Supplementary file 1

Pourseif et al., ***BioImpacts,*** 2021, 11(1), 65-84
doi: 10.34172/bi.2021.11
http://bi.tbzmed.ac.ir/

A domain-based vaccine construct against SARS-CoV-2, the causative agent of COVID-19 pandemic: development of self-amplifying mRNA and peptide vaccines

Mohammad Mostafa Pourseif[1], Sepideh Parvizpour[1], Behzad Jafari[2,1], Jaber Dehghani[1], Behrouz Naghili[3], Yadollah Omidi[4*]

[1] Research Center for Pharmaceutical Nanotechnology, Biomedicine Institute, Tabriz University of Medical Sciences, Tabriz, Iran

[2] Department of Medicinal Chemistry, Faculty of Pharmacy, Urmia University of Medical Sciences, Urmia, Iran

[3] Research Center for Infectious and Tropical Diseases, Tabriz University of Medical Sciences, Tabriz, Iran

[4] Nova Southeastern University, College of Pharmacy, Florida, USA

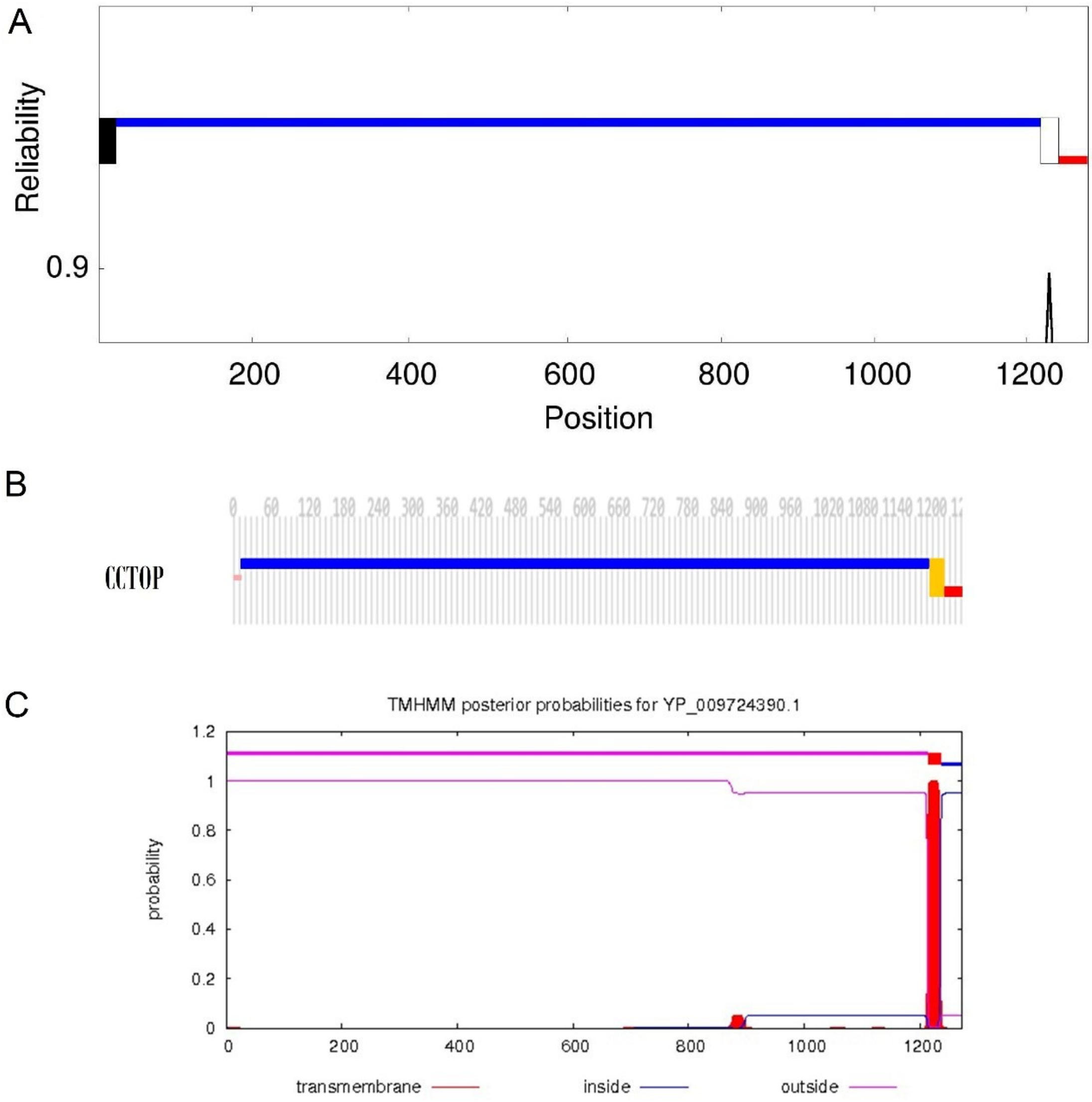
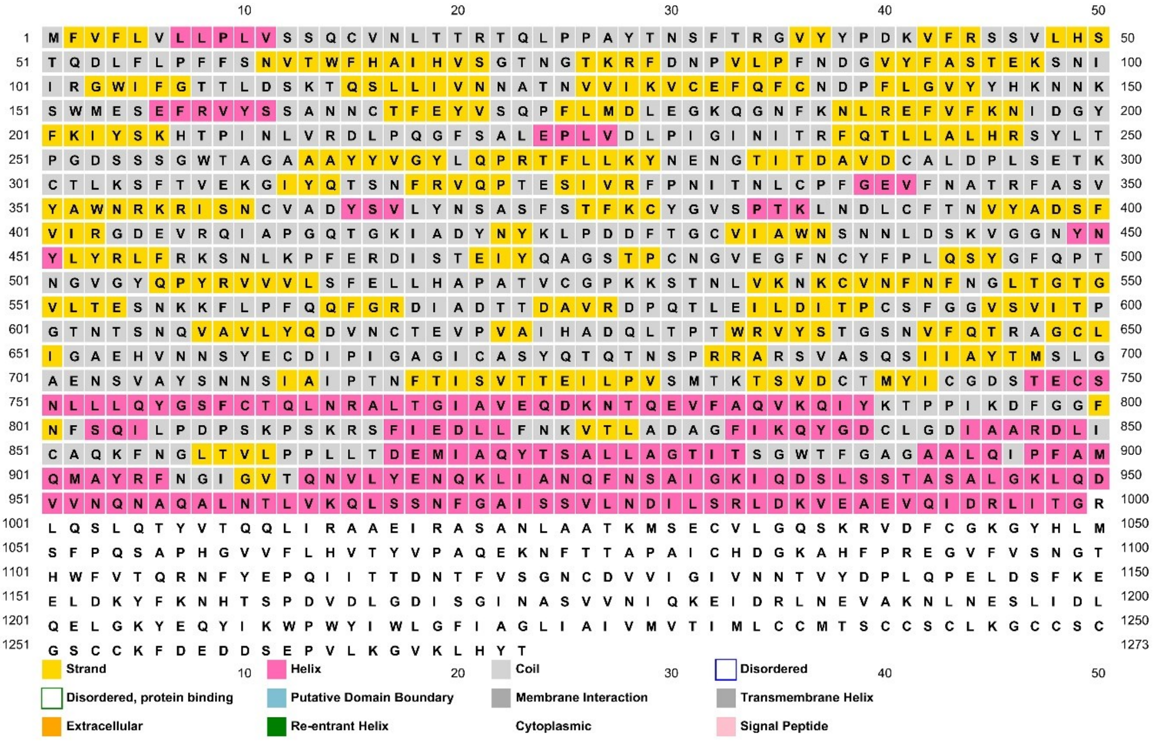


Figure S1. Plots show the potential transmembrane topology and signal peptide in the SARS-CoV-2 spike glycoprotein sequence. **A**) The consensus prediction plot of the potential transmembrane and intracellular regions of the glycoprotein S which is predicted using the TOPCONS server. **B**) the CCTOP plot shows transmembrane topology of the Glycoprotein S. **C**) The residue-based topology prediction plot obtained from TMHMM online server.



1 MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHS 50
51 TQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNI 100
101 IRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNK 150
151 SWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGY 200
201 FKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLT 250
251 PGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK 300
301 CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASV 350
351 YAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSF 400
401 VIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYN 450
451 YLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPT 500
501 NGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTG 550
551 VLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP 600
601 GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCL 650
651 IGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLG 700
701 AENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECS 750
751 NLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGF 800
801 NFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLI 850
851 CAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM 900
901 QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQD 950
951 VVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGR 1000
1001 LQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLM 1050
1051 SFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGT 1100
1101 HWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKE 1150
1151 ELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL 1200
1201 QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSC 1250
1251 GSCCKFDEDDSEPVLKGVKLHYT 1273

Figure S2. The secondary structure of the reference sequence of SARS-CoV-2 spike glycoprotein.

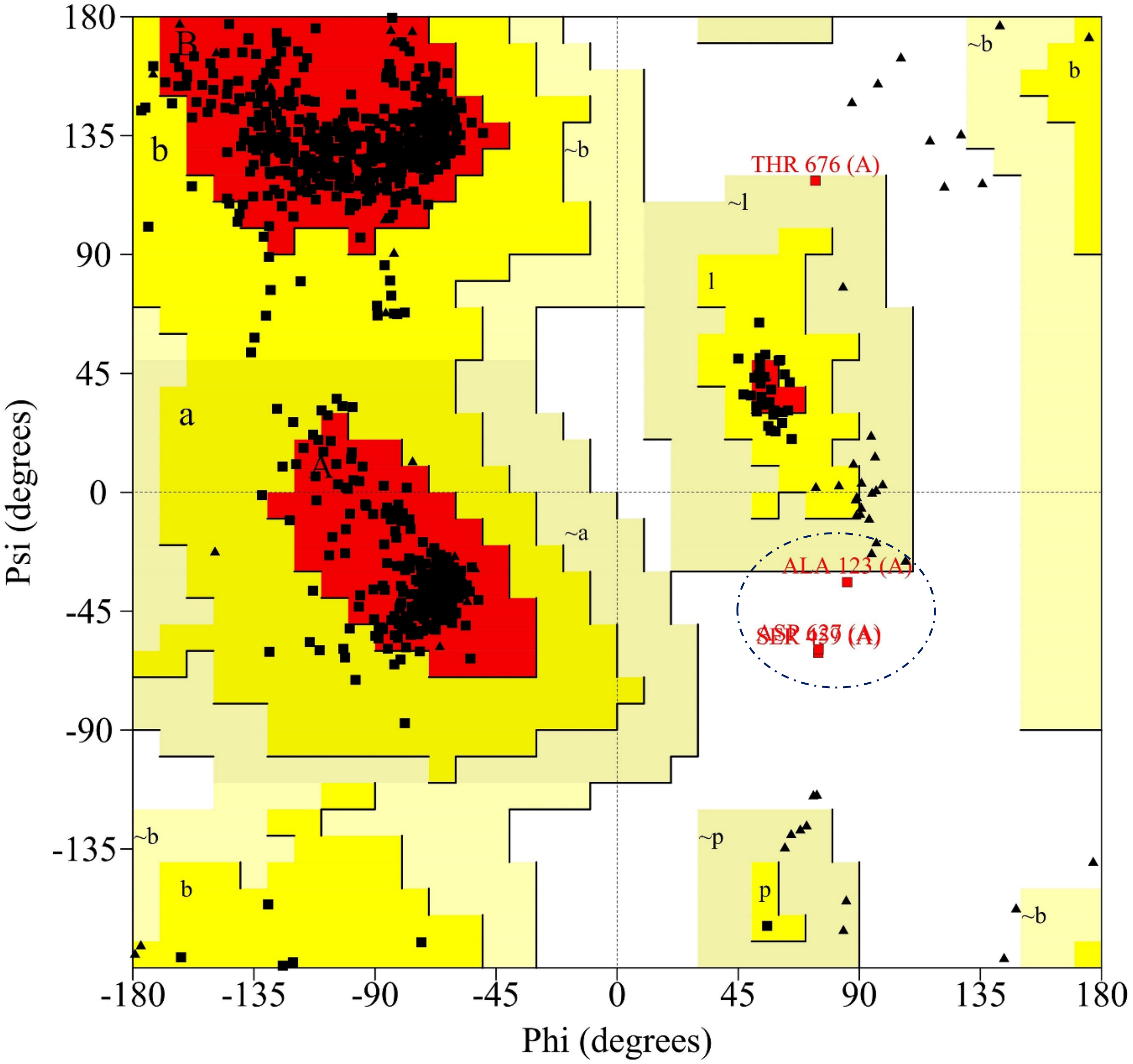
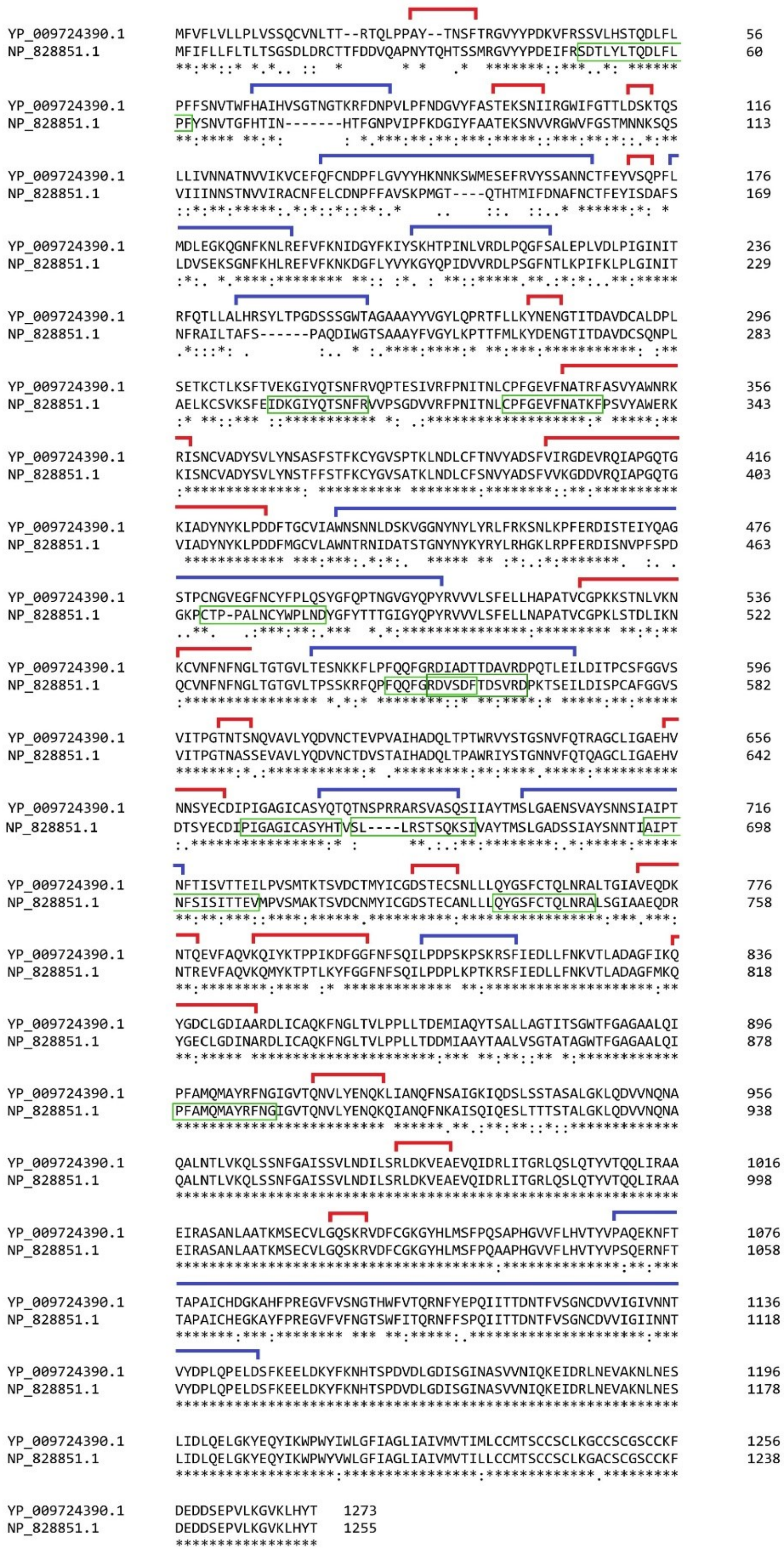


Figure S3. The 3D structure validation of the homology modeled SARS-CoV-2 protein S. Ramachandran plot of the modeled 3D structure. The plot indicates that only three residues (Asp627, Ser459, and Ala123) are located in the disallowed region (marked by a dash dotted circle), while 92.1% of the residues are in the most favorable region, and the rest of them (7.6%) are located in the additionally and generously allowed regions of the map.



```
YP_009724390.1      MFVFLVLLPLVSSQCVNLTT--RTQLPPAY--TNSFTRGVYYPDKVFRSSVLHSTQDLFL      56
NP_828851.1         MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLFL      60
                    **:**::* *.*.. ::  *       * *   .*  *******::***..*: ******

YP_009724390.1      PFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQS      116
NP_828851.1         PFYSNVTGFHTIN-------HTFGNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQS      113
                    **:**** **:*:       : *.***:**:**:***:*****::***:**:*::.*:**

YP_009724390.1      LLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFL      176
NP_828851.1         VIIINNSTNVVIRACNFELCDNPFFAVSKPMGT----QTHTMIFDNAFNCTFEYISDAFS      169
                    ::*:**:*****:.*:*::*::**:.*    ..    ::.  ::..* ******:*: *

YP_009724390.1      MDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINIT      236
NP_828851.1         LDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINIT      229
                    :*:. *.****:******** **:: :*. : **::*****.**.:*:*:..**:*****

YP_009724390.1      RFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPL      296
NP_828851.1         NFRAILTAFS------PAQDIWGTSAAAYFVGYLKPTTFMLKYDENGTITDAVDCSQNPL      283
                    .*:::*: .        :.. * :.****:****:* **:***:***********: :**

YP_009724390.1      SETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRK      356
NP_828851.1         AELKCSVKSFEIDKGIYQTSNFRVVPSGDVVRFPNITNLCPFGEVFNATKFPSVYAWERK      343
                    :* **::*** ::*********** *: .:*******************:* *****:**

YP_009724390.1      RISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTG      416
NP_828851.1         KISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTG      403
                    :**************: ********** ********:********::**:**********

YP_009724390.1      KIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAG      476
NP_828851.1         VIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSPD      463
                     ************ ***:***:.*:*:.  ***** ** :*:.:*:*******.  :. .

YP_009724390.1      STPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKN      536
NP_828851.1         GKPCTP-PALNCYWPLNDYGFYTTTGIGYQPYRVVVLSFELLNAPATVCGPKLSTDLIKN      522
                    ..**.   .:***:**:.***  *.*:****************:********* **:*:**

YP_009724390.1      KCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVS      596
NP_828851.1         QCVNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDSVRDPKTSEILDISPCAFGGVS      582
                    :**************** *.*:* ********::* **:****:* *****:**:*****

YP_009724390.1      VITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHV      656
NP_828851.1         VITPGTNASSEVAVLYQDVNCTDVSTAIHADQLTPAWRIYSTGNNVFQTQAGCLIGAEHV      642
                    *******:*.:***********:*  .*********:**:****.*****:**********

YP_009724390.1      NNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPT      716
NP_828851.1         DTSYECDIPIGAGICASYHTVSL----LRSTSQKSIVAYTMSLGADSSIAYSNNTIAIPT      698
                    :.****************:*  :      **.:.:**:********:.*:*****:*****

YP_009724390.1      NFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK      776
NP_828851.1         NFSISITTEVMPVSMAKTSVDCNMYICGDSTECANLLLQYGSFCTQLNRALSGIAAEQDR      758
                    **:**:***::****:******.**********:*****************:***.***:

YP_009724390.1      NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQ      836
NP_828851.1         NTREVFAQVKQMYKTPTLKYFGGFNFSQILPDPLKPTKRSFIEDLLFNKVTLADAGFMKQ      818
                    **:********:****  :* ************* **:********************:**

YP_009724390.1      YGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQI      896
NP_828851.1         YGECLGDINARDLICAQKFNGLTVLPPLLTDDMIAAYTAALVSGTATAGWTFGAGAALQI      878
                    **:***** ********************:*** **:**::** *:************

YP_009724390.1      PFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNA      956
NP_828851.1         PFAMQMAYRFNGIGVTQNVLYENQKQIANQFNKAISQIQESLTTTSTALGKLQDVVNQNA      938
                    ************************ ******.**.:**:**::*::*************

YP_009724390.1      QALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAA      1016
NP_828851.1         QALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAA      998
                    ************************************************************

YP_009724390.1      EIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFT      1076
NP_828851.1         EIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQAAPHGVVFLHVTYVPSQERNFT      1058
                    **************************************:*************:**:***

YP_009724390.1      TAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNT      1136
NP_828851.1         TAPAICHEGKAYFPREGVFVFNGTSWFITQRNFFSPQIITTDNTFVSGNCDVVIGIINNT      1118
                    *******:***:********* *** **:*****:.*********************:***

YP_009724390.1      VYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNES      1196
NP_828851.1         VYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNES      1178
                    ************************************************************

YP_009724390.1      LIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKF      1256
NP_828851.1         LIDLQELGKYEQYIKWPWYVWLGFIAGLIAIVMVTILLCCMTSCCSCLKGACSCGSCCKF      1238
                    *******************:****************:*************.*********

YP_009724390.1      DEDDSEPVLKGVKLHYT   1273
NP_828851.1         DEDDSEPVLKGVKLHYT   1255
                    *****************
```

Figure S4. Pairwise sequence alignment of the reference sequences of the spike proteins of SARS-CoV (accession ID: NP_828851.1) and SARS-CoV-2 (accession ID: YP_009724390.1). The sequence alignment performed using Clustal Omega server (https://www.ebi.ac.uk/Tools/msa/clustalo/). Total 28 predicted B-cell epitopes are shown by red lines, and the final dominant epitopes are indicated as blue lines. the residues marked by green line are the membrane peptides of SARS-CoV spike protein recognized in SARS convalescent sera by antibodies to IgG, IgM, and IgA. Identical residues are shown by *.

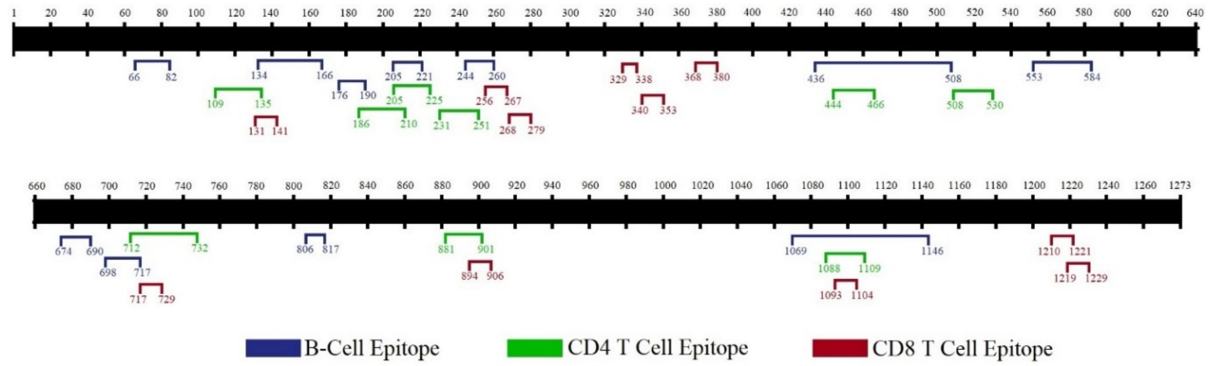


Figure S5. The scaled map shows the position of final chosen immunodominant B- and T-cell epitopes in the spike glycoprotein sequence. We selected three residue fragments as important domains for vaccine design, including I$_{100}$–N$_{280}$, T$_{430}$–C$_{590}$, and V$_{1060}$–E$_{1150}$.

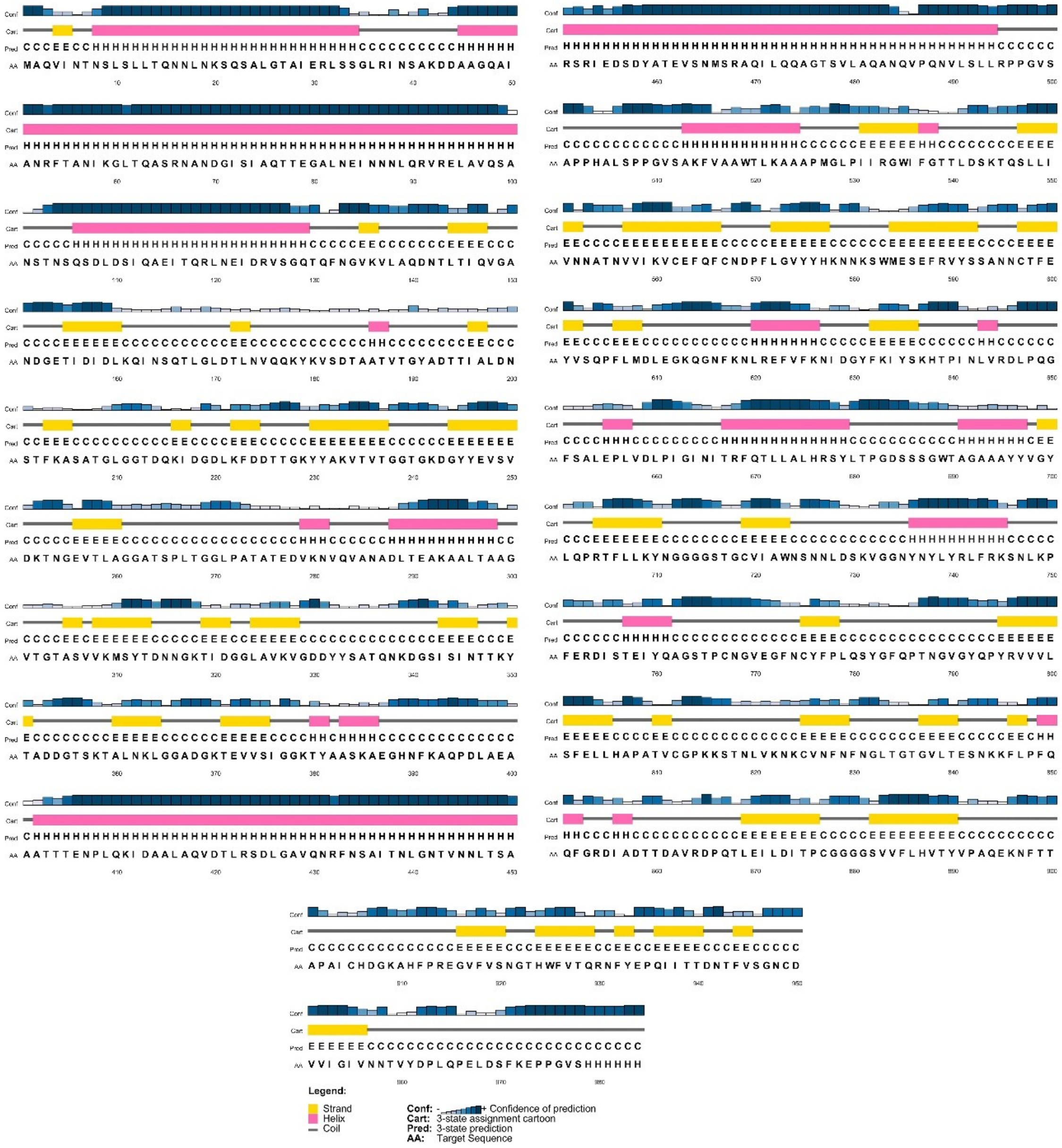


Figure S6. The secondary structure of the designed multi-domain vaccine construct, predicted by PSIPRED web server.

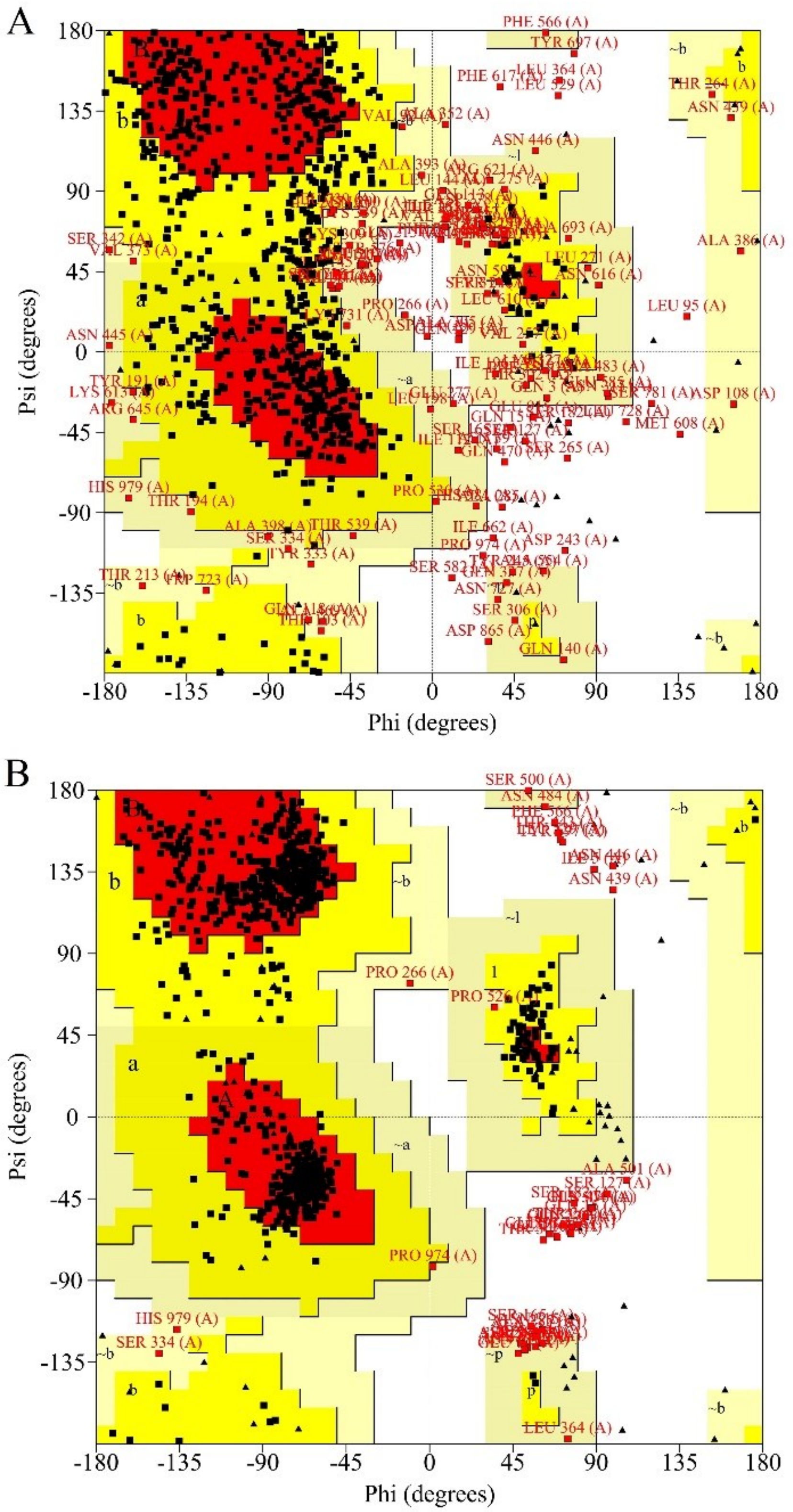


Figure S7. Modeling validation of the 3D structure of the multi-domain vaccine construct. **A**) The Ramachandran plot of the initially modeled vaccine. The Ramachandran plot indicates 26 residues (3.0%) are located in the disallowed region, while 399 of the residues (46.6%) are in the most favorable region, and the rest of them (50.4%) are located in the additionally and generously allowed regions. **B**) The Ramachandran plot of the energy minimized vaccine model. The Ramachandran plot shows 23 residues (2.7%) are placed in the disallowed region, while 710 residues (82.8%) are in the most favorable region, and the rest of them (14.5%) are located in the additionally and generously allowed regions.

Table S1. The VBA code to automate "Addspace" function in Excel.

```
Function AddSpace(Str As String) As String
  Dim i As Long
  For i = 1 To Len(Str)
    AddSpace = AddSpace & Mid(Str, i, 1) & " "
  Next i
  AddSpace = Trim(AddSpace)
End Function
```

Table S2. List of all experimentally (B-cell assay) determined human's Spike glycoprotein SARS-CoV-derived B-cell epitopes obtained from ViPR web-server (Reference sequence accession no. NP_828851).

IEDB ID	Epitope Sequence	Result	Method	Measurement
307	AALVSGTATAGWTFGAG	P	ELISA	qualitative binding
462	AATKMSECVLGQSKRVD	P	ELISA	qualitative binding
525	AAYFVGYLKPTTFMLKY	P	ELISA	qualitative binding
1460	AGCLIGAEHVDTSYECD	P	ELISA	qualitative binding
2053	AIPTNFSISITTEVMPV	P	ELISA	qualitative binding
2092	AISSVLNDILSRLDKVE	P	ELISA	qualitative binding
3983	AQKFNGLTVLPPLLTDD	P	ELISA	qualitative binding
4129	ARDLICAQKFNGLTVLP	N	ELISA	qualitative binding
5773	AYFPREGVFVFNGTSWF	P	ELISA	qualitative binding
5927	AYSNNTIAIPTNFSISI	P	ELISA	qualitative binding
6040	CASYHTVSLLRSTSQKS	P	ELISA	qualitative binding
6110	CDIPIGAGICASYHTVS	P	ELISA	qualitative binding
6332	CGPKLSTDLIKNQCVNF	P	ELISA	qualitative binding
6333	CGPKLSTDLIKNQCVNFNFNGL TGTGVLTPSSKRFQPFQQFG	P	ELISA	qualitative binding
6334	CGPKLSTDLIKNQCVNFNFNGL TGTGVLTPSSKRFQPFQQFGRDVSDFTD	P	western blot	qualitative binding
6562	CLGDINARDLICAQKFN	N	ELISA	qualitative binding
7064	CSQNPLAELKCSVKSFE	P	ELISA	qualitative binding
7066	CSQNPLAELKCSVKSFEIDKGIYQTSNFRVVPSGD	P	ELISA	qualitative binding
7128	CTDVSTAIHADQLTPAW	P	ELISA	qualitative binding
7129	CTDVSTAIHADQLTPAWRIYSTGNNVFQTQAG	P	ELISA	qualitative binding
7193	CTPPALNCYWPLNDYGF	P	ELISA	qualitative binding
7217	CTTFDDVQAPNYTQHTSSMRGVYYPDEIFR	P	ELISA	qualitative binding
7280	CVLAWNTRNIDATSTGN	P	ELISA	qualitative binding
7383	CYGVSATKLNDLCFSNV	P	ELISA	qualitative binding
7452	CYWPLNDYGFYTTTGIG	P	ELISA	qualitative binding
7868	DDSEPVLKGVKLHYT	P	ELISA	qualitative binding
8239	DFCGKGYHLMSFPQAAP	P	ELISA	qualitative binding
8413	DGIYFAATEKSNVVRGW	P	ELISA	qualitative binding
9094	DLGDISGINASVVNIQK	P	ELISA	qualitative binding
9242	DLPSGFNTLKPIFKLPL	P	ELISA	qualitative binding
9365	DMIAAYTAALVSGTATA	P	ELISA	qualitative binding
9956	DRCTTFDDVQAPNYTQH	P	ELISA	qualitative binding
10721	DVSEKSGNFKHLREFVF	P	ELISA	qualitative binding
10778	DVVNQNAQALNTLVKQL	P	ELISA	qualitative binding
11038	EAEVQIDRLITGRLQSL	P	ELISA	qualitative binding
11740	EELDKYFKNHTSPDVDL	N	ELISA	qualitative binding
12417	EIDKGIYQTSNFRVVPS	P	ELISA	qualitative binding
14208	ESLTTTSTALGKLQDVV	P	ELISA	qualitative binding
14626	EVAKNLNESLIDLQELG	P	ELISA	qualitative binding
15783	FFAVSKPMGTQTHTMIF	P	ELISA	qualitative binding
15903	FFSTFKCYGVSATKLND	P	ELISA	qualitative binding
16183	FIEDLLFNKVTLADAGF	P	ELISA	qualitative binding
16376	FKHLREFVFKNKDGFLY	P	ELISA	qualitative binding
16417	FKNKDGFLYVYKGYQPI	P	ELISA	qualitative binding
17132	FNATKFPSVYAWERKKI	N	ELISA	qualitative binding
17644	FRSDTLYLTQDLFLPFY	P	ELISA	qualitative binding
17929	FTDSVRDPKTSEILDIS	P	ELISA	qualitative binding

18161	FVFNGTSWFITQRNFFS	P	ELISA	qualitative binding
18515	GAALQIPFAMQMAYRFN	P	ELISA	qualitative binding
19585	GFMKQYGECLGDINARD	P	ELISA	qualitative binding
19657	GFYTTTGIGYQPYRVVV	P	ELISA	qualitative binding
20288	GIGVTQNVLYENQKQIA	P	ELISA	qualitative binding
23051	GVLTPSSKRFQPFQQFG	P	ELISA	qualitative binding
23053	GVLTPSSKRFQPFQQFGRDVSDFTDSVRDPK	P	ELISA	qualitative binding
23296	GWVFGSTMNNKSQSVII	P	ELISA	qualitative binding
23437	GYQPYRVVVLSFELLNA	P	ELISA	qualitative binding
23520	HADQLTPAWRIYSTGNN	P	ELISA	qualitative binding
24923	HTINHTFGNPVIPFKDG	P	ELISA	qualitative binding
24978	HTSSMRGVYYPDEIFRS	P	ELISA	qualitative binding
25031	HVDTSYECDIPIGAGIC	P	ELISA	qualitative binding
25219	IAAEQDRNTREVFAQVK	P	ELISA	qualitative binding
25250	IADYNYKLPDDFMGCVL	P	ELISA	qualitative binding
25293	IAGLIAIVMVTILLCCM	P	ELISA	qualitative binding
25378	IAPGQTGVIADYNYKLP	P	ELISA	qualitative binding
26610	IINNSTNVVIRACNFEL	P	ELISA	qualitative binding
27357	ILSRLDKVEAEVQIDRL	P	ELISA	qualitative binding
28548	ISITTEVMPVSMAKTSV	N	ELISA	qualitative binding
29108	ITTDNTFVSGNCDVVIG	N	ELISA	qualitative binding
29832	KAISQIQESLTTTSTAL	P	ELISA	qualitative binding
30435	KEIDRLNEVAKNLNESL	N	ELISA	qualitative binding
31581	KKISNCVADYSVLYNST	P	ELISA	qualitative binding
32037	KLNDLCFSNVYADSFVV	P	ELISA	qualitative binding
32070	KLPDDFMGCVLAWNTRN	P	ELISA	qualitative binding
32508	KNHTSPDVDLGDISGIN	P	ELISA	qualitative binding
32912	KPTTFMLKYDENGTITD	P	ELISA	qualitative binding
33032	KQLSSNFGAISSVLNDI	N	ELISA	qualitative binding
33358	KSIVAYTMSLGADSSIA	P	ELISA	qualitative binding
33874	KTSVDCNMYICGDSTEC	P	ELISA	qualitative binding
34217	KVTLADAGFMKQYGECL	N	ELISA	qualitative binding
35083	LCPFGEVFNATKFPSVY	P	ELISA	qualitative binding
36167	LGINITNFRAILTAFSP	P	ELISA	qualitative binding
36579	LIKNQCVNFNFNGLTGT	P	ELISA	qualitative binding
36815	LKCSVKSFEIDKGIYQT	P	ELISA	qualitative binding
36856	LKGACSCGSCCKFDEDD	P	ELISA	qualitative binding
36991	LKPIFKLPLGINITNFR	P	ELISA	qualitative binding
37758	LLRSTSQKSIVAYTMSL	P	ELISA	qualitative binding
38118	LMSFPQAAPHGVVFLHV	P	ELISA	qualitative binding
39023	LQYGSFCTQLNRALSGI	P	ELISA	qualitative binding
39110	LRHGKLRPFERDISNVP	N	ELISA	qualitative binding
40905	LYQDVNCTDVSTAIHAD	P	ELISA	qualitative binding
41177	MAYRFNGIGVTQNVLYE	P	ELISA	qualitative binding
41504	MFIFLLFLTLTSGSDLD	P	ELISA	qualitative binding
41984	MLKYDENGTITDAVDCS	P	ELISA	qualitative binding
42999	MVTILLCCMTSCCSCLK	P	ELISA	qualitative binding
43145	NAFNCTFEYISDAFSLD	P	ELISA	qualitative binding
43264	NAPATVCGPKLSTDLIK	P	ELISA	qualitative binding
43313	NASSEVAVLYQDVNCTD	P	ELISA	qualitative binding
43825	NFELCDNPFFAVSKPMG	P	ELISA	qualitative binding
43834	NFFSPQIITTDNTFVSG	N	ELISA	qualitative binding
43893	NFNGLTGTGVLTPSSKR	P	ELISA	qualitative binding
43924	NFSQILPDPLKPTKRSF	N	ELISA	qualitative binding
44198	NIDATSTGNYNYKYRYL	P	ELISA	qualitative binding
45151	NNKSQSVIIINNSTNVV	P	ELISA	qualitative binding
45506	NPVIPFKDGIYFAATEK	P	ELISA	qualitative binding
46240	NTREVFAQVKQMYKTPT	N	ELISA	qualitative binding
46379	NVFQTQAGCLIGAEHVD	P	ELISA	qualitative binding
46522	NVYADSFVVKGDDVRQI	P	ELISA	qualitative binding
46681	NYNYKYRYLRHGKLRPF	P	ELISA	qualitative binding
76972	P462	P	ELISA, biological activity, surface Plasmon resonance (SPR)	qualitative binding, neutralization, dissociation constant KD
47341	PELDSFKEELDKYFKNH	N	ELISA	qualitative binding
47479	PFAMQMAYRFNGIGVTQ	P	ELISA	qualitative binding
47823	PHGVVFLHVTYVPSQER	P	ELISA	qualitative binding
48383	PLKPTKRSFIEDLLFNK	P	ELISA	qualitative binding
48554	PMGTQTHTMIFDNAFNC	P	ELISA	qualitative binding
49557	PSVYAWERKKISNCVAD	P	ELISA	qualitative binding
49968	PVSMAKTSVDCNMYICGDS	P	ELISA	qualitative binding
50058	PWYVWLGFIAGLIAIVM	P	ELISA	qualitative binding
50311	QALNTLVKQLSSNFGAI	P	ELISA	qualitative binding
50767	QFGRDVSDFTDSVRDPK	P	ELISA	qualitative binding
51043	QIANQFNKAISQIQESL	P	ELISA	qualitative binding
51379	QLIRAAEIRASANLAAT	P	ELISA	qualitative binding

51456	QLNRALSGIAAEQDRNT	P	ELISA	qualitative binding
53116	RAILTAFSPAQDIWGTS	P	ELISA	qualitative binding
53202	RASANLAATKMSECVLG	P	ELISA	qualitative binding
53766	RFQPFQQFGRDVSDFTD	P	ELISA	qualitative binding
54599	RLITGRLQSLQTYVTQQ	P	ELISA	qualitative binding
54989	RNFTTAPAICHEGKAYF	P	ELISA	qualitative binding
55149	RPFERDISNVPFSPDGK	P	ELISA	qualitative binding
58001	SGDVVRFPNITNLCPFG	P	ELISA	qualitative binding
58143	SGNCDVVIGIINNTVYD	P	ELISA	qualitative binding
59120	SLGADSSIAYSNNTIAI	P	ELISA	qualitative binding
59425	SLQTYVTQQLIRAAEIR	P	ELISA	qualitative binding
59944	SNVPFSPDGKPCTPPAL	P	ELISA	qualitative binding
59985	SPAQDIWGTSAAAYFVG	P	ELISA	qualitative binding
60014	SPDGKPCTPPALNCYWP	P	ELISA	qualitative binding
62872	TAGWTFGAGAALQIPFA	P	ELISA	qualitative binding
62908	TALGKLQDVVNQNAQAL	P	ELISA	qualitative binding
63309	TECANLLLQYGSFCTQL	P	ELISA	qualitative binding
64085	THTMIFDNAFNCTFEYI	N	ELISA	qualitative binding
64477	TITDAVDCSQNPLAELK	N	ELISA	qualitative binding
64888	TLKYFGGFNFSQILPDP	P	ELISA	qualitative binding
65110	TLTSGSDLDRCTTFDDV	P	ELISA	qualitative binding
65479	TNVVIRACNFELCDNPF	P	ELISA	qualitative binding
65832	TQDLFLPFYSNVTGFHT	P	ELISA	qualitative binding
66337	TSNFRVVPSGDVVRFPN	P	ELISA	qualitative binding
67220	TVYDPLQPELDSFKEEL	P	ELISA	qualitative binding
68971	VIGIINNTVYDPLQPEL	P	ELISA	qualitative binding
69299	VKQMYKTPTLKYFGGFN	P	ELISA	qualitative binding
69513	VLGQSKRVDFCGKGYHL	N	ELISA	qualitative binding
69699	VLPPLLTDDMIAAYTAA	P	ELISA	qualitative binding
69760	VLSFELLNAPATVCGPK	P	ELISA	qualitative binding
69865	VLYENQKQIANQFNKAI	P	ELISA	qualitative binding
70527	VQAPNYTQHTSSMRGVY	P	ELISA	qualitative binding
70719	VRFPNITNLCPFGEVFN	P	ELISA	qualitative binding
71189	VSVITPGTNASSEVAVL	P	ELISA	qualitative binding
71191	VSVITPGTNASSEVAVLYQDVNCTDVSTAIHADQ LTPAWRIYSTGNNVFQ	P	western blot	qualitative binding
71589	VTYVPSQERNFTTAPAI	P	ELISA	qualitative binding
72205	VYYPDEIFRSDTLYLTQD	P	ELISA	qualitative binding
72519	WGTSAAAYFVGYLKPTT	P	ELISA	qualitative binding
73036	WRIYSTGNNVFQTQAGC	P	ELISA	qualitative binding
74173	YICGDSTECANLLLQYG	P	ELISA	qualitative binding
74367	YISDAFSLDVSEKSGNF	P	ELISA	qualitative binding
75528	YQPIDVVRDLPSGFNTL	P	ELISA	qualitative binding
75843	YSNVTGFHT	P	ELISA	qualitative binding
75920	YSVLYNSTFFSTFKCYG	P	ELISA	qualitative binding
76446	YVYKGYQPIDVVRDLPS	P	ELISA	qualitative binding

P: positive; **N**: negative;

Table S3. Sequence-based features of complete genome of the Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1 (NCBI reference sequence accession No. NC_045512.2).

Genome Parts	**Position**	**Gene**	**Product**	**Note**	**Product**
5' UTR	1...265				
Gene	266...21555	orf1ab			
CDS	266...13468 13468...21555	orf1ab	orf1ab polyprotein	Ribosomal slippage; pp1ab; translated by -1 ribosomal frameshift	orf1ab polyprotein
Mature peptide	266...805	orf1ab	Leader protein	nsp1. Produced by both pp1a and pp1ab	Leader protein
Mature peptide	806...2719	orf1ab	Nsp2	Produced by both pp1a and pp1ab	Nsp2

Mature peptide	2720...8554	orf1ab	Nsp3	Former nsp1; conserved domains are: N-terminal acidic (Ac), predicted phosphoesterase, papain-like proteinase, Y-domain, transmembrane domain 1 (TM1), adenosine diphosphate-ribose 1"-phosphatase (ADRP). produced by both pp1a and pp1ab	Nsp3
Mature peptide	8555...10054	orf1ab	Nsp4	Contains transmembrane domain 2 (TM2); produced by both pp1a and pp1ab	Nsp4
Mature peptide	10055...10972	orf1ab	3C-like proteinase	nsp5; main proteinase (Mpro); mediates cleavages downstream of nsp4. 3D structure has been determined. Produced by both pp1a and pp1ab	3C-like proteinase
Mature peptide	10973…11842	Orf1ab	Nsp6	Putative transmembrane domain. Produced by both pp1a and pp1ab.	Nsp6
Mature peptide	11843…12091	Orf1ab	Nsp7	Produced by both pp1a and pp1ab.	Nsp7
Mature peptide	12092…12685	Orf1ab	Nsp8	Produced by both pp1a and pp1ab.	Nsp8
Mature peptide	12686…13024	Orf1ab	Nsp9	ssRNA-binding protein. Produced by both pp1a and pp1ab	Nsp9
Mature peptide	13025…13441	Orf1ab	Nsp10	Formerly known as growth-factor-like protein (GFL). Produced by both pp1a and pp1ab.	Nsp10
Mature peptide	13442…13468 13468…16236	Orf1ab	RNA-dependent RNA polymerase (RdRp)	nsp12. Produced by pp1ab only.	RNA-dependent RNA polymerase (RdRp)
Mature peptide	16237…18039	Orf1ab	Helicase	nsp13; zinc-binding domain (ZD), NTPase/helicase domain (HEL), RNA 5'-triphosphatase; Produced by pp1ab only.	Helicase
Mature peptide	18040…19620	Orf1ab	3'-to-5' exonuclease	nsp14. Produced by pp1ab only.	3'-to-5' exonuclease
Mature peptide	19621…20658	Orf1ab	endoRNAse	nsp15 Produced by pp1ab only.	endoRNAse
Mature peptide	20659…21552	Orf1ab	2'-O-ribose methyltransferase (2'-o-MT)	nsp16 Produced by pp1ab only	2'-O-ribose methyltransferase (2'-o-MT)
CDS	266…13483	Orf1ab	orf1a polyprotein	pp1a	orf1a polyprotein
Mature peptide	13442…13480	Orf1ab	Nsp11	Produced by pp1a only	Nsp11
Gene	21563…25384	S			
CDS	21563…25384	S	Surface glycoprotein	Structural protein; spike protein	Surface glycoprotein
Gene	25393…26220	Orf3a			

CDS	25393…26220	Orf3a	Orf3a protein		Orf3a protein
Gene	26245…26472	E			
CDS	26245…26472	E	Envelope protein	orf4, structural protein; E protein	Envelope protein
Gene	26523…27191	M			
CDS	26523…27191	M	Membrane glycoprotein	orf5, structural protein	Membrane glycoprotein
Gene	27202…27387	Orf6			
CDS	27202…27387	Orf6	Orf6 protein		Orf6 protein
Gene	27394…27759	Orf7a			
CDS	27394…27759	Orf7a	Orf7a protein		
Gene	27756…27887	Orf7b			
CDS	27756…27887	Orf7b	Orf7b protein		
Gene	27894…28259	Orf8			
CDS	27894…28259	Orf8	Orf8 protein		
Gene	28274…29533	N			
CDS	28274…29533	N	Nucleocapsid phosphoprotein	orf9; structural protein	
Gene	29558…29674	Orf10			
CDS	29558…29674	Orf10	Hypothetical Orf10 protein		
3' UTR	29675…29903				

CDS: coding sequence; **UTR**: untranslated region; **ORF**: open reading frame. **NSP**: non-structural protein.