridae family. This has created major havoc and concern due to the mortality of 250 persons and affecting more than 10,000 people. This virus causes sudden fever, pneumonia and also kidney failure. In this study a computational approach is proposed for drug and vaccine design. The spike protein sequences were collected from a protein database and analysed with various bioinformatics tools to identify suitable natural inhibitors for the N-terminal receptor binding domain of spike protein. Also, it is attempted to identify suitable vaccine candidates by identifying B-Cell and T-cell epitopes. In the drug design, the tanshinone Iia and methyl Tanshinonate were identified as natural inhibitors based on the docking score. In the vaccine design, B-cell epitope VLLPLVSSQCVNLTTRTQLPPAYTN was found to have the highest antigenicity. FVFLVLLPL of MHC class-I allele and FVFLVLLPL of MHC class-II allele were identified as best peptides based on a number of alleles and antigencity scores. The present study identifies natural inhibitors and putative antigenic epitopes which may be useful as effective drug and vaccine candidates for the eradication of novel corona virus.

Keywords: 2019-nCoV, novel corona virus, wuhan virus, drug, vaccine, spike protein, epitope, vaccine design

1. Introduction:

The family of coronavirinae is made up of four genera based on their genetic properties, including Alpha corona virus genus, Beta corona virus genus, Gamma corona virus genus, and Delta corona virus genus. The corona virus family is a positive-stranded RNA virus, which mainly causes respiratory and central nervous system disease in humans and animals [1]. Human coronaviruses namely HCoV-OC43, HCoV-229E, HCoV-NL63 and HCoV-HKU1 circulate in humans, causing mild respiratory diseases [2]. Nevertheless, the SARS-CoV outbreak in 2002 and the MERS-CoV outbreak in 2012 demonstrated that coronaviruses would cross the barrier to species and emerge as highly pathogenic viruses [3]. Coronavirus can infect humans, including swine, cattle, horses, camels, cats, dogs, rats, birds, bats, rabbits, ferrets, mink, snake, and other wildlife species [4].

Recent outbreaks of ongoing viral pneumonia with unknown etiology in Wuhan City, China. A new virus was described as a novel coronavirus (2019-nCov) and named on 7 January 2020 [5]. On 10 January 2020, the viral genome was released on NCBI Genbank (Wuhan-Hu-1, GenBank accession number MN908947), followed on 12 January by the Global Initiative for the Integration of All Influenza Data (GISAID) following the release of genomes. This novel virus is characterized by genome sequencing closely related to members of a viral species Sever Acute Respiratory Syndrome (SARS) and coronavirus Middle East Respiratory Syndrome (MERS-CoV). On 31st december 2020, WHO declared the virus a global health emergency as the total death toll to 213 with almost 10,000 cases worldwide [6]. It is estimated this outbreak surpasses the 2003 severe acute respiratory syndrome (SARS) outbreak that began in china which infected 8,098 people killing 774.

Like all coronaviruses, 2019-nCoV consist of a minimum of three viral proteins namely spike protein (S), a type of glycoprotein, a membrane protein (M) that spans the membrane and an envelope protein (E), a highly hydrophobic protein that covers the entire structure of the coronavirus. The spike (S) glycoprotein in the coronavirus recognise the host cell receptors and causes an important role in viral infection [7].

There is an urgent need for the development of anti-viral drugs and vaccines against the 2019-nCov virus due to the high mortality rate of patients. During the epidemic and pandemic outbreak of new viral pathogens, the conventional method of development of drugs and vaccination is not possible to control as it is a time-consuming process [8,9]. Consequently, the rapid approach based on in-silico informatics has become very popular with recent advances in the sequencing of many pathogen genomes and protein sequence databases [10–14]. The continuous increase of patients and a high mortality rate of 2019-nCov infection highlight the urgent need for the development of a safe and effective vaccine. The aim of the study is to use to computational approach to design both anti-viral drug and vaccine candidates. The spike protein in the novel coronavirus sequence is used to design both anti-viral drug and vaccine candidates. For anti-viral drug design, receptor binding protein of novel coronavirus present in the N-terminal of spike protein is homology modelled and used as a protein receptor. 3C like proteinase (3CLpro) plays a role in viral pathogencity and replication by clevaging the polyprotein. The inhibitors of 3CLpro can block the clevage reaction thereby controling viral replication and pathogenecity. The natural inhibitors were used as ligands and docked with homology modelled coronavirs receptor binding protein. For vaccine design, the full sequence of the spike protein of novel coronavirus is used to predict B-cell and T-cell epitopes. The select best T-cell epitopes based on antigencity used as peptides and docked with human allele protein.

2. Materials & Methods

The overall workflow of computational drug and vaccine design for spike protein summarised in **figure 1**

2.1 Sequence retrieval

Wuhan, china isolate (BetaCoV / Wuhan-Hu-1/2019) primary sequence of Spike (S)-glycoprotein were extracted from the NCBI protein database [15] using the accession number QHD43416.1. For further analysis the sequence was downloaded as the FASTA sequence.

2.2 Physio-chemical, domain and structural analysis

Using the Protparam tool [16] physiochemical characterization such as molecular weight, theoretical PI, instability index and Grand average of hydropathicity (GRAVY) is calculated..
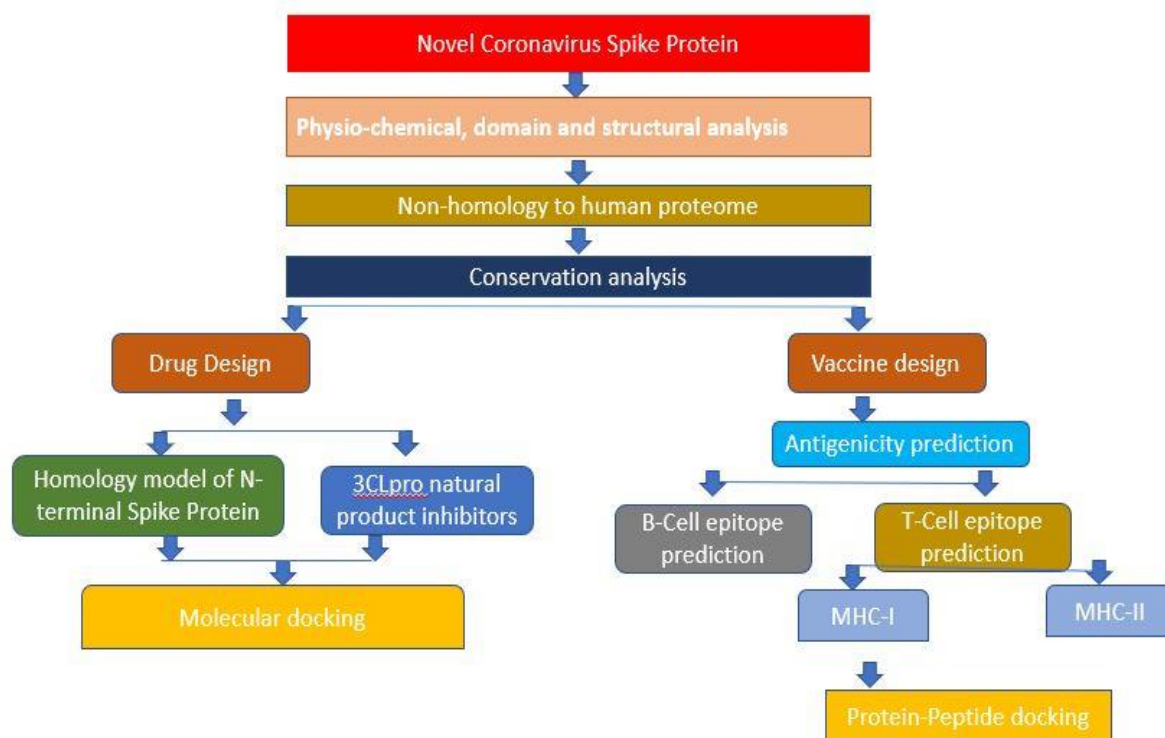

Figure 1: Overall workflow of computational drug and vaccine design by using a novel corona virus (2019-nCov) spike protein.

The spike protein used for domain predictions using the Conserved Domain Database (CDD)[17] and Inteproscan [18]. Signalpeptide, transmembrane helix and coils are predicted using signaP [19], THMMM and signal-TM [20]. The secondary structure was analysed through the SOPMA server with the default parameter. Subcellular localization was predicted using the Virus-mPLoc server [21]. The Spike protein analysed for binding regions and it is 3D structure determined using the phyre2 server [22].

2.3 Non-homology analysis and sequence conservation analysis

The retrieved s-glycoprotein checked to avoid cross-reactivity with the human host, blastp similarity search performed against the human refseq database. Conservancy analysis of the selected s-glycoprotein was searched in virus database available KEGG [23] by using blastP with default parameters. This conservation analysis is essential for determining the broad spectrum of drug target and epitope among all viruses.

2.4 Drug design

2.4.1 Homology modelling of s-protein N-terminal as a drug target

The spike protein was modelled by the Swiss-modeller server [24] with a selected template from the Protein Data Bank (PDB) using the structure of the N-terminal domain (NTD)of SARS-CoV spike protein (PDB ID: 5X4S). The template is a high-resolution structure of the trimeric MERS-CoV and SARS-CoV proteins in its pre-fusion conformation by single particle cryo-electron microscopy [25].
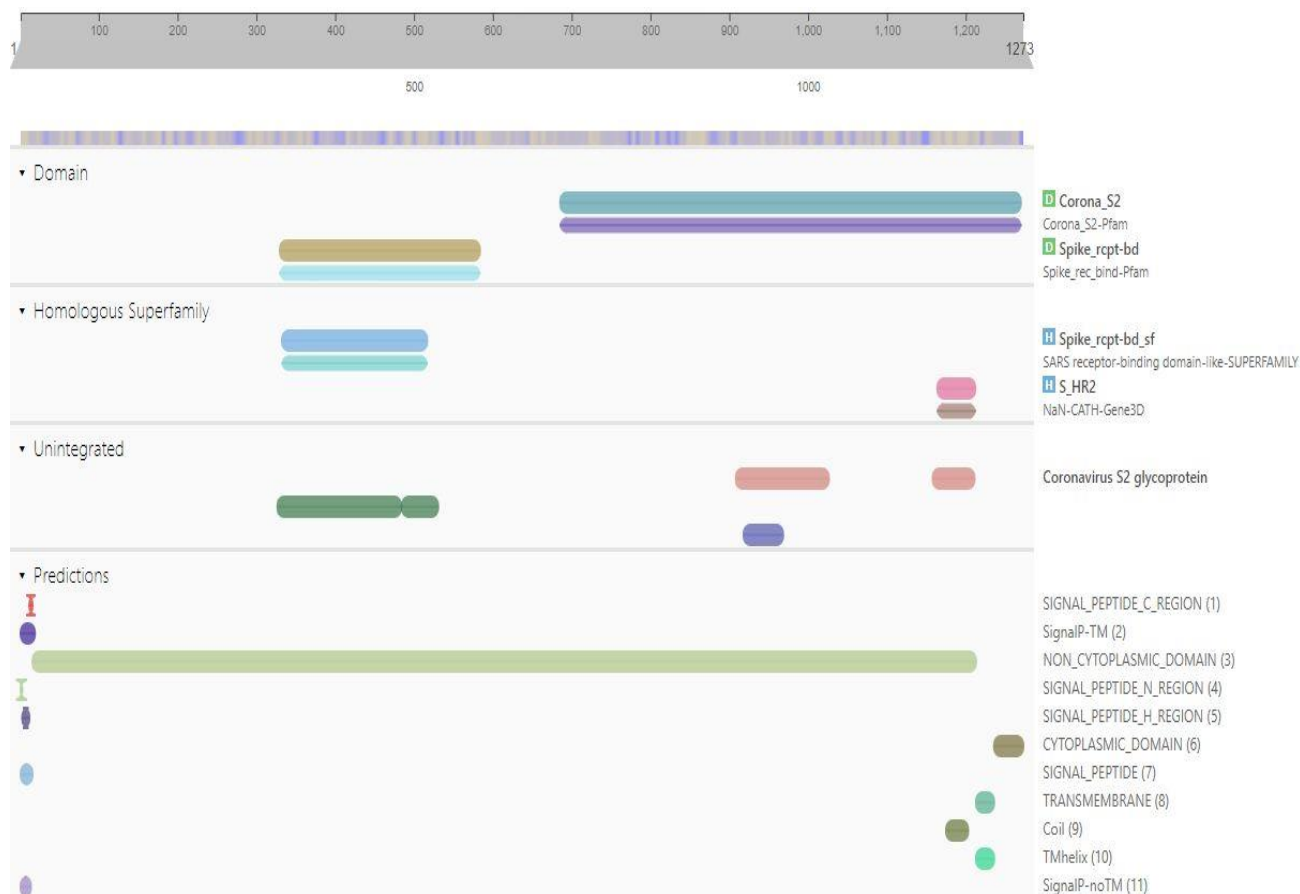

Figure 2: The predicted domain of Spike protein of novel corona virus by Interproscan

2.4.2 Ligand retrieval

The search was carried out 3C-like proteinase inhibitors related to SARS coronavirus [26], since this novel corona virus is genetically closely to the SARS virus and in causing infection. The 9 natural inhibitors ligands were downloaded from pubchem is SDF format. Using openbabel tool all ligand in SDF format are converted to PDB format and optimized for further docking purposes.

2.4.3 Molecular docking of N-terminal Spike protein with natural inhibitors

The molecular docking was carried using autodock [27] available in the MTIopenscreen server. The homology modelled N-terminal spike protein was docked with 2-Thiophenenecarboxaldehyde,Cryptotanshinone, Dihydrotanshinone, Furfural, Methyl

Tanshinonate, Miltirone, Tanshinone I,Tanshinone Iia and Tanshinone Iib ligands. Its server automatically pre-process the N-terminal spike protein by removing all hetero atoms and hydrogen atoms are added to the structure.

2.5. Vaccine design

2.5.1 Antigenic protein identification

To predict the antigenic property of the selected protein, the VaxiJen V2.0 server[28] was used with a threshold greater than equal to 5.

Table 1: Physiochemical characteristics of the highest antigenic surface glycoprotein genbank ID QHD43416.1

Criteria	Assessment
No of amino acids	1273
Molecular Weight	141178.47
Theoretical pI	6.24
Total number of negatively charged residues (Asp + Glu)	110
Total number of positively charged residues (Arg + Lys)	103
Formula	$C_{6336}H_{9770}N_{1656}O_{1894}S_{54}$
Extinction coefficients	148960
Estimated half-life	30 hours (mammalian reticulocytes, in vitro).
Instability index	33.01 (stable)
Aliphatic index	84.67
Grand average of hydropathicity (GRAVY)	-0.079

2.5.2 Prediction of B-cell epitopes

The B cell epitope prediction will enable to find the potential antigen that would interact and initiate an immuroreponse with B lymphocytes. IEDB (Immume-Epitope Database an Analysis Resource) server[29] was used for B-cell eptiope prediction. The B-Cell epitopes were identified based on antigenicity, prediction of the linear epitope, hydrophilicity, accessibility of surface and flexibility. BepiPred linear epitope prediction from the Immune Epitope Database was used to predict linear B cell epitopes with a default threshold. While the

prediction of candidate epitopes antigenicity sites was achieved for the identification using Kolaskar and Tongaonker antigenicity method with default threshold 1.027.

2.5.3 Prediction of T-cell eptiopes

MHC class-I peptides predicted using the Propred-1 server [30] and MHC class-II peptides predicted using the Propred tool. Both servers use a large number of human-leukocyte-antigens (HLA) for predicting epitopes.

Table 2: Blastp similarity search against viruses refseq database in KEGG with default parameters for conservation analysis

NCBI Protein ID	Protein Name	E-value	identity
NP_828851	Severe acute respiratory syndrome-related coronavirus; E2 glycoprotein precursor	0.0	76%
YP_003858584	Bat coronavirus BM48-31/BGR/2008; spike protein	0.0	72%
YP_009072440	Bat Hp-betacoronavirus/Zhejiang2013; spike protein	0.0	41%
YP_001039971	Rousettus bat coronavirus HKU9; spike glycoprotein	0.0	33%
YP_009273005	Rousettus bat coronavirus; spike protein	0.0	36%
YP_001039962	Pipistrellus bat coronavirus HKU5; spike glycoprotein	0.0	37%
YP_009361857	Bat coronavirus; spike glycoprotein	0.0	35%
YP_007188579	Betacoronavirus England 1; S protein	2e-172	35%
YP_009047204	Middle East respiratory syndrome-related coronavirus; spike glycoprotein	2e-171	35%
YP_009513010	Betacoronavirus Erinaceus/VMC/DEU/2012; S	3e-162	34%

2.5.4 Non-allegenicity and non-toxicity prediction of selected T cells epitopes

The selected T-cell epitopes were predicted for allergenicity by using the AllerCatPro server[31] and toxicity prediction using a toxpred tool.

2.5.5 Structural modelling and molecular docking of protein-peptide docking

T-cell epitopes of MHC class-I peptides were modelled using PEPFOLD server available RPBS MOBYL protal. The three-dimensional structure of the HLA-B7 allele (PDB ID: 3VCL) was retrieved from the PDB database and used as receptor protein. The 3D modelled peptide

of the T-cell epitope was docked with HLA-B7 receptor protein using HPEPDOCK webserver[32] with default parameters.

3. Results

3.1. Sequence & Structural analysis of Spike protein

The physiochemical properties calculation by using protparam showed that spike protein contained 1273 amino acids (aa) with a molecular weight of 141178.47 kDa, which represents a strong characteristic for both as a drug target and vaccine candidate. The target protein's theoretical isoelectric point (PI) was 6.24, suggesting its negative existence. An isoelectric point below 7 reveals a protein with negative charges. Of the 1353 residues, 110 aa were found to be negatively charged while others were found to be positively charged. Protparam estimated the instability-index (II) as 33.01, this protein grouped as stable. The other characteristics are shown in **Table 1**.

The domain search using CDD and interproscan spike receptor domain from 330-583 amino acids, S2 domain form 686-1270 and heptad repeat domain is between 1165-1212 amino acids as shown in **Figure 2**. The transmembrane predicted between transmembrane helix between 1214-1236 amino acids and coil predicted between 1176-1203 amino acids **(Figure 2)**. Prediction of secondary structure using SOPMA [33] showed that 28.59 per cent alpha helix, 23.5 per cent extended strand, 3.38 per cent beta turn, 44.78 per cent random coil is present in structure as shown in **Figure 3C**. Gene ontology [34] predicted biological process Membrane fusion (GO:0061025), Receptor-mediated virion attachment to host cell (GO:0046813). Cellular Component predicted as Viral envelope (GO:0019031), Integral component of membrane (GO:0016021) and molecular function none. The whole spike protein was homology modelled by using the phyre2 server **(Figure 3A)**. Based on CASTp [35] analysis for binding pocket, the protein has area 6868.604 and volume 24275.249 **(Figure 3B)**.

The spike protein showed no significant matches with human protein by blast search. conservation analysis of spike protein through blastp search against viral genome showed Severe acute respiratory syndrome-related coronavirus[1] with 76% and with Bat coronavirus BM48-31/BGR/2008[36] with 72% sequence identity **(Table 2)**. The homology modelled of N-terminal spike protein containing receptor binding protein checked for structural quality by using the Ramachandran plot. The homology model of protein shown in **Figure 4A.** The binding pocket of the N-terminal spike protein predicted using the CASTp server showing pocket regions in the receptor binding domain **(Figure 4B)**. The natural inhibitor tanshinone Iia **(Figure 4C)** and methyl tanshinonate **(Figure 4D)** shown the highest binding affinity score based on molecular docking with N-terminal spike protein **(Table 3)**.

3.2. Antigenic prediction

The spike protein of novel coronavirus evaluated for antigenicity by using the Vaxijen 2.0 server with a threshold setting ≥0.5 for virus organisms. The antigenicity prediction of spike protein is 0.4646 showing the characteristic of the expected antigen.
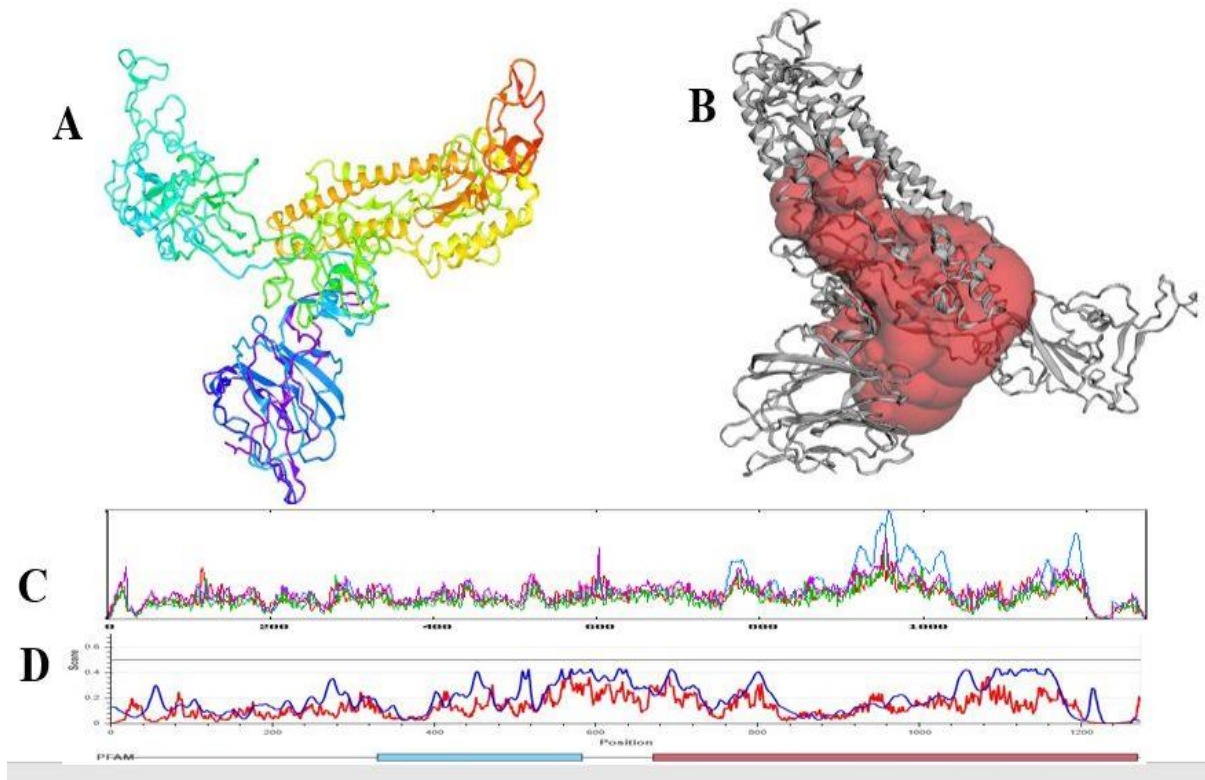
Figure 3: Structure of Spike Protein of the novel corona virus. (A). Homology modelled of Spike protein (B). The binding pocket of spike protein by using the CASTp server (C). Secondary structure prediction by using SOPMA tool (D). Disordered prediction of spike protein by using IUPRED2.
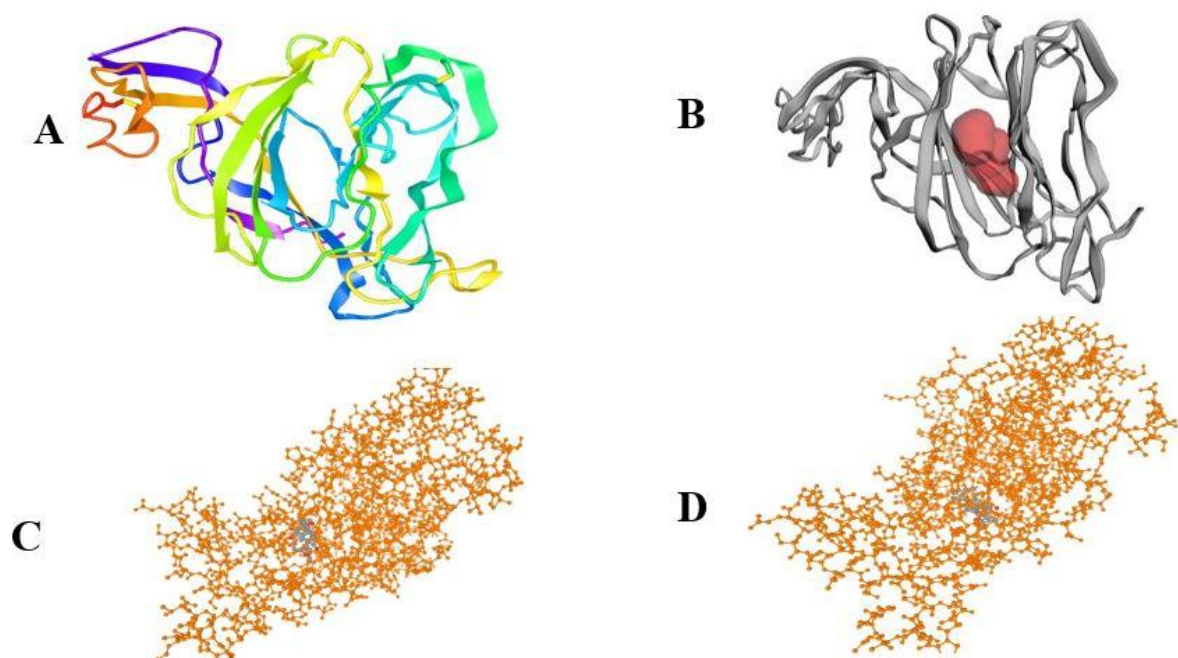
Figure 4: Structure of N-terminal Spike protein (A) Homology model of N-terminal spike protein showing the receptor binding domain (B). Binding pocket prediction of N-terminal spike protein by using CASTp (C). Interaction of tanshinone Iia with receptor binding of N-terminal of spike protein (D). Interaction of methyl tanshinonate with receptor binding of N-terminal of the spike protein

3.3. Prediction of B-cell epitopes

The spike protein was predicted for the B-cell epitope using the IEDB server. over 30 B-cell were predicted and 8 epitopes were selected based on the antigencity score by using Vaxijen 2.0. The predicted epitope VLLPLVSSQCVNLTTRTQLPPAYTN showed a high antigencity score of 1.0555 between the position 6-30 as shown in **Table 4**.

Table 3: Molecular docking of N-terminal spike protein with natural 3CLpro inhibitors

NPASS ID	Pubchem ID	Name	Docking Score
NPC471895	7364	2-Thiophenenecarboxaldehyde	-4.93
NPC77000	160254	Cryptotanshinone	-10.87
NPC988	11425923	Dihydrotanshinone	-10.38
NPC217226	7362	Furfural	-4.54
NPC144010	14610613	Methyl Tanshinonate	-10.88
NPC239185	160142	Miltirone	-8.86
NPC93241	114917	Tanshinone I	-10.36
NPC281398	164676	Tanshinone Iia	-10.91
NPC145830	318797	Tanshinone Iib	-10.47

Table 4: B-cell epitopes present on surface predicted via IEDB analysis resource are shown with antigencity score

sno	Epitope sequence	Vaxigen score
1	VLLPLVSSQCVNLTTRTQLPPAYTN	1.0555
2	RSSVLHSTQD	0.5404
3	VTWFHAIHVSGTNGTKRFDN	0.5485
4	VYFASTEKSNII	0.7795
5	GTTLDSKTQSLLIVNNATNVVIKVC	0.4494
6	YYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLM	0.2569
7	IYSKHTPIN	0.9013
8	DLPQGFSALEPLVDLPIGINITRFQTLLALH RSYLTPGDSSSGWTAGAAAYLLKYNENGTIT DAVDCALDPLSETKCTLKSFTVEKGIYQTSN FRVQPTESVRFPNITNLCPFGE	0.6329

The continuous B-cell epitopes predicted based on parameters such as hydrophilicity, flexibility, accessibility, turns, exposed surface, polarity and antigenic propensity. B-cell epitopes predicted for antigenic determinants based on Kolaskar and Tongaonkar antigenicity measuring tool. This tool predicts antigenic determinants based on physico-chemical properties with an accuracy of 75%. B-cell epitope predicted for hydrophilicity based on Parker's hydrophilicity prediction scale. This method based on retention times during hing-performance liquid chromatography of a reverse-phase column. The mobility of protein predicted based on Karplus and Schulz flexibility scale on the basis of the known temperature B factors. The rationale for predicting turns to predict antibody epitopes is based on the Chou and Fasman beta turn prediction.

Table 5: MHC class-I allele binding peptides predicted via Propred-I with their antigenicity scores

SNO	Peptides	MHC class-I alleles	Vaxijen score
1	MFVFLVLLP	HLA-A1,HLA-A2,HLA-A*0201, HLA-A*0205, HLA-A*1101,HLA-A24	0.4166
2	FVFLVLLPL		0.8601
3	VFLVLLPLV	HLA-A3, HLA-A*3101,HLA-A*3302,HLA-A68.1,HLA-A20 Cattle,HLA-A2.1, HLA-B14,HLA-B*2702,HLA-B*2705,HLA-B*3501,HLA-B*3701, HLA-B*3801, HLA-B*3901, HLA-B*3902, HLA-B40,HLA-B*4403, HLA-B*5101, HLA-B*5102,HLA-B*5103, HLA-B*5201,HLA-B*5301, HLA-B*5401, HLA-B*51, HLA-B*5801,HLA-B60,HLA-B61, HLA-B62, HLA-B7, HLA-B*0702,HLA-B8, HLA-Cw*0301,HLA-Cw*0401,HLA-Cw*0602, HLA-Cw*0702,MHC-Db,MHC-Db revised, MHC-Dd,MHC-Kb, MHC-Kd, MHC-Kk, MHC-Ld	0.4651
4	FLVLLPLVS		0.4266

3.4. Prediction of T-cell epitopes

Propred-I server used for prediction of the T-cell epitope of MHC Class-I. The peptide 'FVFLVLLPL' as best probable based on antigenic score 0.8601 predicted by Vaxijen out of 4 predicted peptides **(Table 5)**. Properd server used for prediction of MHC class-II peptide and predicted 3 peptides, based on antigencity score of 0.8601 and profiling the peptide 'FVFLVLLPL' predicted to be the best probable peptide **(Table 6)**.

3.5. Profiling features of selected T cells epitopes

The selected best T-cell epitopes were checked for allergenicity by using allercatpro, toxicity, hydropathicity, hydrophilicity, hydrophobicity, and charge prediction using toxpred as shown in **Table 7.** Based on the docking of protein-peptide docking **(Figure 5A-D),** the best selected epitope is FVFLVLLPL **(Table 8).**
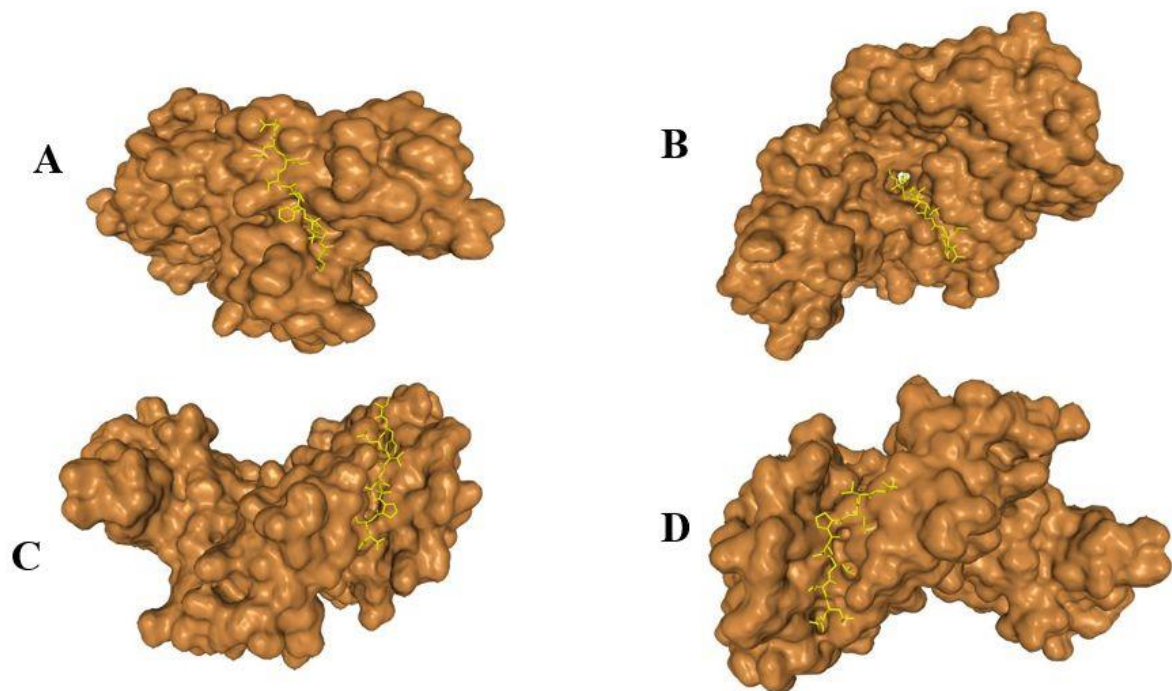
Figure 5: Protein-peptide docking by usinthe g HPEPDOCK server. (A) MFVFLVLLP (B). FVFLVLLPL (C). VFLVLLPLV (D). FLVLLPLVS

4. Discussion

The domain search by CDD and interproscan predicted Spike receptor binding domain from 330-583 amino acids. Spike is a glycoprotein envelope that assists viral entry into the host cell[37]. Coronavirus S2 glycoprotein predicted from 686-1270 amino acids. Transmembrane prediction using TMHMM2.0 predicted outside 1-1213 aa, TMhelix between 1214-1236 aa and inside 1237-1273 aa. SingalP-TM predicted 1-18. Coil predicted between 1176-1203. Coronovirus spike glycoprotein, heptad repeat 2 domain predicted between 1165-1212 amino acids. This superfamily represents one of the two heptad repeat domains found in the S protein which involves in helix-helix interfaces [38]. The secondary structure analysis reveled that a greater number of the random coil is presently followed by helix structure. Through the disorder prediction there are no disordered residues present in the spike protein, the spike protein blast similarity analyses against the viral genome in the KEGG genome database revealed close to spike protein of SARS followed by bat coronavirus BM48-31/BGR/2008.

Table 6: MHC class-II allele binding epitopes predicted using Propred with their antigenicity scores

SNo	Peptides	MHC class-II alleles	Vaxijen score
1	FVFLVLLPL	DRB1_0101,DRB1_0102, DRB1_0701, DRB1_0703,DRB1_0817, DRB1_1101,DRB1_1104 DRB1_1106,DRB1_1128,DRB1_1305,DRB1_1307	0.8601

		DRB1_1311,DRB1_1321, DRB1_1501,DRB1_1502, DBR1_1506,DRB5_0101. DRB5_0105	
2	FLVLLPLVS	DRB1_0101,DRB1_0408,DRB1_1101,DRB1_1104 DRB1_1106,DRB1_1114,DRB1_1120,DRB1_1128 DRB1_1302,DRB1_1305,DRB1_1307,DRB1_1311 DRB1_1321,DRB1_1323	0.4266
3	LVLLPLVSS	DRB1_0102,DRB1_0804,DRB1_1101,DRB1_1102 DRB1_1104,DRB1_1106,DRB1_1114,DRB1_1121 DRB1_1301,DRB1_1304,DRB1_1307,DRB1_1311 DRB1_1322,DRB1_1323,DRB1_1327,DRB1_1328	0.6523

Table 7: Toxicity, allergenicity, hydro and physiochemical profiling of selected peptides

SNO	Peptides	Toxicity	Allergenicity	Hydrophobicity	Hydrophilicity	charge PI	Mol. weight
MHC class-I alleles							
1	MFVFLVLLP	Non-Toxic	NA	0.45	-1.63	0.00	1078.56
2	FVFLVLLPL	Non-Toxic	NA	0.48	-1.69	0.00	1060.53
3	VFLVLLPLV	Non-Toxic	NA	0.48	-1.58	0.00	1012.49
4	FLVLLPLVS	Non-Toxic	NA	0.39	-1.38	0.00	1000.43
MHC class-II alleles							
1	FVFLVLLPL	Non-Toxic	NA	0.48	-1.69	0.00	1060.53
2	FLVLLPLVS	Non-Toxic	NA	0.39	-1.38	0.00	1000.43
3	LVLLPLVSS	Non-Toxic	NA	0.29	-1.07	0.00	940.33

As of now the screening process confined to drugs & vaccines is under development and it will take several months to years to come into the market The virus of 3C-like protease (3CLpro) responsible for pathogencity by replicating the viral genome upon the host cell infection.[39]. The clevage function of 3CLpro can be blocked by using inhibitors thereby inhibiting viral replication. so, it is considered as attractive targets for anti-viral drugs. Accordingly, inhibitors that block the cleavage function of 3CLpro can be expected to inhibit virus replication, making this enzyme one of the most attractive targets for anti-nCOV. To speed up the screening process in view of the ongoing nCov outbreak, the virtual screen process was carried out in search of natural 3CLpro natural inhibitors in NPASS (Natural Product Activity & Species Source Database). NPASS[40] is a comprehensive database that provides unique natural products isolated from source organisms with activity records on targets. By virtual screening of 9 compounds, Tanshinone Iia[41] and Methyl Tanshinonate[42] inhibitor predicted to highest

binding affinity. Tanshinone IIA (Tan IIA) is a lipophilic component showing anti-viral activity extracted from the root of Salvia miltiorrhiza Bunge

In this study using spike protein of noval coronavirus predicted B-cell and T-cell epitopes for vaccine development. The epitopes predicted by using IEDB webserver using parameters like solvent accesibility, flexibility, antigencity. The predicted B-cell epitope is VLLPLVSSQCVNLTTRTQLPPAYTN and T-cell epitope of MHC-I (FVFLVLLPL) and MHC-II (FVFLVLLPL). The predicted epitopes checked for allergencity and toxicity. Based on docking score peptide FVFLVLLPL considered as the best probable T-cell epitope considered for a vaccine candidate. The predicted epitopes from this study can be further tested for therapeutic potency for vaccine development for novel coronavirus.

Table 8: Protein-Peptide docking using HPEPDOCK server of the selected best peptide of MHC-I with Crystal Structure of HLA-B7 (PDB ID: 3VCL) protein

Peptide	Protein PDB ID	Docking score
MFVFLVLLP	3VCL	-182.489
FVFLVLLPL	3VCL	-186.981
VFLVLLPLV	3VCL	-155.515
FLVLLPLVS	3VCL	-162.977

5. Conclusion

In the present study, both drug and vaccine design was applied to identify drug and vaccine candidates. This approach will be cost effective and can save time in the design of drugs and epitopes. The suggested drug inhibitors and vaccine candidates from this study will help to develop anti-viral and vaccine which may help to control this global threat.

References:

1   Perlman, S. *et al.* Coronaviruses post-SARS: update on replication and pathogenesis. *nature.com* at <https://www.nature.com/articles/nrmicro2147>

2   Owusu, M. *et al.* (2014) Human coronaviruses associated with upper respiratory tract infections in three rural areas of Ghana. *PLoS ONE* 9,

3   Perlman, S. (2020) Another Decade, Another Coronavirus. *New England Journal of Medicine* DOI: 10.1056/NEJMe2001126

4   Ji, W. *et al.* (2020) Homologous recombination within the spike glycoprotein of the newly identified coronavirus may boost cross-species transmission from snake to human. *Journal of medical virology* DOI: 10.1002/jmv.25682

5   Corman, V.M. *et al.* (2020) Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance* 25, 2000045

6   Nishiura, H. *et al.* The Extent of Transmission of Novel Coronavirus. *J. Clin. Med*

2020, 330

7   Pillaiyar, T. *et al.* An overview of severe acute respiratory syndrome-coronavirus
        (SARS-CoV) 3CL protease inhibitors: Peptidomimetics and small molecule
        chemotherapy. , *Journal of Medicinal Chemistry*, 59. 28-Jul-(2016) , American
        Chemical Society, 6595–6628

8   Lipinski, C.A. *et al.* Experimental and computational approaches to estimate solubility
        and permeability in drug discovery and development settings. , *Advanced Drug
        Delivery Reviews*. (2012)

9   Hughes, J.P. *et al.* Principles of early drug discovery. , *British Journal of
        Pharmacology*. (2011)

10   Huang, H.J. *et al.* Current developments of computer-aided drug design. , *Journal of
        the Taiwan Institute of Chemical Engineers*. (2010)

11   Kapetanovic, I.M. (2008) Computer-aided drug discovery and development
        (CADDD): In silico-chemico-biological approach. *Chemico-Biological Interactions*
        DOI: 10.1016/j.cbi.2006.12.006

12   Drie, J.H. (2007) Computer-aided drug design: The next 20 years. *Journal of
        Computer-Aided Molecular Design* DOI: 10.1007/s10822-007-9142-y

13   Veselovsky, A. V. and Ivanov, A.S. Strategy of computer-aided drug design. , *Current
        Drug Targets - Infectious Disorders*. (2003)

14   Suresh Kumar (2015) Computational identification and binding analysis of orphan
        human cytochrome P450 4X1 enzyme with substrates. *BMC Research Notes* 8, 10

15   Acland, A. *et al.* (2014) Database resources of the National Center for Biotechnology
        Information. *Nucleic Acids Research* DOI: 10.1093/nar/gkt1146

16   Gasteiger, E. *et al.* (2005) Protein Identification and Analysis Tools on the ExPASy
        Server. In *The Proteomics Protocols Handbook*

17   Marchler-Bauer, A. *et al.* (2011) CDD: A Conserved Domain Database for the
        functional annotation of proteins. *Nucleic Acids Research* DOI: 10.1093/nar/gkq1189

18   Quevillon, E. *et al.* (2005) InterProScan: Protein domains identifier. *Nucleic Acids
        Research* DOI: 10.1093/nar/gki442

19   Bendtsen, J.D. *et al.* (2004) Improved prediction of signal peptides: SignalP 3.0.
        Journal of Molecular Biology DOI: 10.1016/j.jmb.2004.05.028

20   Käll, L. *et al.* (2004) A combined transmembrane topology and signal peptide
        prediction method. *Journal of Molecular Biology* DOI: 10.1016/j.jmb.2004.03.016

21   Chou, K.-C. and Shen, H.-B. (2010) Cell-PLoc 2.0: an improved package of web-
        servers for predicting subcellular localization of proteins in various organisms. *Natural
        Science* DOI: 10.4236/ns.2010.210136

22   Kelley, L.A. *et al.* (2015) The Phyre2 web portal for protein modeling, prediction and
        analysis. *Nature Protocols* DOI: 10.1038/nprot.2015.053

23   Kanehisa, M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids
        Research* DOI: 10.1093/nar/gkh063

24   Reddy, C.S. *et al.* (2006) Homology modeling of membrane proteins: A critical assessment. *Computational Biology and Chemistry* DOI: 10.1016/j.compbiolchem.2005.12.002

25   Yuan, Y. *et al.* (2017) Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nature Communications* DOI: 10.1038/ncomms15092

26   Xiong, B. *et al.* (2003) A 3D model of SARS_CoV 3CL proteinase and its inhibitors design by virtual screening. *Acta Pharmacologica Sinica*

27   Forli, W. *et al.* (2012) AutoDock Version 4.2. *Citeseer*

28   Doytchinova, I.A. and Flower, D.R. (2007) VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* DOI: 10.1186/1471-2105-8-4

29   Vita, R. *et al.* (2019) The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Research* DOI: 10.1093/nar/gky1006

30   Singh, H. and Raghava, G.P.S. (2002) ProPred: Prediction of HLA-DR binding sites. *Bioinformatics* DOI: 10.1093/bioinformatics/17.12.1236

31   Maurer-Stroh, S. *et al.* (2019) AllerCatPro-prediction of protein allergenicity potential from the protein sequence. *Bioinformatics* DOI: 10.1093/bioinformatics/btz029

32   Zhou, P. *et al.* (2018) HPEPDOCK: A web server for blind peptide-protein docking based on a hierarchical algorithm. *Nucleic Acids Research* DOI: 10.1093/nar/gky357

33   Geourjon, C. and Deléage, G. (1995) Sopma: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics* DOI: 10.1093/bioinformatics/11.6.681

34   Gene Ontology Consortium (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research* DOI: 10.1093/nar/gkh036

35   Tian, W. *et al.* (2018) CASTp 3.0: Computed atlas of surface topography of proteins. *Nucleic Acids Research* DOI: 10.1093/nar/gky473

36   Poon, L.L.M. *et al.* (2005) Identification of a Novel Coronavirus in Bats. *JOURNAL OF VIROLOGY* 79, 2001–2009

37   Belouzard, S. *et al.* Mechanisms of coronavirus cell entry mediated by the viral spike protein. , *Viruses*. (2012)

38   Li, F. (2016) Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annual Review of Virology* DOI: 10.1146/annurev-virology-110615-042301

39   Chen, L. *et al.* (2005) Cinanserin Is an Inhibitor of the 3C-Like Proteinase of Severe Acute Respiratory Syndrome Coronavirus and Strongly Reduces Virus Replication In Vitro †. *JOURNAL OF VIROLOGY* 79, 7095–7103

40   Zeng, X. *et al.* (2018) NPASS: Natural product activity and species source database for natural product research, discovery and tool development. *Nucleic Acids Research* DOI: 10.1093/nar/gkx1026

41   Shang, Q. *et al.* Tanshinone IIA: A promising natural cardioprotective agent. ,

Evidence-based Complementary and Alternative Medicine. (2012)

42   Qin, J.K. *et al.* (2006) (S)-Methyl 6,7,8,9,10,11-hexahydro-1,6-dimethyl-10,11-dioxophenanthro[1,2- 6]-furan-6-carboxylate (methyl tashinonate). *Acta Crystallographica Section E: Structure Reports Online* DOI: 10.1107/S1600536806007203