ARTICLE

Design of a possible subunit vaccine against SARS-CoV-2

Cecilia P. Mikita[1] and Eduardo A. Padlan*[,2]

[1]Walter Reed National Military Medical Center, Department of Medicine, Allergy/Immunology/Immunizations Service, 4954 North Palmer Road, Bethesda, MD 20889-5600, USA
[2]4006 Simms Drive, Kensington, MD 20895-1336, USA

ABSTRACT

A possible subunit vaccine that focuses the immune response to the receptor-binding site of the spike glycoprotein of SARS-CoV-2 is designed. Sequence and three-dimensional structural data are used to map the epitopes of neutralizing antibodies on a suitable spike protein for use as a subunit vaccine. The chemical reactivity of the residues in the spike protein is reduced by judicious amino acid replacements, except for the residues in the epitopes. This results in a molecule that would elicit an immune response directed at the known epitopes. The proposed subunit vaccine is expected to work regardless of the existence of mutations in one or more, but not all, of the epitopes, as well as elsewhere in the spike protein. The amino acid sequence of the proposed vaccine may serve as the starting point for even more efficacious vaccines if more neutralizing antibody epitopes are included.

KEYWORDS

SARS-CoV-2, spike glycoprotein, receptor-binding site, subunit vaccine

*Corresponding author
Email Address: eduardo.padlan@gmail.com
Date received: March 2, 2021
Date revised: April 20, 2021
Date accepted: April 23, 2021

INRODUCTION

The global scientific community is unified in their effort to combat the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus, which causes coronavirus disease 2019 (COVID-19). Since its discovery in December 2019, COVID-19 has affected over 134 million people and is associated with the more than 2.9 million deaths worldwide (WHO Coronavirus Disease (COVID-19) Dashboard. Data last updated: 2021/4/10, 2.52pm CET). Worldwide, the strategy to combat the COVID-19 pandemic includes vaccination, physical distancing, personal protective equipment, and targeted medications. There are approximately two hundred different vaccine candidates in all stages of development worldwide. All of the vaccine candidates in development target a portion of the S protein and a few, including two mRNA vaccines, have received emergency use authorization (EUA) by the US Food and Drug Administration and other agencies, to administer to at-risk populations. Recombinant viral-vectored, inactivated or killed virus, protein subunit, live attenuated virus, virus-like particles, and plasmid DNA based vaccines are also in development to combat SARS-CoV-2 and some are expected to apply for EUA approval in early 2021.

SARS-CoV-2 is a single stranded RNA virus which belongs to the *Coronaviridae* family. SARS-CoV-2 enters the host through bronchial and alveolar epithelial cells through the S, or spike, protein. An envelope glycoprotein, the S protein is expressed on the virus surface and has two distinct subunits. The S1 region of the S protein binds to the angiotensin converting enzyme 2 (ACE2) receptor on host cells, and the S2 region is responsible for membrane fusion. Ideally, SARS-CoV-2 vaccine candidates should produce antibodies targeting the S protein, which can neutralize the virus and prevent entry into host cells.

Table 1

```
          10        20        30        40        50        60        70
          |         |         |         |         |         |         |
MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHV
             QCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHV
             ACVSLTTATALPPAATSSTTRGVYAPTTVAASSVLTSTTTLFLPFTSSVTWFTAITV mut
         80        90        100       110       120       130       140
          |         |         |         |         |         |         |
SGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPF
SGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPF
SGTSGTTAFSNPVLPFTTGVYFASTEKSTIIRGWIFGTTLTSTTTSLLIVTTATTVVIKVCTFTFCTTPF mut
         150       160       170       180       190       200       210
          |         |         |         |         |         |         |
LGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPI
LGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPI
LGVTTTTTTTSVMESETTVTSSATTCTTTTVSTPTLMSLAGTAGSFTGLAETVFKTITGTFTIYSTTTPI mut
         220       230       240       250       260       270       280
          |         |         |         |         |         |         |
NLVRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYN
NLVRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYN
TLVASLPTGASALAPLVTLPIGISITAFQTLLALTRSTLTPGGSSSGWTAGAAAYYVGTLTPRTFLLTYT mut
         290       300       310       320       330       340       350
          |         |         |         |         |         |         |
ENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASV
ENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASV
TTGTITGAVDCALTPLSATACTLASATVATGIATTSTFAVAPTASIVRFPTITSLCPAGAVAAATRFASV mut
         360       370       380       390       400       410       420
          |         |         |         |         |         |         |
YAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
YAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
YAWARARISTCVAGTSVLYNSASFSTFKCYGVSPTKLAALCFTSVTADSFVIRGDEVRQIAPGQTGKIAD mut
         430       440       450       460       470       480       490
          |         |         |         |         |         |         |
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYF
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYF
YNYALPDDFTGCVIATTSTNLDSKVGGNYNYLYRLFRKSNLTPAERGISTEIYQAGSTPCNGVEGFNCYF mut
         500       510       520       530       540       550       560
          |         |         |         |         |         |         |
PLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL
PLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL
PLQSYGFQPTNGVGYQPYRVVVLSFTLLTAPATVCGPTTSTSLVTSTCVGFNFNGLTGTGVLTTSTTTFL mut
         570       580       590       600       610       620       630
          |         |         |         |         |         |         |
PFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQGVNCTEVPVAIHADQLT
PFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCT-----------
PAAQFGASIASTTGAVRDPTTLTILSITPCSTGGVSVITPGTTTSGTVAVLYTGVTCTAVPVAIAAATLT mut
         640       650       660       670       680       690       700
          |         |         |         |         |         |         |
PTWRVYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLG
-TWRVYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQ----------SQSIIAYTMSLG
PTWAVASTGSGVATTTAGCLIGATTVSTSTTCTIPIGAGICASTTTTTTSPAAAASVASTSIIATTMSLG mut
         710       720       730       740       750       760       770
          |         |         |         |         |         |         |
AENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGI
AENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGI
AASSVAASSSSIAIPTTFTISVTTTILPVSMTKTSVSCTMYICGASTTCSALLLAAGSACTALAAALTGI mut
         780       790       800       810       820       830       840
          |         |         |         |         |         |         |
AVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC
AVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC
AVAQAANTAAVFAAVAAIATTPPITSFGGFSFSAILPGPSTPSTRSAIETLLAAAVTLAGAGAITTTGSC mut
         850       860       870       880       890       900       910
          |         |         |         |         |         |         |
LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIG
LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIG
```

```
LGAIAAAALICAAAAAGLTVLPPLLTSTMIAAATSALLAGTITSGATFGAGAALTIPFAMQMAARFAGIG mut
        920       930       940       950       960       970       980
         |         |         |         |         |         |         |
VTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDI
VTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDI
VTAGVLAAAAALIAAAFASAIGAIAASLSSTASALGALAAVVAANAAALATLVAALSSAAGAISSVLSGI mut
        990      1000      1010      1020      1030      1040      1050
         |         |         |         |         |         |         |
LSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLM
LSRLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLM
LSALAAVTAAVQIAALITGRLASLATYVTAQLIAAAAIAASAALAATKMSACVLGQSTAVSACGTGTHLM mut
       1060      1070      1080      1090      1000      1100      1120
         |         |         |         |         |         |         |
SFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT
SFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT
SFPQSAPTGVVFLHVTYVPATTTTFTTAPAICTSGTAHAPTTGVFVSTGTTAAVTTTTTATPTIITTSST mut
       1130      1140      1150      1160      1170      1180      1190
         |         |         |         |         |         |         |
FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVA
FVSGNCDVVIGIVNNTVYDPLQPELD
AVSGTCTVVIGIVGGTVTSPLTPTLA                                              mut
       1200      1210      1220      1230      1240      1250      1260
         |         |         |         |         |         |         |
KNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD
       1270
         |
SEPVLKGVKLHYT
```

Shown is the sequence of NCBI Entry QIK50427, the most common spike glycoprotein of the SARS-CoV-2 virus in the NCBI database. The spike protein has two main parts: S1 comprised by residues 14 to 685 and S2 comprised by residues 686 to 1273. The receptor binding domain (RBD) is comprised by residues 319 to 541 (shown bold and underlined) and is responsible for binding to the angiotensin-converting enzyme 2 (ACE2) of the target cell. The sequence of the spike structure in PDB Entry 7jji, which extends only to residue 1146, is shown in green. Note the existence of the two gaps in 7jji that were "repaired" using PDB Entry 6xm0 for the first gap and by modeling using SWISS-MODEL for the second gap. The T-cell epitopes (from IEDB) are underlined in the "mut" sequence.

The proposed protein subunit vaccine in this paper targets the S1 region of the S protein, specifically the receptor binding domain (RBD) of the S protein, and is designed to induce antibodies to the epitopes of known neutralizing antibodies. We used a method, proposed earlier by one of us (Padlan 2010) that enhances the antigenicity of a chosen region of a protein molecule, to direct the immune response to the epitopes of the neutralizing antibodies. We propose that our designed subunit vaccine is ideal for the SARS-CoV-2 virus, regardless of anticipated virus mutations.

MATERIALS AND METHODS

From the available sequence and structural data on the spike glycoprotein of SARS-CoV-2, a suitable structure is constructed for the design of a possible subunit vaccine that focuses the immune response to the receptor-binding site of the molecule. The spike glycoprotein has 1273 amino acids which are divided into two structural regions: S1 comprised by residues 14 to 685 and S2 comprised by residues 686 to 1273. The region that binds the receptor angiotensin converting enzyme 2 (ACE2) on cells is in the S1 region. The receptor binding domain (RBD) of the spike glycoprotein is comprised by residues 319 to 541. RBD is the focus of our analysis.

Sequence data

Sequences for the SARS-CoV-2 spike glycoprotein were obtained from the National Center for Biological Information (NCBI) database (www.ncbi.nlm.nih.gov). Only complete sequences in which all the residues had been identified were included in our collection. In all, 3143 sequences were obtained from NCBI. Comparing the sequences revealed numerous duplications. A total of 1804 sequences are identical (represented by (NCBI Entry Code) QIK50427), 993 are identical (represented by QHD43416), 36 are identical (represented by QJQ84676), 20 are identical (represented by QIZ54578), 6 sequences are identical (represented by QJA17268), and another 6 sequences are identical (represented by QJE38426). The sequence of QIK50427 is shown in Table 1. The differences between these more popular sequences are minor: QHD43416 differs from QIK50427 only at position 614 (D in QHD43416 vs G in QIK50427); QJQ84676 differs from QIK50427 at position 614 (D in QJQ84676 vs G in QIK50427) and at position 829 (T in QJQ84676 vs A in QIK50427); QIZ54578 differs from QIK50427 at position 146 (Y in QIZ54578 vs H in QIK50427); QJA17268 differs from QIK50427 at position 5 (F in QJA17268 vs L in QIK50427); and QJE38426 differs from QIK50427 at position 614 (D in QJE38426 vs G in QIK50427) and at position 845 (S in QJE38426 vs A in QIK50427). None of these differences occur in the RBD region of the molecule (see, Table 1). In view of this close similarity in the most popular sequences of the spike protein, we will confine our analysis to structures with the QIK50427 sequence.

Structural data

Three-dimensional structures for the spike protein obtained by X-ray crystallography and electron microscopy (EM) are available from the Protein Data Bank (PDB) (www.rcsb.org). We collected structures of complexes of the spike protein with its cellular receptor ACE2 and with CR3022 (Yuan et al. (2020)

Table 2

PDB Code	Structure analysis by	Resolution/R-value	Bound structure
6m0j	X-ray	2.45A/0.227	ACE2
6lzg	X-ray	2.50A/0.216	ACE2
7a91	EM	3.60A	ACE2
6yla	X-ray	2.42A/0.237	CR3022
6xc3	X-ray	2.70A/0.225	CR3022 (+ 1 neut. Ab)
6z2m	X-ray	2.71A/0.241	CR3022
6xc7	X-ray	2.88A/0.259	CR3022 (+ 1 neut. Ab)
6yor	EM	3.30A	CR3022
7jmp	X-ray	1.71A/0.209	neutralizing Ab
7bz5	X-ray	1.84A/0.191	neutralizing Ab
6xc4	X-ray	2.34A/0.219	neutralizing Ab
7jmo	X-ray	2.38A/0.238	neutralizing Ab
7chb	X-ray	2.40A/0.194	neutralizing Ab
6xkq	X-ray	2.55A/0.207	neutralizing Ab
7ch5	X-ray	2.70A/0.229	neutralizing Ab
7chc	X-ray	2.71A/0.216	neutralizing (2)Abs
6xkp	X-ray	2.72A/0.222	neutralizing Ab
6xe1	X-ray	2.75A/0.213	neutralizing Ab
7bwj	X-ray	2.85A/0.221	neutralizing Ab
6zcz	X-ray	2.85A/0.260	neutralizing Ab
7c01	X-ray	2.88A/0.265	neutralizing Ab
7jmw	X-ray	2.89A/0.241	neutralizing Ab
7k9z	X-ray	2.95A/0.287	neutralizing (2)Abs
7k43	EM	2.60A	neutralizing Ab
7jv6	EM	3.00A	neutralizing Ab

Shown are the X-ray and EM structures from the PDB that were included in this study. 6xc3 and 6xc7 each have one other neutralizing antibody bound in addition to CR3022 and are subscripted 1 and 2; 7chc, 7k9z, and 3xdg each have two neutralizing antibodies bound and are similarly subscripted.

and other neutralizing antibodies. Only the structures obtained at relatively high resolution were considered for this study. Our study included X-ray structures obtained at resolutions around 3 Angstroms (3A). The EM structures, which are of lower resolution, were used to support the X-ray results. The structures analyzed in this study are listed in Table 2.

All of the structures that we examined have missing residues and none were complete to the end of the spike sequence. We chose the structure described in PDB Entry 7jji (Bangaru et al. 2020), which has the fewest missing residues, and the QIK50427 sequence as the basis of our design. The residues missing in 7jji are between positions 619 and 631 and between positions 678 and 688 (numbering follows that in Table 1). Both of those gaps are outside the RBD and are exposed (see, Figure 1). To complete the 7jji structure between residues 619 and 631, we incorporated the corresponding peptide in PDB Entry 6xm0 (Zhou et al. 2020) after aligning the bordering structures (residues 600 to 650) using TM-Align (Zhang and Skolnick 2005). None of the PDB structures we examined could be used to model the gap between residues 678 and 688. That part of 7jji was modeled using SWISS-MODEL (Waterhouse et al. 2018) (https://swissmodel.expasy.org) and the QIK50427 sequence. All subsequent calculations used this composite structure, which has the first 1122 residues of the 1273 known to constitute the spike sequence. The subunit vaccine that we are proposing is based on this modeled structure.

Design of a possible subunit vaccine against SARS-CoV-2

In view of the primary function of the RBD in the spike's binding to target cells, we concentrated on RBD, in particular the region

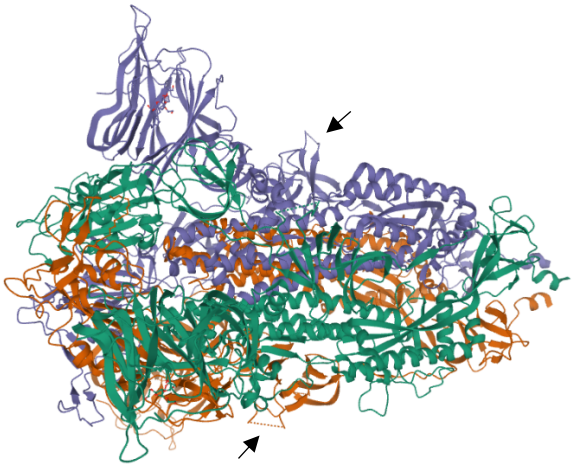

Figure 1: This representation of the structure of spike glycoprotein (PDB Entry 7jji) shows the two gaps (indicated by dots and pointed to by arrows) in the structure: residues 619-631 (lower arrow) in one monomer and residues 678-688 (upper arrow) in a neighboring monomer.

Table 3

```
  320       330       340       350       360       370       380       390       400       410       420
   |         |         |         |         |         |         |         |         |         |         |
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD

RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 6m0j
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 6lzg
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 7a91

        ETGPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 6yla
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 6xc31
           ITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 6z2m
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 6xc71
          PNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 6yor

RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 7jmp
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 7bz5
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 6xc4
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 7jmo
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 7chb
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 6xkq
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 7ch5
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 6xc32
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 7chc1
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 7chc2
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 6xkp
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 6xe1
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 7bwj
             TNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 6zcz
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 6xc72
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 7c01
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 7jmw
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 7k9z1
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 7k9z2
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 7k43

RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD 7jv6


        430       440       450       460       470       480       490       500       510       520
         |         |         |         |         |         |         |         |         |         |
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA

YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 6m0j
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 6lzg
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 7a91

YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 6yla
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 6xc31
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 6z2m
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 6xc71
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 6yor

YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 7jmp
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 7bz5
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 6xc4
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 7jmo
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 7chb
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 6xkq
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 7ch5
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 6xc32
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 7chc1
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 7chc2
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 6xkp
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 6xe1
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 7bwj
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 6zcz
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 6xc72
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 7c01
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 7jmw
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 7k9z1
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 7k9z2
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 7k43
YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPA 7jv6
```

The residues comprising the receptor-binding domain (RBD) in NCBI Entry QIK50427 are shown underlined. The residues in the various ligands (ACE2, CR3022 and other neutralizing antibodies) that are involved in the binding to RBD are shown bold and underlined in the ligand sequences. (Two residues, one in a ligand and one in RBD, are said to be in contact if the residues have atoms that are within the sum of their van der Waals radii plus half an Angstrom.) No ligand residues in contact with RBD are found in the segment 522-541 so that this segment is omitted simply for space considerations.

that has been shown to bind ACE2 and neutralizing antibodies. Here, we use the method described earlier by one of us (Padlan 2010). Briefly, the method enhances the antigenicity of a chosen region by reducing the antigenicity (de-Antigenization) of all the other regions. This should focus the immune response to the chosen region of the antigen. Here, the chosen region is that part of the RBD that has been shown to be involved in the binding to target cells.

Table 4: Replacement rules

Amino acid	If in Helix	Sheet	Coil	Turn
	Change to:			
Arg	Ala	Thr	Ala	Ala
Asn	Ala	Thr	Ser	Gly
Asp	Ala	Thr	Ser	Gly
Gln	Ala	Thr	Ala	Thr
Glu	Ala	Thr	Ala	Thr
His	Ala	Thr	Thr	Thr
Lys	Ala	Thr	Thr	Thr
Phe	Ala	Thr	Ala	Ala
Trp	Ala	Thr	Ala	Val
Tyr	Ala	Thr	Ala	Thr

Ala, Cys, Gly, Ile, Leu, Met, Pro, Ser, Thr, and Val are not replaced.

The amino-acid replacement rules designed to reduce the antigenicity of protein epitopes (extracted from Table 1 in Padlan 2008).

The method of Padlan (2010) assigns chemical reactivities (Sandberg et al. 1998, De Genst et al. 2002) to each amino acid in the structure; the chemical reactivity represents the contribution of an amino acid to the antigenicity of a particular epitope. The chemical reactivities used here are the ZZ3 values from Sandberg et al. (1998) (see also De Genst et al. 2002): the value of -3.5 is assigned to Arg, 1.93 to Asp, -0.11 to Glu, -2.49 to Lys, 1.15 to Ser, 1.04 to Asn, -1.44 to Gln, 0.36 to Gly, 1.84 to Pro, -1.12 to Thr, 0.60 to Ala, 0.26 to His, 3.71 to Cys, 0.47 to Met, -1.54 to Val, -1.71 to Ile, -1.49 to Leu, 0.43 to Tyr, 1.06 to Phe, and 0.59 to Trp. To reduce the contribution of an amino acid to antigenicity, referred to here as de-Antigenization, the amino acid is replaced by another of lower reactivity while ensuring the preservation of the overall structure of the antigen. The structure preservation is achieved by the replacement rules of Padlan (2008) shown in Table 4.

The residues that have been shown to be involved in those interactions are summarized in Table 3. What is shown in Table 3 is a collection of possible neutralizing epitopes. Instead of concentrating on just one, we choose to target all of those shown.

Calculation of antigenicities

The method developed by Padlan (2008) was used for the characterization of a putative antibody epitope centered at a particular residue. Briefly, the method identifies all the residues within a certain radius of the Cα (alpha-carbon) of each residue of the protein and assigns an antigenicity value for that putative epitope site based on the summed reactivities of those residues weighted by their exposure to solvent. The radius that we chose from past experience was 17 A (Angstroms).

Exposure to solvent, or solvent accessibility, of the individual residues was calculated using the computer algorithms developed by Connolly (1983) using programs developed by Sheriff et al. (1985). Fractional solvent accessibilities are used here to describe the degree of exposure of individual residues; those can be estimated by simply dividing the solvent accessibility by the total surface area of the amino acid (obtained, for example, from http://prowl.rockefeller.edu/aainfo/volume.htm). There is a great variety in size and shape of antibody epitopes, but in this case the results of the structural analyses of the binding of RBD to its cellular receptor (ACE2) and of neutralizing antibodies to RBD provide the identification and exact locations of the epitope residues of the spike protein that we could use for developing an effective subunit vaccine.

Trial calculations on the degree of de-Antigenization were performed to try to find the one that is probably best for the design of the desired subunit vaccine. The effect of lowering the reactivity of all the residues in the molecule, except for those in the RBD epitopes, was calculated with the side-chains whose fractional exposures to solvent are 0.50 (half buried, half exposed), 0.3333 (one-third exposed), 0.25 (one-fourth exposed), and 0.20 (one-fifth exposed). Mutations to lower the reactivity at the individual positions followed the replacement rules shown in Table 4. None of the residues in the RBD that were found to be involved in ligand binding were changed (compare Tables 1 and 3). The antigenicity values were expressed in terms of standard deviations above the mean for all the residues of the structure. The results of our analyses are summarized in Tables 1, 2, and 3.

At exposure levels of 0.3333 (one-third exposed), 0.25 (one-fourth exposed) and 0.20 (one-fifth exposed), high antigenicity values were found for residues within the RBD of the molecule, while none of the positions outside the RBD showed comparable antigenicity values. As an example, the result of de-Antigenization at the level of one-fourth (0.25) exposed is shown in Table 1 (with the label "mut").

Since it is the RBD that initially binds to the target cell, all the resulting changes in this fragment are incorporated in vaccine(s) that we propose by this method. (Other exposure levels could also result in potentially useful molecules.)

We examined the ability of "mut" to elicit a T-cell response. We referred to the immune-epitope-database (IEDB), maintained by the National Institute of Allergy and Infectious Diseases [https://www.iedb.org/], which contains the known SARS-CoV-2 T-cell epitopes. We found three regions: residues 448-458 (NYNYLYRLFRK), residues 481-489 (NGVEGFNCY), and residues 489-497 (YFPLQSYGF) that are currently listed in IEDB. These sequences in "mut" are underlined in Table 1.

The de-Antigenized sequence that we propose could be used in the development of a potentially useful subunit vaccine, or as the basis for a transformed-viral vaccine. Additional residues like the transmembrane region might need to be added to the end of this proposed subunit vaccine if it is used as the basis for a transformed-viral vaccine. Further, amino acid substitutions to help stabilize the trimeric structure of the spike protein might need to be done, although an isolated monomer might suffice

since the structure of the ACE2 binding region does not appear to be influenced by the state of the stem region.

The advantage of our de-Antigenized subunit vaccine over one that uses an unaltered subunit is the fact that the latter can be expected to produce antibodies against all the exposed parts of the protein, whereas our proposed subunit would predominantly promote the production of antibodies against the part that is critically important to the binding of the virus to its target cell - the receptor-binding site. Moreover, since all the exposed residues in the spike monomer had been de-Antigenized, any mutation of the virus that does not involve its ACE2 binding region should not affect the efficacy of this type of subunit vaccine that we are proposing. The efficacy of the subunit vaccine that we are proposing would be even better and longer lasting when more neutralizing antibody epitopes are added to those that we have included above.

CONFLICT OF INTEREST

There are no conflicts of interest. The views expressed in this article are those of the authors and do not reflect the official policy of the Department of Army/Navy/Air Force, Department of Defense, or U.S. Government.

CONTRIBUTIONS OF INDIVIDUAL AUTHORS

C. P. Mikita provided the medical significance of the work and researched the early literature. E. A. Padlan did the sequence and structural analyses.

REFERENCES

Bangaru S, Ozorowski G, Turner HL, Antanasijevic A, Hunag D, Wang X, Torres JL, Diedrich JK, Tian JH, Portnoff AD, Patel N, Massare MJ, Yates JR, Nemazee D, Paulson JC, Glenn G, Smith G, Ward AB. Structural analysis of full-length SARS-CoV-2 spike protein from an advanced vaccine candidate. Science 2020; 370(6520):1089-94

Connolly ML. Analytical molecular surface calculation. J Appl Crystallogr 1983; 16:548-58.

De Genst E, Areskoug D, Decanniere K, Muyldermans S, Andersson K. Kinetic and affinity predictions of a protein-protein interaction using multivariate experimental design. J Biol Chem 2002; 277:29897-907.

Padlan EA. A novel method for designing vaccines against constantly mutating pathogens. Phil J Sci 2008; 137:39-51.

Padlan EA. A method for designing molecules for use in directing the antibody response to a chosen region of a protein antigen. Phil Sci Letts 2010; 3(2):36-47.

Sandberg M, Eriksson L, Jonsson J, Sjostrom M, Wold S. New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. J Med Chem 1998; 41:2481-91.

Sheriff S, Hendrickson WA, Stenkamp RE, Sieker LC, Jensen LH. Influence of solvent accessibility and intermolecular contacts on atomic mobilities in hemerythrin. Proc Natl Acad Sci USA 1985; 82:1104-07.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 2018; 46(W1):W296-W303.

Yuan M, Wu NC, Zhu X, Lee CCD, So RTY, Lv H, Mok CKP, Wilson IA. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV Science 2020; 368(6491):630–3.

Zhang Y, Skolnick J. TM-align: A protein structure alignment algorithm based on TM-score. Nucleic Acids Res 2005; 33:2302-9.

Zhou T, Tsybovsky Y, Olia AS, Gorman J, Rapp M, Cerutti G, ChuangG-Y, Katsamba PS, Nazzar A, Sampson JM, Schön A, Wang P, Bimela J, Shi W, Teng I-T, Zhang B, Boyington JC, Sastry M, Stephens T, Stuckey J, Wang S, Friesner RA, Ho DD, Mascola JR, Shapiro L, Kwong PD. Cryo-EM structures delineate a pH-dependent switch that mediates endosomal positioning of SARS-CoV-2 spike receptor-binding domains. (2020) Biorxiv.