(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 July 2022 (21.07.2022)

WIPO | PCT

(10) International Publication Number
WO 2022/155476 A1

(51) International Patent Classification:
A61K 39/215 (2006.01)

(21) International Application Number:
PCT/US2022/012530

(22) International Filing Date:
14 January 2022 (14.01.2022)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/138,221 15 January 2021 (15.01.2021) US

(71) Applicant: THE UNITED STATES OF AMERICA, AS REPRESENTED BY THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES [US/US]; National Institutes of Health, Office of Technology Transfer, National Institutes of Health, 6701 Rockledge Drive, Suite 700, MSC 7788 Bethesda, Maryland 20892-7788 (US).

(72) Inventor: CONNORS, Mark; NIAID, Bldg 10, Magnuson Clinical Center, 11B09, Bethesda, Maryland 20814 (US).

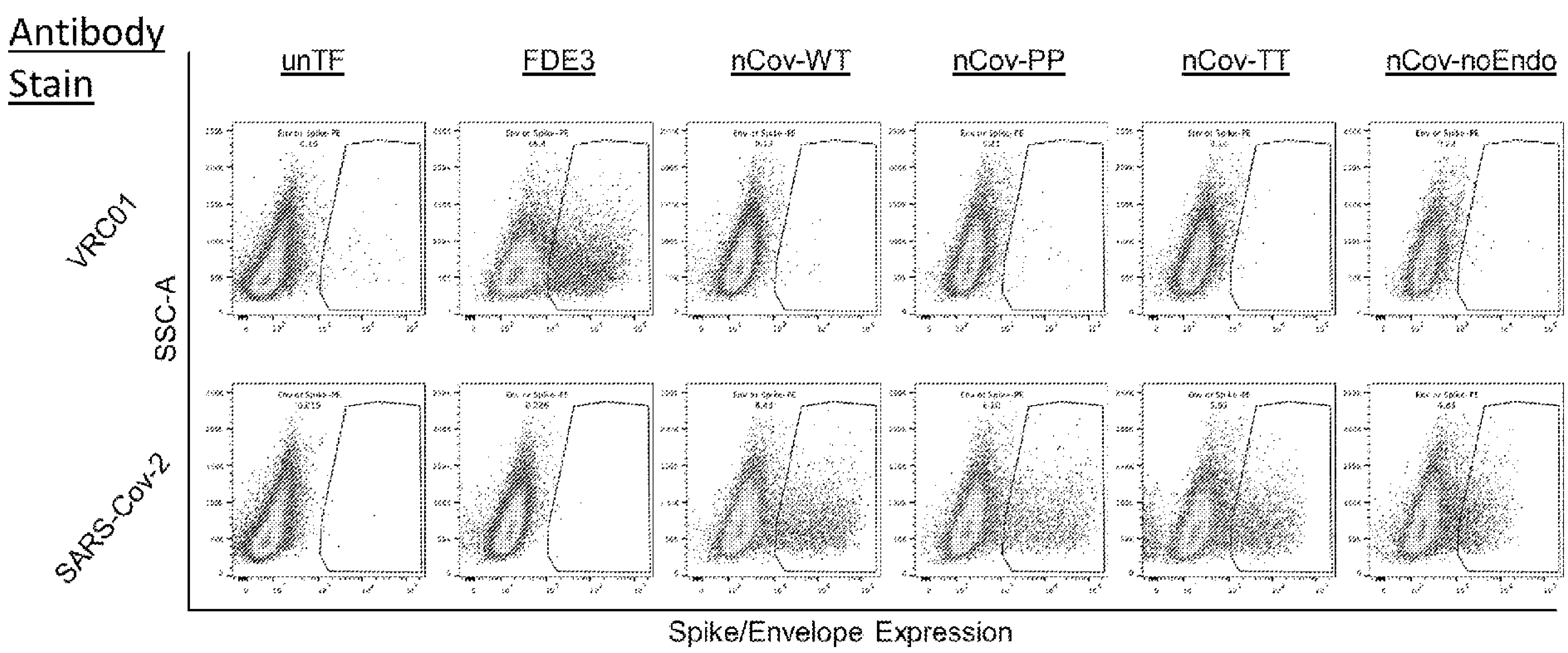
(74) Agent: CONNOLLY, Jodi L. et al.; Klarquist, One World Trade Center, Suite 1600, 121 SW Salmon Street, Portland, Oregon 97204 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) **Title:** REPLICATION-COMPETENT ADENOVIRUS TYPE 4 SARS-COV-2 VACCINES AND THEIR USE

FIG. 1



(57) **Abstract:** A replication-competent adenovirus type 4 (Ad4) modified to express the SARS-CoV-2 spike protein is described. The genome of the recombinant Ad4 is modified to have a deletion of at least a portion of the adenovirus E3 region to accommodate insertion of the spike protein coding sequence. Administration of the recombinant Ad4 to the upper respiratory tract elicits mucosal immunity, which is important for protection against SARS-CoV-2 infection and for preventing transmission of the virus.

WO 2022/155476 A1

[Continued on next page]

Declarations under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*
— *of inventorship (Rule 4.17(iv))*

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

REPLICATION-COMPETENT ADENOVIRUS TYPE 4 SARS-COV-2 VACCINES AND THEIR USE

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No. 63/138,221, filed January 15, 2021, which is herein incorporated by reference in its entirety.

FIELD

This disclosure concerns a recombinant replication-competent adenovirus type 4 (Ad4) expressing a SARS-CoV-2 spike protein and its use as an immunogenic composition for inhibiting SARS-CoV-2 infection and transmission.

BACKGROUND

Coronaviruses are a large family of viruses that typically cause mild to moderate upper respiratory tract disease; however, some members of this family can cause severe disease and death in humans. In the last 20 years, coronaviruses have caused three major outbreaks in humans resulting from severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), and SARS-CoV-2, the latter of which first emerged in Wuhan, China in December 2019. As of January 2021, SARS-CoV-2 had infected more than 84 million people worldwide, leading to nearly 2 million deaths. Although several SARS-CoV-2 vaccines have been approved for use in the U.S. and other countries, a need remains for an effective SARS-CoV-2 vaccine that induces mucosal immunity and can be rapidly produced in large quantities.

SUMMARY

Disclosed herein are immunogenic compositions comprised of a replication-competent adenovirus type 4 (Ad4) expressing a SARS-CoV-2 spike (S) protein (“Ad4-Spike”), such as a wild-type or modified version of the S protein from the original Wuhan strain or from a SARS-CoV-2 variant, such as the beta (B.1.351) variant, the delta (B.1.617.2) variant, the gamma (P.1) variant, the delta plus variant, or the omicron (B.1.1.529) variant. In the disclosed Ad4 vector, the gene encoding the SARS-CoV-2 S protein is cloned into the E3 region of an Ad4 vaccine strain. To accommodate insertion of the S protein, at least a portion of the E3 region is deleted. The disclosed Ad4-Spike vaccines possess several important advantages over other proposed and licensed SARS-CoV-2 vaccine platforms. In particular, as a replicating vector, Ad4-Spike is

capable of inducing a durable immune response, including mucosal immunity, which is an important factor for inhibiting both infection and transmission of the virus. Furthermore, Ad4-Spike vaccines can be rapidly produced to high titers at a relatively low cost.

Provided herein is a recombinant, replication-competent Ad4 expressing a SARS-CoV-2 S protein. The genome of the recombinant Ad4 includes a deletion in the adenovirus E3 region and an insertion of a coding sequence for the SARS-CoV-2 S protein. The SARS-CoV-2 S protein can be a native S protein or a modified S protein, such as a stabilized or truncated S protein. Additionally, the S protein can be from the Wuhan strain of SARS-CoV-2 or a variant thereof, such as a variant of concern (VOC).

Also provided is a recombinant, replication-competent Ad4 vector having a deletion in the adenovirus E3 region and an insertion of a coding sequence for the SARS-CoV-2 S protein. The SARS-CoV-2 S protein can be a native S protein or a modified S protein, such as a stabilized or truncated S protein, derived from either the Wuhan strain or a SARS-CoV-2 variant, such as a VOC.

Further provided are immunogenic compositions that include a recombinant Ad4 or a recombinant Ad4 vector disclosed herein, and a pharmaceutically acceptable carrier.

Also provided are methods of eliciting an immune response against SARS-CoV-2 in a subject and methods of immunizing a subject against SARS-CoV-2 infection by administering to the subject a therapeutically effective amount of a recombinant Ad4, a recombinant Ad4 vector, or an immunogenic composition disclosed herein. In some embodiments, the recombinant Ad4, recombinant Ad4 vector or immunogenic composition is administered to the upper respiratory tract, such as intranasally.

The foregoing and other objects and features of the disclosure will become more apparent from the following detailed description, which proceeds with reference to the accompanying figures.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1: SARS-CoV-2 spike expression of stabilized and truncated designs in transfected A549 Cells. A549 cells were transfected with a shuttle vector plasmid containing the gene for the SARS-CoV-2 spike protein from the Wuhan strain (nCoV). Four spike protein constructs were made: wild-type (WT), stabilized (PP), tail truncated (TT), and endocytosis motif truncated (noEndo). Controls included untransfected (unTF) cells and cells transfected with a plasmid expressing an HIV-1 envelope (Env) protein (FDE3). Expression of spike and Env was measured

by flow cytometry using a SARS-CoV-2 spike protein-specific antibody and an HIV-1 Env-specific antibody (VRC01), respectively. SARS-CoV-2 spike protein expression in transfected A549 cells diminished with stabilizing mutations, truncation of the tail, and truncation of the endocytosis motif, relative to wild-type spike protein.

FIGS. 2A-2B: SARS-CoV-2 spike expression of stabilized and truncated designs in infected A549 Cells. Replicating adenovirus carrying a SARS-CoV-2 protein gene was used to infect A549 cells. Three spike protein designs based on the Wuhan strain were tested for expression on the surface of A549 cells: wild-type (nCoV-WT), PP-stabilized (nCoV-PP), and tail-truncated (nCoV-TT) spike protein. A replicating adenovirus expressing an HIV-1 Env protein (FDE3) was used as a positive control of infection and uninfected (unIF) cells were used a negative control. Expression of spike protein was measured by flow cytometry using a SARS-CoV-2 spike protein-specific antibody. Antibody VRC01 was used to detect expression of HIV Env. Expression of spike by nCoV-WT is shown in FIG. 2A; expression of spike by FDE3, nCoV-PP and nCoV-TT is shown in FIG. 2B. As shown in FIGS. 2A-2B, expression of spike protein was high from both the nCoV-WT and nCoV-PP constructs.

FIG. 3: Immunization with replicating Ad4 containing SARS-CoV-2 spike protein gene induces neutralization in rabbits. New Zealand white rabbits were immunized on day 0 and day 28 (indicated by the arrows) with 1.29 x 10^9 infectious units (IFU) of purified replicating Ad4 nCoV-WT. Using a luciferase assay, serum neutralization against Wuhan SARS-CoV-2 pseudovirus was detected starting at 4 weeks post-immunization (prior to the second dose), and continued to increase up to 12 weeks post-immunization.

FIG. 4: Amino acid alignment of nCoV-PP, nCoV-WT, nCoV-Tail-Truncation, and nCoV-No-Endo spike proteins. Alignment displays locations of three mutations introduced to the SARS-Cov-2 wild-type (Wuhan) spike protein. nCoV-PP contains double proline stabilization substitutions at amino acid position 986 and 987; nCoV-Tail-Truncation includes a deletion of the terminal 24 amino acids at the cytoplasmic tail; and nCoV-No-Endo contains a deletion of the terminal endocytosis signaling motif (terminal five residues). Amino acid numbering is with reference to wild-type spike protein set forth herein as SEQ ID NO: 2.

FIGS. 5A-5B: Serum neutralization against Wuhan pseudovirus in a dose titration of intranasal Ad4-SARS-CoV-2_{WuPP} in hamsters. Syrian golden hamsters were intranasally administered 10^2 to 10^7 infection forming units (IFU) of Ad4-SARS-CoV-2 Wuhan spike with PP stabilization (Ad4-SARS-CoV-2_{WuPP}). Serum neutralization against Wuhan pseudovirus was measured at week 4 (FIG. 5A) and week 8 (FIG. 5B). Strong neutralization was observed at both timepoints for the highest doses of Ad4-SARS-CoV-2_{WuPP}.

FIGS. 6A-6E: Serum neutralization of intranasal Ad4-SARS-CoV-2 expressing the indicated VOC spike in hamsters. Syrian golden hamsters were immunized with intranasal Ad4 expressing stabilized spike proteins from either the Wuhan strain (Ad4-CoV2-Wuhan), the beta variant (Ad4-CoV2-SA), the delta variant (Ad4-CoV2-Indian) or the gamma variant (Ad4-CoV2-Brazil), or a stabilized chimeric spike protein having the beta variant RBD (Ad-CoV2-Wu/RBD-SA). An Ad4 expressing an influenza virus H5 hemagglutinin (Ad4-H5) and sham inoculation were included as negative controls. Serum neutralization against Wuhan pseudovirus (FIG. 6A) or delta pseudovirus (FIG. 6B) was determined 28 days following intranasal administration. In addition, serum neutralization against Wuhan pseudovirus (FIG. 6C), delta pseudovirus (FIG. 6D) and omicron pseudovirus (FIG. 6E) was determined 56 days following intranasal administration.

SEQUENCE LISTING

The nucleic and amino acid sequences listed in the accompanying sequence listing are shown using standard letter abbreviations for nucleotide bases, and three letter code for amino acids, as defined in 37 C.F.R. 1.822. Only one strand of each nucleic acid sequence is shown, but the complementary strand is understood as included by any reference to the displayed strand. The Sequence Listing is submitted as an ASCII text file, created on January 14, 2022, 199 KB, which is incorporated by reference herein. In the accompanying sequence listing:

SEQ ID NO: 1 is the nucleotide sequence of the Ad4-SARS-CoV-2 spike vector.

TAAATTTAAATGAATTCCGTCAAGGGCGACACAAAAGGTATTCTAAATGCATAATAAATACTGATAACATCTTATAGTTT
GTATTATATTTTGTATTATCGTTGACATGTATAATTTTGATATCAAAAACTGATTTTCCCTTTATTATTTTCGAGATTTA
TTTTCTTAATTCTCTTTAACAAACTAGAAATATTGTATATACAAAAAATCATAAATAATAGATGAATAGTTTAATTATAG
GTGTTCATCAATCGAAAAAGCAACGTATCTTATTTAAAGTGCGTTGCTTTTTTCTCATTTATAAGGTTAAATAATTCTCA
TATATCAAGCAAAGTGACAGGCGCCCTTAAATATTCTGACAAATGCTCTTTCCCTAAACTCCCCCCATAAAAAAACCCGC
CGAAGCGGGTTTTTACGTTATTTGCGGATTAACGATTACTCGTTATCAGAACCGCCCAGGATGCCTGGCAGTTCCCTACT
CTCGCCGCTGCGCTCGGTCGTTCGGCTGCGGGACCTCAGCGCTAGCGGAGTGTATACTGGCTTACTATGTTGGCACTGAT
GAGGGTGTCAGTGAAGTGCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATATGTGATACAGGATA
TATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCGAGCGGAAATGGCTTACGAACGGGGCGG
AGATTTCCTGGAAGATGCCAGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCG
CCCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGT
TTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATTCCGCTGTTATGGCCGCGT
TTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCACGAACCCCCCGTTC
AGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAAGACATGCAAAAGCACCACTGGCAGCA
GCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTG
ACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGCG
GTTTTTTCGTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAAGCTTAGAAAAAC
TCATCGAGCATCAAATGAAATTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAA
TGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACAT
CAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGT
GAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGC
ATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGGCGAAATACGCGATCGCTGTTAAAAGGACAATTACAAA
CAGGAATCGAGTGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAAT
ACCTGGAACGCTGTTTTTCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGT
CGGAAGTGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCAT
GTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAAGCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGA
GCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTCGACGTTTCCCGTTGAATATGGCTCAT
ATTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGA

AAAATAAACAAATAGGGGTCAGTGTTACAACCAATTAACCAATTCTGAACATTATCGCGAGCCCATTTATACCTGAATAT
GGCTCATAACACCCCTTGTTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAAC
GCCGTAGCGCCGATGGTAGTGTGGGGACTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCA
GTCGAAAGACTGGGCCTTTCGCCCGGGCTAATTAGGGGGTGTCGCCCTTATCGCTGAGGATCCATTTAAATTTAATTAAC
ATCATCAATAATATACCTTATTTTTTTTGTGTGAGTTAATATGCAAATAAGGCGTGAAAATTTGGGGATGGGGCGCGCTG
ATTGGCTGTGACAGCGGCGTTCGTTAGGGGCGGGGCAGGTGACGTTTTGATGACGCGACTATGAGGAGGAGTTAGTTTGC
AAGTTCTGGTGGGAAAAGTGACGTCAAACGAGGTGTGGTTTAAACACGGAAATACTCAATTTTCCCACGCTGTCTAACA
GGAAATGAGGTGTTTTTGGGCGGATGCAAGTGAAAACGGACCATTTTCGCGCGAAAACTGAATGAGGAAGTGAAATCTGA
GTAATTTAGTGTTTATGACAGGGAGGAGTATTTGCCGAGGGCCGAGTAGACTTTGACCGTTTACGTGGGGGTTTCGATTA
CCGTGTTTTTCACCTAAAGTTCCGCGTACGGTGTCAAAGTCCGGTGTTTTTACGTAGGTGTCAGCTGATCGTCAGGGTAT
TTAAACCTGCGCTCTGCAGTCAAGAGGCCACTCTTGAGTGCCAGCGAGAAGAGTTTTCTCCTCCGCGCCGCGAGTCAGAT
CTACACTTTGAAATATGAGGCACCTAAGAGACCTGCCCGATGAGGAAATTATCATCGCTTCCGGGAGCGAGATTCTGGAA
CTGGTGGTAAATGCTATGATGGGCGACGACCATCCGGAACCCCCCACCCCATTTGAGACACCTTCGCTGCACGATTTGTA
TGATCTGGAGGTGGATGTGCCCGAGGACGACCCCAACGAGAAGGCGGTAAATGATTTATTTAGCGATGCCGCGCTGCTAG
CTGCCGAGGAGGCTTCAAGCCCTAGCTCAGACAGCGACTCTTCACTGCATACCCCTAGACACGACAGAGGTGAGAAAGAG
ATCCCCGGGCTTAAATGGGAAAAGATGGACTTGCGTTGCTATGAGGAATGCCTGCCCCCAAGCGATGATGAGGACGAGCA
GGCGATTCAGAACGCAGCGAGCCATGGAGTGCAAGCCGTCAGCGAGAGCTTTGCACTGGACTGCCCGCCTTTGCCCGGAC
ACGGCTGTAAGTCTTGTGAATTTCATCGCATCAATACTGGAGATAAAGCTGTGTTATGTGCACTTTGCTATATGAGAGCG
TACAACCATTGTGTTTACAGTAAGTGTGATTAAGTGAACTTTAAAGGGAGGCAAAGAGTAGGGTGACTGGGTGATGACTG
GTTTATTTATGTATATCTGTTTTTTATATAGGTCCCGTTTCTGACGCAGATGATGAGACCCCCACTACAGAGTCCACTTT
GTCACCCCCTGAAATTGGCACGTCTCCATCTGACAATATTGTTAGACCAGTTCCTGTAAGAGCCACTGGGAGGAGAGCAG
CTGTAGAATGTTTGGATGATTTGCTTCAGGGTGGAGATGAACCTTTGGACTTGTGTACCCGGAAACGCCCCAGGCATTAA
GTGCCACACATGTGTGTTTACTTGAGGTGATGTCAGTATTTATAGGGTGTGGAGTGCAATAAAATATGTGTTGACTTTAA
GTGCGTGGTTTATGACTCAGGGGAGGGGACTTTGGGTATATAAGCAGGTGCAGACCTGTGTGGTTAGCTCAGAGCGGTAT
GGAGATTTGGACGGTTTTGGAAGACTTTCACAAGACTAGGCAGCTGCTAGAGAACGCCTCGAACGGAGTCTCTTACCTGT
GGAGATTCTGCTTCGGCGGTGACCTAGCTAAGCTAGTCTATAGGGCCAAACAGGATTATAGGGAACAATTTGAGGATATT
TTGAGAGAGTGTCCTGGTCTTTTTGACGCTCTTAACTTGGGCCATCAGTCTCACTTTAACCAGAGAATTTCAAGAGCCCT
TGACTTTACTACTCCTGGCAGAACCACTGCAGCAGTAGCCTTTTTTGCTTTTATTTTTGACAAATGGAGTCAAGAAACCC
ATTTCAGCAGGGATTACCAGCTGGATTTCTTAGCAGTAGCTTTGTGGAGAACATGGAAGTGCCAGCGCCTGAATGCAATC
TCCGGCTACTTGCCGGTACAGCCGCTAGACACTCTGAGGATCCTGAGTCTCCAGCAGCAGGAGGATCAAGAAGAGAATCC
GAGAGCCGGCCTGGACCCTCCGGCGGAGGAGTAGCTGACCTGTTTCCTGAACTGCACCGGGTGCTGACTAGGTCTTCGAG
TGGTCGGGAGAGGGGTATTAAGCGGGAGAGGCATGATGAGACTAATCACAGAATTGAACTGACTGTGGGTCTGATGAGCC
GCAAGCGTCCAGAAACAGTGTGGTGGTATGAGGTGCAGTCAACTGGCACAGATGAGGTGTCAGTCATGCATGAGAGATTT
TCCCTAGAACAAGTCAAGACTTGTTGGTTGGAGCCTGAGGATGATTGGGAGGTAGCCATCAGGAATTATGCCAAGCTGGC
TCTGAGGCCAGATAGAAAGTACAAGATTACTAAGCTGATAAATATCAGAAATGCCTGCTACATCTCAGGGAATGGGGCTG
AAGTGGAGATCTGTCTCCAGGATAGAGTGGCTTTCAGATGCTGCATGATGAATATGTACCCGGGAGTGGTGGACATGGAT
GGGGTCACCTTTATGAACATGAGGTTCAGGGGAGATGGGTATAATGGGACGGTCTTTATGGCCAATACCAAGCTGACAGT
GCATGGATGCTCCTTCTTTGGGTTTAATAACACCTGCATCGAGGCTTGGGGTCAGGTCGGTGTTAAGGGGTGCAGTTTTT
CAGCCAACTGGATGGGGGTAGTGGGCAGGACCAAGAGTATGCTGTCTGTGAAGAAATGCTTGTTTGAGAGGTGCCACCTG
GGGGTGATGAGCGAGGGCGAAGCCAGAATCCGCCACTGTGCCTCTACCGAGACGGGCTGTTTTGTGCTGTGCAAGGGCAA
TGCCAAGATCAAGCATAATATGATCTGTGGAGCCTCGGACGAGCGCGGCTACCAGATGCTGACCTGCGCCGGTGGGAACA
GTCATATGCTGGCCGCCGTGCATGTGGCTTCCCATTCCCGCAAGCCCTGGCCTGAGTTCGAGCACAATGTCATGACCAGG
TGCAATATGCATCTGGGGGCTCGCCGAGGCATGTTTATGCCCTACCAGTGCAACCTGAATTATGTAAAGGTGCTCCTGGA
GCCCGATGTCATGTCCAGAGTGAGCCTGACGGGGGTGTTTGACATGAATGTGGAAGTGTGGAAGATTCTAAGATATGATG
AATACAAGACCAGGTGTCGAGCCTGCGAGTGCGGAGGGAAGCATGCCAGGTTCCAGCCCGTGTGTGTGGATGTGACGGAG
GACCTGCGACCCGATCATTTGGTGTTGTCCTGCACCGGGACGGAGTTCGGCTCCAGTGGGGAAGAATCTGACTAGAGTGA
GTAGTGTTTTGGGGAGGGAGAGGACCTGCATAAGGGGCAGAATGATTAAAATCTGTGCTTTTCTGTGTGTTGCAGCAGCA
TGAGCGGAAACGGCTCCTTTGAGGGAGGGGTATTCAGCCCTTATCTGACGGGGCGTCTCCCCTCCTGGGCGGGAGTGCGT
CAAAATGTGATGGGATCCACGGTGGACGGCCGGCCCGTACAGCCCGCGAACTCTTCAACCCTGACCTATGCAACCCTGAG
CTCCTCGTCGGTGGACGCAGCTGCCGCCGCAGCTGCTGCTTCTGCCGCCAGCGCCGTGCGCGGAATGGCCATGGGCGCCG
GCTATTACGGCACTCTGGTGGCCAACTCGAGTTCCACTAATAATCCCGCCAGCCTGAACGAGGAGAAGCTGCTGCTGTTG
ATGGCCCAGCTCGAGGCCTTGACCCAGCGCCTGGGCGAGCTGACCCAGCAGGTGGCTCAGCTGCAGGAGCAGACGCGGGC
CGCGGTTGCCACGGTGAAATCCAAATAAAAAATGAATCAATAAATAAACGGAGACGGTTGTTGATTTTAAAAATCAGAGT
CTGAATCTTTATTTGATTTTTCGCGCACGGTAGGCCCTGGACCACCGGCCTCGATCATTGAGCACCCGGTGGATCTTTTC
CAAGACCCGGTAGAGGTGGGATTGGATATTGAGGTACATGGGCATGAGCCCGTCCCGGGGGTGAAGGTAGCTCCATTGCA
GGGCCTCGTGCTCGGGGGTGGTGTTGTAAATCACCCAGTCATAGCAGGGACGCAGGGCGTGGTGTTGCACAATATCTTTG
AGGAGGAGACTGATGGCCACGGGCAGCCCTTTGGTGTAGGTGTTTACAAACCTGTTGAGCTGGGAGGGATGCATGCGGGG
GGAGATGAGGTGCATCTTAGCCTGGATCTTCAGATTGGCGATGTTACCGCCCAGATCCCGCCTGGGATTCATGTTGTGCA
GGACCACCAGCACGGTGTATCCGGTGCACTTGGGGAATTTGTCATGCAACTTGGAAGGGAAGGCATGAAAGAATTTGGAG
ACGCCCTTGTGGCCGCCCAGGTTTTCCATGCACTCATCCATGATAATGGCTATGGGCCCGTGGGCGGCGGCTTGGGCAAA
GACGTTTCGGGGGTCGGACACATCGTAGTTGTGGTCCTGGGTGAGATCTTCATAGGCCATTTTAATGAATTTGGGGCGGA
GGGTGCCCGATTGGGGGACGAAGGTACCCTCAATCCCGGGGGCGTAGTTTCCCTCACAGATCTGCATCTCCCAGGCCTTA

AGCTCCGAGGGGGGGATCATGTCCACCTGCGGGGCGATAAAGAAAACGGTTTCCGGGGCGGGGGAGATGAGCTGGGCGGA
AAGCAGGTTGCGGAGTAGCTGGGACTTACCGCAGCCGGTGGGGCCGTAGATAACCCCAATGACCGGCTGCAGGTGGTAGT
TGAGGGAGACACAGCTGCCGTCCTCCCTAAGAAGGGGGGCCACCTCGTTCATCATTTGGCGCACGTGCATGTTCTCGCGC
ACCAGTTCCGCCAGGAGTCGCTCTCCGCCCAGCGAGAGGAGCTCCTGGAGCGAGGCGAAGTTTTTCAGCGGCTTGAGCCC
GTCGGCCATGGGCATTTTGGAAAGGGTCTGTTGCAGGAGTTCCAAGCGGTCCCAGAGCTCGGTGATGTGCTCTACGGCAT
CTCGATCCAGCAGACCTCCTCGTTTCGCGGGTTGGGGCGACTGCGGGAGTAGGGCGCCAGACGATGGGCGTCCAGCGCGG
CCAGGGTCCGGTCCTTCCAGGGTCGCAGCGTCCGCGTCAGGGTGGTCTCCGTCACGGTAAAGGGGTGCGCGCCGGGCTGG
GCGCTTGCGAGGGTGCGCTTCAGGCTCATCCGGCTGGTCGAGAACCGCTCCCGATCGGCGCCCTGTGCGTCGGCCAGGTA
GCAATTGACCATGAGTTCGTAGTTGAGCGCCTCGGCCGCGTGGCCTTTGGCGCGGAGCTTACCTTTGGAAGTCTGCCCAC
AGGCGGGACAGAGGAGGGACTTGAGGGCGTAGAGCTTGGGGGCGAGGAAGACGGACTCGGGGGCGTAGGCGTCCGCGCCG
CAGTGGGCGCAGACGGTCTCGCACTCCACGAGCCAGGTGAGGTCGGGCTGATTGGGATCAAAAACCAGTTTTCCGCCGTT
CTTTTTGATGCGTTTCTTACCTCTGGTCTCCATGAGCTCGTGTCCCCGCTGGGTGACAAAGAGGCTGTCCGTGTCCCCGT
AAACCGACTTTATGGGTCGGTCCTCGAGTGGGACGCCGCGGTCCTCGTCGTAGAGGAAACCCGACCACTCTGAGACGAAG
GCCCGGGTCCAAGCCAGCACGAAGGAGGCCACGTGGGAGGGATAGCGGTCGTTATCCACCAGCGGGTCCACCTTCTCCAG
TGTATGCAAACACATGTCCCCCTCGTCCACATCCAGGAAGGTGATTGGCTTGTAAGTGTAGGCCACGTGACCGGGGGTCC
CGGCCGGGGGGGTATAAAAGGGGGCGGGCCGCTGCTCGTCTTCACTGTCTTCCGGATCGCTGTCCAGGAGCGCCAGCTGT
TGGGGTAGGTATTCCCTCTCAAAGGCGGGCATGACCTCCGCACTCAGGTTGTCAGTTTCTAGAAACGAGGAGGATTTGAT
ATTGACGGTGCCGGCGGAGATGCCTTTCAAGAGCCCCTCGTCCATCTGGTCAGAAAAGACAATCTTTTTGTTGTCGAGTT
TGGTGGCGAAGGAGCCGTAGAGGGCGTTGGAGAGGAGCTTGGCGATGGAGCGCATGGTCTGGTTCTTTTCCTTGTCGGCG
CGCTCCTTGGCGGCGATGTTGAGCTGCACGTACTCGCGCGCCACGCACTTCCATTCGGGGAAGACGGTGGTTAGCTCGTC
TGGCACGATTCTGACCTGCCAGCCCCGGTTATGCAGGGTGATGAGGTCAACGCTGGTGGCCACCTCGCCGCGCAGGGGCT
CGTTGGTCCAGCAGAGGCGGCCGCCCTTGCGCGAGCAGAAGGGGGGCAGGGGGTCCAGCATAAGCTCGTCGGGGGGGTCA
GCATCGATGGTGAAGATGCCTGGCAGGAGGTCGGGGTCGAAGTAGCTTATGCAGGTGCCCAGATCGTCCAGAGAAGCTTG
CCATTCGCGCACGGCCAGCGCGCGCTCGTAGGGACTAAGGGGCGTGCCCCAGGGCATGGGGTGGGTGAGCGCGGAGGCGT
ACATGCCGCAGATGTCGTAGACGTAGAGGGGCTCATCAAGGATGCCAATGTAGGTGGGGTAGCAGCGGCCCCCGCGGATG
CTGGCGCGCACGTAGTCATACAACTCGTGCGAGGGGGCGAGGAGCCCGGCTCCGAGATTGGCGCGGCTGGGTTTTTCGGC
GCGGTAGACGATCTGACGGAAGATGGCGTGGGAGTTGGAGGAGATGGTGGGTCTTTGGAAGATGTTGAAGTGGGCGTGGG
GCAGGCCGACCGAGTCGCGGATGAAGTGGGCGTAGGAGTCTTGCAGCTTGGCGACAAGCTCGGCGGTGACGAGGACGTCC
AGGGCGCAGTAGTCAAGGGTCTCTTGGATGATGTCATACTTGAGCTGGCCCTTTTGTTTCCACAGCTCGCGGTTGAGAAG
GAACTCTTCGCGGTCCTTCCAGTACTCTTCAAGGGGGAACCCGTCCTGGTCGGCACGGTAAGAGCCTAGCATGTAGAACT
GGTTAACGGCCTTGTAGGCGCAGCAGCCCTTCTCCACGGGGAGGGCATAGGCCTGGGCGGCCTTGCGCAGGGAGGTGTGC
GTGAGGGCGAAGGTGTCCCTGACCATGACCTTTAGGAACTGGTGCTTGAAGTCGATATCGTCGCAGCCCCCCTGCTCCCA
GAGCTGGAAGTCCGTGCGCTTCTTGTAGGCGGGGTTGGGCAAAGCGAAAGTAACATCGTTGAAGAGGATCTTGCCCGCGC
GGGGCATAAAGTTGCGAGTGATGCGGAAAGGCTGGGGCACCTCGGCCCGGTTGTTGATGACCTGGGCGGCGAGCACGATC
TCGTCGAAGCCGTTAATGTTGTGGCCCACAATGTATAGTTCCACGAACCGCGGGCGGCCCTTGACGTGGGGCAGTTTCTT
GAGCTCCTCGTAGGTGAGCTCGTCGGGGTCGCTGAGCCCGTGCTGCTCGAGGGCCCAGTCGGCGAGATGGGGGTTGGCGC
GGAGGAAGGAAGTCCAGAGATCCACGGCCAGGGCGGTTTGCAGACGATCCCGGTACTGGCGGAACTGCTGACCCACGGCC
ATTTTTTCGGGGGTGACGCAGTAGAAGGTGCGGGGGTCGCCGTGCCAACGGTCCCATTTTAGCTGGAGGGCGAGATCAAG
GGCGAGCTCAACGAGCCGGTCGTCCCCGGAGAGTTTCATGACCAGCATGAAGGGGACGAGCTGCTTGCCGAAGGACCCCA
TCCAGGTGTAGGTTTCCACATCGTAGGTGAGGAAGAGCCTTTCGGTGCGAGGATGCGAGCCGATGGGGAAGAACTGGATC
TCCTGCCACCAGTTGGAGGAATGGCTGTTGATGTGATGGAAGTAGAAATGCCGACGGCGCGCCGAACATTCGTGCTTGTG
TTTATACAAGCGGCCACAGTGCTCGCAACGCTGCACGGGATGCACGTGCTGCACGAGCTGTACCTGGGTTCCTTTGACGA
GGAATTTCAGTGGGAAGTGGAGTCGTGGCGCCTGCATCTGGTGCTGTACTACGTCGTGGTGGTCGGCCTGGCCCTCTTCT
GCCTCGATGGTGGTCATGCTGACGAGCCCGCGCGGGAGGCAGGTCCAGACCTCGGCGCGAACGGGTCGGAGAGCGAGGAC
GAGGGCGCGCAGGCCGGAGCTGTCCAGGGTCCTGAGACGCTGCGGAGTCAGGTCAGTGGGCAGCGGCGGCGCGCGGTTGA
CTTGCAGGAGTTTTTCAAGGGCGCGCGGGAGGTCCAGATGGTACTTGATCTCCACCGCGCCGTTGGTGGCGACGTCGATG
GCTTGCAGTGTCCCGTGCCCCTGGGGAGTGACCACCGTCCCCCGTTTCTTCTTGGCGGGCGGAAGCGGTTTGGCTTCCAT
GGTTAAAAGCGGCGGCGAGGACGCGCGCCGGGCGGTAGGGGCGGCTCGGGACCCGGAGGCAGTGGTGGCAGGGGCACGTC
GGCGCCGCGCGCGGGCAGGTTCTGGTACTGCGCCCGGAGAAGACTGGCGTGAGCGACGACGCGACGGTTGACGTCCTGGA
TCTGACGCCTCTGGGTGAAGGCCACGGGACCCGTGAGTTTGAACCTGAAAGACAGTTCGACAGAATCAATCTCGGTATCA
TTGACGGCGGCCTGCCGCAGAATCTCTTGCACGTCGCCCGAGTTGTCCTGGTAGGCAATCTCGGTCATGAACTGCTCGAT
CTCCTCCTCCTGAAGGTCTCCGCGGCCGGCGCGCTCCACGGTGGCCGCGAGGTCGTTGGAGATGCGGCCCATGAGCTGCG
AGAAGGCGTTCATGCCCGCCTCGTTCCAGACGCGGCTGTAAACCACGGCGCCCTCGGGATCGCGGGCGCGCATGACCACC
TGGGCGAGGTTGAGCTCCACGTGGCGCGCAAAAACCGCGTAGTTGCAGAGGCGCTGGTAGAGGTAGTTGAGCGTGGTGGC
AATGTGCTCAGTGACAAAGAAGTACATAATCCAGCGGCGGAGCGGCATTTCGCTGACGTCGCCCAGGGCTTCCAAGCGCT
CCATGGCCTCGTAAAAGTCCACGGCGAAGTTGAAAAACTGGGAGTTGCGTGCAGATACGGTCAAGTCCTCCTCCAGAAGA
CGGATGAGCTCGGCGATGGTGGCGCGCACCTCGCGCTCGAAGGCTCCCGTGAGTTCCTCCACTTCCTCCTCTTCATCCAC
TAACATCTCTTCTACTTCCTCCTCAGGCGGTGGTGGCGGGGGAGGGGGCCTGCGTCGCCGGCGGCGCACGGGCAGACGGT
CGATGAAACGCTCGATGGTCTCGCCGCGCCGGCGTCGCATGGTCTCGGTGACGGCGCGCCCGTCCTCGCGGGGTCGCAGC
GTAAAGACGCCGCCGCGCATCTCCAGGTGGCCCGGGGGGTCCCCGTTGGGCAGGGAGAGTGCGCTGACGATGCATCTTAT
CAATTGCCCCGTAGGGACTCCGCGCAAGGACCTAAGCGTCTCTAGATCCACGGGATCTGAAAACCGTTGAACGAAGGCTT
CGAGCCAGTCGCAGTCGCAAGGTAGGCTGAGCACGGTTTCTTCTGGCGGCGGTGGGGTGTGGGCGGGGGCGATGCTGCTG

GTGATGAAGTTGAAATAGGCGGTTCTGAGACGGCGGATGGTGGCGAGGAGCACCAGGTCTTTGGGCCCGGCTTGCTGGAT
GCGCAGACGGTCGGCCATGCCCCAGGCGTGGTCCTGACACCTGGCCAGGTCCTTGTAGTAGTCCTGCATGAGCCGCTCCA
CGGGCACCTCCTCCTCGCCCGCGCGGCCGTGCATACGCGTGAGCCCAAACCCGCGCTGCGGCTGGACGAGCGCCAGGTCA
GCGACGACGCGCTCGGCGAGGATGGCCTGCTGGATCTGGGTGAGGGTGGTCTGGAAGTCGTCAAAGTCGACGAAGCGGTG
GTAGGCTCCGGTGTTAATGGTGTAGGAGCAGTTGGCCATGACGGACCAGTTGACAGTCTGGTGACCGGGCCGCGCGAGCT
CGTGGTACTTGAGGCGCGAGTAGGCGCGCGAGTCGAAGATGTAGTCGTTGCAGGTGCGCACCAGGTACTGGTAGCCGATG
AGGAAGTGCGGCGGCGGCTGGCGGTAGAGCGGCCATCGCTCGGTGGCGGGGGCGCCGGGCGCTAGGTCCTCGAGCATGGT
GCGGTGGTAGCCGTAGATGTACCTTGACATCCAGGTGATGCCGGCGGCGGTGGTGGAGGCGCGAGGGAACTCGCGGACGC
GGTTCCAGATGTTGCGCAGCGGCAGGAAGTAGTTCATGGTGGGCACGGTCTGGCCCGTGAGGCGCGCGCAGTCGTTGATG
CTCTAGACATACGGGCAAAAACGAAAGCGGTCAGCGGCTCGACTCCGTGGCCTGGAGGCTAAGCGAACGGGTTGGGCTGC
GCGTGTACCCCGGTTCGAATCTCGAATCAGGCTGGAGCCGCAGCTAACGTGGTACTGGCACTCCCGTCTCGACCCAGGCC
TGCACAAAACCTCCAGGATACGGAGGCGGGTCGTTTTGCAAATTTTTGGCGGTCGAAAAAAGCTAGTAAGCGCGGAAAGC
GGCCGACCGCAATGGCTCACTGCCGTAGATTGGAGAAGAATCGCCAGGGTTGCGTTGCGGTGTGCCCCGGTTCGAGACCG
CTCGGGTCGGCCGAATTCCGCGGCTAACGAGGGCGTGGCTGCCCCGTCGTTTCCAAGACCCCATAAGCCAGCCGACTTCT
CCAGTTACGGAGCGAGCCCCTCTTTTGTTTTGTTTTTTGCCAGATGCATCCCGTACTGCGGCAGATGCGCCCCCACCCTC
CACCGCAACAGCAGCCCCCTCCTACGCAACAGCCGGCGCTTCTGCCTCCGCCCCAGCAGCAGCAACTTCCAGCCACTACC
GCCGCGGCCGCCGTGAGCGGGGCCGGGCAGAGTCAGTATGACCTGGCTTTGGAAGAGGGCGAGGGGCTGGCGCGCCTGGG
GGCGTCGTCGCCGGAGCGGCACCCGCGCGTGCAGATGAAAAGGGACGCTCGCGAGGCCTACGTGCCCAAGCAGAACCTGT
TCAGAGACAGGAGCGGCGAGGAGCCCGAGGAGATGCGCGCAGCCCGTTTCCACGCGGGGCGGGAGCTGCGGCGCGGCCTG
GACAGAAAGAGGGTGCTGAGGGACGAGGATTTCGAGGCGGACGAGCTGACGGGGATCAGCCCTGCGCGCGCGCACGTGGC
CGCGGCCAACCTGGTCACGGCGTACGAGCAGACCGTGAAGGAGGAGAGCAACTTCCAAAAATCCTTCAACAACCACGTGC
GCACCCTGATCGCGCGCGAGGAGGTGACCCTGGGCCTGATGCACCTGTGGGACCTGCTGGAGGCCATTGTGCAGAACCCC
ACCAGCAAACCGCTGACGGCGCAGCTGTTCCTGGTGGTGCAGCACAGTCGGGACAACGAGACTTTTAGGGAGGCGCTGCT
GAATATCACCGAGCCCGAGGGCCGCTGGCTTCTGGACCTGGTGAATATTCTGCAGAGCATCGTGGTGCAGGAGCGCGGGC
TGCCGCTGTCCGAGAAGCTGGCGGCCATCAACTTTTCGGTGCTGAGTTTGGGCAAGTACTACGCTAGGAAGATCTACAAG
ACCCCGTACGTGCCCATAGACAAGGAGGTGAAGATCGACGGGTTTTACATGCGCATGACCCTGAAAGTGCTGACCCTGAG
CGACGATCTGGGGGTGTACCGCAACGACAGGATGCGCCGCGCGGTAAGCGCCAGCAGGCGGCGCGAGCTGAGCGATCAGG
AGCTGATGCACAGCCTGCAGCGGGCCCTGACCGGGGCCGGGACCGAGGGGGAGAGCTACTTTGACATGGGCGCGGACCTG
CACTGGCAGCCCAGCCGCCGGGTCTTGGAAGCCGCGGCGGTCCCTTACGTAGAAGAGGTGGACGATGAGGATGAGGGCGA
GTACCTGGAAGACTGATGGCGCGACCGTATTTTTGCTAGATGCAGCAACAGCCACCTCCTGATCCCGCAATGCGGGCGGC
GCTGCAGAGCCAGCCGTCCGGCATTAACTCCTCGGACGATTGGACCCAGGCCATGCAACGCATCATGGCGCTGACGACCC
GCAACCCCGAAGCCTTTAGACAGCAGCCCCAGGCCAACCGGCTCTCGGCCATCCTGGAGGCCGTGGTGCCCTCGCGCTCC
AACCCCACGCACGAGAAGGTGCTGGCCATCGTGAACGCGCTGGTTGAGAACAAGGCCATTCGCGGCGACGAGGCCGGGCT
GGTGTACAACGCACTGCTGGAGCGCGTGGCCCGCTACAACAGCACCAACGTGCAGACCAACCTGGACCGCATGGTGACCG
ACGTGCGCGAAGCCGTGGCCCAGCGCGAACGGTTCCACCGCGAGTCCAACCTGGGATCCATGGTGGCACTGAACGCCTTC
CTCAGCACGCAGCCCGCCAACGTGCCCCGGGGCCAGGAGGACTACACCAACTTCATTAGCGCCCTGCGGCTAATGGTGAC
CGAGGTGCCCCAGAGCGAGGTGTACCAGTCGGGCCCGGACTACTTCTTCCAGACCAGTCGCCAGGGCTTGCAGACCGTGA
ACCTGAGTCAGGCTTTCAAGAACTTGCAGGGACTGTGGGGCGTGCAGGCTCCGGTCGGGGACCGCGCGACGGTGTCGAGC
CTGCTGACGCCGAACTCGCGCCTGCTGCTGCTGCTGGTGGCGCCCTTCACGGACAGCGGTAGTATCAACCGCAACTCGTA
CCTGGGCTACCTGATTAACCTGTACCGCGAGGCCATTGGCCAGGCGCACGTGGACGAGCAGACCTACCAGGAGATTACCC
ACGTGAGCCGCGCCCTTGGCCAGGACGACCCGGGCAATCTGGAAGCCACCCTGAACTTCTTGCTGACCAACCGGTCGCAG
AAGATCCCGCCCCAGTACGCGCTGAGCGCCGAGGAGGAGCGTATATTGAGATACGTGCAGCAAAGTGTGGGACTGTTCCT
GATGCAGGAGGGGGCCACCCCCAGCGCCGCGCTCGACATGACCGCGCGCAACATGGAGCCCAGCATGTACGCCAGTAATC
GCCCGTTTATTAATAAGCTGATGGACTACCTGCATCGGGCGGCCGCCATGAACTCTGACTATTTCACCAACGCCATCCTG
AACCCCCACTGGCTCCCGCCGCCGGGGTTCTACACGGGCGAGTACGACATGCCCGACCCCAATGACGGGTTTCTGTGGGA
CGACGTGGACAGCAGCGTGTTCTCCCCCCGACCGGGTGCTAACGAGCGCCCCTTGTGGAAGAAAGAGGGCAGCGACCGGC
GCCCGTCCTCGGCGCTGTCCGGCCGCACGGGTGCTGCCGCAGCGGTGCCCGAGGCCGCCAGTCCCTTTCCGAGCTTGTCA
CTGAACAGCGTCCGCAGTAGCGAGCTGGGCAGGATCACGCGCCCGCGCTTGCTGGGCGAGGAGGAGTACTTAAATAACTC
GCTGTTGAGGCCCGAGCGGGAGAAGAACTTCCCCAATAACGGGATAGAGAGTCTGGTGGATAAGATGAGCCGCTGGAAGA
CGTACGCGCATGAGCACAGGGACGATCCCCGGGCAACGCAGGGGGCCACCAGCCGGGGCAGTGCCGCCCGTAAACGCCGC
TGGCACGACAGGCAGCGGGGACTGATGTGGGACGATGAGGATTCCGCCGACGACAGCAGCGTGTTGGACTTGGGCGGGAG
TGGTGGTGGTAACCCGTTCGCTCACCTGCGCCCCCGCGTCGGGCGCCTGATGTAAAAAGAAACCAAAAATAAATGGTACT
CACCAAGGCCATGGCGACCAGCGTGCGTTCGTTTCTTCTCTGTTGTATCTAGTATGATGAGGCGTGCGTACCCGGAGGGT
CCTCCTCCCTCGTACGAGAGCGTGATGCAGCAGGCAATGGCGGCGGCGGCGGCGATGCAGCCCCCGCTGGAGGCTCCTTA
CGTGCCACCGCGGTACCTGGCGCCTACGGAGGGGCGAAACAGCATTCGTTACTCGGAGCTGGCACCCTTGTACGATACCA
CCCGGTTGTACCTGGTGGACAACAAGTCGGCGGACATCGCCTCGCTGAACTACCAGAACGACCACAGCAACTTTCTGACC
ACCGTGGTGCAGAACAACGATTTCACCCCCACGGAGGCCAGCACCCAGACCATCAACTTTGACGAGCGCTCGCGGTGGGG
CGGTCAGCTGAAAACCATCATGCATACCAACATGCCCAACGTGAACGAGTTCATGTACAGCAACAAGTTCAAGGCGCGGG
TCATGGTCTCCCGCAAGACCCCCAACGGGGTGACAGTAGGGGATGATTATGATGGTAGTCAGGATGAGCTGAAATACGAG
TGGGTGGAGTTTGAGCTGCCCGAAGGCAACTTCTCGGTGACCATGACCATTGACCTGATGAACAACGCCATCATCGACAA
TTACTTGGCAGTGGGGCGGCAGAACGGGGTGCTGGAGAGCGACATCGGCGTGAAGTTCGACACCCGGAACTTCAGGCTGG
GTTGGGACCCCGTGACCGAGCTGGTCATGCCCGGGGTGTACACCAACGAGGCCTTCCACCCCGACATCGTGCTGTTGCCC

GGCTGCGGGGTGGACTTTACCGAGAGCCGCCTCAGTAATATGCTGGGCATCCGCAAGAGGCAGCCCTTCCAGGAGGGTTT
CCAGATCATGTACGAGGACCTGGATGGAGGTAACATCCCCGCGCTCTTGGATGTCGAGGCCTATGAGAAAAGCAAGGAGG
AGAGCGTCGCCGCGTCAACCGCAGCCGTAGCCACCGCCTCTACCGAGGTCCGGGGCGATAATTTTGCTAGCGCCGCAGCA
GTGGCGGCGGCCAAGGCTGATGAAACCGAAAGTAAGATAGTTATTCAGCCGGTGGAGAAGGATAGCAAGGATAGGAGCTA
CAACGTGCTCTCGGACAAGAAAAACACCGCCTACCGCAGCTGGTACCTGGCCTACAACTATGGCGACCACGAGAAGGGCG
TGCGCTCCTGGACGCTGCTCACCACCTCGGACGTCACCTGCGGCGTGGAGCAAGTCTACTGGTCGCTGCCCGACATGATG
CAAGACCCGGTCACCTTCCGCTCCACGCGTCAAGTTAGCAACTACCCGGTGGTGGGCGCCGAGCTCATGCCCGTCTACTC
CAAGAGCTTCTTCAACGAGCAGGCCGTCTACTCGCAGCAGCTGCGCGCCTTCACCTCGCTCACGCACGTCTTCAACCGCT
TCCCTGAGAACCAGATCCTCGTCCGCCCGCCCGCGCCCACCATTACCACCGTCAGTGAAAACGTTCCTGCTCTCACAGAT
CACGGGACCCTGCCGCTGCGCAGCAGTATCCGGGGAGTCCAGCGCGTGACCGTTACTGACGCCAGACGCCGCACCTGCCC
CTACGTCTACAAGGCCCTGGGCATAGTCGCGCCGCGCGTCCTCTCGAGCCGCACCTTCTAAAAAATGTCCATTCTCATCT
CGCCCAGTAATAACACCGGTTGGGGTCTGCGCGCGCCCAGCAAGATGTACGGAGGCGCTCGCCAACGCTCCACGCAACAC
CCCGTGCGCGTGCGCGGGCACTTCCGCGCTCCCTGGGGCGCCCTCAAGGGCCGCGTGCGGTCGCGCACCACCGTCGACGA
CGTGATCGACCAGGTGGTGGCCGACGCTCGCAACTACACCCCCGCCGCCGCGCCCGTCTCCACCGTGGACGCCGTCATTG
ACAGCGTGGTGTCCGACGCGCGCCGGTACGCCCGCGCCAAGAGCCGGCGGCGGCGCATCGCCCGGCGGCACCGTAGCACC
ACCGCCATGCGTGCGGCGCGAGCCTTGCTGCGCAGGGCCAGGCGCACGGGACGCAGGGCCATGCTCAGGGCGGCCAGACG
CGCGGCTTCAGGCGCCAGCGCCGGCAGGACTCGGAGACGCGCGGCCACGGCGGCGGCAGCGGCCATAGCCAGCATGTCCC
GCCCGCGGCGAGGGAACGTGTACTGGGTGCGCGACGCCGCCACCGGTGTGCGCGTGCCCGTGCGCACCCGCCCCCCTCGC
ACTTGAAGATGTTCACTTCGCGATGTTGATGTGTCCCAGCGGCGAGGAGAAGGATGTCCAAGCGCAAATTCAAGGAAGAG
ATGCTCCAGGTCATCGCGCCTGAGATCTACGGCCCCGCGGCGGCGGTGAAGGATGAAAGAAATCCCCGCAAAATCAAGCG
GGTCAAAAAGGACAAAAAGGAAGAAGATGATGTGGACGATATGGTAGAGTTTGTGCGCGAGTTTGCCCCCCGGAGGCGCG
TGCAGTGGCGCGGGCGGAAAGTGCGTCCGGTGCTGAGACCCGGCACCACGGTGGTTTTCGCGCCTGGCGAGCGGTCCGGC
ACGACATCCAAGCGCTCCTACGATGAGGTGTACGGGGACGAGGATATTCTCGAGCAGGCGGCCGAGCGCCTGGGCGAGTT
TGCTTACGGCAAGCGCAACCGCCTTGCGCCCCTGAAGGAAGAGGTGGTGTCCATCCCGCTGGACCACGGCAACCCCACGC
CGAGTCTTAAGCCCGTGACCCTGCAGCAGGTGCTGCCGAGCGCGGCGCCGCGTCGGGGCTTGAAGCGCGAGGGCGAGGAT
GTGTACCCCACCATGCAGCTGATGGTGCCCAAGCGCCAGAAGCTGGAAGACGTGCTGGAGACCATGAAGGTGGACCCGGA
CGTGCAGCCCGAGGTCAAGGTGAGGCCCATCAAGCAGGTGGCCCCGGGCCTTGGCGTGCAGACCGTGGACATCAAGATCC
CCACGGAGCCCATGGAAACGCAGACCGAGGTCGTGAAGCCCATCACCAGCACCATGGAGGTGCAGACGGATCCTTGGATG
CCGGCGGCGCCCCGAAAACCCCGGCGCAAGTACGGCGCGGCCAGCCTGCTGATGCCCAACTACGCGCTGCATCCTTCCAT
CATCCCCACGCCGGGCTACCGCGGCACGCGCTTCTACCACGGCTATACCGGCTCCCGCCGCCGCAAGACCACCACCCGCC
GCCGTCGTCGCCGCACAGCTGCAACTCCCGCTGCCGCCCTGGTGCGGAGAGTGTACCGCCGCGGCCGCGCGCCTCTGACC
CTGCCGCGGGCGCGCTACCACCCGAGCATTACCATTTAACTTTGCCGTCGCCTTTGCAGATATGGCTCTCACATGCCGCA
TTCGCGTCCCCATTACGGGCTACCGAGGAAGAAAACCGCGCCGTAGAAGGCTGGCGGGAAGCGGGATGCGCCGCCACCCC
CACCGGCGGCGGCGCGCCATCAGCAAGCGGTTGGGGGGAGGCTTCCTGCCCGCGCTGATCCCCATCATCGCCGCGGCGAT
CGGGGCGATCCCCGGCATTGCTTCCGTGGCGGTGCAGGCCTCTCAGCGCCACTGAGACACACACTTGGAAATTGTAATAA
ACCCGAATGGACTCTGACGCTCCTGGTCCTGTGATGTGTTTTTGTAGACAGATGGAAGACATCAATTTTTCGTCCCTGGC
TCCGCGACACGGCACGCGGCCGTTTATGGGCACCTGGAGCGACATCGGCACCAGCCAACTGAACGGGGGCGCCTTCAATT
GGAGCAGTCTCTGGAGCGGGCTTAAGAATTTTGGGTCCACGCTTAAAACCTATGGCAGCAAGGCGTGGAACAGCACCACA
GGGCAGGCGCTGAGAGATAAGCTGAAAGAGCAGAACTTCCAGCAGAAGGTAGTCGATGGCCTCGCCTCAGGCATCAACGG
GGTGGTGGACCTGGCCAATCAGGCCGTGCAGCGGCAGATCAACAGCCGCCTGGACCCGGTTCCCCCCGCCGGCTCCGTGG
AGATGCCGCAGGTGGAGGAGGAGCTGCCTCCCCTGGACAAGCGGGGCGACAAGCGTCCCCGTCCCGACGCGGAGGAGACG
CTGCTGACGCACACGGACGAACCGCCCCCGTACGAGGAGGCGGTGAAACTGGGCCTGCCCACCACGCGTCCCATTGCGCC
TCTAGCTACCGGGGTGCTGAAACCCGAGAGTAGTAAGCCCGCGACCTTGGACTTGCCTCCTCCGCCCACTCCCCGCCCCT
CCACAGTGGCTAAGCCCCTGCCGCCGGTGGCCGTGGCCCGCGCGCGACCGGGGGCTCGCCCTCAGGCGAACTGGCAGAGC
ACTCTGAACAGCATCGTGGGTCTGGGAGTGCAGAGTGTGAAGCGCCGCCGCTGTTATTAAAAAACACTGTAGCGCTTAAC
TTGCTTGTCTGTGTATATGTGTATGTCCGCCGCCGCTGCTGTCCAGAAGGAGGAGTGAAGAGAAAGGCGCGTCGTCGAGT
TGCAAGATGGCCACCCCATCGATGCTGCCCCAGTGGGCGTACATGCACATCGCCGGACAGGACGCTTCGGAGTACCTGAG
TCCGGGTCTGGTGCAGTTCGCCCGCGCCACAGACACCTACTTCAGTCTGGGGAACAAGTTTAGGAACCCCACGGTGGCGC
CTACCCACGATGTGACCACCGACCGCAGCCAGCGGCTGACGCTGCGCTTTGTGCCCGTGGACCGGGAGGACAACACCTAC
TCGTACAAAGTGCGCTACACGCTGGCCGTGGGCGACAACCGCGTGCTGGACATGGCCAGCACCTACTTTGACATCCGCGG
CGTGCTGGATCGGGGCCCTAGCTTCAAACCCTACTCCGGCACTGCCTACAACAGCCTGGCTCCCAAGGGAGCGCCCAACA
CCTGCCAGTGGAAGGATTCTGACAGCAAAATGCATACCTTTGGGGCAGCTGCCATGCCCGGTGTTACTGGGAAAAAGATA
GAAGCTGATGGGCTGCCTATTAGAATAGATTCAACTTCTGGAACTGACACAGTAATTTATGCTGATAAAACTTTCCAACC
AGAACCACAAGTTGGAAATGACAGTTGGGTTGACACCAATGGTGCAGAGGAAAAATATGGAGGCAGAGCTCTAAAGGACA
CTACAAAAATGAAACCCTGTTATGGTTCATTCGCCAAGCCTACCAACAAAGAAGGTGGTCAGGCTAACTTAAAAGATTCA
GAACCCGCCGCCACCACTCCTAACTATGATATAGACCTGGCTTTCTTTGACAGCAAAACTATTGTTGCTAACTACGATCC
AGATATTGTAATGTACACAGAAAATGTTGACTTGCAGACTCCAGATACTCATATTGTATACAAACCTGGAACAGAGGACA
CCAGCTCTGAATCCAATTTGGGTCAGCAGGCCATGCCTAACAGACCCAACTACATTGGCTTCAGAGACAATTTTATCGGG
CTCATGTACTACAACAGCACTGGCAATATGGGGGTGCTGGCCGGTCAGGCCTCTCAGCTGAATGCTGTGGTTGACTTGCA
AGACAGAAACACTGAACTGTCCTACCAGCTCTTGCTTGACTCTCTGGGTGACAGAACCCGGTATTTCAGTATGTGGAATC
AGGCGGTGGACAGCTATGATCCTGATGTGCGCATTATTGAAAACCATGGTGTGGAGGATGAATTGCCAAACTATTGCTTT
CCGTTGAATGGTGTGGGATTGACAGACACTTACCAGGGTGTTAAAGTTAAAACAGATGCAGGTTCTGAAAAGTGGGACAA

AGATGACACCACAGTTAGTAATGCTAATGAAATCCATGTAGGCAATCCTTTTGCCATGGAAATCAACATCCAAGCCAACCTGTGGAGGAACTTCCTCTATGCCAATGTTGCCCTCTATTTGCCTGATAAATACAAATACACACCGGCCAACATCACCCTGCCCACCAACACCAACACCTACGAGTACATGAACGGCCGGGTGGTGGCGCCCTCGCTGGTGGACGCCTACATTAACATTGGGGCGCGCTGGTCGCTGGACCCCATGGACAACGTAAATCCCTTCAACCACCACCGCAATGCGGGCTTGCGCTACCGCTCCATGCTCCTGGGCAACGGGCGCTACGTGCCATTCCACATCCAGGTGCCCCAGAAATTTTTTGCCATTAAGAGCCTCCTGCTCCTGCCCGGGTCCTACACCTACGAGTGGAACTTCCGCAAGGACGTCAACATGATCCTGCAGAGTTCCCTTGGCAACGACCTGCGCACAGACGGGGCCTCCATCACCTTCACCAGCATTAACCTCTACGCCACCTTCTTCCCCATGGCGCACAACACCGCCTCCACGCTTGAGGCCATGCTGCGCAACGACACCAATGACCAATCCTTCAACGACTACCTCTCGGCGGCCAACATGCTCTATCCCATCCCGGCCAACGCCACCAACGTGCCCATCTCCATCCCCTCGCGCAACTGGGCCGCCTTTCGCGGCTGGTCCTTCACGCGTCTCAAGACCAAAGAGACGCCCTCGCTGGGCTCCGGGTTCGACCCCTACTTCGTCTACTCGGGCTCCATCCCCTACCTCGACGGCACCTTCTACCTCAACCACACCTTCAAGAAGGTCTCCATCACCTTCGACTCTTCCGTCAGCTGGCCCGGCAACGACCGGCTCCTGACGCCCAACGAGTTCGAAATCAAGCGCACCGTCGACGGCGAGGGATACAACGTGGCCCAGTGCAACATGACCAAGGACTGGTTCCTGGTCCAGATGCTGGCCCACTACAACATCGGCTACCAGGGCTTCTACGTGCCCGAGGGCTACAAGGACCGCATGTACTCCTTCTTCCGCAACTTCCAGCCCATGAGCCGCCAGGTGGTGGACGAGGTTAACTACAAGGACTACCAGGCCGTCACCCTGGCCTACCAACACAACAACTCGGGCTTCGTTGGATACCTCGCGCCCACTATGCGCCAGGGCCAGCCCTACCCCGCCAACTACCCCTACCCGCTCATCGGCAAGAGCGCCGTTACCAGCGTCACCCAGAAAAAGTTCATCTGCGACAGGGTCATGTGGCGCATCCCCTTCTCCAGCAACTTCATGTCCATGGGCGCGCTCACCGACCTCGGCCAGAACATGCTCTATGCTAACTCCGCCCACGCGCTAGACATGAATTTCGAAGTCGACCCCATGGATGAGTCCACCCTTCTCTATGTTGTCTTCGAAGTCTTCGACGTCGTCCGAGTGCACCAGCCCCACCGCGGCGTCATTGAGGCCGTCTACCTGCGCACCCCCTTCTCAGCCGGTAACGCCACCACATAAATTCTTGCTTCTTGCAAGAAGCCATGGCCGCGGGCTCCGGCGAGCAGGAGCTCAGGGCCATCATCCGCGACCTGGGGTGCGGGCCCTACTTCCTGGGCACCTTCGATAAGCGATTCCCGGGATTCATGGCCCCGCACAAGGTGGCCTGCGCCATCGTCAACACGGCCGGCCGCGAGACCGGGGGCGAGCATTGGCTGGCCTTCGCCTGGAACCCGCGCTCGAACACCTGCTACCTCTTCGACCCCTTCGGGTTCTCGGACCAGCGCCTCAAGCAAATCTACCAGTTCGAGTACGAGGGACTGCTGCGCCGCAGCGCCCTGGCCACCAAGGACCGCTGCGTTACCCTGGAAAAGTCCACCCAGACCGTGCAGGGTCCGCGTTCGGCCGCCTGCGGGCTTTTCTGCTGCATGTTCCTACACGCCTTCGTGCACTGGCCCAACCGCCCCATGGACAAAAATCCCACCATGAACTTGCTGACGGGGGTGCCCAACGGCATGCTCCAGTCGCCCCAGGTGGAACCTACCCTGCGCCGCAACCAGGAGGCACTCTACCGCTTCCTCAACTCCCACTCTGCATACTTTCGCTCTCACCGCGCGCGCATTGAGAAGGCCACCGCCTTCGACCGCATGAATCAAGACATGTAACAGTGTGTTTTAAAATATGTTTAATAAACAGCACTTTTTATGTGACACATGCATTTGAGATAATTTTATTCTTAAAAATCGAAGGGGTTCTGCCGGGAGGTTTCGGCATGGCCCGCGGGCAGGGACACGTTGCGGAACTGGTACTTGGCCAGCCACTTGAACTCGGGGATCAGCAGTTTCGGCAGCAGGGTGTCGGGGAACGAGTCGGTCCACAGCTTCCGCGTCAGTTGCAGGGCGCCCAGCAGGTCGGGCGCGGAGATCTTGAAATCGCAGTTGGGACCCGCGTTTTGCGCGCGAGAGTTGCGGTACACAGGGTTGCAGCACTGGAACACCATCAGGGCCGGATGCTTCACGCTCGCCAGCACCGTAGCGTCGGTGATCCCGTCCACGTCGAGGTCTTCGGCGTTGGCCATCCCGAAGGGGGTCATCTTGCAGGTCTGCCGGCCCATGGTGGGCACGCAGCCGGGCTTGTGGTTGCAATCGCAGTGCAGGGGGATCAGCATCATCTGGGCCTGGTCGGCGTTCATCCCCGGGTACATGGCCTTCATGAAAGCCTCCAGCTGCTTAAACGCCTGCTGGGCCTTGGCTCCCTCGGTGAAGAAGACCCCGCAGGACTTGCTAGAAAACTGGTTGGTAGCGCACCCGGCGTCGTGCACGCAGCAGCGCGCGTCGTTGTTGGCCAGCTGCACCACGCTGCGCCCCCAGCGGTTCTGGGTAATCTTGGCCCGGTCGGGGTTCTCCTTTAGCGCGCGTTGCCCGTTCTCGCTTGCCACATCCATCTCGATCATGTGCTCCTTCTGGATCATGGTGGTCCCGTGCAGGCACCGCAGCTTGCCCTCGACTTCGGTACAGCCGTGCAGCCACAGCGCGCACCCCGTGCTCTCCCAGTTCTTGTGGGCGATCTGGGAATGCGCATGCACGAACCCCTGCAGGAAGCGGCCCATCATGGTCGTCAGGGTCTTGTTACTGGTAAAGGTCAGCGGAATGCCGCGGTGCTCCTCGTTGATGTACAGGTGGCAGATGCGGCGATACACCTCGCCCTGCTCGGGCATCAGTTGGAAGTTGGATTTTAGGTCGCTTTCCACACGGTAGCGCTCCATCAGCATATTCATGATTTCCATGCCCTTCTCCCAGGCCGATACAATGGGCAGGCTCAGGGGGTTCGTCACCGCCATCTTAGCGCTAGCAGCCTTCGTCAGCGGGTCGTTCTCATTGAGAGTCTCAAAGCTCCGCTTGCCGTCCTTCTCGGTGATCCGCACGGGGGGGTAGCTGAAGCCCACGGCCGCCAGCTCCTCCTCGGCCTCTCTTTCGTCCTCGCTGTCCTGGCTGACGTCCTGCAGGGGCACATGCTTCGTTTTGCGGGGTTTCTTTTTGGGCGGCTGCTGCGGCGGCGGTGGTTGTTCCTGAGGCGAGGGGGAGCGCGAGTTCTCGCTCACCACTACTATCTCTTCTTCTTGGTCCGAGGCCACGCGGCGGTAGGTATGTCTCTTCAGGGGCAGAGGCGGAGGCGACGGGCTCTCGCGGCCCGGCGGGTGGCTGGCAGAGCCCCTTCCGCGATCGGGGGTGCGCTCCCGGCGGCGCTCTAACTGACTTCCTCCGCGGCCGGCCATTGTGTTCTCCTAGGGAACAACAACAAGCATGGAGACTCAGCCATCGTCGCCAACCTCGCCATCTGCCCCCACCGCCGACAAGAAGCAGCAGCAGAATGAGAGCTTAACCGCCCCGCCGCCCAGCCCCGCCACCTTTGTCGCGGCCCCAGACATGCAAGAGATGGAGGAATCCATTCAGATTGACCTGGGCTATGTGACGCCCGCGGAGCACGAGGAGGAGCTTGCAGTGCGCTTTTCAACCCAGGAAGAGATACACCAAGAACAGCCAGAGCAGGAAGCAAAGAGCGAGCATGACTACCTCCACCAGAGCGGGGGGGAGGACGCCCTCATCAAGCATCTGGCCCGGCAGGCCATCATCGTCAAGGACGCGCTGCTTGACCGCACCGAGGTGCCCCTCAGCGTGGAGGAGCTCAGCCGCGCCTACGAGCTCAACCTCTTCTCGCCGCGCGTGCCCCCCAAGCGCCAGCCCAACGGCACCTGCGAGCCCAACCCACGCCTCAACTTCTACCCGGTCTTCGCGGTGCCCGAGGCCCTGGCCACCTACCACATCTTTTTCAAGAACCAAAGGATCCCTGTCTCCTGTCGCGCCAACCGCACCCGCGCCGACTCCCTTTTCAACCTGGGCCCCGGTGCCCGCCTACCTGATATCGCCTCCTTGGAAGAGGTTCCCAAGATCTTCGAGGGTCTGGGCAGCGACGAGACTCGGGCCGCAAACGCTCTGCAAGGAGAAGGAGGAGATCATGAGCACCACAGCGCCCTGGTGGAGTTGGAAGGCGACAACGCGCGTCTGGCGGTGCTCAAGCGCACGATCGAGCTGACCCATTTCGCCTACCCGGCGCTTAACCTGCCCCCCAAAGTCATGAGCACGGTTATGGATCAGGTGCTCATCAAGCGCGCGTCGCCCATCTCCAAGGAGATGCAAGACCCCGAGAGCTCCGAGGAGGGCAAGCCCGTGGTCAGCGACGAGCAGCTGGCGCGGTGGCTGGGACCCCAAGCTAGTCCCCAGAGCTTGGAAGAGCGGCGCAAGCTCATAATGGCCGTGGTCCTGGTGACCGCGGAGCTGGAGTGTCTGCGCCGCTTCTTCGCCGACGCAGAAATTCTGCGCAAGGTCGAGGAGAACCTGCACTACATCTTCAGGCACGGGTTCGTACGCCAGGCCTGCAAGATCT

CCAACGTGGAGCTGACCAACCTGGTCTCCTACATGGGCATCTTGCACGAGAACCGCCTGGGGCAGAACGTGCTGCACACC
ACCCTGCGCGGGGAGGCCCGCCGCGACTACATCCGCGACTGCGTTTACCTCTACCTCTGCCACACCTGGCAGACAGCCAT
GGGCGTGTGGCAGCAGTGTCTGGAGGAGCAGAACCTAAAAGAGCTCTGCAAGCTCCTGCAGAAGAACCTCAAGGCCCTGT
GGACCGGGTTCGACGAGCGCACCACCGCCTCGGACCTGGCAGACCTCATTTTCCCCGAGCGTCTCAGGCTGACGCTGCGC
AACGGTTTGCCCGACTTTATGAGTCAAAGCATGTTGCAAAACTTTCGCTCTTTCATCCTCGAACGCTCCGGGATCCTGCC
GGCCACCTGCTCCGCGCTGCCCTCGGACTTCGTGCCGCTGACCTTCCGCGAGTGCCCCCCGCCGCTGTGGAGCCACTGCT
ACCTGCTGCGCTTGGCCAACTACCTGGCCTACCACTCGGACGTGATCGAGGACGTCAGCAGCGAGGGCCTGCTCGAGTGC
CACTGCCGCTGCAACCTCTGCACGCCGCACCGCTCCCTGGCCTGCAACCCCCAGCTGCTGAGCGAGACCCAGATCATCGG
CACCTTCGAGTTGCAAGGGCCCGGCGATGAGGGTTCTGCCGCCAAGGGGGGTCTGAAACTCACCCCGGGGCTGTGGACCT
CGGCCTACTTGCGCAAGTTCGTGCCCGAGGACTACCATCCCTTCGAGATCAGGTTCTACGAGGACCAATCCCAGCCGCCC
AAGGCCGAGCTGTCGGCCTGCGTCATCACCCAGGGGGCGATCCTGGCCCAATTGCAAGCTATCCAGAAATCCCGCCAAGA
ATTCTTGCTGAAAAAGGGCCGCGGGGTCTACCTTGATCCCCAGACCGGTGAGGAGCTTAACCCCGGCTTCCCCCAGGATG
CCCCGAGGAAGCAGCAAGAAGCTGAAAGTGGAGCTGCCGCCCGTGGAGGATTTGGAGGAAGACTGGGAGAGCAGTCAGGC
AGAGGAGGAGGAGATGGAAGACTGGGACAGCACTCAGGCAGAGGACAGCCTGCAAGACAGTCTGGAAGACGAGGAGGAGG
CAGAGGAGGTGGAAGAAGTAGCCGCCGCCGCCAGACCGTCGTCCTCGGCGGAGAAAGCAAGCAGCACGGATACCATCTCC
GCTCCGGGTCGGGGTCCCGCTCGACCCCACAGTAGATGGGACGAGACCGGGCGATTCCCGAACCCCACCACCCAGACCGG
TAAGAAGGAGCGGCAGGGATACAAGTCCTGGCGGGGGCACAAAAACGCCATCGTCTCCTGCTTGCAAGCTTGCGGGGGCA
ACATCTCATTCACCCGGCGCTACCTGCTCTTTCACCGCGGGGTGAACTTCCCCCGCAACATCTTGCATTACTACCGTCAC
CTCCACAGCCCCTACTACTTCCAAGAAGAGGCAGAAAAAGACAAAACCAGCAGCTAGAAAATCCACAGCGGCGGCGGCGG
CAGGTGGACTGAGGATCGCGGCGAACGAGCCGGCGCAGACCCGGGAACTGAGGAACCGGATCTTTCCCACCCTCTATGCC
ATCTTCCAGCAGAGTCGGGGGCAGGAGCAGGAACTGAAAGTCAAGAACCGTTCTCTGCGCTCGCTCACCCGCAGTTGTCT
GTATCACAAGAGCGAAGACCAACTTCAGCGCACGCTTGAGGACGCCGAGGCTCTCTTCAACAAGTACTGCGCACTCACTC
TTAAAGAGTAGCCCGCGCCCGCCCACACACGGAAAAAGGCGGGAATTACGTCACCTGTGCACCCCCACCCAGCACCGCTA
TGAGCAAAGAAATTCCCACGCCTTACATGTGGAGCTACCAGCCCCAGATGGGCCTGGCCGCCGGCGCCGCCCAGGACTAC
TCCACCCGCATGAATTGGCTCAGCGCCGGGCCCGGGATGATCTCACGGGTGAATGACATCCGCGCCCACCGAAACCAGAT
ACTCCTAGAACAGTCAGCGCTCACCGCCACGCCCCGCAATCACCTCAATCCGCGTAATTGGCCCGCCGCCCTAGTGTACC
AGGAAATTCCCCAGCCCACGACCGTACTACTTCCGCGAGACGCCCAGGCCGAAGTCCAGCTGACTAACTCAGGTGTCCAG
CTGGCGGGCGGCGCCACCCTGTGTCGTCACCACCCCGCTCAGGGTATAAAGCGGCTGGTGATCCGGGGCAGAGGCACACA
GCTCAACGACGAGGTGGTGAGCTCTTCACTGGGTTTGCGACCTGACGGAGTCTTCCAACTCGCCGGATCGGGAAGATCTT
TTCGGGGCAACATCTCATTCACCCGGCGCTACCTGCTCTTTCACCGCGGGGTGAACTTCCCCCGCAACATCTTGCATTAC
TACCGTCACCTCCACAGCCCCTACTACTTCCAAGAAGAGGCAGAAAAAGACAAAACCAGCAGCTAGAAAATCCACAGCGG
CGGCGGCGGCAGGTGGACTGAGGATCGCGGCGAACGAGCCGGCGCAGACCCGGGAACTGAGGAACCGGATCTTTCCCACC
CTCTATGCCATCTTCCAGCAGAGTCGGGGGCAGGAGCAGGAACTGAAAGTCAAGAACCGTTCTCTGCGCTCGCTCACCCG
CAGTTGTCTGTATCACAAGAGCGAAGACCAACTTCAGCGCACGCTTGAGGACGCCGAGGCTCTCTTCAACAAGTACTGCG
CACTCACTCTTAAAGAGTAGCCCGCGCCCGCCCACACACGGAAAAAGGCGGGAATTACGTCACCTGTGCACCCCCACCCA
GCACCGCTATGAGCAAAGAAATTCCCACGCCTTACATGTGGAGCTACCAGCCCCAGATGGGCCTGGCCGCCGGCGCCGCC
CAGGACTACTCCACCCGCATGAATTGGCTCAGCGCCGGGCCCGGGATGATCTCACGGGTGAATGACATCCGCGCCCACCG
AAACCAGATACTCCTAGAACAGTCAGCGCTCACCGCCACGCCCCGCAATCACCTCAATCCGCGTAATTGGCCCGCCGCCC
TAGTGTACCAGGAAATTCCCCAGCCCACGACCGTACTACTTCCGCGAGACGCCCAGGCCGAAGTCCAGCTGACTAACTCA
GGTGTCCAGCTGGCGGGCGGCGCCACCCTGTGTCGTCACCACCCCGCTCAGGGTATAAAGCGGCTGGTGATCCGGGGCAG
AGGCACACAGCTCAACGACGAGGTGGTGAGCTCTTCACTGGGTTTGCGACCTGACGGAGTCTTCCAACTCGCCGGATCGG
GAAGATCTTCCTTCACGCCTCGTCAGGCCGTGCTGACTTTGGAGAGTTCTTCCTCGCAACCTCGCTCGGGCGGCATCGGC
ACTCTCCAGTTTGTGGAGGAGTTCACTCCCTCGGTCTACTTCAACCCCTTCTCCGGCTCCCCCGGCCACTATCCGGACGA
GTTCATCCCGAACTTCGATGCCATCAGCGAATCGGTAGACGGCTACGATTGAATGTCCCATGGTGGCGCGGCTGACCTAG
CTCGGCTTCGACACCTGGACCACTGCCGCCGCTTTCGCTGCTTCGCTCGGGACCTCGCCGAGTTTACCTACTTTGAGCTG
TCCGAGGAGCACCCTCAGGGCCCGGCCCACGGAGTGCGGATCGTCGTCGAAGGGGGCCTAGACTCCCACCTGCTTCGTAT
CTTCAGCCAGCGCCCGATCCTGGTCCAGCGCCAACAGGGCAACACCCTCCTGACCCTTTACTGCATCTGCAACCACCCCG
GCCTGCACGAAAGTCTTTGTTGTCTGCTGTGTACTGAGTATAATAAAAGCTGAGATCAGCGACTACTCCGGACTCGATTG
TGTTCCAGCAGTCTGGCGATACCAAGGGTTGCATCCACTGCTCCTGCGACTCCCCCGAGTGCGTTCACACCCTCATCAAG
ACCCTATGCGGCCTCCGCGACCTCCTCCCCATGAACTAATCAACTAACCCCTTACCCCATTACCCATCCAGTAAAAAAAA
TAAAGATTAAAGAGACGATGATTTTGAATTACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATA
TATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAA
TAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCC
CACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCA
TTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGA
TGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTC
AATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGG
CGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCACTGTCTTCCGGATCGCTGTCCAGGAGCGCCAGCTGTTG
GGCTCGCGGTTGAGAAGGAACTCTTCGCGGTCCTTCCAGTACTCTTCAAGGGGGAACCCGTCCTGGTCGGCACGGGACTC
CGCGCAAGGACCTAAGCGTCTCCAGATCCACGGGATCTGAAAACCGTTGAACGAAGGCTTCGAGCCAGTCGCAGTCGCAA
GTCTAGAGCCACCATGTTCGTCTTCCTGGTCCTGCTGCCCCTGGTCTCATCTCAGTGCGTGAATCTGACTACAAGAACTC
AGCTGCCTCCCGCCTACACCAATTCCTTCACCCGGGGCGTGTACTATCCTGACAAGGTGTTTAGAAGCTCCGTGCTGCAC

TCTACACAGGATCTGTTTCTGCCATTCTTTAGCAACGTGACCTGGTTCCACGCCATCCACGTGAGCGGCACCAATGGCAC
AAAGCGGTTCGACAATCCCGTGCTGCCTTTTAACGATGGCGTGTACTTCGCCTCTACCGAGAAGAGCAACATCATCAGAG
GCTGGATCTTTGGCACCACACTGGACTCCAAGACACAGTCTCTGCTGATCGTGAACAATGCCACCAACGTGGTCATCAAG
GTGTGCGAGTTCCAGTTTTGTAATGATCCCTTCCTGGGCGTGTACTATCACAAGAACAATAAGAGCTGGATGGAGTCCGA
GTTTAGAGTGTATTCTAGCGCCAACAATTGCACATTTGAGTACGTGTCCCAGCCTTTCCTGATGGACCTGGAGGGCAAGC
AGGGCAATTTCAAGAACCTGAGGGAGTTCGTGTTTAAGAATATCGATGGCTACTTCAAGATCTACTCTAAGCACACCCCC
ATCAACCTGGTGCGCGACCTGCCTCAGGGCTTCAGCGCCCTGGAGCCACTGGTGGATCTGCCTATCGGCATCAACATCAC
CCGGTTTCAGACACTGCTGGCCCTGCACAGAAGCTACCTGACACCCGGCGACTCCTCTAGCGGATGGACCGCAGGAGCAG
CAGCCTACTATGTGGGCTATCTGCAGCCTAGGACCTTCCTGCTGAAGTACAACGAGAATGGCACCATCACAGACGCAGTG
GATTGCGCCCTGGACCCCCTGAGCGAGACAAAGTGTACACTGAAGTCCTTTACCGTGGAGAAGGGCATCTATCAGACATC
CAATTTCAGGGTGCAGCCAACCGAGTCTATCGTGCGCTTTCCTAATATCACAAACCTGTGCCCATTTGGCGAGGTGTTCA
ACGCAACCAGGTTCGCAAGCGTGTACGCATGGAATAGGAAGCGCATCTCTAACTGCGTGGCCGACTATAGCGTGCTGTAC
AACTCCGCCTCTTTCAGCACCTTTAAGTGCTATGGCGTGTCCCCCACAAAGCTGAATGACCTGTGCTTTACCAACGTGTA
CGCCGATTCTTTCGTGATCAGGGGCGACGAGGTGCGCCAGATCGCACCTGGACAGACAGGCAAGATCGCCGACTACAATT
ATAAGCTGCCAGACGATTTCACCGGCTGCGTGATCGCCTGGAACAGCAACAATCTGGATTCCAAAGTGGGCGGCAACTAC
AATTATCTGTACCGGCTGTTTAGAAAGAGCAATCTGAAGCCCTTCGAGAGGGACATCTCTACAGAGATCTACCAGGCCGG
CAGCACCCCTTGCAATGGCGTGGAGGGCTTTAACTGTTATTTCCCACTGCAGTCCTACGGCTTCCAGCCCACAAACGGCG
TGGGCTATCAGCCTTACCGCGTGGTGGTGCTGAGCTTTGAGCTGCTGCACGCACCAGCAACAGTGTGCGGACCCAAGAAG
TCCACCAATCTGGTGAAGAACAAGTGCGTGAACTTCAACTTCAACGGCCTGACCGGAACAGGCGTGCTGACCGAGTCCAA
CAAGAAGTTCCTGCCATTTCAGCAGTTCGGCAGGGACATCGCAGATACCACAGACGCCGTGCGCGACCCACAGACCCTGG
AGATCCTGGATATCACACCCTGCTCTTTCGGCGGCGTGAGCGTGATCACACCAGGAACCAATACAAGCAACCAGGTGGCC
GTGCTGTATCAGGACGTGAATTGTACCGAGGTGCCTGTGGCCATCCACGCCGATCAGCTGACCCCAACATGGCGGGTGTA
CAGCACCGGCTCCAACGTGTTCCAGACAAGAGCAGGATGCCTGATCGGAGCAGAGCACGTGAACAATTCCTATGAGTGCG
ACATCCCAATCGGCGCCGGCATCTGTGCCTCTTACCAGACCCAGACAAACTCTCCAAGGAGAGCACGGAGCGTGGCATCC
CAGTCTATCATCGCCTATACCATGTCCCTGGGCGCCGAGAATTCTGTGGCCTACTCTAACAATAGCATCGCCATCCCAAC
CAACTTCACAATCTCTGTGACCACAGAGATCCTGCCCGTGTCCATGACCAAGACATCTGTGGACTGCACAATGTATATCT
GTGGCGATTCTACCGAGTGCAGCAACCTGCTGCTGCAGTACGGCAGCTTTTGTACCCAGCTGAATAGAGCCCTGACAGGC
ATCGCCGTGGAGCAGGATAAGAACACACAGGAGGTGTTCGCCCAGGTGAAGCAGATCTACAAGACCCCCCCTATCAAGGA
CTTTGGCGGCTTCAATTTTTCCCAGATCCTGCCTGATCCATCCAAGCCTTCTAAGCGGAGCTTTATCGAGGACCTGCTGT
TCAACAAGGTGACCCTGGCCGATGCCGGCTTCATCAAGCAGTATGGCGATTGCCTGGGCGACATCGCAGCACGGGACCTG
ATCTGTGCCCAGAAGTTTAATGGCCTGACCGTGCTGCCACCCCTGCTGACAGATGAGATGATCGCACAGTACACAAGCGC
CCTGCTGGCAGGAACCATCACATCCGGATGGACCTTCGGCGCAGGAGCCGCCCTGCAGATCCCCTTTGCCATGCAGATGG
CCTATAGGTTCAACGGCATCGGCGTGACCCAGAATGTGCTGTACGAGAACCAGAAGCTGATCGCCAATCAGTTTAACTCC
GCCATCGGCAAGATCCAGGACAGCCTGTCCTCTACAGCCTCCGCCCTGGGCAAGCTGCAGGATGTGGTGAATCAGAACGC
CCAGGCCCTGAATACCCTGGTGAAGCAGCTGAGCTCCAACTTCGGCGCCATCTCTAGCGTGCTGAATGATATCCTGAGCC
GGCTGGACAAGGTGGAGGCAGAGGTGCAGATCGACCGGCTGATCACAGGCAGACTGCAGTCTCTGCAGACCTATGTGACA
CAGCAGCTGATCAGGGCAGCAGAGATCAGGGCAAGCGCCAATCTGGCAGCAACCAAGATGTCCGAGTGCGTGCTGGGCCA
GTCTAAGAGAGTGGACTTTTGTGGCAAGGGCTATCACCTGATGTCCTTCCCTCAGTCTGCCCCACACGGCGTGGTGTTTC
TGCACGTGACCTACGTGCCCGCCCAGGAGAAGAACTTCACCACAGCCCCTGCCATCTGCCACGATGGCAAGGCCCACTTT
CCAAGGGAGGGCGTGTTCGTGTCCAACGGCACCCACTGGTTTGTGACACAGCGCAATTTCTACGAGCCCCAGATCATCAC
CACAGACAATACCTTCGTGAGCGGCAACTGTGACGTGGTCATCGGCATCGTGAACAATACCGTGTATGATCCACTGCAGC
CCGAGCTGGACAGCTTTAAGGAGGAGCTGGATAAGTACTTCAAGAATCACACCTCCCCTGACGTGGATCTGGGCGACATC
AGCGGCATCAATGCCTCCGTGGTGAACATCCAGAAGGAGATCGACCGCCTGAACGAGGTGGCCAAGAATCTGAACGAGAG
CCTGATCGATCTGCAGGAGCTGGGCAAGTATGAGCAGTACATCAAGTGGCCATGGTACATCTGGCTGGGCTTCATCGCCG
GCCTGATCGCCATCGTGATGGTGACCATCATGCTGTGCTGTATGACATCCTGCTGTTCTTGCCTGAAGGGCTGCTGTAGC
TGTGGCTCCTGCTGTAAGTTTGATGAGGACGATTCCGAACCCGTGCTGAAGGGAGTGAAGCTGCATTACACCTGAGGATC
CCTCGAGCTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCAC
TCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGG
TGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGTGATCAAT
AAAGAATCACTTACTTGAAATCTGAAACCAGGTCTCTGTCCATGTTTTCTGTCAGCAGCACTTCGCTCCCCTCTTCCCAG
CTCTGGTACTGCAGGCCCCGGCGGGCTGCAAACTTCCTCCACACTCTGAAGGGGATGTCAAATTCCTCCTGTCCCTCAAT
CTTCATTTTTTATTTCTATTAGATGTCCAAAAAGCGCGCGCGGGTGGATGATGGCTTCGACCCCGTGTATCCCTACGATG
CAGACAACGCACCGACCGTGCCCTTCATCAACCCTCCCTTCGTCTCTTCAGATGGATTCCAAGAAAAGCCCCTGGGGGTG
TTGTCCCTTAGGCTGGCCGACCCTGTCACCACCAAGAATGGGGAAATTACCCTCAAGCTGGGGGAGGGGGTGGACCTTGA
CGACTCGGGAAAACTCATTGCAAACACAGTAAACAAGGCCATTGCCCCTCTCAGTTTTTCCAACAACACCATTTCCCTTA
ACATGGATACCCCTTTATACACCAAAGATGGAAAACTATCCTTACAAGTTTCTCCACCATTAAGTATATTAAAATCAACA
ATTTTGAATACATTAGCTCTAGCTTTTGGCTCAGGTTTAGGACTCAGTGGCAGCGCCCTGGCAGTACAGTTAGCCTCTCC
ACTTACATTTGATGATAAAGGGAATATAAAGATTACCCTAAACAGGGGATTGCATGTTACAACAGGAGATGCAATTGAAA
GCAACATCAGTTGGGCTAAAGGTATAAAATTTGAAGATGGTGCCATAGCTACAAACATTGGTAAGGGGCTAGAGTTCGGA
ACCAGTAGTACAGAAACAGGAGTTAATAATGCTTATCCAATCCAAGTTAAACTTGGCTCTGGTCTCAGCTTTGACAGCAC
AGGAGCCATAATGGCTGGCAATAAAGACTATGATAAATTAACTTTGTGGACAACGCCTGACCCATCACCAAACTGTCAAA
TACTTGCAGAAAATGATGCAAAACTAACACTTTGCTTAACTAAGTGTGACAGTCAAATACTGGCCACTGTATCAGTTTTG

GTTGTTAGAAGTGGAAACTTAAACCCAATTACTGGCACAGTAAGCAGTGCTCAAGTTTTTCTACGTTTTGATGCAAATGG
TGTTCTTTTAACAGAACACTCTACACTAAAAAAATACTGGGGCTACAAGCAAGGAGATAGCATAGATGGCACTCCATACA
CCAATGCTGTTGGTTTTATGCCAAATTCAACAGCTTATCCAAAGACCCAAAGTTCTACTACTAAAAATAATATAGTGGGT
CAAGTATACATGAATGGAGATGTTTCAAAACCCATGCTTCTTACTATAACTCTTAATGGTACTGATGACACCACCAGTGC
ATACTCAATGTCATTTTCATACACCTGGACTAACGGAAGCTATATCGGAGCAACATTTGGAGCTAACTCATACACCTTCT
CCTACATAGCCCAACAATAATCCCACCCTGCATGCCAACCCACCTTTTCCCTCTATTTATAAATGGAAACTGAAACAAAA
ATAAAGTTCAAGTGTTTTATTGATTCAACAGTTTTCACAGGATTCGAGTAGTTATTTCCCTCCACCCTCCCATCTCAT
GGAATACACTATCCTCTCCCCACGCACAGCCTTAAACATCTGAATGCTATTGGTAATGGACATGGTTTTGATCTCCACAT
TCCACACAGTTTCAGAGCGAGACAGTCTCGGGTCGGTCAAGGAGATGAAACCCTCCGGGCACTCCTGCATCTGCACCTCA
CAGTTCAACAGCTGAGGGCTGTCCTCGGTGATTGGAATCACAGTTATCTGGAATAAGAGCGATGAGAATCATAATCCGCA
AACGGGATCGGGCGGTTGTGGCGCATCAGGCCCCGCAGCAGTCGCTGTCTGCGCCGCTCCGTCAAGCTGCTACTCAAGGG
GTCCGGGTCCAGGGACTCCCTGCGCATGATGCCAATGGCCCTGAGCATCAGTCGCCTGGTACGGCGGGCGCAGCAGCGGA
TGCGGATCTCACTCAGGTCGGAGCAGTACGTGCAGCACAGCACCACCAAGTTGTTCAACAGTCCATAGTTCAACGTGCTC
CAGCCAAAACTCATTTGTGGAACTATGCTGCCCACATGTCCATCGTACCAGATCCTGATGTAAATCAGGTGGCGTCCCCT
CCAGAACACACTGCCCATGTACATGATCTCCTTGGGCATGTGCAGGTTCACCACCTCCCGGTACCACATCACCCGCTGGT
TGAACATGCAGCCCTGGATAATTCTGCGGAACCAGATGGCAAGTACCGTCCCGCCCGCCATGCAGCGCAGGGACCCCGGG
TTCTGGCAATGGCAGTGGATCACCCACCGCTCGCGACCGTGGATCAACTGGGAACTAAACAAGTCTATGTTGGCACAGCA
CAGGCACACGCTCATGCATGTCTTCAGCACTCTCAATTCCTCGGGGGTCAGGACCATATCCCAGGGCACAGGGAACTCTT
GCAGGACAGTGAACCCGGCCGAACAGGGCAATCCTCGCACGGAACTTACATTGTGCATGGACAGGGTATCGCAATCAGGC
AGCACCGGATGATCCTCCACCAGAGAAGCGCGGCTCTCGGTCTCCTCACAGCGAGGTAAGGTGGCCGGCGGTTGGTACGG
ATGATGGCGAGATAACGCTAATCGTGTTCTGGATCGTGTCATGATGGAGCTGTTTCCGGACATTTTCGTATTTCACAAAG
CAGAACCTGGTCCGGGCACTGCACACCGCTCGTCGGCGACGGTCTCGGCGCTTCGAGCGCTCAATGTTGAAGTTATAGAA
CAGCCACTCCCTCAGAACGTGCAGTATCTCCTGAGCCTCTTGGGTGATGAAAATCCCATCCGCCCTGATGGCTCTGATTA
CATCAACCACGGTGGAATGGGCCAAACCCAGCCAGATGATGCAATTTTGTTGGGTTTCGGTGACGGCGGGGGAGGGAAGA
ACAGGAAGAACCATGATTAACTTTATTCCAAACGGTCTCGGAACACTTCAAAATGCAGGTCCCGGAGGTGGCACCTCTCG
CCCCCACTGTGTTGGTGGAAAATAACAGCCAGGTCAAAGGTAACACGGTTCTCGAGATGTTCCACGGTGGCTTCCAGCAA
AGCCTCCACGCGCACATCCAGAAACAAGAGGACAGCGAAAGCGGGAGCGTTTTCTAATTCCTCAATCATCATATTACACT
CCTGCACCATGCCTAGATAATTTTCATTTTTCCAGCCTTGAATGATTCGTATTAGTTCCTGAGGTAAATCCAAGCCAGCC
ATGATAAAAAGCTCGCGCAGAGCGCCCTCCACCGGCATTCTTAAGCACACCCTCATAATTCCAACAGATTCTGCTCCTGG
TTCACCTGTAGTAGATTAACAAGTGGAATATCAATTGCTCTGCCGCAATCCCTAAGCTCCTCCCTTAGCAGTAACTGTAT
GTACTCATTCATATCTTCTCCGAAATTTTTAGCCATAGGACCACCAGGAACAAGAGAAGGGCAAGCCACATTACAGATAA
AGCGAAGTCCTCCCCAGTGAGCATTGCCAAATGTAAGATTGAAATAAGCATGCTGGCTAGACCCGGTGATATCTTCCAGA
TAACTGGACAGAAAATCAGGCAAGCAATTTTTAAGAAAATTAACAAAAGAAAAGTCGTCTAGGTGCACGTTTAGAGCCTC
AGGAACAACGATGGAATAAGTGCAAGGAGTACGTTCCAGCATGGTTAGTGTTTTTGGTGATCTGTAGAACAAAAAATAAA
CATGCAATATTAAACCATGCTAGCCTGGCGAACAGGTGGATAAATCACTCTTTCCAACACCAGGCAGGCTACAGGGTCTC
CGGCGCGACCATTGTAGAAGCTGACATTATGATTAAAAAGCATCACCGACAGACCTTCCCGGTGGCCGGCATGGATGATT
CGAGAAGAAGCATACACTCCGGGAACATTGGCGTCCGTGAGTGAAAAAAAGCGACCTATAAAGCCTTGAGGCACTACAAT
GCTTAATCTTAATTCCAGCAAAGCGACCCCATGCGGATGAAGCACAAAATTGGCAGGTGCGTAAAAAATGTAATTACTCC
CCTTCTGCACAGGCAGCAAAGCCCCCGCTCCCTCCAGAAACACATACAAAACCTGAGCGTCCATAGCTTACCGAGCACGG
CAGGCGCAAGAGTCAGAGAAAAAGCTGAGCTCTAACCTAACTGCCCGCTTCTGTACTCAATATATAGCCCTAACCTCACT
GACGTAAAGGCCAAGGTCTAAAAATACCCGCCAACACGCCCAGAAACCGGTGACACACTAAAAAAATACGTGCACTTCCT
CAAACGCCCAAACTGGCGTCATTTCCGGTTTCCCACGCTACGTCACCTCTCAACGACTTTCAAATTCCGTCGACCGTTAA
ACACATCAGTTACCCCGCCCCTAACGAACGCCGCTGTCACAGCCAATCAGCGCGCCCCATCCCCAAATTTTCACGCCTTA
TTTGCATATTAACTCACACAAAAAAAATAAGGTATATTATTGATGATGAAGCTTTTAAT

SEQ ID NO: 2 is the amino acid sequence of a wild-type SARS-CoV-2 (Wuhan strain) spike protein deposited under GenBank Accession No. YP_009724390.1.

MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFD
NPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVY
SSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT
LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRV
QPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSF
VIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL
PFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGS
NVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGF
NFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAG
TITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN
TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRV

DFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT
FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL
QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

SEQ ID NO: 3 is the amino acid sequence of a stabilized SARS-CoV-2 spike protein with a double proline substitution (nCoV-PP).

MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFD
NPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVY
SSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT
LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRV
QPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSF
VIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL
PFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGS
NVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGF
NFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAG
TITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN
TLVKQLSSNFGAISSVLNDILSRLD**PP**EAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRV
DFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT
FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL
QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

SEQ ID NO: 4 is the amino acid sequence of a tail-truncated SARS-CoV-2 spike protein (nCoV-TT).

MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFD
NPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVY
SSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT
LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRV
QPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSF
VIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL
PFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGS
NVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGF
NFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAG
TITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN
TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRV
DFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT
FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL
QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCS

SEQ ID NO: 5 is the amino acid sequence of a SARS-CoV-2 spike protein lacking the C-terminal endocytosis motif (nCoV-noEndo).

MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFD
NPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVY
SSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT
LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRV
QPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSF
VIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL
PFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGS
NVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGF
NFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAG
TITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN

TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRV
DFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT
FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL
QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGV

SEQ ID NO: 6 is a nucleic acid sequence encoding a SARS-CoV-2 spike protein.

ATGTTTGTTTTTCTTGTTTTATTGCCACTAGTCTCTAGTCAGTGTGTTAATCTTACAACCAGAACTCAATTACCCCCTGC
ATACACTAATTCTTTCACACGTGGTGTTTATTACCCTGACAAAGTTTTCAGATCCTCAGTTTTACATTCAACTCAGGACT
TGTTCTTACCTTTCTTTTCCAATGTTACTTGGTTCCATGCTATACATGTCTCTGGGACCAATGGTACTAAGAGGTTTGAT
AACCCTGTCCTACCATTTAATGATGGTGTTTATTTTGCTTCCACTGAGAAGTCTAACATAATAAGAGGCTGGATTTTTGG
TACTACTTTAGATTCGAAGACCCAGTCCCTACTTATTGTTAATAACGCTACTAATGTTGTTATTAAAGTCTGTGAATTTC
AATTTTGTAATGATCCATTTTTGGGTGTTTATTACCACAAAAACAACAAAAGTTGGATGGAAAGTGAGTTCAGAGTTTAT
TCTAGTGCGAATAATTGCACTTTTGAATATGTCTCTCAGCCTTTTCTTATGGACCTTGAAGGAAAACAGGGTAATTTCAA
AAATCTTAGGGAATTTGTGTTTAAGAATATTGATGGTTATTTTAAAATATATTCTAAGCACACGCCTATTAATTTAGTGC
GTGATCTCCCTCAGGGTTTTTCGGCTTTAGAACCATTGGTAGATTTGCCAATAGGTATTAACATCACTAGGTTTCAAACT
TTACTTGCTTTACATAGAAGTTATTTGACTCCTGGTGATTCTTCTTCAGGTTGGACAGCTGGTGCTGCAGCTTATTATGT
GGGTTATCTTCAACCTAGGACTTTTCTATTAAAATATAATGAAAATGGAACCATTACAGATGCTGTAGACTGTGCACTTG
ACCCTCTCTCAGAAACAAAGTGTACGTTGAAATCCTTCACTGTAGAAAAAGGAATCTATCAAACTTCTAACTTTAGAGTC
CAACCAACAGAATCTATTGTTAGATTTCCTAATATTACAAACTTGTGCCCTTTTGGTGAAGTTTTTAACGCCACCAGATT
TGCATCTGTTTATGCTTGGAACAGGAAGAGAATCAGCAACTGTGTTGCTGATTATTCTGTCCTATATAATTCCGCATCAT
TTTCCACTTTTAAGTGTTATGGAGTGTCTCCTACTAAATTAAATGATCTCTGCTTTACTAATGTCTATGCAGATTCATTT
GTAATTAGAGGTGATGAAGTCAGACAAATCGCTCCAGGGCAAACTGGAAAGATTGCTGATTATAATTATAAATTACCAGA
TGATTTTACAGGCTGCGTTATAGCTTGGAATTCTAACAATCTTGATTCTAAGGTTGGTGGTAATTATAATTACCTGTATA
GATTGTTTAGGAAGTCTAATCTCAAACCTTTTGAGAGAGATATTTCAACTGAAATCTATCAGGCCGGTAGCACACCTTGT
AATGGTGTTGAAGGTTTTAATTGTTACTTTCCTTTACAATCATATGGTTTCCAACCCACTAATGGTGTTGGTTACCAACC
ATACAGAGTAGTAGTACTTTCTTTTGAACTTCTACATGCACCAGCAACTGTTTGTGGACCTAAAAAGTCTACTAATTTGG
TTAAAAACAAATGTGTCAATTTCAACTTCAATGGTTTAACAGGCACAGGTGTTCTTACTGAGTCTAACAAAAAGTTTCTG
CCTTTCCAACAATTTGGCAGAGACATTGCTGACACTACTGATGCTGTCCGTGATCCACAGACACTTGAGATTCTTGACAT
TACACCATGTTCTTTTGGTGGTGTCAGTGTTATAACACCAGGAACAAATACTTCTAACCAGGTTGCTGTTCTTTATCAGG
ATGTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAGATCAACTTACTCCTACTTGGCGTGTTTATTCTACAGGTTCT
AATGTTTTTCAAACACGTGCAGGCTGTTTAATAGGGGCTGAACATGTCAACAACTCATATGAGTGTGACATACCCATTGG
TGCAGGTATATGCGCTAGTTATCAGACTCAGACTAATTCTCCTCGGCGGGCACGTAGTGTAGCTAGTCAATCCATCATTG
CCTACACTATGTCACTTGGTGCAGAAAATTCAGTTGCTTACTCTAATAACTCTATTGCCATACCCACAAATTTTACTATT
AGTGTTACCACAGAAATTCTACCAGTGTCTATGACCAAGACATCAGTAGATTGTACAATGTACATTTGTGGTGATTCAAC
TGAATGCAGCAATCTTTTGTTGCAATATGGCAGTTTTTGTACACAATTAAACCGTGCTTTAACTGGAATAGCTGTTGAAC
AAGACAAAAACACCCAAGAAGTTTTTGCACAAGTCAAACAAATTTACAAAACACCACCAATTAAAGATTTTGGTGGTTTT
AATTTTTCACAAATATTACCAGATCCATCAAAACCAAGCAAGAGGTCATTTATTGAAGATCTACTTTTCAACAAAGTGAC
ACTTGCAGATGCTGGCTTCATCAAACAATATGGTGATTGCCTTGGTGATATTGCTGCTAGAGACCTCATTTGTGCACAAA
AGTTTAACGGCCTTACTGTTTTGCCACCTTTGCTCACAGATGAAATGATTGCTCAATACACTTCTGCACTGTTAGCGGGT
ACAATCACTTCTGGTTGGACCTTTGGTGCAGGTGCTGCATTACAAATACCATTTGCTATGCAAATGGCTTATAGGTTTAA
TGGTATTGGAGTTACACAGAATGTTCTCTATGAGAACCAAAAATTGATTGCCAACCAATTTAATAGTGCTATTGGCAAAA
TTCAAGACTCACTTTCTTCCACAGCAAGTGCACTTGGAAAACTTCAAGATGTGGTCAACCAAAATGCACAAGCTTTAAAC
ACGCTTGTTAAACAACTTAGCTCCAATTTTGGTGCAATTTCAAGTGTTTTAAATGATATCCTTTCACGTCTTGACAAAGT
TGAGGCTGAAGTGCAAATTGATAGGTTGATCACAGGCAGACTTCAAAGTTTGCAGACATATGTGACTCAACAATTAATTA
GAGCTGCAGAAATCAGAGCTTCTGCTAATCTTGCTGCTACTAAAATGTCAGAGTGTGTACTTGGACAATCAAAAAGAGTT
GATTTTTGTGGAAAGGGCTATCATCTTATGTCCTTCCCTCAGTCAGCACCTCATGGTGTAGTCTTCTTGCATGTGACTTA
TGTCCCTGCACAAGAAAAGAACTTCACAACTGCTCCTGCCATTTGTCATGATGGAAAAGCACACTTTCCTCGTGAAGGTG
TCTTTGTTTCAAATGGCACACACTGGTTTGTAACACAAAGGAATTTTTATGAACCACAAATCATTACTACAGACAACACA
TTTGTGTCTGGTAACTGTGATGTTGTAATAGGAATTGTCAACAACACAGTTTATGATCCTTTGCAACCTGAATTAGACTC
ATTCAAGGAGGAGTTAGATAAATATTTTAAGAATCATACATCACCAGATGTTGATTTAGGTGACATCTCTGGCATTAATG
CTTCAGTTGTAAACATTCAAAAAGAAATTGACCGCCTCAATGAGGTTGCCAAGAATTTAAATGAATCTCTCATCGATCTC
CAAGAACTTGGAAAGTATGAGCAGTATATAAAATGGCCATGGTACATTTGGCTAGGTTTTATAGCTGGCTTGATTGCCAT
AGTAATGGTGACAATTATGCTTTGCTGTATGACCAGTTGCTGTAGTTGTCTCAAGGGCTGTTGTTCTTGTGGATCCTGCT
GCAAATTTGATGAAGACGACTCTGAGCCAGTGCTCAAAGGAGTCAAATTACATTACACATAA

SEQ ID NO: 7 is the amino acid sequence of a stabilized SARS-CoV-2 beta variant spike protein with a double proline substitution.

MFVFLVLLPLVSSQCVNFTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFA
NPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVY

SSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRGLPQGFSALEPLVDLPIGINITRFQT
LHISYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPT
ESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIR
GDEVRQIAPGQTGNIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGV
KGFNCYFPLQSYGFQPTYGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQ
QFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGSNVF
QTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGVENSVAYSNNSIAIPTNFTISVT
TEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFS
QILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTIT
SGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALNTLV
KQLSSNFGAISSVLNDILSRLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFC
GKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVS
GNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQEL
GKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

SEQ ID NO: 8 is the amino acid sequence of a stabilized, double proline-substituted, chimeric SARS-CoV-2 spike protein comprising the RBD of the beta variant and remaining sequence from the Wuhan strain.

MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFD
NPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVY
SSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT
LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRV
QPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSF
VIRGDEVRQIAPGQTGNIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
NGVKGFNCYFPLQSYGFQPTYGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL
PFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGS
NVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGF
NFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAG
TITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN
TLVKQLSSNFGAISSVLNDILSRLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRV
DFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT
FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL
QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

SEQ ID NO: 9 is the amino acid sequence of a stabilized SARS-CoV-2 delta variant spike protein with a double proline substitution.

MFVFLVLLPLVSSQCVNLTTTTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFD
NPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLDVYYHKNNKSWMKSEFRVY
SSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPHGFSALEPLVDLPIGINITRFQT
LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRV
QPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSF
VIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYRYRLFRKSNLKPFERDISTEIYQAGSTPC
NGVQGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL
PFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGS
NVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSRRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGF
NFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAG
TITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN
TLVKQLSSNFGAISSVLNDILSRLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRV
DFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTDWFVTQRNFYEPQIITTDNT
FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL
QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

SEQ ID NO: 10 is the amino acid sequence of a stabilized SARS-CoV-2 gamma variant spike protein with a double proline substitution.

MFVFLVLLPLVSSQCVNFTNRTQLPSAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFD
NPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNYPFLGVYYHKNNKSWMESEFRVY
SSANNCTFEYVSQPFLMDLEGKQGNFKNLSEFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT
LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRV
QPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSF
VIRGDEVRQIAPGQTGTIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
NGVKGFNCYFPLQSYGFQPTYGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL
PFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGS
NVFQTRAGCLIGAEYVNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGF
NFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAG
TITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN
TLVKQLSSNFGAISSVLNDILSRLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAAIKMSECVLGQSKRV
DFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT
FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASFVNIQKEIDRLNEVAKNLNESLIDL
QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

SEQ ID NO: 11 is the amino acid sequence of a stabilized SARS-CoV-2 delta plus variant spike protein with a double proline substitution.

MFVFLVLLPLVSSQCVNLRTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFD
NPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLDVYYHKNNKSWMESGVYSS
ANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLL
ALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQP
TESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVI
RGDEVRQIAPGQTGNIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYRYRLFRKSNLKPFERDISTEIYQAGSKPCNG
VEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPF
QQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGSNV
FQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSRRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISV
TTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNF
SQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTI
TSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQNVVNQNAQALNTL
VKQLSSNFGAISSVLNDILSRLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDF
CGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFV
SGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQE
LGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

SEQ ID NO: 12 is the amino acid sequence of a stabilized SARS-CoV-2 omicron variant spike protein with a double proline substitution.

MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHVISGTNGTKRFDNP
VLPFNDGVYFASIEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLDHKNNKSWMESEFRVYSSANN
CTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPIIVEPERDLPQGFSALEPLVDLPIGINITRFQTLLA
LHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPT
ESIVRFPNITNLCPFDEVFNATRFASVYAWNRKRISNCVADYSVLYNLAPFFTFKCYGVSPTKLNDLCFTNVYADSFVIR
GDEVRQIAPGQTGNIADYNYKLPDDFTGCVIAWNSNKLDSKVSGNYNYLYRLFRKSNLKPFERDISTEIYQAGNKPCNGV
AGFNCYFPLRSYSFRPTYGVGHQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLKGTGVLTESNKKFLPFQ
QFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGSNVF
QTRAGCLIGAEYVNNSYECDIPIGAGICASYQTQTKSHRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVT
TEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLKRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKYFGGFNFS
QILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFKGLTVLPPLLTDEMIAQYTSALLAGTIT
SGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNHNAQALNTLV
KQLSSKFGAISSVLNDIFSRLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFC
GKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVS
GNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQEL
GKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

SEQ ID NO: 13 is a codon-optimized nucleic acid sequence encoding a stabilized SARS-CoV-2 beta variant spike protein with a double proline substitution.

ATGTTCGTGTTTCTGGTGCTGCTGCCTCTGGTGAGCTCCCAGTGCGTGAACTTCACCACAAGAACCCAGCTGCCCCCTGC
CTACACCAATTCCTTCACAAGGGGCGTGTACTATCCCGACAAGGTGTTTCGCTCTAGCGTGCTGCACTCCACACAGGATC
TGTTTCTGCCTTTCTTTTCTAACGTGACCTGGTTCCACGCCATCCACGTGAGCGGCACCAATGGCACAAAGCGGTTCGCC
AATCCAGTGCTGCCCTTTAACGACGGCGTGTACTTCGCCTCCACCGAGAAGTCTAACATCATCAGAGGCTGGATCTTTGG
CACCACACTGGATAGCAAGACACAGTCCCTGCTGATCGTGAACAATGCCACCAACGTGGTCATCAAGGTGTGCGAGTTCC
AGTTTTGTAATGACCCATTCCTGGGCGTGTACTATCACAAGAACAATAAGTCTTGGATGGAGAGCGAGTTTAGGGTGTAC
TCCTCTGCCAACAATTGCACATTTGAGTACGTGAGCCAGCCCTTCCTGATGGACCTGGAGGGCAAGCAGGGCAATTTCAA
GAACCTGCGCGAGTTCGTGTTTAAGAATATCGATGGCTACTTCAAGATCTACTCCAAGCACACCCCAATCAACCTGGTGA
GGGGACTGCCACAGGGCTTCTCTGCCCTGGAGCCACTGGTGGACCTGCCCATCGGCATCAACATCACCCGCTTTCAGACA
CTGCACATCAGCTACCTGACACCAGGCGATAGCTCCTCTGGATGGACCGCAGGAGCAGCAGCCTACTATGTGGGCTACCT
GCAGCCCAGGACCTTCCTGCTGAAGTATAACGAGAATGGCACCATCACAGACGCAGTGGATTGCGCCCTGGACCCCCTGT
CTGAGACCAAGTGTACACTGAAGAGCTTTACCGTGGAGAAGGGCATCTACCAGACAAGCAATTTCCGGGTGCAGCCTACC
GAGTCCATCGTGAGATTTCCCAATATCACAAACCTGTGCCCTTTTGGCGAGGTGTTCAACGCCACCCGCTTCGCCAGCGT
GTATGCCTGGAATAGGAAGCGCATCTCCAACTGCGTGGCCGACTATTCTGTGCTGTACAACAGCGCCTCCTTCTCTACCT
TTAAGTGCTACGGCGTGAGCCCCACAAAGCTGAATGACCTGTGCTTTACCAACGTGTATGCCGATTCCTTCGTGATCAGG
GGCGACGAGGTGCGCCAGATCGCACCAGGCCAGACAGGCAATATCGCCGACTACAACTATAAGCTGCCTGACGATTTCAC
CGGCTGCGTGATCGCCTGGAACAGCAACAATCTGGATAGCAAAGTGGGCGGCAACTACAATTATCTGTACCGGCTGTTTA
GAAAGTCTAACCTGAAGCCATTCGAGAGGGACATCTCCACAGAGATCTACCAGGCCGGCTCTACCCCCTGCAATGGCGTG
AAGGGCTTTAACTGTTATTTCCCTCTGCAGAGCTACGGCTTCCAGCCAACCTACGGCGTGGGCTATCAGCCCTACCGCGT
GGTGGTGCTGTCTTTTGAGCTGCTGCACGCACCTGCAACAGTGTGCGGCCCAAAGAAGAGCACCAATCTGGTGAAGAACA
AGTGCGTGAACTTCAACTTCAACGGACTGACCGGCACAGGCGTGCTGACCGAGTCCAACAAGAAGTTCCTGCCTTTTCAG
CAGTTCGGCCGGGACATCGCCGATACCACAGACGCCGTGAGAGACCCTCAGACCCTGGAGATCCTGGATATCACACCATG
CTCCTTCGGCGGCGTGTCTGTGATCACACCAGGCACCAATACAAGCAACCAGGTGGCCGTGCTGTACCAGGGCGTGAATT
GTACCGAGGTGCCCGTGGCAATCCACGCAGACCAGCTGACCCCTACATGGAGGGTGTATTCTACCGGCAGCAACGTGTTC
CAGACACGCGCCGGATGCCTGATCGGAGCAGAGCACGTGAACAATAGCTACGAGTGCGATATCCCTATCGGCGCCGGCAT
CTGTGCCTCCTATCAGACCCAGACAAACTCCCCACGGAGAGCCCGGTCTGTGGCAAGCCAGTCCATCATCGCCTACACCA
TGAGCCTGGGCGTGGAGAACAGCGTGGCCTATTCCAACAATTCTATCGCCATCCCTACCAACTTCACAATCTCCGTGACC
ACAGAGATCCTGCCAGTGAGCATGACCAAGACATCCGTGGACTGCACAATGTACATCTGTGGCGATTCCACCGAGTGCTC
TAACCTGCTGCTGCAGTATGGCTCTTTTTGTACCCAGCTGAATAGAGCCCTGACAGGCATCGCCGTGGAGCAGGACAAGA
ACACACAGGAGGTGTTCGCCCAGGTGAAGCAGATCTACAAGACCCCACCCATCAAGGACTTTGGCGGCTTCAACTTCAGC
CAGATCCTGCCCGATCCTAGCAAGCCATCCAAGCGGTCTTTTATCGAGGACCTGCTGTTCAACAAGGTGACCCTGGCCGA
TGCCGGCTTCATCAAGCAGTACGGCGATTGCCTGGGCGACATCGCAGCCAGAGACCTGATCTGTGCCCAGAAGTTTAATG
GCCTGACCGTGCTGCCTCCACTGCTGACAGATGAGATGATCGCCCAGTATACATCTGCCCTGCTGGCAGGAACCATCACA
AGCGGATGGACCTTCGGCGCAGGAGCCGCCCTGCAGATCCCCTTTGCCATGCAGATGGCCTACAGGTTCAACGGCATCGG
CGTGACCCAGAATGTGCTGTATGAGAACCAGAAGCTGATCGCCAATCAGTTTAACTCCGCCATCGGCAAGATCCAGGACT
CTCTGAGCTCCACAGCAAGCGCCCTGGGCAAGCTGCAGGATGTGGTGAATCAGAACGCCCAGGCCCTGAATACCCTGGTG
AAGCAGCTGTCTAGCAACTTCGGCGCCATCTCCTCTGTGCTGAATGATATCCTGAGCCGGCTGGACCCTCCTGAGGCAGA
GGTGCAGATCGACCGGCTGATCACAGGCAGACTGCAGTCCCTGCAGACCTACGTGACACAGCAGCTGATCAGGGCAGCAG
AGATCAGGGCATCTGCCAATCTGGCCGCCACCAAGATGAGCGAGTGCGTGCTGGGCCAGTCCAAGAGAGTGGACTTTTGT
GGCAAGGGCTACCACCTGATGAGCTTCCCACAGTCCGCCCCTCACGGCGTGGTGTTTCTGCACGTGACCTATGTGCCAGC
CCAGGAGAAGAACTTCACCACAGCACCAGCCATCTGCCACGATGGCAAGGCACACTTTCCTCGGGAGGGCGTGTTCGTGA
GCAACGGCACCCACTGGTTTGTGACACAGAGAAATTTCTACGAGCCACAGATCATCACCACAGACAATACCTTCGTGAGC
GGCAACTGTGACGTGGTCATCGGAATCGTGAACAATACCGTGTACGATCCTCTGCAGCCAGAGCTGGACTCTTTTAAGGA
GGAGCTGGATAAGTATTTCAAGAATCACACCAGCCCCGACGTGGATCTGGGCGACATCTCTGGCATCAATGCCAGCGTGG
TGAACATCCAGAAGGAGATCGACCGCCTGAACGAGGTGGCCAAGAATCTGAACGAGTCCCTGATCGATCTGCAGGAGCTG
GGCAAGTATGAGCAGTACATCAAGTGGCCCTGGTACATCTGGCTGGGCTTCATCGCCGGCCTGATCGCCATCGTGATGGT
GACCATCATGCTGTGCTGTATGACAAGCTGCTGTTCCTGCCTGAAGGGCTGCTGTTCTTGTGGCAGCTGCTGTAAGTTTG
ATGAGGACGATAGCGAGCCTGTGCTGAAGGGCGTGAAGCTGCACTATACCTGA

SEQ ID NO: 14 is a codon-optimized nucleic acid sequence encoding a stabilized, double proline-substituted, chimeric SARS-CoV-2 spike protein comprising the RBD of the beta variant and remaining sequence from the Wuhan strain.

ATGTTCGTGTTTCTGGTGCTGCTGCCTCTGGTGAGCTCCCAGTGCGTGAACCTGACCACAAGGACCCAGCTGCCCCCTGC
CTACACCAATTCCTTCACACGGGGCGTGTACTATCCCGACAAGGTGTTTAGATCTAGCGTGCTGCACTCCACACAGGATC
TGTTTCTGCCTTTCTTTTCTAACGTGACCTGGTTCCACGCCATCCACGTGAGCGGCACCAATGGCACAAAGCGGTTCGAC

AATCCAGTGCTGCCCTTTAACGATGGCGTGTACTTCGCCTCCACCGAGAAGTCTAACATCATCAGAGGCTGGATCTTTGG
CACCACACTGGACAGCAAGACACAGTCCCTGCTGATCGTGAACAATGCCACCAACGTGGTCATCAAGGTGTGCGAGTTCC
AGTTTTGTAATGATCCATTCCTGGGCGTGTACTATCACAAGAACAATAAGTCTTGGATGGAGAGCGAGTTTCGCGTGTAC
TCCTCTGCCAACAATTGCACATTTGAGTACGTGAGCCAGCCCTTCCTGATGGACCTGGAGGGCAAGCAGGGCAATTTCAA
GAACCTGAGGGAGTTCGTGTTTAAGAATATCGATGGCTACTTCAAGATCTACTCCAAGCACACCCCAATCAACCTGGTGC
GCGACCTGCCACAGGGCTTCTCTGCCCTGGAGCCACTGGTGGATCTGCCCATCGGCATCAACATCACCCGGTTTCAGACA
CTGCTGGCCCTGCACAGAAGCTACCTGACACCAGGCGACAGCTCCTCTGGATGGACCGCAGGAGCAGCAGCCTACTATGT
GGGCTACCTGCAGCCCAGGACCTTCCTGCTGAAGTATAACGAGAATGGCACCATCACAGACGCAGTGGATTGCGCCCTGG
ACCCCCTGTCTGAGACCAAGTGTACACTGAAGAGCTTTACCGTGGAGAAGGGCATCTACCAGACAAGCAATTTCAGGGTG
CAGCCTACCGAGTCCATCGTGCGCTTTCCCAATATCACAAACCTGTGCCCTTTTGGCGAGGTGTTCAACGCCACCCGCTT
CGCCAGCGTGTATGCCTGGAATAGGAAGCGCATCTCCAACTGCGTGGCCGACTATTCTGTGCTGTACAACAGCGCCTCCT
TCTCTACCTTTAAGTGCTACGGCGTGAGCCCCACAAAGCTGAATGACCTGTGCTTTACCAACGTGTATGCCGATTCCTTC
GTGATCAGGGGCGACGAGGTGCGCCAGATCGCACCAGGCCAGACAGGCAATATCGCCGACTACAACTATAAGCTGCCTGA
CGATTTCACCGGCTGCGTGATCGCCTGGAACAGCAACAATCTGGATAGCAAAGTGGGCGGCAACTACAATTATCTGTACC
GGCTGTTTAGAAAGTCTAACCTGAAGCCATTCGAGAGGGACATCTCCACAGAGATCTACCAGGCCGGCTCTACCCCCTGC
AATGGCGTGAAGGGCTTTAACTGTTATTTCCCTCTGCAGAGCTACGGCTTCCAGCCAACCTACGGCGTGGGCTATCAGCC
CTACCGCGTGGTGGTGCTGTCTTTTGAGCTGCTGCACGCACCTGCAACAGTGTGCGGCCCAAAGAAGAGCACCAATCTGG
TGAAGAACAAGTGCGTGAACTTCAACTTCAACGGACTGACCGGCACAGGCGTGCTGACCGAGTCCAACAAGAAGTTCCTG
CCTTTTCAGCAGTTCGGCAGGGACATCGCAGATACCACAGACGCCGTGCGCGACCCTCAGACCCTGGAGATCCTGGATAT
CACACCATGCTCCTTCGGCGGCGTGTCTGTGATCACACCAGGCACCAATACAAGCAACCAGGTGGCCGTGCTGTACCAGG
GCGTGAATTGTACCGAGGTGCCCGTGGCAATCCACGCAGACCAGCTGACCCCTACATGGCGGGTGTATTCTACCGGCAGC
AACGTGTTCCAGACAAGAGCCGGATGCCTGATCGGAGCAGAGCACGTGAACAATAGCTACGAGTGCGATATCCCTATCGG
CGCCGGCATCTGTGCCTCCTATCAGACCCAGACAAACTCCCCACGGAGAGCCCGGTCTGTGGCAAGCCAGTCCATCATCG
CCTACACCATGAGCCTGGGCGCCGAGAACAGCGTGGCCTATTCCAACAATTCTATCGCCATCCCTACCAACTTCACAATC
TCCGTGACCACAGAGATCCTGCCAGTGAGCATGACCAAGACATCCGTGGACTGCACAATGTACATCTGTGGCGATTCCAC
CGAGTGCTCTAACCTGCTGCTGCAGTATGGCTCTTTTTGTACCCAGCTGAATAGAGCCCTGACAGGCATCGCCGTGGAGC
AGGACAAGAACACACAGGAGGTGTTCGCCCAGGTGAAGCAGATCTACAAGACCCCACCCATCAAGGACTTTGGCGGCTTC
AACTTCAGCCAGATCCTGCCCGATCCTAGCAAGCCATCCAAGCGGTCTTTTATCGAGGACCTGCTGTTCAACAAGGTGAC
CCTGGCCGATGCCGGCTTCATCAAGCAGTACGGCGATTGCCTGGGCGACATCGCAGCCAGAGACCTGATCTGTGCCCAGA
AGTTTAATGGCCTGACCGTGCTGCCTCCACTGCTGACAGATGAGATGATCGCCCAGTATACATCTGCCCTGCTGGCAGGA
ACCATCACAAGCGGATGGACCTTCGGCGCAGGAGCCGCCCTGCAGATCCCCTTTGCCATGCAGATGGCCTACAGATTCAA
CGGCATCGGCGTGACCCAGAATGTGCTGTATGAGAACCAGAAGCTGATCGCCAATCAGTTTAACTCCGCCATCGGCAAGA
TCCAGGACTCTCTGAGCTCCACAGCAAGCGCCCTGGGCAAGCTGCAGGATGTGGTGAATCAGAACGCCCAGGCCCTGAAT
ACCCTGGTGAAGCAGCTGTCTAGCAACTTCGGCGCCATCTCCTCTGTGCTGAATGATATCCTGAGCCGGCTGGACCCACC
AGAGGCAGAGGTGCAGATCGACCGGCTGATCACAGGCAGACTGCAGTCCCTGCAGACCTACGTGACACAGCAGCTGATCA
GGGCAGCAGAGATCAGGGCATCTGCCAATCTGGCCGCCACCAAGATGAGCGAGTGCGTGCTGGGCCAGTCCAAGAGAGTG
GACTTTTGTGGCAAGGGCTACCACCTGATGAGCTTCCCACAGTCCGCCCCTCACGGCGTGGTGTTTCTGCACGTGACCTA
TGTGCCAGCCCAGGAGAAGAACTTCACCACAGCACCAGCCATCTGCCACGATGGCAAGGCACACTTTCCCCGGGAGGGCG
TGTTCGTGAGCAACGGAACCCACTGGTTTGTGACACAGCGCAATTTCTACGAGCCACAGATCATCACCACAGACAATACA
TTCGTGTCCGGCAACTGTGACGTGGTCATCGGAATCGTGAACAATACCGTGTACGATCCTCTGCAGCCAGAGCTGGACTC
TTTTAAGGAGGAGCTGGATAAGTATTTCAAGAATCACACCAGCCCCGACGTGGATCTGGGCGACATCTCTGGCATCAATG
CCAGCGTGGTGAACATCCAGAAGGAGATCGACAGGCTGAACGAGGTGGCCAAGAATCTGAACGAGTCCCTGATCGATCTG
CAGGAGCTGGGCAAGTATGAGCAGTACATCAAGTGGCCCTGGTACATCTGGCTGGGCTTCATCGCCGGCCTGATCGCCAT
CGTGATGGTGACCATCATGCTGTGCTGTATGACAAGCTGCTGTTCCTGCCTGAAGGGCTGCTGTTCTTGTGGCAGCTGCT
GTAAGTTTGATGAGGACGATAGCGAGCCTGTGCTGAAGGGCGTGAAGCTGCACTATACCTGA

SEQ ID NO: 15 is a codon-optimized nucleic acid sequence encoding a stabilized SARS-CoV-2 delta variant spike protein with a double proline substitution.

ATGTTCGTGTTTCTGGTGCTGCTGCCTCTGGTGAGCTCCCAGTGCGTGAACCTGACCACAACCACACAGCTGCCCCCTGC
CTATACCAATTCCTTCACACGCGGCGTGTACTATCCTGACAAGGTGTTTCGGTCTAGCGTGCTGCACTCCACACAGGATC
TGTTTCTGCCATTCTTTTCTAACGTGACCTGGTTCCACGCCATCCACGTGAGCGGCACCAATGGCACAAAGCGGTTCGAC
AATCCAGTGCTGCCCTTTAACGATGGCGTGTACTTCGCCTCCACCGAGAAGTCTAACATCATCCGGGGCTGGATCTTTGG
CACCACACTGGACAGCAAGACACAGTCCCTGCTGATCGTGAACAATGCCACCAACGTGGTCATCAAGGTGTGCGAGTTCC
AGTTTTGTAATGATCCCTTCCTGGACGTGTACTATCACAAGAACAATAAGTCTTGGATGAAGAGCGAGTTTAGAGTGTAT
TCCTCTGCCAACAATTGCACATTTGAGTACGTGTCCCAGCCTTTCCTGATGGACCTGGAGGGCAAGCAGGGCAATTTCAA
GAACCTGAGAGAGTTCGTGTTTAAGAATATCGATGGCTACTTCAAGATCTACTCCAAGCACACCCCAATCAACCTGGTGA
GGGACCTGCCACACGGCTTCTCTGCCCTGGAGCCACTGGTGGATCTGCCCATCGGCATCAACATCACCAGATTTCAGACA
CTGCTGGCCCTGCACAGGAGCTACCTGACACCCGGCGACAGCTCCTCTGGATGGACCGCCGGCGCTGCCGCCTACTATGT
GGGCTATCTGCAGCCTCGCACCTTCCTGCTGAAGTACAACGAGAATGGCACCATCACAGACGCAGTGGATTGCGCCCTGG
ACCCCCTGTCTGAGACCAAGTGTACACTGAAGAGCTTTACCGTGGAGAAGGGCATCTATCAGACAAGCAATTTCCGCGTG

CAGCCAACCGAGTCCATCGTGCGGTTTCCCAATATCACAAACCTGTGCCCTTTTGGCGAGGTGTTCAACGCAACCAGGTT
CGCAAGCGTGTACGCATGGAATCGCAAGCGGATCTCCAACTGCGTGGCCGACTATTCTGTGCTGTACAACAGCGCCTCCT
TCTCTACCTTTAAGTGCTATGGCGTGAGCCCAACAAAGCTGAATGACCTGTGCTTTACCAACGTGTACGCCGATTCCTTC
GTGATCCGGGGCGACGAGGTGCGGCAGATCGCACCAGGACAGACAGGCAAGATCGCAGACTACAATTATAAGCTGCCTGA
CGATTTCACCGGCTGCGTGATCGCCTGGAACTCTAACAATCTGGATAGCAAAGTGGGCGGCAACTACAATTATAGATACA
GGCTGTTTAGAAAGTCTAATCTGAAGCCATTCGAGAGGGACATCTCCACAGAGATCTACCAGGCCGGCTCTACCCCCTGC
AATGGCGTGCAGGGCTTTAACTGTTATTTCCCTCTGCAGAGCTACGGCTTCCAGCCAACCAACGGCGTGGGCTATCAGCC
CTACCGGGTGGTGGTGCTGTCTTTTGAGCTGCTGCACGCACCTGCAACAGTGTGCGGACCAAAGAAGAGCACCAATCTGG
TGAAGAACAAGTGCGTGAACTTCAACTTCAACGGACTGACCGGAACAGGCGTGCTGACCGAGTCCAACAAGAAGTTCCTG
CCATTTCAGCAGTTCGGCAGAGACATCGCCGATACCACAGACGCCGTGAGGGACCCTCAGACCCTGGAGATCCTGGATAT
CACACCATGCTCCTTCGGCGGCGTGTCTGTGATCACACCCGGCACCAATACAAGCAACCAGGTGGCCGTGCTGTATCAGG
GCGTGAATTGTACCGAGGTGCCAGTGGCAATCCACGCAGACCAGCTGACCCCTACATGGCGCGTGTACTCTACCGGCAGC
AACGTGTTCCAGACAAGGGCAGGATGCCTGATCGGAGCAGAGCACGTGAACAATAGCTATGAGTGCGATATCCCCATCGG
CGCCGGCATCTGTGCCTCCTACCAGACCCAGACAAACTCCCGGAGAAGGGCCAGATCTGTGGCCAGCCAGTCCATCATCG
CCTATACCATGAGCCTGGGCGCCGAGAACAGCGTGGCCTACTCCAACAATTCTATCGCCATCCCTACCAACTTCACAATC
TCCGTGACCACAGAGATCCTGCCAGTGAGCATGACCAAGACATCCGTGGACTGCACAATGTATATCTGTGGCGATTCCAC
CGAGTGCTCTAACCTGCTGCTGCAGTACGGCTCTTTTTGTACCCAGCTGAATAGGGCCCTGACAGGAATCGCAGTGGAGC
AGGACAAGAACACACAGGAGGTGTTCGCCCAGGTGAAGCAGATCTACAAGACCCCACCCATCAAGGACTTTGGCGGCTTC
AACTTCAGCCAGATCCTGCCCGATCCTAGCAAGCCCTCCAAGCGGAGCTTCATCGAGGACCTGCTGTTCAACAAGGTGAC
CCTGGCCGATGCCGGCTTCATCAAGCAGTATGGCGATTGCCTGGGCGACATCGCAGCAAGGGACCTGATCTGTGCCCAGA
AGTTTAATGGCCTGACCGTGCTGCCTCCACTGCTGACAGATGAGATGATCGCCCAGTACACATCTGCCCTGCTGGCAGGA
ACCATCACAAGCGGATGGACCTTCGGCGCAGGAGCCGCCCTGCAGATCCCTTTTGCCATGCAGATGGCCTATCGCTTCAA
CGGCATCGGCGTGACCCAGAATGTGCTGTACGAGAACCAGAAGCTGATCGCCAATCAGTTTAACTCCGCCATCGGCAAGA
TCCAGGACTCTCTGAGCTCCACAGCAAGCGCCCTGGGCAAGCTGCAGGATGTGGTGAATCAGAACGCCCAGGCCCTGAAT
ACCCTGGTGAAGCAGCTGTCTAGCAACTTCGGCGCCATCTCCTCTGTGCTGAATGATATCCTGAGCAGACTGGACCCCCC
CGAGGCCGAGGTGCAGATCGACAGACTGATCACAGGCAGGCTGCAGTCCCTGCAGACCTACGTGACACAGCAGCTGATCA
GGGCCGCCGAGATCAGGGCCTCTGCCAATCTGGCCGCCACCAAGATGAGCGAGTGCGTGCTGGGCCAGTCCAAGAGGGTG
GATTTTTGTGGCAAGGGCTATCACCTGATGAGCTTCCCACAGTCCGCCCCTCACGGAGTGGTGTTTCTGCACGTGACCTA
CGTGCCAGCCCAGGAGAAGAACTTCACCACAGCACCAGCAATCTGCCACGACGGCAAGGCACACTTTCCAAGAGAGGGCG
TGTTCGTGAGCAACGGCACCGATTGGTTTGTGACACAGAGGAATTTCTACGAGCCCCAGATCATCACCACAGACAATACA
TTCGTGTCCGGCAACTGTGACGTGGTCATCGGCATCGTGAACAATACCGTGTATGATCCTCTGCAGCCAGAGCTGGACTC
TTTTAAGGAGGAGCTGGATAAGTACTTCAAGAATCACACCAGCCCCGACGTGGATCTGGGCGACATCTCTGGCATCAATG
CCAGCGTGGTGAACATCCAGAAGGAGATCGACCGGCTGAACGAGGTGGCCAAGAATCTGAACGAGTCCCTGATCGATCTG
CAGGAGCTGGGCAAGTATGAGCAGTACATCAAGTGGCCTTGGTATATCTGGCTGGGCTTCATCGCCGGCCTGATCGCCAT
CGTGATGGTGACCATCATGCTGTGCTGTATGACAAGCTGCTGTTCCTGCCTGAAGGGCTGCTGTTCTTGTGGCAGCTGCT
GTAAGTTTGATGAGGACGATAGCGAGCCAGTGCTGAAGGGCGTGAAGCTGCACTACACCTGA

SEQ ID NO: 16 is a codon-optimized nucleic acid sequence encoding a stabilized SARS-CoV-2 gamma variant spike protein with a double proline substitution.

ATGTTCGTGTTTCTGGTGCTGCTGCCTCTGGTGAGCTCCCAGTGCGTGAATTTCACCAACAGAACACAGCTGCCTTCTGC
CTACACCAATAGCTTCACACGGGGCGTGTACTATCCAGACAAGGTGTTTAGATCTAGCGTGCTGCACAGCACACAGGATC
TGTTTCTGCCATTCTTTTCCAACGTGACCTGGTTCCACGCCATCCACGTGTCCGGCACCAATGGCACAAAGCGGTTCGAC
AATCCCGTGCTGCCTTTTAACGATGGCGTGTACTTCGCCTCCACCGAGAAGTCTAACATCATCAGAGGCTGGATCTTTGG
CACCACACTGGACAGCAAGACACAGTCCCTGCTGATCGTGAACAATGCCACCAACGTGGTCATCAAGGTGTGCGAGTTCC
AGTTTTGTAATTATCCCTTCCTGGGCGTGTACTATCACAAGAACAATAAGTCTTGGATGGAGAGCGAGTTTAGGGTGTAC
TCCTCTGCCAACAATTGCACATTTGAGTATGTGAGCCAGCCTTTCCTGATGGACCTGGAGGGCAAGCAGGGCAATTTCAA
GAACCTGAGCGAGTTCGTGTTTAAGAATATCGATGGCTACTTCAAGATCTACTCCAAGCACACCCCCATCAACCTGGTGC
GCGACCTGCCTCAGGGCTTCTCTGCCCTGGAGCCCCTGGTGGATCTGCCTATCGGCATCAACATCACCCGGTTTCAGACA
CTGCTGGCCCTGCACAGAAGCTACCTGACACCCGGCGACAGCTCCTCTGGATGGACCGCCGGCGCTGCCGCCTACTATGT
GGGCTACCTGCAGCCTAGGACCTTCCTGCTGAAGTATAACGAGAATGGCACCATCACAGACGCAGTGGATTGCGCCCTGG
ACCCCCTGTCCGAGACCAAGTGTACACTGAAGTCTTTTACCGTGGAGAAGGGCATCTACCAGACATCTAATTTCAGGGTG
CAGCCAACCGAGAGCATCGTGCGCTTTCCTAATATCACAAACCTGTGCCCATTTGGCGAGGTGTTCAACGCCACCCGCTT
CGCCAGCGTGTATGCCTGGAATAGGAAGCGCATCAGCAACTGCGTGGCCGACTATTCCGTGCTGTACAACAGCGCCTCCT
TCTCTACCTTTAAGTGTTACGGCGTGTCTCCTACAAAGCTGAATGACCTGTGCTTTACCAACGTGTATGCCGATAGCTTC
GTGATCAGGGGCGACGAGGTGCGCCAGATCGCACCAGGACAGACCGGAACAATCGCAGACTACAATTATAAGCTGCCTGA
CGATTTCACCGGCTGCGTGATCGCCTGGAACTCCAACAATCTGGATTCTAAAGTGGGCGGCAACTACAATTATCTGTACC
GGCTGTTTAGAAAGTCCAACCTGAAGCCATTCGAGCGGGACATCAGCACAGAGATCTACCAGGCAGGCTCCACCCCATGC
AATGGAGTGAAGGGCTTTAACTGTTATTTCCCACTGCAGAGCTACGGCTTCCAGCCCACATATGGCGTGGGCTATCAGCC
TTACAGAGTGGTGGTGCTGTCCTTTGAGCTGCTGCACGCACCAGCAACAGTGTGCGGACCCAAGAAGTCTACCAATCTGG
TGAAGAACAAGTGCGTGAACTTCAACTTCAACGGACTGACCGGAACAGGCGTGCTGACCGAGTCCAACAAGAAGTTCCTG

CCATTTCAGCAGTTCGGCAGGGACATCGCAGATACCACAGACGCCGTGCGCGACCCACAGACCCTGGAGATCCTGGATAT
CACACCCTGCAGCTTCGGCGGCGTGTCCGTGATCACACCAGGAACCAATACAAGCAACCAGGTGGCCGTGCTGTACCAGG
GCGTGAATTGTACCGAGGTGCCTGTGGCAATCCACGCAGACCAGCTGACCCCAACATGGCGGGTGTATTCTACCGGCAGC
AACGTGTTCCAGACAAGAGCCGGCTGCCTGATCGGCGCCGAGTATGTGAACAATTCTTACGAGTGCGATATCCCTATCGG
CGCCGGCATCTGTGCCAGCTACCAGACCCAGACAAACAGCCCACGGAGAGCACGGTCCGTGGCAAGCCAGTCCATCATCG
CCTACACCATGTCTCTGGGCGCCGAGAATAGCGTGGCCTATTCCAACAATTCTATCGCCATCCCAACCAACTTCACAATC
TCCGTGACCACAGAGATCCTGCCCGTGTCTATGACCAAGACAAGCGTGGACTGCACAATGTACATCTGTGGCGATTCCAC
CGAGTGCTCTAACCTGCTGCTGCAGTATGGCAGCTTTTGTACCCAGCTGAATAGAGCCCTGACAGGCATCGCCGTGGAGC
AGGACAAGAACACACAGGAGGTGTTCGCCCAGGTGAAGCAGATCTACAAGACCCCCCCTATCAAGGACTTTGGCGGCTTC
AACTTCAGCCAGATCCTGCCTGATCCAAGCAAGCCATCCAAGAGGTCTTTTATCGAGGACCTGCTGTTCAACAAGGTGAC
CCTGGCCGATGCCGGCTTCATCAAGCAGTACGGCGATTGCCTGGGCGACATCGCAGCAAGGGACCTGATCTGTGCCCAGA
AGTTTAATGGCCTGACCGTGCTGCCACCCCTGCTGACAGATGAGATGATCGCCCAGTATACATCCGCCCTGCTGGCCGGC
ACCATCACATCTGGATGGACCTTCGGCGCAGGAGCCGCCCTGCAGATCCCCTTTGCCATGCAGATGGCCTACAGGTTCAA
CGGCATCGGCGTGACCCAGAATGTGCTGTATGAGAACCAGAAGCTGATCGCCAATCAGTTTAACTCCGCCATCGGCAAGA
TCCAGGACTCCCTGAGCTCCACAGCCTCTGCCCTGGGCAAGCTGCAGGATGTGGTGAATCAGAACGCCCAGGCCCTGAAT
ACCCTGGTGAAGCAGCTGTCTAGCAACTTCGGCGCCATCTCCTCTGTGCTGAATGATATCCTGAGCCGGCTGGACCCCCC
CGAGGCAGAGGTGCAGATCGACCGGCTGATCACCGGCAGACTGCAGAGCCTGCAGACCTACGTGACACAGCAGCTGATCA
GGGCCGCCGAGATCAGGGCATCCGCCAATCTGGCCGCCATCAAGATGTCTGAGTGCGTGCTGGGCCAGAGCAAGAGAGTG
GACTTTTGTGGCAAGGGCTACCACCTGATGAGCTTCCCTCAGTCCGCCCCACACGGAGTGGTGTTTCTGCACGTGACCTA
TGTGCCCGCCCAGGAGAAGAACTTCACCACAGCCCCTGCCATCTGCCACGATGGCAAGGCCCACTTTCCAAGGGAGGGCG
TGTTCGTGTCCAACGGCACCCACTGGTTTGTGACACAGCGCAATTTCTACGAGCCCCAGATCATCACCACAGACAATACC
TTCGTGAGCGGCAACTGTGACGTGGTCATCGGCATCGTGAACAATACCGTGTACGATCCACTGCAGCCCGAGCTGGACTC
CTTTAAGGAGGAGCTGGATAAGTATTTCAAGAATCACACCTCTCCCGACGTGGATCTGGGCGACATCTCCGGCATCAATG
CCTCTTTCGTGAACATCCAGAAGGAGATCGACCGCCTGAACGAGGTGGCCAAGAATCTGAACGAGTCCCTGATCGATCTG
CAGGAGCTGGGCAAGTATGAGCAGTACATCAAGTGGCCCTGGTACATCTGGCTGGGCTTCATCGCCGGCCTGATCGCCAT
CGTGATGGTGACCATCATGCTGTGCTGTATGACAAGCTGCTGTTCCTGCCTGAAGGGCTGCTGTTCTTGTGGCAGCTGCT
GTAAGTTTGATGAGGACGATAGCGAGCCTGTGCTGAAGGGCGTGAAGCTGCACTATACCTGA

SEQ ID NO: 17 is a codon-optimized nucleic acid sequence encoding a stabilized SARS-CoV-2 delta plus variant spike protein with a double proline substitution.

ATGTTTGTGTTTCTGGTGCTGCTGCCACTGGTGAGTAGCCAGTGTGTGAACCTGAGAACCCGAACACAGCTGCCTCCTGC
CTATACCAACAGCTTCACCAGAGGCGTGTACTACCCTGACAAGGTGTTCCGATCTAGCGTGCTCCATAGCACCCAGGACC
TGTTCTTGCCTTTTTTCTCTAACGTGACATGGTTCCACGCCATTCACGTGTCTGGCACCAACGGAACAAAAAGATTCGAC
AACCCTGTGCTGCCCTTCAACGACGGTGTCTATTTTGCCAGCACCGAGAAGAGCAACATCATCAGAGGCTGGATCTTCGG
AACCACCCTGGACAGCAAGACCCAGAGCCTGCTGATCGTCAATAACGCAACAAATGTGGTGATCAAGGTGTGCGAGTTCC
AATTTTGCAACGATCCTTTCCTGGATGTGTACTACCACAAGAACAACAAAAGCTGGATGGAAAGTGGAGTTTATAGCAGC
GCCAACAACTGCACCTTCGAGTACGTGAGCCAACCTTTCCTGATGGACCTCGAAGGGAAACAGGGCAACTTCAAGAACCT
TAGAGAGTTCGTCTTTAAGAACATCGACGGCTACTTTAAAATCTACTCCAAGCACACCCCCATCAACCTGGTGCGGGACC
TGCCTCAGGGCTTTAGCGCGCTGGAACCCTTGGTTGACCTGCCCATCGGCATCAACATCACTAGATTCCAGACCCTTCTG
GCCCTCCACCGGTCTTACCTGACACCTGGCGACAGTAGTTCTGGCTGGACAGCCGGCGCCGCTGCCTACTACGTGGGCTA
TCTGCAGCCTAGAACCTTCCTGCTGAAGTACAACGAGAACGGCACCATCACCGACGCTGTGGATTGCGCCCTGGACCCTC
TGTCCGAAACCAAGTGCACACTGAAGTCCTTCACCGTGGAAAAGGGCATCTACCAGACCTCTAACTTCCGGGTGCAGCCT
ACTGAAAGCATCGTGCGGTTCCCAAACATTACAAACCTGTGCCCTTTCGGAGAAGTTTTCAACGCCACTCGCTTCGCCTC
TGTCTATGCCTGGAACAGAAAGCGGATCAGCAATTGTGTGGCCGATTACAGCGTGCTGTACAACAGCGCCAGCTTTTCTA
CATTCAAGTGCTACGGCGTGTCTCCCACCAAGCTGAATGATCTGTGCTTCACCAACGTGTACGCCGACTCGTTTGTGATC
CGGGGAGACGAAGTGCGCCAGATCGCCCCTGGGCAGACAGGAAACATCGCCGATTACAATTACAAACTGCCTGACGATTT
TACAGGATGTGTGATAGCTTGGAACTCCAACAACCTCGACAGCAAAGTGGGCGGCAACTACAATTACCGGTACAGACTGT
TTAGAAAGAGCAACCTAAAACCCTTCGAGAGAGATATCTCTACCGAGATCTACCAGGCCGGCAGCAAGCCTTGTAATGGC
GTTGAGGGCTTCAACTGTTACTTCCCTCTGCAGAGCTACGGCTTCCAGCCCACCAACGGCGTCGGGTACCAGCCTTACAG
AGTTGTGGTTCTGAGCTTCGAGCTGCTCCACGCTCCTGCCACCGTGTGTGGTCCTAAGAAAAGCACCAACCTGGTGAAGA
ACAAGTGCGTGAATTTCAATTTCAACGGCCTGACAGGCACAGGCGTGCTGACCGAGAGCAACAAAAAGTTCCTGCCCTTC
CAGCAGTTCGGCAGAGATATTGCCGATACCACAGACGCCGTGCGGGACCCTCAAACCCTGGAAATCTTGGACATCACACC
TTGCAGCTTCGGCGGAGTGTCTGTGATCACTCCCGGGACCAACACCAGCAACCAGGTTGCCGTGCTGTACCAGGGCGTCA
ACTGCACCGAAGTGCCAGTGGCTATACACGCCGACCAGCTGACCCCTACATGGCGGGTGTACAGCACCGGCAGCAACGTG
TTCCAGACCAGAGCCGGCTGCCTGATCGGCGCAGAGCACGTGAACAACTCTTATGAATGCGACATCCCCATCGGAGCCGG
CATTTGCGCCAGCTACCAGACACAGACCAATAGCAGAAGACGGGCTAGAAGCGTGGCCTCGCAGAGCATAATCGCATACA
CAATGAGCCTGGGAGCCGAGAACAGCGTGGCCTACAGCAACAATAGTATCGCCATCCCCACAAATTTTACCATCAGCGTG
ACAACCGAAATCCTGCCAGTGAGCATGACAAAGACCAGCGTCGACTGCACAATGTACATATGTGGCGATAGCACGGAGTG
CAGCAATCTGCTGCTCCAATACGGCAGCTTCTGCACCCAGCTGAATCGGGCACTGACCGGCATCGCCGTGGAACAGGATA
AAAATACCCAGGAGGTGTTTGCCCAGGTGAAGCAGATATATAAGACCCCTCCGATCAAGGACTTCGGAGGCTTCAATTTC

AGCCAGATCCTGCCCGATCCAAGCAAGCCTAGCAAGCGGTCCTTCATCGAGGATCTGCTGTTCAATAAGGTGACCCTGGC
CGACGCCGGATTCATCAAACAGTACGGCGACTGCCTGGGCGACATCGCCGCCAGAGATCTGATCTGTGCTCAAAAGTTCA
ACGGACTGACAGTCCTGCCACCTCTGTTGACAGATGAAATGATCGCTCAGTACACCTCCGCCCTCCTGGCCGGGACGATC
ACCTCTGGATGGACCTTCGGCGCCGGCGCTGCACTGCAGATCCCTTTCGCCATGCAGATGGCCTACAGATTCAACGGCAT
CGGAGTGACCCAAAACGTCCTGTACGAGAACCAGAAGCTGATCGCCAACCAGTTCAACTCTGCTATCGGCAAGATCCAGG
ACAGCCTCAGCAGCACCGCCAGCGCCCTGGGCAAACTCCAGAACGTGGTGAACCAGAACGCACAGGCCCTGAATACCCTG
GTGAAGCAGCTGAGCAGCAACTTCGGCGCTATCAGCTCTGTGCTGAACGACATCCTGAGCAGACTGGACCCTCCCGAGGC
CGAGGTGCAGATTGACAGGCTGATCACAGGCAGACTGCAGTCGCTGCAAACTTACGTGACCCAGCAACTGATCCGGGCCG
CCGAAATCAGGGCCAGCGCCAACCTGGCTGCTACAAAGATGTCCGAATGCGTGTTGGGCCAGTCCAAGAGAGTGGACTTC
TGCGGCAAGGGATACCACCTGATGAGCTTCCCTCAGTCCGCTCCCCACGGCGTCGTGTTCCTGCATGTGACATACGTGCC
CGCCCAGGAGAAGAATTTCACCACCGCCCCTGCCATCTGCCACGACGGCAAGGCCCACTTCCCCAGAGAGGGCGTGTTCG
TGTCCAACGGCACCCACTGGTTCGTGACCCAGCGGAACTTCTACGAGCCTCAGATCATCACCACCGATAACACATTCGTG
TCCGGCAACTGCGACGTGGTTATCGGCATCGTGAACAATACCGTGTACGACCCTCTGCAGCCAGAACTGGATTCTTTTAA
GGAAGAGCTGGACAAATACTTTAAGAACCACACATCTCCTGATGTGGACCTGGGCGACATCAGCGGCATCAACGCCTCCG
TGGTCAACATCCAAAAGGAGATCGATAGACTGAACGAGGTGGCCAAGAACCTCAACGAGTCTCTGATTGACCTGCAGGAG
CTGGGCAAGTACGAGCAGTACATCAAGTGGCCTTGGTACATCTGGCTGGGCTTCATCGCCGGCCTGATCGCTATCGTCAT
GGTGACCATCATGCTGTGCTGTATGACCTCCTGCTGCAGCTGTCTGAAAGGCTGCTGTTCTTGCGGCAGCTGTTGCAAGT
TTGACGAGGACGACTCCGAGCCCGTGCTGAAGGGGGTGAAGCTGCACTACACGTGA

SEQ ID NO: 18 is a codon-optimized nucleic acid sequence encoding a stabilized SARS-CoV-2 omicron variant spike protein with a double proline substitution.

ATGTTCGTGTTCCTGGTGCTGCTGCCCCTGGTGTCTAGCCAATGTGTGAACCTGACAACAAGGACCCAGCTTCCCCCAGC
TTACACCAATTCATTTACAAGAGGCGTGTATTACCCCGATAAGGTGTTCCGAAGCAGCGTGCTGCACAGCACCCAGGATC
TCTTCCTGCCTTTTTTCAGCAATGTGACTTGGTTCCACGTGATCAGCGGAACCAACGGCACCAAGCGGTTTGACAATCCT
GTGCTGCCCTTCAACGACGGCGTGTACTTCGCCAGCATCGAGAAGAGCAACATTATCCGGGGCTGGATCTTCGGCACCAC
CCTCGATAGCAAGACCCAGAGCTTACTGATCGTAAACAACGCCACCAATGTCGTAATCAAGGTCTGTGAATTTCAGTTCT
GCAACGACCCCTTTCTGGACCACAAGAACAACAAGTCGTGGATGGAAAGCGAGTTCAGAGTGTACAGCTCCGCTAACAAT
TGTACATTCGAGTACGTGTCTCAGCCTTTCCTGATGGACCTGGAAGGCAAGCAGGGAAACTTCAAGAATCTGAGGGAGTT
CGTGTTCAAAAACATCGACGGCTACTTCAAGATCTACAGCAAGCATACCCCCATCATCGTTGAACCTGAGAGAGACCTGC
CACAGGGTTTCAGCGCTCTGGAGCCTCTGGTTGACCTGCCCATCGGCATCAACATCACCCGGTTTCAGACACTGTTAGCC
CTGCATAGATCTTACCTGACCCCAGGCGATTCTTCCTCTGGCTGGACCGCCGGAGCCGCAGCCTACTACGTGGGATATCT
GCAGCCCAGAACCTTCCTGCTGAAATACAACGAGAACGGAACCATCACCGATGCCGTGGACTGCGCCCTGGACCCTCTGT
CTGAAACCAAGTGCACCCTGAAGAGCTTCACCGTGGAAAAGGGCATCTACCAGACCAGCAACTTTCGGGTGCAGCCCACC
GAGAGCATCGTGAGATTTCCAAACATCACCAACCTGTGTCCTTTCGACGAGGTGTTTAATGCCACAAGATTCGCCAGCGT
GTACGCCTGGAATAGAAAAAGAATCTCCAACTGCGTGGCTGATTACTCAGTGCTTTACAACCTGGCCCCATTCTTCACCT
TCAAGTGCTACGGCGTTAGCCCTACCAAGCTCAATGATCTGTGCTTCACGAACGTGTACGCCGACAGCTTCGTGATCCGG
GGCGACGAAGTCAGACAGATCGCCCCTGGACAGACCGGTAATATCGCCGACTACAATTACAAGCTGCCTGATGATTTCAC
AGGTTGCGTGATCGCCTGGAACTCCAACAAGCTGGACAGCAAGGTGTCCGGCAACTACAACTACCTGTATAGACTTTTCA
GAAAGTCCAACCTGAAGCCATTCGAGCGGGACATCAGCACTGAGATCTACCAGGCCGGCAACAAACCCTGCAACGGAGTT
GCCGGATTCAACTGCTATTTCCCTCTGAGATCTTACTCCTTCAGACCTACATACGGCGTGGGACACCAGCCTTACAGAGT
AGTGGTGCTCAGCTTCGAGCTTCTGCACGCTCCTGCCACCGTGTGCGGCCCTAAGAAGAGCACGAACCTGGTGAAGAACA
AATGTGTTAATTTTAACTTCAACGGCCTGAAGGGCACAGGAGTCCTGACCGAGAGCAATAAAAAATTCTTGCCCTTCCAG
CAGTTCGGAAGAGACATCGCCGACACCACAGATGCTGTGAGAGACCCTCAGACCCTGGAAATCCTCGACATCACCCCTTG
CAGCTTCGGCGGCGTCAGCGTGATCACCCCGGGCACCAACACCTCTAACCAGGTGGCCGTGCTGTACCAGGGCGTGAATT
GCACCGAGGTTCCTGTGGCCATCCACGCGGACCAGCTGACACCAACATGGCGGGTGTACAGCACCGGCTCCAACGTGTTT
CAGACCAGAGCCGGCTGTCTGATCGGCGCCGAATATGTGAACAACAGCTACGAATGCGACATCCCAATCGGCGCCGGCAT
TTGCGCCAGCTACCAGACACAGACCAAAAGTCACCGGAGAGCTCGGAGCGTGGCCTCTCAGAGCATTATCGCCTATACCA
TGAGCCTGGGGGCCGAGAACAGCGTGGCCTATTCCAACAACAGCATCGCCATCCCTACCAATTTCACCATCTCTGTGACC
ACCGAGATCCTGCCAGTGTCCATGACAAAGACAAGCGTGGACTGCACCATGTACATCTGCGGCGACTCTACCGAGTGCAG
CAACCTGCTGCTGCAGTACGGCAGCTTTTGCACACAGCTGAAACGGGCGCTGACAGGAATTGCCGTTGAGCAGGACAAGA
ACACTCAGGAGGTGTTTGCCCAAGTGAAGCAGATATATAAGACCCCTCCTATCAAATACTTCGGCGGCTTTAACTTCAGC
CAGATCCTCCCTGATCCTTCTAAGCCTAGCAAGCGCAGCTTCATCGAGGACCTGCTGTTCAACAAGGTAACCCTGGCTGA
CGCCGGCTTCATCAAGCAGTACGGTGATTGCCTGGGCGACATCGCAGCCCGGGACCTGATCTGTGCCCAAAAATTCAAGG
GCCTGACTGTTCTGCCTCCTCTGCTGACAGATGAAATGATCGCCCAGTACACCTCCGCCCTGCTGGCTGGCACAATCACC
AGCGGCTGGACATTCGGCGCCGGCGCCGCGCTGCAGATCCCTTTCGCCATGCAGATGGCCTACAGATTCAACGGCATCGG
AGTGACTCAGAACGTGCTGTACGAAAACCAGAAACTGATTGCAAATCAGTTTAACAGCGCAATCGGCAAGATCCAGGATA
GCCTGTCCAGCACCGCCTCCGCTCTGGGCAAGCTGCAAGACGTGGTGAACCACAATGCCCAGGCTCTGAACACCTTGGTG
AAGCAGCTGAGCAGCAAGTTCGGCGCCATTTCTTCCGTGCTGAACGACATCTTCAGCAGACTCGATCCTCCCGAGGCCGA
GGTGCAGATCGACAGACTGATCACGGGCAGACTGCAGTCTCTGCAGACATACGTGACACAGCAACTGATCAGAGCCGCTG
AAATCAGGGCCTCTGCCAACCTGGCCGCCACCAAGATGTCTGAGTGCGTGCTCGGCCAGTCTAAAAGAGTGGACTTCTGC

GGCAAAGGCTACCACCTGATGAGCTTCCCCCAGAGCGCCCCCCACGGCGTGGTGTTCCTACACGTTACCTACGTGCCGGC
TCAAGAAAAGAACTTTACCACCGCCCCTGCCATCTGCCACGACGGAAAGGCCCACTTCCCTCGGGAGGGTGTGTTTGTCA
GCAACGGCACACACTGGTTCGTGACACAGCGGAACTTCTACGAGCCCCAAATCATCACAACAGATAACACCTTCGTCAGC
GGCAACTGTGACGTGGTGATCGGCATCGTGAACAACACCGTGTATGACCCTCTGCAGCCTGAGCTGGACAGCTTTAAGGA
AGAGCTGGACAAGTACTTCAAGAATCACACAAGTCCTGACGTGGATCTGGGCGATATCAGTGGCATCAACGCCTCTGTGG
TGAACATACAAAAGGAGATCGACAGACTGAACGAGGTGGCAAAGAACCTGAATGAAAGCCTGATCGACCTGCAAGAACTG
GGCAAGTACGAGCAGTACATCAAGTGGCCTTGGTACATTTGGCTGGGATTTATCGCAGGCCTCATCGCCATCGTGATGGT
GACAATCATGCTGTGTTGCATGACCAGCTGTTGCAGCTGCCTGAAAGGCTGTTGTAGCTGCGGCAGCTGCTGCAAGTTCG
ATGAGGACGACAGCGAGCCTGTCCTGAAGGGGGTGAAGCTGCACTACACATGA

SEQ ID NO: 19 is a codon-optimized nucleic acid sequence encoding a stabilized SARS-CoV-2 Wuhan strain spike protein with a double proline substitution.

ATGTTCGTCTTCCTGGTCCTGCTGCCCCTGGTCTCATCTCAGTGCGTGAATCTGACTACAAGAACTCAGCTGCCTCCCGC
CTACACCAATTCCTTCACCCGGGGCGTGTACTATCCTGACAAGGTGTTTAGAAGCTCCGTGCTGCACTCTACACAGGATC
TGTTTCTGCCATTCTTTAGCAACGTGACCTGGTTCCACGCCATCCACGTGAGCGGCACCAATGGCACAAAGCGGTTCGAC
AATCCCGTGCTGCCTTTTAACGATGGCGTGTACTTCGCCTCTACCGAGAAGAGCAACATCATCAGAGGCTGGATCTTTGG
CACCACACTGGACTCCAAGACACAGTCTCTGCTGATCGTGAACAATGCCACCAACGTGGTCATCAAGGTGTGCGAGTTCC
AGTTTTGTAATGATCCCTTCCTGGGCGTGTACTATCACAAGAACAATAAGAGCTGGATGGAGTCCGAGTTTAGAGTGTAT
TCTAGCGCCAACAATTGCACATTTGAGTACGTGTCCCAGCCTTTCCTGATGGACCTGGAGGGCAAGCAGGGCAATTTCAA
GAACCTGAGGGAGTTCGTGTTTAAGAATATCGATGGCTACTTCAAGATCTACTCTAAGCACACCCCCATCAACCTGGTGC
GCGACCTGCCTCAGGGCTTCAGCGCCCTGGAGCCACTGGTGGATCTGCCTATCGGCATCAACATCACCCGGTTTCAGACA
CTGCTGGCCCTGCACAGAAGCTACCTGACACCCGGCGACTCCTCTAGCGGATGGACCGCAGGAGCAGCAGCCTACTATGT
GGGCTATCTGCAGCCTAGGACCTTCCTGCTGAAGTACAACGAGAATGGCACCATCACAGACGCAGTGGATTGCGCCCTGG
ACCCCCTGAGCGAGACAAAGTGTACACTGAAGTCCTTTACCGTGGAGAAGGGCATCTATCAGACATCCAATTTCAGGGTG
CAGCCAACCGAGTCTATCGTGCGCTTTCCTAATATCACAAACCTGTGCCCATTTGGCGAGGTGTTCAACGCAACCAGGTT
CGCAAGCGTGTACGCATGGAATAGGAAGCGCATCTCTAACTGCGTGGCCGACTATAGCGTGCTGTACAACTCCGCCTCTT
TCAGCACCTTTAAGTGCTATGGCGTGTCCCCCACAAAGCTGAATGACCTGTGCTTTACCAACGTGTACGCCGATTCTTTC
GTGATCAGGGGCGACGAGGTGCGCCAGATCGCACCTGGACAGACAGGCAAGATCGCCGACTACAATTATAAGCTGCCAGA
CGATTTCACCGGCTGCGTGATCGCCTGGAACAGCAACAATCTGGATTCCAAAGTGGGCGGCAACTACAATTATCTGTACC
GGCTGTTTAGAAAGAGCAATCTGAAGCCCTTCGAGAGGGACATCTCTACAGAGATCTACCAGGCCGGCAGCACCCCTTGC
AATGGCGTGGAGGGCTTTAACTGTTATTTCCCACTGCAGTCCTACGGCTTCCAGCCCACAAACGGCGTGGGCTATCAGCC
TTACCGCGTGGTGGTGCTGAGCTTTGAGCTGCTGCACGCACCAGCAACAGTGTGCGGACCCAAGAAGTCCACCAATCTGG
TGAAGAACAAGTGCGTGAACTTCAACTTCAACGGCCTGACCGGAACAGGCGTGCTGACCGAGTCCAACAAGAAGTTCCTG
CCATTTCAGCAGTTCGGCAGGGACATCGCAGATACCACAGACGCCGTGCGCGACCCACAGACCCTGGAGATCCTGGATAT
CACACCCTGCTCTTTCGGCGGCGTGAGCGTGATCACACCAGGAACCAATACAAGCAACCAGGTGGCCGTGCTGTATCAGG
ACGTGAATTGTACCGAGGTGCCTGTGGCCATCCACGCCGATCAGCTGACCCCAACATGGCGGGTGTACAGCACCGGCTCC
AACGTGTTCCAGACAAGAGCAGGATGCCTGATCGGAGCAGAGCACGTGAACAATTCCTATGAGTGCGACATCCCAATCGG
CGCCGGCATCTGTGCCTCTTACCAGACCCAGACAAACTCTCCAAGGAGAGCACGGAGCGTGGCATCCCAGTCTATCATCG
CCTATACCATGTCCCTGGGCGCCGAGAATTCTGTGGCCTACTCTAACAATAGCATCGCCATCCCAACCAACTTCACAATC
TCTGTGACCACAGAGATCCTGCCCGTGTCCATGACCAAGACATCTGTGGACTGCACAATGTATATCTGTGGCGATTCTAC
CGAGTGCAGCAACCTGCTGCTGCAGTACGGCAGCTTTTGTACCCAGCTGAATAGAGCCCTGACAGGCATCGCCGTGGAGC
AGGATAAGAACACACAGGAGGTGTTCGCCCAGGTGAAGCAGATCTACAAGACCCCCCCTATCAAGGACTTTGGCGGCTTC
AATTTTTCCCAGATCCTGCCTGATCCATCCAAGCCTTCTAAGCGGAGCTTTATCGAGGACCTGCTGTTCAACAAGGTGAC
CCTGGCCGATGCCGGCTTCATCAAGCAGTATGGCGATTGCCTGGGCGACATCGCAGCACGGGACCTGATCTGTGCCCAGA
AGTTTAATGGCCTGACCGTGCTGCCACCCCTGCTGACAGATGAGATGATCGCACAGTACACAAGCGCCCTGCTGGCAGGA
ACCATCACATCCGGATGGACCTTCGGCGCAGGAGCCGCCCTGCAGATCCCCTTTGCCATGCAGATGGCCTATAGGTTCAA
CGGCATCGGCGTGACCCAGAATGTGCTGTACGAGAACCAGAAGCTGATCGCCAATCAGTTTAACTCCGCCATCGGCAAGA
TCCAGGACAGCCTGTCCTCTACAGCCTCCGCCCTGGGCAAGCTGCAGGATGTGGTGAATCAGAACGCCCAGGCCCTGAAT
ACCCTGGTGAAGCAGCTGAGCTCCAACTTCGGCGCCATCTCTAGCGTGCTGAATGATATCCTGAGCCGGCTGGACCCCCC
CGAGGCAGAGGTGCAGATCGACCGGCTGATCACAGGCAGACTGCAGTCTCTGCAGACCTATGTGACACAGCAGCTGATCA
GGGCAGCAGAGATCAGGGCAAGCGCCAATCTGGCAGCAACCAAGATGTCCGAGTGCGTGCTGGGCCAGTCTAAGAGAGTG
GACTTTTGTGGCAAGGGCTATCACCTGATGTCCTTCCCTCAGTCTGCCCCACACGGCGTGGTGTTTCTGCACGTGACCTA
CGTGCCCGCCCAGGAGAAGAACTTCACCACAGCCCCTGCCATCTGCCACGATGGCAAGGCCCACTTTCCAAGGGAGGGCG
TGTTCGTGTCCAACGGCACCCACTGGTTTGTGACACAGCGCAATTTCTACGAGCCCCAGATCATCACCACAGACAATACC
TTCGTGAGCGGCAACTGTGACGTGGTCATCGGCATCGTGAACAATACCGTGTATGATCCACTGCAGCCCGAGCTGGACAG
CTTTAAGGAGGAGCTGGATAAGTACTTCAAGAATCACACCTCCCCTGACGTGGATCTGGGCGACATCAGCGGCATCAATG
CCTCCGTGGTGAACATCCAGAAGGAGATCGACCGCCTGAACGAGGTGGCCAAGAATCTGAACGAGAGCCTGATCGATCTG
CAGGAGCTGGGCAAGTATGAGCAGTACATCAAGTGGCCATGGTACATCTGGCTGGGCTTCATCGCCGGCCTGATCGCCAT
CGTGATGGTGACCATCATGCTGTGCTGTATGACATCCTGCTGTTCTTGCCTGAAGGGCTGCTGTAGCTGTGGCTCCTGCT
GTAAGTTTGATGAGGACGATTCCGAACCCGTGCTGAAGGGAGTGAAGCTGCATTACACCTGA

DETAILED DESCRIPTION

I. Abbreviations

Ad	adenovirus
CoV	coronavirus
COVID-19	coronavirus disease 2019
Env	envelope
GI	gastrointestinal
HIV	human immunodeficiency virus
IFU	infection forming units
IM	intramuscular
IN	intranasal
OPV	oral poliovirus
PP	double protein substitution
S	spike protein
SARS	severe acute respiratory syndrome
TT	tail truncated
URT	upper respiratory tract
VOC	variant of concern
Wu	Wuhan strain

II. Terms

Unless otherwise noted, technical terms are used according to conventional usage. Definitions of common terms in molecular biology may be found in Benjamin Lewin, *Genes X*, published by Jones & Bartlett Publishers, 2009; and Meyers *et al.* (eds.), *The Encyclopedia of Cell Biology and Molecular Medicine*, published by Wiley-VCH in 16 volumes, 2008; and other similar references.

As used herein, the singular forms "a," "an," and "the," refer to both the singular as well as plural, unless the context clearly indicates otherwise. For example, the term "an antigen" includes single or plural antigens and can be considered equivalent to the phrase "at least one antigen." As used herein, the term "comprises" means "includes." It is further to be understood that any and all base sizes or amino acid sizes, and all molecular weight or molecular mass values, given for nucleic acids or polypeptides are approximate, and are provided for descriptive purposes, unless otherwise indicated. Although many methods and materials similar or equivalent to those described herein can be used, particular suitable methods and materials are described herein. In case of conflict, the present specification, including explanations of terms, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting. To facilitate review of the various embodiments, the following explanations of terms are provided:

Adenovirus: A non-enveloped virus with a liner, double-stranded DNA genome and an icosahedral capsid. There are at least 68 known serotypes of human adenovirus, which are divided into seven species (species A, B, C, D, E, F and G). Different serotypes of adenovirus are associated with different types of disease, with some serotypes causing respiratory disease (primarily species B and C), conjunctivitis (species B and D) and/or gastroenteritis (species F and G). Adenovirus type 4 (Ad4) is a species E virus that can cause acute respiratory disease and ocular disease. Adenovirus-based vectors are commonly used for a variety of therapeutic applications, including vaccine and gene therapy vectors. In some embodiments herein, the adenovirus vector is a human replication-competent Ad4 with a complete or partial deletion in the E3 region.

Adjuvant: A component of an immunogenic composition used to enhance antigenicity. In some embodiments, an adjuvant can include a suspension of minerals (alum, aluminum hydroxide, or phosphate) on which antigen is adsorbed; or water-in-oil emulsion, for example, in which antigen solution is emulsified in mineral oil (Freund incomplete adjuvant), sometimes with the inclusion of killed mycobacteria (Freund's complete adjuvant) to further enhance antigenicity (inhibits degradation of antigen and/or causes influx of macrophages). In some embodiments, the adjuvant used in a disclosed immunogenic composition is a combination of lecithin and carbomer homopolymer (such as the ADJUPLEX™ adjuvant available from Advanced BioAdjuvants, LLC; see also Wegmann, *Clin Vaccine Immunol* 22(9): 1004-1012, 2015). Additional adjuvants for use in the disclosed immunogenic compositions include the QS21 purified plant extract, Matrix M, AS01, MF59, and ALFQ adjuvants. Immunostimulatory oligonucleotides (such as those including a CpG motif) can also be used as adjuvants. Adjuvants include biological molecules (a "biological adjuvant"), such as costimulatory molecules. Exemplary adjuvants include IL-2, RANTES, GM-CSF, TNF-α, IFN-γ, G-CSF, LFA-3, CD72, B7-1, B7-2, OX-40L, 4-1BBL and toll-like receptor (TLR) agonists, such as TLR-9 agonists. The person of ordinary skill in the art is familiar with adjuvants (see, *e.g.*, Singh (ed.) Vaccine Adjuvants and Delivery Systems. Wiley-Interscience, 2007).

Administration: The introduction of a composition into a subject by a chosen route. Administration can be local or systemic. For example, if the chosen route is intravenous, the composition is administered by introducing the composition into a vein of the subject. Exemplary routes of administration include, but are not limited to, intranasal, inhalation, oral, injection (such as subcutaneous, intramuscular, intradermal, intraperitoneal, and intravenous), sublingual, rectal, transdermal (for example, topical) and vaginal routes.

Codon-optimized: A nucleic acid sequence that has been altered such that the codons are optimal for expression in a particular system (such as a particular species or group of species). For example, a nucleic acid sequence can be optimized for expression in mammalian cells or in a particular mammalian species (such as human cells). Codon optimization does not alter the amino acid sequence of the encoded protein.

Conservative variant: A protein containing conservative amino acid substitutions that do not substantially affect or decrease the function of a protein, such as a coronavirus spike protein. "Conservative" amino acid substitutions are those substitutions that do not substantially affect or decrease a function of a protein, such as the ability of the protein to elicit an immune response when administered to a subject. The term conservative variation also includes the use of a substituted amino acid in place of an unsubstituted parent amino acid. Furthermore, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (for instance less than 5%, in some embodiments less than 1%) in an encoded sequence are conservative variations where the alterations result in the substitution of an amino acid with a chemically similar amino acid.

The following six groups are examples of amino acids that are considered to be conservative substitutions for one another:

1) Alanine (A), Serine (S), Threonine (T);
2) Aspartic acid (D), Glutamic acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Non-conservative substitutions are those that reduce an activity or function of a protein, such as a recombinant Env protein, such as the ability to elicit an immune response when administered to a subject. For instance, if an amino acid residue is essential for a function of the protein, even an otherwise conservative substitution may disrupt that activity. Thus, a conservative substitution does not alter the basic function of a protein of interest.

Coronavirus: A large family of positive-sense, single-stranded RNA viruses that can infect humans and non-human animals. Coronaviruses get their name from the crown-like spikes on their surface. The viral envelope is comprised of a lipid bilayer containing the viral membrane (M), envelope (E) and spike (S) proteins. Most coronaviruses cause mild to moderate upper respiratory tract illness, such as the common cold. However, three coronaviruses have emerged that can cause more serious illness and death: severe acute respiratory syndrome coronavirus (SARS-CoV), SARS-CoV-2, and Middle East respiratory syndrome coronavirus (MERS-CoV). Other coronaviruses that infect humans include human coronavirus HKU1 (HKU1-CoV), human coronavirus OC43 (OC43-CoV), human coronavirus 229E (229E-CoV), and human coronavirus NL63 (NL63-CoV).

COVID-19: The disease caused by the coronavirus SARS-CoV-2.

Degenerate variant: A polynucleotide encoding a polypeptide that includes a sequence that is degenerate as a result of the genetic code. There are 20 natural amino acids, most of which are specified by more than one codon. Therefore, all degenerate nucleotide sequences are included as long as the amino acid sequence of the polypeptide is unchanged.

E3 region: Refers to the adenovirus early region 3 (E3) gene, which contains multiple open reading frames (ORFs). The E3 region of human adenovirus type 4 (Ad4) includes the following ORFs: 12.1K, 23.3K, 19K, 24.8K, 6.3K, 29.7K, 10.4K, 14.5K and 14.7K. In some embodiments herein, the deletion in the E3 region comprises a deletion of the 23.3K, 19K, 24.8K, 6.3K, 29.7K, 10.4K, 14.5K and 14.7K ORFs. In other embodiments, the deletion in the E3 region is a deletion of only the 24.8K, 6.3K and 29.7K ORFs.

Heterologous: Originating from a separate genetic source or species. For example, a heterologous polypeptide or polynucleotide refers to a polypeptide or polynucleotide derived from a different source or species.

Immune response: A response of a cell of the immune system, such as a B cell, T cell, or monocyte, to a stimulus. In some embodiments, the response is specific for a particular antigen (an "antigen-specific response"), such as a SARS-CoV-2 spike protein. In some embodiments, the immune response is a T cell response, such as a CD4+ response or a CD8+ response. In other embodiments, the response is a B cell response, and results in the production of specific antibodies. "Priming an immune response" refers to treatment of a subject with a "prime" immunogen/immunogenic composition to induce an immune response that is subsequently "boosted" with a boost immunogen/immunogenic composition. Together, the prime and boost immunizations produce the desired immune response in the subject.

Immunogenic composition: A composition that includes an immunogen or a nucleic acid molecule or vector encoding an immunogen (such as SARS-CoV-2 spike protein), that elicits a measurable CTL response against the immunogen, and/or elicits a measurable B cell response (such as production of antibodies) against the immunogen, when administered to a subject. It further refers to isolated nucleic acids encoding an immunogen, such as a nucleic acid that can be used to express the immunogen (and thus be used to elicit an immune response against this immunogen). For *in vivo* use, the immunogenic composition can include the protein or nucleic acid molecule in a pharmaceutically acceptable carrier and may also include other agents, such as an adjuvant.

Immunize: To render a subject protected from infection by a particular infectious agent, such as SARS-CoV-2. Immunization does not require 100% protection. In some examples, immunization provides at least 50%, at least 60%, at least 70%, at least 80%, at least 90% or at least 95% protection against infection compared to infection in the absence of immunization.

Isolated: An "isolated" biological component has been substantially separated or purified away from other biological components, such as other biological components in which the component naturally occurs, such as other chromosomal and extrachromosomal DNA, RNA, and proteins. Proteins, peptides, nucleic acids, and viruses that have been "isolated" include those purified by standard purification methods. Isolated does not require absolute purity, and can include protein, peptide, nucleic acid, or virus molecules that are at least 50% isolated, such as at least 75%, 80%, 90%, 95%, 98%, 99%, or even 99.9% isolated.

Neutralizing antibody: An antibody that reduces the infectious titer of an infectious agent by binding to a specific antigen on the infectious agent, such as a virus (*e.g.*, a coronavirus). In some embodiments, an antibody that is specific for a SARS-CoV-2 spike protein neutralizes the infectious titer of SARS-CoV-2. For example, an antibody that neutralizes SARS-CoV-2 may interfere with the virus by binding it directly and limiting entry into cells. Alternately, a neutralizing antibody may interfere with one or more post-attachment interactions of the pathogen with a receptor, for example, by interfering with viral entry using the receptor. In some embodiments, a SARS-CoV-2 neutralizing antibody inhibits SARS-CoV-2 infection of cells, for example, by at least 50%, by at least 60%, by at least 70%, by at least 80% or by at least 90%, compared to a control antibody.

Pharmaceutically acceptable carriers: The pharmaceutically acceptable carriers of use are conventional. *Remington's Pharmaceutical Sciences*, by E. W. Martin, Mack Publishing Co., Easton, PA, 19th Edition, 1995, describes compositions and formulations suitable for pharmaceutical delivery of the disclosed immunogens (such as recombinant Ad4 expressing SARS-CoV-2 S protein) and immunogenic compositions.

In general, the nature of the carrier will depend on the particular mode of administration being employed. For instance, parenteral formulations usually comprise injectable fluids that include pharmaceutically and physiologically acceptable fluids such as water, physiological saline, balanced salt solutions, aqueous dextrose, glycerol or the like as a vehicle. For solid compositions (*e.g.*, powder, pill, tablet, or capsule forms), conventional non-toxic solid carriers can include, for example, pharmaceutical grades of mannitol, lactose, starch, or magnesium stearate. In addition to biologically neutral carriers, pharmaceutical compositions to be administered can contain minor amounts of non-toxic auxiliary substances, such as wetting or emulsifying agents, preservatives, and pH buffering agents and the like, for example, sodium acetate or sorbitan monolaurate. In particular embodiments, suitable for administration to a subject the carrier may be sterile, and/or suspended or otherwise contained in a unit dosage form containing one or more measured doses of the composition suitable to elicit the desired anti-SARS-CoV-2 immune response. It may also be accompanied by medications for its use for treatment purposes. The unit dosage form may be, for example, in a sealed vial that contains sterile contents or a syringe for injection into a subject, or lyophilized for subsequent solubilization and administration or in a solid or controlled release dosage.

Preventing, treating or ameliorating a disease: "Preventing" a disease refers to inhibiting the full development of a disease. "Treating" refers to a therapeutic intervention that ameliorates a sign or symptom of a disease or pathological condition after it has begun to develop, such as a reduction in viral load. "Ameliorating" refers to the reduction in the number or severity of signs or symptoms of a disease, such as a coronavirus infection.

Recombinant: A recombinant nucleic acid, vector or virus is one that has a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two otherwise separated segments of sequence. This artificial combination can be accomplished, for example, by the artificial manipulation of isolated segments of nucleic acids, for example, using genetic engineering techniques.

Replication-competent virus: A virus capable of undergoing genome replication and protein synthesis to produce progeny virus.

Sequence identity: The similarity between amino acid or nucleotide sequences is expressed in terms of the similarity between the sequences, otherwise referred to as sequence identity. Sequence identity is frequently measured in terms of percentage identity; the higher the percentage, the more similar the two sequences are. Homologs, orthologs, or variants of a polypeptide or polynucleotide will possess a relatively high degree of sequence identity when aligned using standard methods.

Methods of alignment of sequences for comparison are known. Various programs and alignment algorithms are described in: Smith & Waterman, *Adv. Appl. Math.* 2:482, 1981; Needleman & Wunsch, *J. Mol. Biol.* 48:443, 1970; Pearson & Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444, 1988; Higgins & Sharp, *Gene*, 73:237-44, 1988; Higgins & Sharp, *CABIOS* 5:151-3, 1989; Corpet *et al.*, *Nuc. Acids Res.* 16:10881-90, 1988; Huang *et al. Computer Appls. In the*

Biosciences 8, 155-65, 1992; and Pearson *et al.*, *Meth. Mol. Bio.* 24:307-31, 1994. Altschul *et al.*, *J. Mol. Biol.* 215:403-10, 1990, presents a detailed consideration of sequence alignment methods and homology calculations.

Variants of a polypeptide or nucleic acid sequence are typically characterized by possession of at least about 75%, for example, at least about 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity counted over the full length alignment with the amino acid or nucleotide sequence of interest. Sequences with even greater similarity to the reference sequences will show increasing percentage identities when assessed by this method, such as at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% sequence identity. When less than the entire sequence is being compared for sequence identity, homologs and variants will typically possess at least 80% sequence identity over short windows of 10-20 amino acids (or 30-60 nucleotides), and may possess sequence identities of at least 85% or at least 90% or 95% depending on their similarity to the reference sequence. Methods for determining sequence identity over such short windows are available at the NCBI website on the internet.

As used herein, reference to "at least 90% identity" (or similar language) refers to "at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or even 100% identity" to a specified reference sequence.

SARS-CoV-2: A coronavirus of the genus betacoronavirus that first emerged in humans in 2019. This virus is also known as Wuhan coronavirus, 2019-nCoV, or 2019 novel coronavirus. The term "SARS-CoV-2" includes variants thereof, such as, but not limited to, alpha (B.1.1.7 and Q lineages); beta (B.1.351 and descendent lineages); delta (B.1.617.2 and AY lineages); gamma (P.1 and descendent lineages); epsilon (B.1.427 and B.1.429); eta (B.1.525); iota (B.1.526); kappa (B.1.617.1); 1.617.3; mu (B.1.621, B.1.621.1), zeta (P.2) and omicron (B.1.1.529 and BA lineages). Symptoms of SARS-CoV-2 infection include fever, chills, dry cough, shortness of breath, fatigue, muscle/body aches, headache, new loss of taste or smell, sore throat, nausea or vomiting, and diarrhea. Patients with severe disease can develop pneumonia, multi-organ failure, and death. The time from exposure to onset of symptoms is approximately 2 to 14 days. The SARS-CoV-2 virion includes a viral envelope with large spike glycoproteins. The SARS-CoV-2 genome, like most coronaviruses, has a common genome organization with the replicase gene included in the 5'-two thirds of the genome, and structural genes included in the 3'-third of the genome. The SARS-CoV-2 genome encodes the canonical set of structural protein genes in the order 5' - spike (S) - envelope (E) - membrane (M) and nucleocapsid (N) - 3'.

SARS Spike (S) protein: A class I fusion glycoprotein initially synthesized as a precursor protein of approximately 1256 amino acids for SARS-CoV, and 1273 amino acids for SARS-CoV-

2. Individual precursor S polypeptides form a homotrimer and undergo glycosylation within the Golgi apparatus as well as processing to remove the signal peptide, and cleavage by a cellular protease between approximately position 679/680 for SARS-CoV, and 685/686 for SARS-CoV-2, to generate separate S1 and S2 polypeptide chains, which remain associated as S1/S2 protomers within the homotrimer, thereby forming a trimer of heterodimers. The S1 subunit is distal to the virus membrane and contains the receptor-binding domain (RBD) that is believed to mediate virus attachment to its host receptor. The S2 subunit is believed to contain the fusion protein machinery, such as the fusion peptide. S2 also includes two heptad-repeat sequences (HR1 and HR2) and a central helix typical of fusion glycoproteins, a transmembrane domain, and a cytosolic tail domain. An exemplary wild-type (Wuhan strain) SARS-CoV-2 spike protein sequence is set forth herein as SEQ ID NO: 2. Exemplary modified Wuhan SARS-CoV-2 spike protein sequences are set forth herein as SEQ ID NOs: 3-5. In addition, exemplary SARS-CoV-2 variant spike protein sequences are set forth herein as SEQ ID NOs: 7-12.

Subject: Living multicellular vertebrate organisms, a category that includes human and non-human mammals. In some embodiments, the subject is a human. In some examples, a subject who is in need of inhibiting or preventing a SARS-CoV-2 infection is selected. For example, the subject can be uninfected and at risk of SARS-CoV-2 infection.

Therapeutically effective amount: A quantity of a specific substance, such as a disclosed immunogen (*e.g.*, a recombinant Ad4 expressing SARS-CoV-2 S protein) or immunogenic composition, sufficient to achieve a desired effect in a subject being treated, such as a protective immune response. A "therapeutically effective amount" can be the amount necessary to inhibit SARS-CoV-2 replication or treat COVID-19 in a subject with an existing SARS-CoV-2 infection. A "prophylactically effective amount" refers to administration of an agent or composition that inhibits or prevents establishment of an infection, such infection by SARS-CoV-2. It is understood that to obtain a protective immune response against an antigen of interest, multiple administrations of a disclosed immunogen/immunogenic composition can be required, and/or administration of a disclosed composition as the "prime" in a prime boost protocol wherein the boost immunogen can be different from the prime immunogenic composition. Accordingly, an effective amount of a disclosed immunogen/immunogenic composition can be the amount of the immunogen or immunogenic composition sufficient to elicit a priming immune response in a subject that can be subsequently boosted with the same or a different immunogen to elicit a protective immune response.

In one example, a desired response is to elicit an immune response that inhibits or prevents SARS-CoV-2 infection. The SARS-CoV-2 infected cells do not need to be completely eliminated or prevented for the composition to be effective. For example, administration of an effective amount of an immunogen or immunogenic composition can elicit an immune response that decreases the number of SARS-CoV-2 infected cells (or prevents the infection of cells) by a desired amount, for example, by at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, or even at least 100% (elimination or prevention of detectable SARS-CoV-2 infected cells), as compared to the number of SARS-CoV-2 infected cells in the absence of the immunization.

Unit dosage form: A physically discrete unit, such as a capsule, tablet, or solution, that is suitable as a unitary dosage for a human patient, each unit containing a predetermined quantity of one or more active ingredient(s) calculated to produce a therapeutic effect, in association with at least one pharmaceutically acceptable diluent or carrier, or combination thereof.

Vaccine: A pharmaceutical composition that elicits a prophylactic or therapeutic immune response in a subject. In some cases, the immune response is a protective immune response. Typically, a vaccine elicits an antigen-specific immune response to an antigen of a pathogen, for example a viral pathogen, or to a cellular constituent correlated with a pathological condition. A vaccine may include a polynucleotide (such as a nucleic acid encoding a disclosed antigen), a peptide or polypeptide (such as a disclosed antigen), a virus, a cell or one or more cellular constituents. In one specific, non-limiting example, a vaccine reduces the severity of the symptoms associated with SARS-CoV-2 infection and/or decreases the viral load compared to a control. In another non-limiting example, a vaccine reduces SARS-CoV-2 infection and/or transmission compared to a control.

Vector: An entity containing a DNA or RNA molecule bearing a promoter(s) that is operationally linked to the coding sequence of a protein (such as an immunogenic protein) of interest and can express the coding sequence. Non-limiting examples include a naked or packaged (lipid and/or protein) DNA, a naked or packaged RNA, a subcomponent of a virus or bacterium or other microorganism that may be replication-incompetent, or a virus or bacterium or other microorganism that may be replication-competent. A vector is sometimes referred to as a construct. Recombinant DNA vectors are vectors having recombinant DNA. A vector can include nucleic acid sequences that permit it to replicate in a host cell, such as an origin of replication. A vector can also include one or more selectable marker genes and other genetic elements. Viral vectors are recombinant nucleic acid vectors having at least some nucleic acid sequences derived from one or more viruses. Non-limiting examples of viral vectors include adenovirus vectors, adeno-associated virus (AAV) vectors, and poxvirus vectors (e.g., vaccinia, fowlpox).

III. Introduction

Of the available vaccine platforms for presenting viral glycoproteins to the immune system, replicating vectors have several important advantages over most non-replicating vectors (Robert-Guroff, *Curr Opin Biotechnol* 18(6):546-556, 2007). Replication-competent vectors can express viral surface proteins such that the total dose of antigen vastly exceeds those of non-replicating vectors. Replicating mucosal vaccines induce mucosal immunity, including IgA and IgG antibodies, and a balanced T cell response including resident memory T cells. In addition, replicating vectors, such as replication-competent adenovirus (Ad) vectors, express viral glycoproteins over a prolonged period of time, similar to live virus infections. This feature is thought to be important for the loading of dendritic cells in the lymph node and the induction of a durable antibody response (Cirelli *et al.*, *Cell* 177(5): 1153-1171, 2019; Tam *et al.*, *Proc Natl Acad Sci USA* 113(43): E6639-E6648, 2016; Mueller *et al.*, *Mol Pharm* 12(5): 1356-1365, 2015). Each of these features contributes to the magnitude and durability of immune responses observed after replicating viral vaccinations.

The vaccine constructs disclosed herein are replication-competent Ad4 encoding a SARS-CoV-2 spike (S) protein. In the disclosed Ad4 vector, which is derived from an Ad4 vaccine strain, the gene encoding a SARS-CoV-2 spike protein is cloned into an E3 region having a deletion of multiple E3 ORFs. The parent Ad4 vaccine vector has been given to over 10 million people with an excellent safety record. Ad4-recombinants have been developed for both influenza virus H5 and human immunodeficiency virus (HIV) envelope (Env) and Gag proteins. These Ad4-based vaccines have been through pre-clinical testing in rabbits for immunogenicity and human testing in phase 1 clinical trials.

The replication-competent Ad4-based vaccine platform has several distinct advantages compared to other proposed and licensed SARS-CoV-2 vaccines. For example, the efficacy of Ad4 vaccines has already been established as they have been administered routinely as a single dose enteric capsule in the U.S. military and found to prevent respiratory disease with an efficacy of greater than 95%. In addition, when administered intranasally or onto the tonsils, replication-competent Ad4-based vaccines induce a neutralizing antibody response in human subjects. Upper respiratory tract administration also bypasses pre-existing Ad4 immunity in most people. By inducing mucosal immunity, the Ad4-based vaccine platform not only provides protection for vaccinated subjects, but also has the potential to interrupt transmission of SARS-CoV-2 to others. In contrast to non-replicating viral vaccines, the replication-competent Ad4-based system produces a durable immune response. Furthermore, unlike mRNA-based SARS-CoV-2 vaccines, Ad4 vaccines can be stored long term at 4-8°C. Moreover, the disclosed vaccine platform is unmatched in terms of scalability and cost. It is estimated that the disclosed SARS-CoV-2 vaccine can be produced for less than 1 cent per dose.

IV. Overview of Embodiments

Disclosed herein is a recombinant adenovirus type 4 (Ad4) expressing a SARS-CoV-2 spike (S) protein (in some embodiments, referred to herein as "Ad4-SARS-CoV-2-spike" or "Ad4-Spike"), a recombinant Ad4 nucleic acid vector encoding the recombinant Ad4-Spike, and immunogenic compositions thereof.

In one aspect, provided herein is a recombinant Ad4 expressing a SARS-CoV-2 S protein. The recombinant Ad4 is replication-competent and the genome of the Ad4 includes a deletion in the adenovirus E3 region and an insertion of a coding sequence for the SARS-CoV-2 S protein. In some embodiments, the amino acid sequence of the S protein is at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to the amino acid sequence of a native S protein, such as the S protein of the Wuhan SARS-CoV-2 strain set forth herein as SEQ ID NO: 2. In specific examples, the amino acid sequence of the S protein comprises or consists of SEQ ID NO: 2.

The amino acid numbering used herein for residues of the SARS-CoV-2 S protein is with reference to the wild-type Wuhan strain SARS-CoV-2 S sequence provided as SEQ ID NO: 2. With reference to the SARS-CoV-2 S protein sequence provided as SEQ ID NO: 2, the ectodomain of the SARS-CoV-2 S protein includes about residues 16-1208. Residues 1-15 are the signal peptide, which is removed during cellular processing. The S1/S2 cleavage site is located at position 685/686. The HR1 is located at about residues 915-983. The central helix is located at about residues 988-1029. The HR2 is located at about 1162-1194. The C-terminal end of the S2 ectodomain is located at about residue 1208. The position numbering of the S protein may vary between SARS-CoV-2 stains, but the sequences can be aligned to determine relevant structural domains and cleavage sites (see, *e.g.*, FIG. 4).

In some embodiments, the recombinant Ad4 comprises a coding sequence for a SARS-CoV-2 S protein comprising one or more (such as two, for example two consecutive) proline substitutions at or near the boundary between a HR1 domain and a central helix domain that stabilize the S protein in the prefusion conformation. In some such embodiments, the one or more (such as two, for example two consecutive) proline substitutions that stabilize the S protein in the prefusion conformation are located between a position 15 amino acids N-terminal of a C-terminal residue of the HR1 and a position 5 amino acids C-terminal of a N-terminal residue of the central helix. In some embodiments, the one or more (such as two, for example two consecutive) proline substitutions that stabilize the SARS-CoV-2 S protein in the prefusion conformation are located between residues 975 to 995 (such as 981-992). In some embodiments, the SARS-CoV-2 S protein is stabilized in the prefusion conformation by K986P and V987P substitutions ("PP" or "2P"). In some embodiments, the SARS-CoV-2 S protein is stabilized in the prefusion conformation by one or two proline substitutions at positions D985, K986, or V987 of the S ectodomain protomers in the trimer. In some examples, the SARS-CoV-2 S protein stabilized in the prefusion conformation by the one or more proline substitutions (such as K986P and V987P substitutions) comprises one or more additional modifications for stabilization in the prefusion conformation.

In some embodiments, the SARS-CoV-2 S protein encoded by the recombinant Ad4 genome comprises an amino acid sequence at least 90% (such as at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%) identical to SEQ ID NO: 3 (Wuhan-PP), wherein the SARS-CoV-2 S protein is stabilized in the prefusion conformation with one or more of the modifications provided herein (such as the K986P and V987P substitutions). In other embodiments, the stabilized, proline substituted S protein is derived from a SARS-CoV-2 variant. In some examples, stabilized S protein derived from a SARS-CoV-2 variant comprises an amino acid sequence at least 90% (such as at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%) identical to SEQ ID NO: 7 (beta-PP), SEQ ID NO: 8 (Wuhan/RDB-beta-PP), SEQ ID NO: 9 (delta-PP), SEQ ID NO: 10 (gamma-PP), SEQ ID NO: 11 (delta plus-PP) or SEQ ID NO: 12 (omicron-PP). In particular examples, the amino acid sequence of the stabilized SARS-CoV-2 S protein comprises or consists of SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11 or SEQ ID NO: 12.

In other embodiments, the SARS-CoV-2 S protein encoded by the recombinant Ad4 genome comprises a C-terminal truncation, such as a truncation of the cytoplasmic tail or a truncation of the endocytosis motif. In specific examples, the truncated SARS-CoV-2 S protein comprises or consists of the amino acid sequence of SEQ ID NO: 4 or SEQ ID NO: 5.

An exemplary nucleic acid sequence encoding a SARS-CoV-2 S protein is provided as SEQ ID NO: 6. In some examples, the nucleic acid sequence encoding the S protein is at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 6. In specific non-limiting examples, the nucleic acid sequence encoding the S protein comprises or consists of SEQ ID NO: 6.

The DNA sequence of the exemplary SARS-CoV-2 S protein provided above can be modified to introduce the amino acid substitutions and deletions disclosed herein for prefusion stabilization. In some embodiments, this DNA sequence (with or without modification to introduce amino acid substitutions) can be included in the recombinant Ad4 vector as the sequence encoding the SARS-CoV-2 S protein. In some embodiments, the S protein is encoded by a codon-optimized nucleic acid sequence. In some examples, the nucleic acid sequence encoding the S protein is at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 13 (beta-PP), SEQ ID NO: 14 (Wuhan/RBD beta-PP), SEQ ID NO: 15 (delta-PP), SEQ ID NO: 16 (gamma-PP), SEQ ID NO: 17 (delta plus-PP), SEQ ID NO: 18 (omicron-PP) or SEQ ID NO: 19 (Wuhan-PP). In specific examples, the nucleic acid sequence encoding the S protein comprises or consists of any one of SEQ ID NOs: 13-19.

In some embodiments, the deletion in the E3 region is a deletion of at least two, at least three, at least four, at least five, at least six, or at least seven E3 open reading frame (ORFs). In some examples, the deletion includes at least two, at least three, at least four, at least five, at least six, or at least seven of the 23.3K, 19K, 24.8K, 6.3K, 29.7K, 10.4K, 14.5K and 14.7K ORFs. In particular non-limiting examples, the deletion in the E3 region includes a deletion of each of the 23.3K, 19K, 24.8K, 6.3K, 29.7K, 10.4K, 14.5K and 14.7K ORFs.

In some embodiments, the coding sequence for the SARS-CoV-2 S protein is inserted in place of the deleted portion of the E3 region.

In some embodiments, the nucleotide sequence of the genome of the recombinant Ad4 is at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 1. In some examples, the nucleotide sequence of the genome of the recombinant Ad4 comprises or consists of SEQ ID NO: 1.

Also provided herein is a recombinant, replication-competent Ad4 nucleic acid vector. In some embodiments, the recombinant Ad4 vector includes a deletion in the adenovirus E3 region and an insertion of a coding sequence for the SARS-CoV-2 S protein. In some embodiments, the amino acid sequence of the S protein is at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to the amino acid sequence of a native S protein, such as the S protein of the Wuhan SARS-CoV-2 strain set forth herein as SEQ ID NO: 2. In specific examples, the amino acid sequence of the S protein comprises or consists of SEQ ID NO: 2.

In some embodiments, the SARS-CoV-2 S protein is stabilized in the prefusion conformation by K986P and V987P substitutions (“PP” or “2P”). In some embodiments, the SARS-CoV-2 S protein is stabilized in the prefusion conformation by one or two proline substitutions at positions D985, K986, or V987 of the S ectodomain protomers in the trimer. In some examples, the SARS-CoV-2 S protein stabilized in the prefusion conformation by the one or more proline substitutions (such as K986P and V987P substitutions) comprises one or more additional modifications for stabilization in the prefusion conformation.

In some embodiments, the SARS-CoV-2 S protein encoded by the recombinant Ad4 nucleic acid vector comprises an amino acid sequence at least 90% (such as at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%) identical to SEQ ID NO: 3 (Wuhan-PP), wherein the SARS-CoV-2 S protein is stabilized in the prefusion conformation with one or more of the modifications provided herein (such as the K986P and V987P substitutions). In other embodiments, the stabilized, proline substituted S protein is derived from a SARS-CoV-2 variant. In some embodiments, the S protein is encoded by a codon-optimized nucleic acid sequence. In some examples, stabilized S protein derived from a SARS-CoV-2 variant comprises an amino acid sequence at least 90% (such as at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%) identical to SEQ ID NO: 7 (beta-PP), SEQ ID NO: 8 (Wuhan/RDB-beta-PP), SEQ ID NO: 9 (delta-PP), SEQ ID NO: 10 (gamma-PP), SEQ ID NO: 11 (delta plus-PP) or SEQ ID NO: 12 (omicron-PP). In particular examples, the amino acid sequence of the stabilized SARS-CoV-2 S protein comprises or consists of SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11 or SEQ ID NO: 12.

In other embodiments, the SARS-CoV-2 S protein encoded by the recombinant Ad4 nucleic acid vector comprises a C-terminal truncation, such as a truncation of the cytoplasmic tail or a truncation of the endocytosis motif. In specific examples, the truncated SARS-CoV-2 S protein comprises or consist of the amino acid sequence of SEQ ID NO: 4 or SEQ ID NO: 5.

In some embodiments of the disclosed Ad4 vector, the deletion in the E3 region is a deletion of at least two, at least three, at least four, at least five, at least six, or at least seven E3 ORFs. In some examples, the deletion includes at least two, at least three, at least four, at least five, at least six, or at least seven of the 23.3K, 19K, 24.8K, 6.3K, 29.7K, 10.4K, 14.5K and 14.7K ORFs. In particular non-limiting examples, the deletion in the E3 region includes a deletion of each of the 23.3K, 19K, 24.8K, 6.3K, 29.7K, 10.4K, 14.5K and 14.7K ORFs.

In some embodiments of the disclosed Ad4 vector, the coding sequence for the SARS-CoV-2 S protein is inserted in place of the deleted portion of the E3 region. In some examples, the coding sequence for the S protein is at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to any one of SEQ ID NOs: 2-5 and 7-12. In specific non-limiting examples, the coding sequence for the S protein comprises or consists of any one of SEQ ID NOs: 2-5 and 7-12.

In some embodiments, the nucleotide sequence of the Ad4 vector is at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 1. In some examples, the nucleotide sequence of the Ad4 vector comprises or consists of SEQ ID NO: 1.

Further provided herein are immunogenic compositions that include a recombinant Ad4 or a recombinant Ad4 vector, and a pharmaceutically acceptable carrier. In some embodiments, the immunogenic composition further includes an adjuvant. In other embodiments, the immunogenic composition does not include an adjuvant.

Methods of eliciting an immune response against SARS-CoV-2 in a subject are also provided. In some embodiments, the method includes administering to the subject a therapeutically effective amount of a recombinant Ad4, a recombinant Ad4 (nucleic acid) vector, or an immunogenic composition disclosed herein. Also provided are methods of immunizing a subject against SARS-CoV-2 infection. In some embodiments, the method includes administering to the subject a therapeutically effective amount of a recombinant Ad4, a recombinant Ad4 vector, or an immunogenic composition disclosed herein.

In some embodiments of the disclosed methods, the recombinant Ad4, recombinant Ad4 vector, or immunogenic composition is administered intranasally or onto the tonsils. In some examples, intranasal administration includes administration of an aerosol. The particle size of the aerosol should allow for delivery to the upper respiratory tract, but not the lower respiratory tract. In specific examples, the aerosol contains particles greater than 10 microns in diameter, such as greater than 20 microns, greater than 30 microns, greater than 40 microns or greater than 50 microns. In particular examples, the aerosol contains particles of about 10 to about 150 microns, such as about 20 to about 125 microns or about 30 to about 100 microns. One of skill in the art is capable of selecting an appropriate device for intranasal delivery of the disclosed recombinant Ad4, recombinant Ad4 vector, or immunogenic composition to the upper respiratory tract. Non-limiting examples of devices include Accuspray™ (Becton-Dickinson) and the MAD Nasal™ (Teleflex ®) atomizer.

In some embodiments, the method includes administering a dose of about 10^4 to about 10^6 recombinant Ad4 particles, such as about 5 x 10^4 to about 5 x 10^5 viral particles or about 1 x 10^5 viral particles. In some examples, the dose is about 1 x 10^4, 2 x 10^4, 3 x 10^4, 4 x 10^4, 5 x 10^4, 6 x 10^4, 7 x 10^4, 8 x 10^4, 9 x 10^4, 1 x 10^5, 2 x 10^5, 3 x 10^5, 4 x 10^5, 5 x 10^5, 6 x 10^5, 7 x 10^5, 8 x 10^5, 9 x 10^5, or 1 x 10^6 recombinant Ad4 particles.

In some embodiments, the recombinant Ad4, the recombinant Ad4 vector, or the immunogenic composition is administered in a single dose.

In some embodiments, the recombinant Ad4, the recombinant Ad4 vector, or the immunogenic composition is administered as part of a prime-boost immunization protocol. In some examples, the recombinant Ad4, the recombinant Ad4 vector, or the immunogenic composition is the prime dose. In other examples, the recombinant Ad4, the recombinant Ad4 vector, or the immunogenic composition is the boost dose.

V. Preclinical and Clinical Studies Relevant to COVID-19 Vaccine Development

By studying the vaccine-induced mucosal neutralizing antibody responses in a series of live oral poliovirus (OPV) challenge studies, investigators have robustly demonstrated the remarkable separation of the systemic and mucosal antibody systems (Brickley *et al.*, *Clin Infect Dis.* 2018;67(suppl_1):S42-S50). This research demonstrates that, despite inducing high levels of serum antibody and providing individual protection from paralytic polio, inactivated Salk vaccines fail to induce the intestinal IgA responses that are critical for inhibiting enteric poliovirus replication and preventing fecal-oral transmission. In contrast, primary vaccination with live attenuated Sabin OPV induces robust mucosal IgA responses and sterilizing immunity upon challenge with live OPV. This observation emphasizes the critical nature of inducing mucosal immunity to prevent infection and transmission of COVID-19. It is believed that the lack of mucosal immunogenicity seen with OPV will be echoed by subunit or replication-incompetent systemically administered SARS-CoV-2 vaccines.

In pre-clinical testing of SARS-CoV-2 vaccines, a similar advantage to mucosal immunization in blocking infection has been observed. In ferrets, IM or mucosal immunization with a replication-defective Ad5-spike recombinant induced similar levels of spike-specific antibodies in the serum, yet only mucosal immunization induced sterilizing protection of the upper respiratory tract (URT) (Wu *et al.*, *Nat Commun* 11(1): 4081, 2020). A similar advantage of intranasal administration over intramuscular administration in inducing mucosal immunity and sterilizing protection of the URT has been observed using lentiviral- or chimp adenoviral-spike recombinants in mouse models permissive to SARS-CoV-2 infection (Ku *et al.*, *Cell Host Microbe* S1931-3128(20)30672-7, 2020; Hassan *et al.*, *Cell* 183(1): 169-184, 2020; King *et al.*, King *et al.*, *bioRxiv* 2020.10.10.331348, 2020). It has been observed that local specific IgA is highly associated with terminating viral shedding in humans after challenge with coronavirus 229E (Callow *et al.*, *J Hyg* 95(1): 173-189, 1985).

Prior attempts to protect against a viral mucosal infection for which the host is naïve using a parenterally administered non-replicating vaccine have failed or produced enhanced disease. Examples include respiratory syncytial virus (RSV), parainfluenza virus (PIV)-3, Ad4, rotavirus, and measles virus. The reasons for these failures lie in part in the difficulty in protecting mucosal surfaces coated on their apical surfaces with viral receptors, 100-1000-fold lower antibodies on these surfaces compared to serum, and distorted and short-lived immune responses generated by non-replicating vectors. Clinical trials of the disclosed Ad4-SARS-CoV-2-spike vaccine will evaluate in detail the humoral and mucosal responses to the SARS-CoV-2 spike protein and the adenovirus vector. It is expected that the disclosed Ad4-SARS-CoV-2-spike vaccine will produce mucosal antibodies in the respiratory tract and most closely mimic the immune profile observed following natural SARS-CoV-2 infection. Furthermore, it is believed that the disclosed vaccine offers the best possibility for durably interrupting transmission during the COVID-19 pandemic.

Among the recombinant viral vectors available for human use, replicating adenoviruses offer several important advantages. Replicating Ad4 has been given to more than 10 million people in the military as a vaccine against Ad4 respiratory disease and has an extraordinary safety and efficacy record (Gaydos and Gaydos, *Mil Med.* 1995;160(6):300-304). This recombinant Ad4 is attenuated by administration to the gastrointestinal tract in the form of an enteric coated tablet, and does not cause respiratory disease (Choudhry *et al.*, *Vaccine* 2016:34(38) 4558-4564). Using an enteric capsule delivery, a phase 3 study was undertaken with 4,000 volunteers entering basic military training. The results demonstrated a vaccine efficacy of 99.3% and seroconversion in 94.5% against respiratory disease caused by Ad4 (Kuschner *et al.*, *Vaccine* 2013:31 2963-2971).

In one trial in humans, replicating recombinant adenoviral vectors expressing influenza virus H5 delivered enterically were only modestly immunogenic. This is most likely related to the attenuation of replication by administration to the gastrointestinal tract (Gurwith *et al.*, *Lancet Infect Dis.* 2013;13(3):238-50) coupled with the E3 deletion. The introduction of a large gene such as that coding for the coronavirus spike protein into an adenovirus vector involves the removal of most early (in this case E3) genes and conveys at least a 10-fold attenuation to the parent adenovirus in tissue culture, chimpanzees, and humans (Lubeck *et al.*, *Nat Med.* 1997;3(6):651-8).

In another clinical trial, high and remarkably durable levels of influenza-specific neutralizing antibodies were observed when a replication-competent Ad4 expressing the influenza virus hemagglutinin type 5 Vietnam (Ad4-H5-Vtn) was administered to the URT compared to the gastrointestinal (GI) tract (Matsuda *et al.*, *Sci Immunol.* 2019;4(34):eaau2710; Matsuda *et al.*, *J Clin Invest* 131(5):e140794, 2021). The vaccine delivered into the URT was very safe (nasal congestion or throat discomfort in 25% of participants, none above grade 2) up to a dose of 10^8. This level of reactogenicity is at approximately the same level as seen in placebos, and with some parenterally administered non-replicating platforms now being tested against SARS-CoV-2, and below that of a currently licensed varicella zoster (Shingrix) vaccine. URT administration of adenoviruses to Ad4-seropositive humans did result in reinfection. URT administration uses the difficulties in protecting the upper respiratory tree to its advantage to overcome vector-specific immunity. An example of that is the ability of an adenovirus expressing Ebola glycoprotein to induce protective immunity on Ebola challenge by the intranasal route in adeno-immune primates while no protection was observed after IM administration of the Ebola construct in previously adeno immune animals.

Prior results with Ad4-H5-Vtn and Ad4-HIV recombinants indicated that nearly all human participants developed a response to the transgene. After a single intranasal or tonsillar administration of the vaccine, increases in H5-specific B cells, H5-specific antibody somatic hypermutation, and potency were observed. The vaccines also induced a very durable response. The response to the licensed split influenza vaccine typically wanes by 5-10-fold within 2-6 months following immunization. However, when Ad4-H5-Vtn participants were asked to return for boosting 3-5 years later, neutralizing antibodies were still at the level that one observes at the peak response after immunization with the licensed vaccine. The Ad4-SARS-CoV-2-spike vaccine construct disclosed herein could be used to generate mucosal immunity after a systemic vaccination. Alternatively, a subunit vaccine could be administered following immunization with the disclosed vaccine to boost mucosal and systemic antibody, which has been shown to occur with the H5-Vtn vaccine construct.

VI. Immunogenic Compositions

Immunogenic compositions that include a disclosed immunogen (*e.g.*, a recombinant Ad expressing a SARS-CoV-2 S protein, or a recombinant Ad4 nucleic acid vector comprising a SARS-CoV-2 S protein coding sequence), and a pharmaceutically acceptable carrier are also provided. Such compositions can be administered to subjects by a variety of administration modes, for example, intranasal, onto the tonsils, inhalation, oral, intramuscular, subcutaneous, intravenous, intra-arterial, intra-articular, intraperitoneal, or parenteral routes. Methods for preparing administrable compositions are described in more detail in such publications as *Remingtons Pharmaceutical Sciences*, 19th Ed., Mack Publishing Company, Easton, Pennsylvania, 1995.

Thus, an immunogen described herein can be formulated with pharmaceutically acceptable carriers to help retain biological activity while also promoting increased stability during storage within an acceptable temperature range. Potential carriers include, but are not limited to, physiologically balanced culture medium, phosphate buffer saline solution, water, emulsions (*e.g.,* oil/water or water/oil emulsions), various types of wetting agents, cryoprotective additives or stabilizers such as proteins, peptides or hydrolysates (*e.g.,* albumin, gelatin), sugars (*e.g.,* sucrose, lactose, sorbitol), amino acids (*e.g.,* sodium glutamate), or other protective agents. The resulting aqueous solutions may be packaged for use as is or lyophilized. Lyophilized preparations are combined with a sterile solution prior to administration for either single or multiple dosing.

Formulated compositions, especially liquid formulations, may contain a bacteriostat to prevent or minimize degradation during storage, including but not limited to effective concentrations (usually ≦1% w/v) of benzyl alcohol, phenol, m-cresol, chlorobutanol, methylparaben, and/or propylparaben. A bacteriostat may be contraindicated for some patients; therefore, a lyophilized formulation may be reconstituted in a solution either containing or not containing such a component.

The immunogenic compositions of the disclosure can contain as pharmaceutically acceptable vehicles substances as required to approximate physiological conditions, such as pH adjusting and buffering agents, tonicity adjusting agents, wetting agents and the like, for example, sodium acetate, sodium lactate, sodium chloride, potassium chloride, calcium chloride, sorbitan monolaurate, and triethanolamine oleate.

The pharmaceutical composition may optionally include an adjuvant to enhance an immune response of the host. Suitable adjuvants are, for example, toll-like receptor agonists, alum, $AlPO_4$, alhydrogel, Lipid-A and derivatives or variants thereof, oil-emulsions, saponins, neutral liposomes, liposomes containing the vaccine and cytokines, non-ionic block copolymers, and chemokines. Non-ionic block polymers containing polyoxyethylene (POE) and polyxylpropylene (POP), such as POE-POP-POE block copolymers, MPL™ (3-O-deacylated monophosphoryl lipid A; Corixa, Hamilton, IN) and IL-12 (Genetics Institute, Cambridge, MA), may be used as an adjuvant (Newman *et al.*, 1998, *Critical Reviews in Therapeutic Drug Carrier Systems* 15:89-142). These adjuvants have the advantage in that they help to stimulate the immune system in a non-specific way, thus enhancing the immune response to a pharmaceutical product. In some embodiments, an adjuvant is not required and is thus not administered with the Ad4-Spike vaccine.

In some embodiments, the composition can be provided as a sterile composition. The pharmaceutical composition typically contains an effective amount of a disclosed immunogen and can be prepared by conventional techniques. Typically, the amount of immunogen in each dose of the immunogenic composition is selected as an amount which elicits an immune response without significant, adverse side effects. In some examples, the dose is about 1 x 10^4 to about 10^6 viral particles, such as about 5 x 10^4 to about 5 x 10^5 viral particles or about 1 x 10^5 viral particles.

In some embodiments, the composition can be provided in unit dosage form for use to elicit an immune response in a subject, for example, to prevent SARS-CoV-2 infection in the subject. A unit dosage form contains a suitable single preselected dosage for administration to a subject, or suitable marked or measured multiples of two or more preselected unit dosages, and/or a metering mechanism for administering the unit dose or multiples thereof. In some examples, the unit dosage is about 1 x 10^4 to about 10^6 viral particles, such as about 5 x 10^4 to about 5 x 10^5 viral particles. In specific examples, the unit dosage is about 1 x 10^5 viral particles.

VII. Methods of Eliciting an Immune Response

The disclosed immunogens (*e.g.,* a recombinant replication-competent adenovirus expressing a SARS-CoV-2 spike protein), polynucleotides and vectors encoding the disclosed immunogens, and compositions including same, can be used in methods of inducing an immune response to SARS-CoV-2 to prevent, inhibit (including inhibiting transmission), and/or treat a SARS-CoV-2 infection.

Provided herein are methods of eliciting an immune response against SARS-CoV-2 in a subject. In some embodiments, the method includes administering to the subject an effective amount of a recombinant adenovirus, adenovirus vector or immunogenic composition disclosed herein. In some examples, the recombinant adenovirus, vector or immunogenic composition is administered intranasally (such as in a spray) or orally (such as by using enteric-coated tablets).

When inhibiting, treating, or preventing SARS-CoV-2 infection, the methods can be used either to avoid infection in an SARS-CoV-2 seronegative subject (*e.g.,* by inducing an immune response that protects against SARS-CoV-2 infection), or to treat existing infection in a SARS-CoV-2 seropositive subject.

To identify subjects for prophylaxis or treatment according to the methods of the disclosure, accepted screening methods are employed to determine risk factors associated with a targeted or suspected disease or condition, or to determine the status of an existing disease or condition in a subject. These screening methods include, for example, conventional work-ups to determine environmental, familial, occupational, and other such risk factors that may be associated with the targeted or suspected disease or condition, as well as diagnostic methods, such as various ELISA and other immunoassay methods to detect and/or characterize SARS-CoV-2 infection. These and other routine methods allow the clinician to select patients in need of therapy using the methods and immunogenic compositions of the disclosure. In accordance with these methods and principles, a composition can be administered according to the teachings herein, or other conventional methods, as an independent prophylaxis or treatment program, or as a follow-up, adjunct or coordinate treatment regimen to other treatments.

The disclosed immunogens can be used in coordinate (or prime-boost) immunization protocols or combinatorial formulations. In certain embodiments, novel combinatorial immunogenic compositions and coordinate immunization protocols employ separate immunogens or formulations, each directed toward eliciting an anti- SARS-CoV-2 immune response, such as an immune response to SARS-CoV-2 spike protein. Separate immunogenic compositions that elicit the anti- SARS-CoV-2 immune response can be combined in a polyvalent immunogenic composition administered to a subject in a single immunization step, or they can be administered separately (in monovalent immunogenic compositions) in a coordinate immunization protocol.

In one embodiment, a suitable immunization regimen includes at least two separate inoculations with one or more immunogenic compositions including a disclosed Ad4-Spike with a second inoculation being administered more than about two, about three to eight, or about four weeks following the first inoculation. A third inoculation can be administered several months after the second inoculation, and in specific embodiments, more than about five months after the first inoculation, more than about six months to about two years after the first inoculation, or about eight months to about one year after the first inoculation. Periodic inoculations beyond the third are also desirable to enhance the subject's "immune memory." The adequacy of the vaccination parameters chosen, *e.g.*, formulation, dose, regimen and the like, can be determined by taking aliquots of serum from the subject and assaying antibody titers during the course of the immunization program. Alternatively, the T cell populations can be monitored by conventional methods. In addition, the clinical condition of the subject can be monitored for the desired effect, *e.g.*, prevention of SARS-CoV-2 infection, improvement in disease state (*e.g.*, reduction in viral load), or reduction in transmission frequency. If such monitoring indicates that vaccination is sub-optimal, the subject can be boosted with an additional dose of immunogenic composition, and the vaccination parameters can be modified in a fashion expected to potentiate the immune response. Thus, for example, a dose of a disclosed immunogen can be increased or the route of administration can be changed.

It is contemplated that there can be several boosts, and that each boost can be a different immunogen. It is also contemplated in some examples that the boost may be the same immunogen as another boost, or the prime.

The prime and the boost can be administered as a single dose or multiple doses, for example, two doses, three doses, four doses, five doses, six doses or more can be administered to a subject over days, weeks or months. Multiple boosts can also be given, such one to five, or more. Different dosages can be used in a series of sequential inoculations. For example, a relatively large dose in a primary inoculation and then a boost with relatively smaller doses. The immune response against the selected antigenic surface can be elicited by one or more inoculations of a subject.

In several embodiments, a disclosed immunogen can be administered to the subject simultaneously with the administration of an adjuvant. In other embodiments, the immunogen can be administered to the subject after the administration of an adjuvant and within a sufficient amount of time to elicit the immune response. In other embodiments, no adjuvant is administered.

SARS-CoV-2 infection does not need to be completely inhibited for the methods to be effective. For example, elicitation of an immune response to SARS-CoV-2 can reduce or inhibit SARS-CoV-2 infection by a desired amount, for example, by at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, or even at least 100% (elimination or prevention of detectable SARS-CoV-2 infected cells), as compared to SARS-CoV-2 infection in the absence of immunization. In additional examples, SARS-CoV-2 replication can be reduced or inhibited by the disclosed methods. SARS-CoV-2 replication does not need to be completely eliminated for the method to be effective. For example, the immune response elicited using one or more of the disclosed immunogens can reduce SARS-CoV-2 replication by a desired amount, for example, by at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, or even at least 100% (elimination or prevention of detectable SARS-CoV-2 replication), as compared to SARS-CoV-2 replication in the absence of the immune response.

Following immunization of a subject, serum can be collected from the subject at appropriate time points, frozen, and stored for neutralization testing. Methods to assay for neutralization activity, include, but are not limited to, plaque reduction neutralization (PRNT) assays, microneutralization assays, flow cytometry based assays, single-cycle infection assays, and pseudovirus neutralization assays.

In some embodiments, immunization is achieved by administration of recombinant Ad4 vector DNA. Immunization by nucleic acid constructs is taught, for example, in U.S. Patent No. 5,643,578 (which describes methods of immunizing vertebrates by introducing DNA encoding a desired antigen to elicit a cell-mediated or a humoral response), U.S. Patent No. 5,593,972 and U.S. Patent No. 5,817,637 (which describe operably linking a nucleic acid sequence encoding an antigen to regulatory sequences enabling expression), and broadly described in Janeway & Travers, *Immunobiology: The Immune System In Health and Disease,* page 13.25, Garland Publishing, Inc., New York, 1997; and McDonnell & Askari, *N. Engl. J. Med.* 334:42-45, 1996.

The following examples are provided to illustrate certain particular features and/or embodiments. These examples should not be construed to limit the disclosure to the particular features or embodiments described.

EXAMPLES

Example 1: Expression of Wild-Type and Modified SARS-CoV-2 Spike Proteins

The following studies evaluated cell-surface expression of wild-type Wuhan strain SARS-CoV-2 spike protein (SEQ ID NO: 2) and three modified versions of the Wuhan strain spike protein: stabilized (PP), tail truncated (TT), and endocytosis motif truncated (no-Endo). PP contains double proline stabilization substitutions at amino acid positions 986 and 987 (SEQ ID NO: 3); TT includes a deletion of the terminal 24 amino acids of the cytoplasmic tail (SEQ ID NO: 4); and no-Endo contains a deletion of the C-terminal endocytosis signaling motif (SEQ ID NO: 5) (see FIG. 4).

Expression of SARS-CoV-2 WT, PP, TT and no-Endo spike proteins was evaluated in A549 cells. Cells were transfected with a shuttle vector plasmid containing the gene for a WT or modified SARS-CoV-2 spike protein. Untransfected cells served as negative controls and cells transfected with a plasmid expressing an HIV-1 Env protein was used as a positive control for transfection. Expression of spike and Env was measured by flow cytometry using a SARS-CoV-2 spike protein-specific antibody and an HIV Env-specific antibody (VRC01), respectively. As shown in FIG. 1, SARS-CoV-2 spike protein expression in transfected A549 cells diminished with truncation of the tail, and truncation of the endocytosis motif, relative to wild-type spike protein.

Nucleic acid sequence encoding the WT, PP or TT SARS-CoV-2 spike protein was inserted into the E3 region of a replication-competent Ad4 vector having a deletion of the E3 23.3K, 19K, 24.8K, 6.3K, 29.7K, 10.4K, 14.5K and 14.7K ORFs. The nucleotide sequence of the recombinant Ad4 containing the WT spike protein coding sequence is set forth herein as SEQ ID NO: 1. Expression of the WT, stabilized and truncated spike protein in recombinant Ad4-infected A549 cells was evaluated. Replicating Ad4 carrying the WT spike nucleic acid sequence (nCoV-WT), the PP-stabilized spike nucleic acid sequence (nCov-PP) or the tail-truncated spike nucleic acid sequence (nCov-TT) was used to infect A549 cells. A replicating adenovirus expressing an HIV-1 Env protein (FDE3) was used as a positive control of infection and uninfected (unIF) cells were used as a negative control. Expression of spike protein was measured by flow cytometry using a SARS-CoV-2 spike protein-specific antibody. Antibody VRC01 was used to detect expression of HIV-1 Env. Spike protein expression from the Ad4-Spike after 2 days of infection is shown in FIG. 2A. In FIG. 2B, expression of the PP-stabilized and truncated Spike proteins is shown. As shown in FIGS. 2A-2B, expression of spike protein was high from both the nCoV-WT and nCoV-PP constructs.

Example 2: Immunogenicity of Ad4-Spike (WT) in rabbits

Immunogenicity of Ad4-Spike (expressing the WT spike protein sequence of SEQ ID NO: 2) was tested in New Zealand white rabbits. Rabbits and other experimental animals do not replicate the Ad4 virus, however intramuscular administration (IM) is commonly used as a screen for immunogenicity. Rabbits were immunized IM on day 0 and day 28 with 1.29 x 10^9 infectious units (IFU) of purified replicating Ad4-Spike. Using a luciferase assay, serum neutralization against Wuhan SARS-CoV-2 pseudovirus was detected at 4 weeks (prior to the second immunization), and continued to increase through the 12-week study period.

Example 3: Immunogenicity studies in hamsters

Human adenoviruses are capable of infecting Syrian golden hamsters (van der Lubbe *et al.*, *NPJ Vaccines* 6(1):39, 2021). Thus, immunogenicity studies were performed in these animals. A dose titration from 10^2-10^7 infection forming units (IFU) of intranasal Ad4-SARS-CoV-2 Wuhan spike with PP stabilization (Ad4-SARS-CoV-2_{WuPP}) was conducted. Strong serum neutralization was observed at week 4 (FIG. 5A) and week 8 (FIG. 5B) in a lentivirus pseudotype assay at the highest doses of Ad4-SARS-CoV-2_{WuPP}.

These results suggested that the hamster is semi-permissive for Ad4, but replicates the virus sufficiently to induce serum neutralizing antibodies. Spike-specific IgA and IgG were also observed in the nasal wash on day 60.

Hamsters were then immunized with intranasal Ad4 expressing stabilized (double proline substituted – PP) spike proteins from variants of concern (VOC). Included in this study were Ad4-CoV2-Wuhan, Ad4-CoV2-SA (beta), Ad-CoV2-Wu/RBD-SA, Ad4-CoV2-Indian (delta) and Ad4-CoV2-Brazil (gamma). An Ad4 expressing an influenza virus H5 hemagglutinin (Ad4-H5) and sham inoculation were included as negative controls.

Serum neutralization against Wuhan, delta and omicron pseudovirus was determined 28 days and 56 days following intranasal administration. The results are shown in FIGS. 6A-6E. Ad4 expressing the Wuhan-PP (SEQ ID NO: 3) or Delta-PP (SEQ ID NO: 9) were the most immunogenic.

Example 4: Challenge study in hamsters

This example describes a study to test candidate vaccines in the Syrian golden hamster model.

In this study, Syrian golden hamsters are intranasally administered an immunogenic candidate identified in Example 3 (Candidate 1 or Candidate 2) at a dose of 10^7 IFU and subsequently challenged with SARS-CoV-2 by co-habitation with SARS-CoV-2 Delta- or SARS-CoV-2 Omicron-infected animals (van Doremalen *et al.*, *Sci Transl Med* 13(607):eabh0755, 2021). Table 1 shows the groups of animals that are used. Animals in Group A are challenged at day 60, while animals in Group B are challenged 6 months after immunization. Hamsters receiving intranasal administration of Ad4-H5-Vtn are included as negative controls. Pfizer mRNA or Ad26-Spike is administered intramuscularly as a comparator.

Table 1. Challenge study in hamsters

Group A	**Vaccine candidate**	**Dose**
1a	Ad4-H5-Vtn	1x10^7 IFU IN
2a	Candidate 1	1x10^7 IFU IN
3a	Candidate 2	1x10^7 IFU IN
4a	Pfizer mRNA BNT162b2	5 µg x 2 IM
5a	Ad26-Spike	1x10^9 VPU IM
Group B	**Vaccine candidate**	**Dose**
1b	Ad4-H5-Vtn	1x10^7 IFU IN
2b	Variant 1	1x10^7 IFU IN
3b	Variant 2	1x10^7 IFU IN
4a	Pfizer mRNA BNT162b2	5 µg x 2 IM
4b	Ad26-Spike	1x10^9 VPU IM

It is expected that intranasal Ad4-Spike vaccine will give systemic neutralizing antibodies that are of the same order of magnitude as mRNA or Ad26 but is more durable. It is also expected that the Ad4-Spike will cause greater restriction of the challenge virus compared to parenterally administered vaccines.

Example 4: Human Clinical Study

A Phase 1/2 open-label study of a single dose of intranasally administered Ad4-Spike in healthy volunteers is conducted. Enrollment begins with volunteers who may or may not have had prior coronavirus disease 2019 (COVID-19) or vaccination. The international setting chosen is one where supplies of COVID-19 vaccines are limited and SARS-CoV-2-naïve volunteers may be more easily enrolled. All SARS-CoV-2-naïve participants are offered an emergency use authorization (EUA) vaccine at the completion of the study or following the 6-month timepoint if their neutralization titer is below ~40 (which is the lower boundary of the interquartile range for the Moderna mRNA 1272 vaccine). Each study participant receives a single dose of an intranasal Ad4-SARS-CoV-2 vaccine or an intramuscular (IM) immunization with an authorized or licensed booster. Study participants are monitored for adverse events (AEs), and blood and respiratory secretions are collected for immunogenicity and safety testing periodically throughout the study period. Nasal swabs are collected to monitor adenovirus shedding, and nasal washes are collected to monitor mucosal immune responses. Household and intimate contacts willing to participate are also enrolled and monitored for transmission of the vaccine virus by serology.

The primary endpoints are for safety measured by the frequency and grade of solicited and unsolicited adverse events in the first 28 days after vaccination. Safety is evaluated by separately assessing the incidence, severity, and type of adverse events in the candidate vaccine arms of the trial over the duration of follow-up. It is expected that 21% (N=10/48) of vaccine recipients may experience vaccine-related signs and symptoms (*e.g.*, headache, fatigue, myalgia, rhinorrhea, nausea, diarrhea). Vaccine virus shedding is evaluated by describing the presence, quantity, and duration of shed virus in serially collected nasal wash samples.

A second endpoint is immunogenicity. Immunogenicity is evaluated in serially collected serum, nasal, and stool samples. Immunogenicity is determined by a lentivirus-based pseudovirus neutralization assay. The assay includes functional antibodies as measured by characterization of B-cell clones, complement-enhancement and antibody dependent enhancement, mucosal and T cell immunity. Respiratory mucosal responses are being seen after COVID-19 infection and are thus expected to be a distinguishing hallmark of the Ad4-Spike vaccine. If the Ad4-vectored SARS-CoV-2 vaccine 'takes' in 95% of recipients and is immunogenic to adenovirus 4 and SARS-CoV-2 spike protein in 90% of these recipients, it is expected that systemic immune responses will be induced in 85% (N=44/52) of vaccine recipients and mucosal responses will be induced in 90-100% of volunteers.

A second dose at 60 days is administered in the rare instance of no evidence of vaccine take at 30 days. However, the primary analysis is after 1 dose as this vaccine is expected to be a single dose regimen. Most participants in prior Ad4-based vaccine trials did not develop a higher response after a second immunization, a second dose would only induce a response in the infrequent case that a participant is not infected on the first dose.

As volunteers will not be pre-screened for serum antibodies, a subset of the volunteers will be seropositive at baseline for Ad4 (~30%, N=20/60) as a result of exposure to circulating wild-type adenoviruses. The response of those with pre-existing Ad4 immunity in the previous vectored vaccine trials has suggested that Ad4 immunity may modulate the response to the vector and limit virus shedding, but vector specific immunity will still be induced.

Participants are monitored for safety and immunogenicity for one year. The Phase 1 trial optionally includes parallel exploratory arms designed into the clinical trial to permit using Ad4-Spike in conjunction with other SARS-CoV-2 Spike immunogens such as DNA, mRNA, or protein vaccines. It is expected that Ad4-Spike will contribute greater durability and mucosal T and B cell responses compared to non-replicating, parenterally administered protein or nucleic acid vaccines.

The target study population excludes only those who may be negatively impacted by respiratory viral infections, such as pregnant women or those with severe immunodeficiencies. The symptoms of recombinant Ad4 vaccination, when they occur, tend to be mild and self-limited. Those persons without difficulties in handling upper respiratory infections should not experience severe symptoms with the Ad4-Spike vaccine. Although pre-existing immunity to Ad4 is not uncommon (30%), it is largely overcome by intranasal vaccination. The degree to which vector-specific immunity is overcome will be assessed and is expected to be a function of the replication of the vaccine virus and the immunogenicity of the spike protein. The prevalence of Ad4 antibodies in persons under 16 is extremely low, making this vaccine a very attractive mode to induce durable immunity in school aged children. The primary endpoints are safety and immunogenicity. Safety is definitively addressed in phase 2 of the trial if the primary endpoint is reached.

When prior Ad4 recombinant virus vaccines were given intranasally, the virus replicated at a low level for 2-4 weeks. However, shedding of the virus detected by viral culture was at a low level and for a median of one day. Participants are counselled to avoid intimate contact for 14 days after vaccination. For these reasons, transmission of the vaccine virus to household or intimate contacts has not been observed. Most vaccinees are asymptomatic. However, the most common adverse events (AEs) are throat discomfort and nasal congestion in 25% of participants, none above grade 2. It is expected that a recombinant Ad4 that includes the SARS-CoV-2 Spike protein will yield results similar to prior Ad4-based, intranasally administered vaccines.

A phase 3 study and/or challenge study is conducted following phase 2.

In view of the many possible embodiments to which the principles of the disclosed subject matter may be applied, it should be recognized that the illustrated embodiments are only preferred examples of the disclosure and should not be taken as limiting the scope of the disclosure. Rather, the scope of the disclosure is defined by the following claims. We therefore claim all that comes within the scope and spirit of these claims.

CLAIMS

1. A recombinant adenovirus type 4 (Ad4) expressing a SARS-CoV-2 spike (S) protein, wherein:

the amino acid sequence of the S protein is at least 95% identical to SEQ ID NO: 2;

the recombinant Ad4 is replication-competent; and

the genome of the recombinant Ad4 comprises a deletion in the adenovirus E3 region and an insertion of a coding sequence for the SARS-CoV-2 S protein.

2. The recombinant Ad4 of claim 1, wherein the amino acid sequence of the S protein is at least 99% identical to SEQ ID NO: 2.

3. The recombinant Ad4 of claim 1 or claim 2, wherein the amino acid sequence of the S protein comprises or consists of SEQ ID NO: 2.

4. The recombinant Ad4 of claim 1, wherein the amino acid sequence of the S protein comprises at least one modification to stabilize the protein in the prefusion conformation.

5. The recombinant Ad4 of claim 4, wherein the at least one modification comprises K986P and V987P substitutions.

6. The recombinant Ad4 of claim 4 or claim 5, wherein the amino acid sequence of the S protein comprises or consists of SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11 or SEQ ID NO: 12.

7. The recombinant Ad4 of any one of claims 1-6, wherein the deletion in the E3 region comprises a deletion of the 23.3K, 19K, 24.8K, 6.3K, 29.7K, 10.4K, 14.5K and 14.7K open reading frames (ORFs).

8. The recombinant Ad4 of any one of claims 1-7, wherein the coding sequence for the SARS-CoV-2 S protein is inserted in place of the deleted E3 region.

9. The recombinant Ad4 of any one of claims 1-8, wherein the S protein is encoded by a codon-optimized nucleic acid sequence.

10. The recombinant Ad4 of claim 9, wherein the codon-optimized nucleic acid sequence comprises or consists of SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18 or SEQ ID NO: 19.

11. The recombinant Ad4 of any one of claims 1-3, wherein the nucleotide sequence of the genome is at least 95% identical to SEQ ID NO: 1.

12. The recombinant Ad4 of any one of claims 1-3, wherein the nucleotide sequence of the genome is at least 99% identical to SEQ ID NO: 1.

13. The recombinant Ad4 of any one of claims 1-3, wherein the nucleotide sequence of the genome comprises or consists of SEQ ID NO: 1.

14. A recombinant adenovirus type 4 (Ad4) vector, comprising a deletion in the adenovirus E3 region and an insertion of a coding sequence for the SARS-CoV-2 S protein, wherein the amino acid sequence of the S protein is at least 95% identical to SEQ ID NO: 2.

15. The recombinant Ad4 vector of claim 14, wherein the amino acid sequence of the S protein is at least 99% identical to SEQ ID NO: 2.

16. The recombinant Ad4 vector of claim 14 or claim 15, wherein the amino acid sequence of the S protein comprises or consists of SEQ ID NO: 2.

17. The recombinant Ad4 vector of claim 14, wherein the amino acid sequence of the S protein comprises at least one modification to stabilize the protein in the prefusion conformation.

18. The recombinant Ad4 of claim 17, wherein the at least one modification comprises K986P and V987P substitutions.

19. The recombinant Ad4 of claim 17 or claim 18, wherein the amino acid sequence of the S protein comprises or consists of SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11 or SEQ ID NO: 12.

20. The recombinant Ad4 vector of any one of claims 14-19, wherein the deletion in the E3 region comprises a deletion of the 23.3K, 19K, 24.8K, 6.3K, 29.7K, 10.4K, 14.5K and 14.7K open reading frames (ORFs).

21. The recombinant Ad4 vector of any one of claims 14-20, wherein the coding sequence for the SARS-CoV-2 S protein is inserted in place of the deleted E3 region.

22. The recombinant Ad4 vector of any one of claims 14-21, wherein the S protein is encoded by a codon-optimized nucleic acid sequence.

23. The recombinant Ad4 vector of claim 22, wherein the codon-optimized nucleic acid sequence comprises of consists of SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18 or SEQ ID NO: 19.

24. The recombinant Ad4 vector of any one of claims 14-16, wherein the nucleotide sequence of the vector is at least 95% identical to SEQ ID NO: 1.

25. The recombinant Ad4 vector of any one of claims 14-16, wherein the nucleotide sequence of the vector is at least 99% identical to SEQ ID NO: 1.

26. The recombinant Ad4 vector of any one of claims 14-16, wherein the nucleotide sequence of the vector comprises or consists of SEQ ID NO: 1.

27. An immunogenic composition comprising the recombinant Ad4 of any one of claims 1-13 or the recombinant Ad4 vector of any one of claims 14-26, and a pharmaceutically acceptable carrier.

28. A method of eliciting an immune response against SARS-CoV-2 in a subject, comprising administering to the subject a therapeutically effective amount of the recombinant Ad4 of any one of claims 1-13, the recombinant replication-competent Ad4 vector of any one of claims 14-26, or the immunogenic composition of claim 27, thereby eliciting an immune response against SARS-CoV-2 in the subject.

29. A method of immunizing a subject against SARS-CoV-2 infection, comprising administering to the subject a therapeutically effective amount of the recombinant Ad4 of any one of claims 1-13, the recombinant replication-competent Ad4 vector of any one of claims 14-26, or the immunogenic composition of claim 27, thereby immunizing the subject against SARS-CoV-2 infection.

30. The method of claim 28 or claim 29, wherein administration comprises intranasal administration.

31. The method of claim 30, wherein intranasal administration comprises administration of an aerosol comprising particles greater than 10 microns in diameter.

32. The method of any one of claims 28-31, comprising administering a dose of about 10^4 to about 10^6 recombinant Ad4 particles.

33. The method of claim 32, comprising administering a dose of about 10^5 recombinant Ad4 particles.

34. The method of any one of claims 28-33, wherein the recombinant Ad4, the recombinant Ad4 vector, or the immunogenic composition is administered in a single dose.

FIG. 1

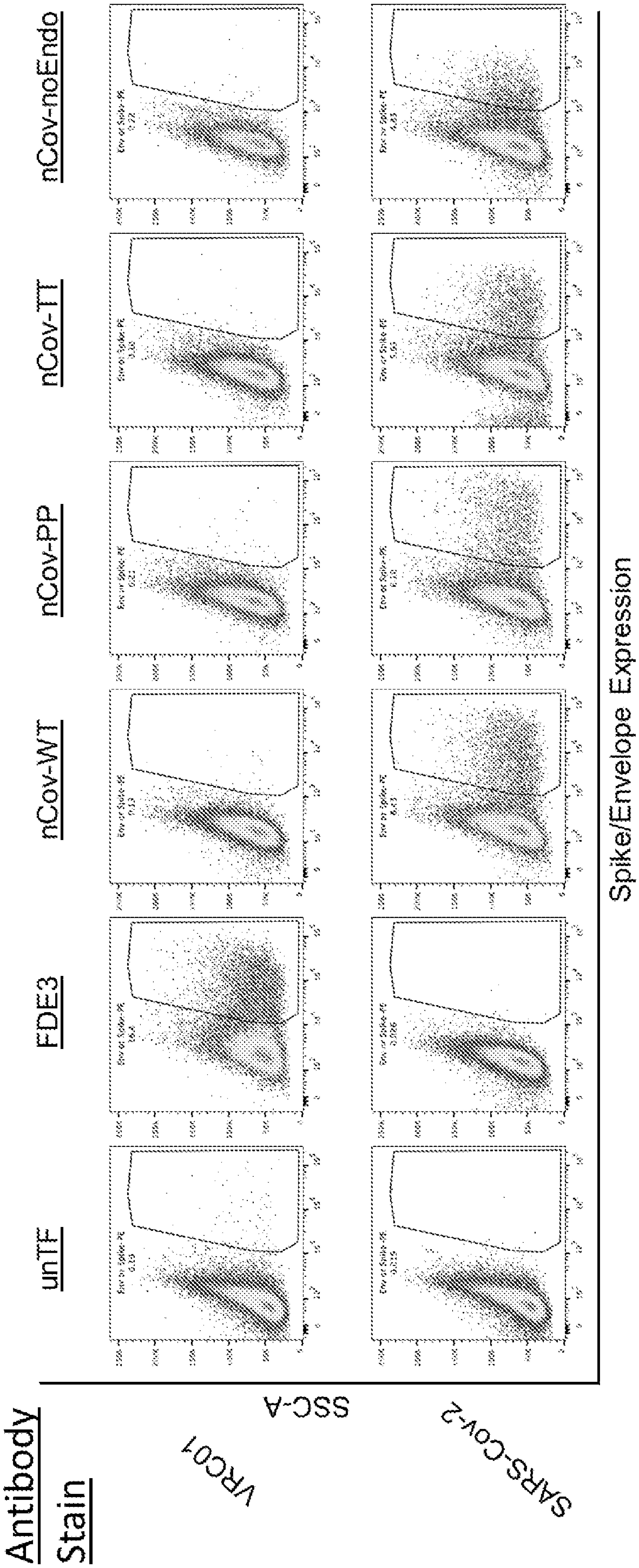

unTF
FDE3
nCov-WT
nCov-PP
nCov-TT
nCov-noEndo
Antibody Stain
VRC01
SARS-Cov-2
SSC-A
Spike/Envelope Expression

FIG. 2A

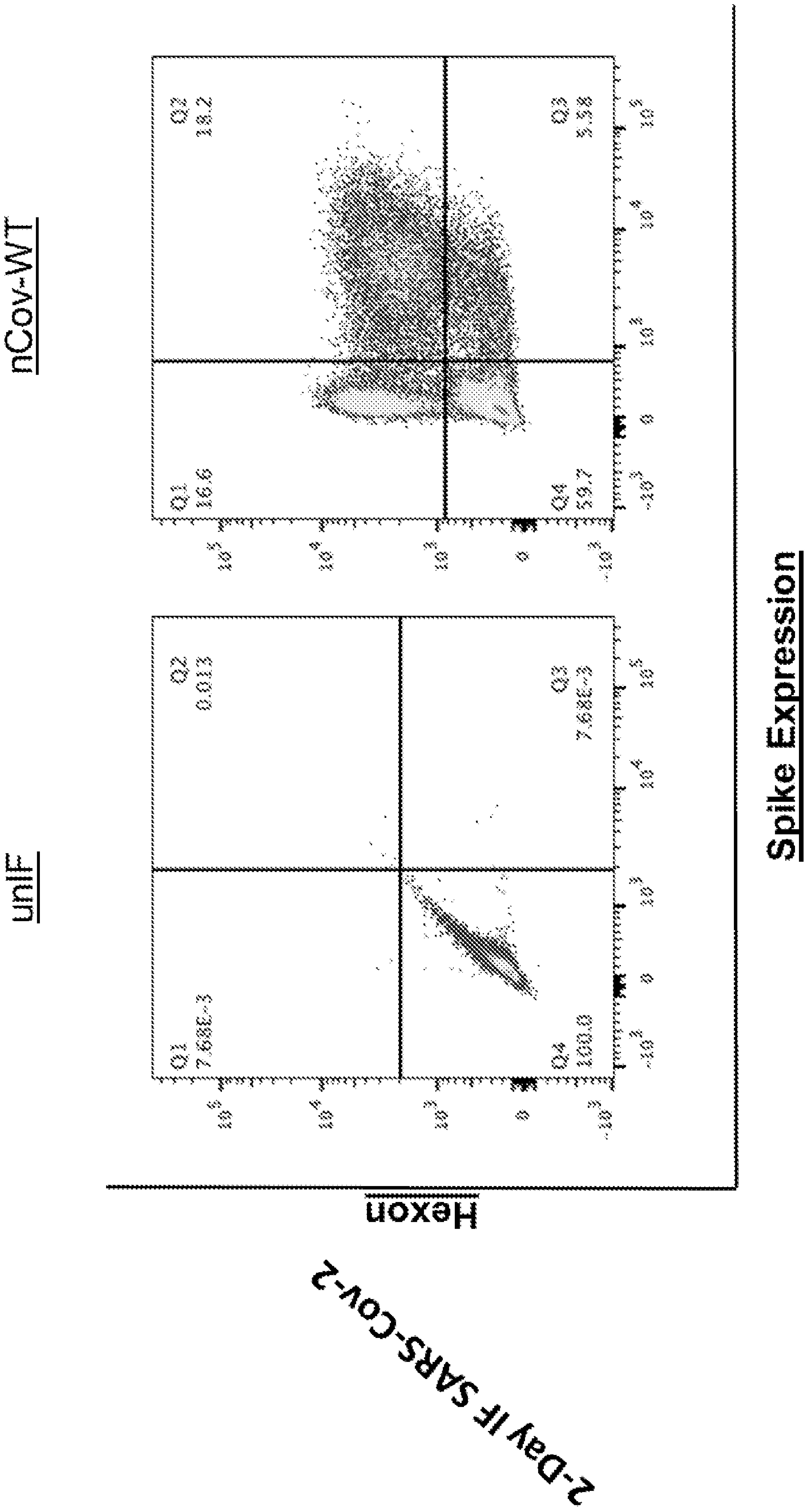

unIF
nCov-WT
Q1 7.68E-3
Q2 0.013
Q3 7.68E-3
Q4 100.0
Q1 16.6
Q2 18.2
Q3 5.58
Q4 59.7
Spike Expression
Hexon
2-Day IF SARS-Cov-2

FIG. 2B

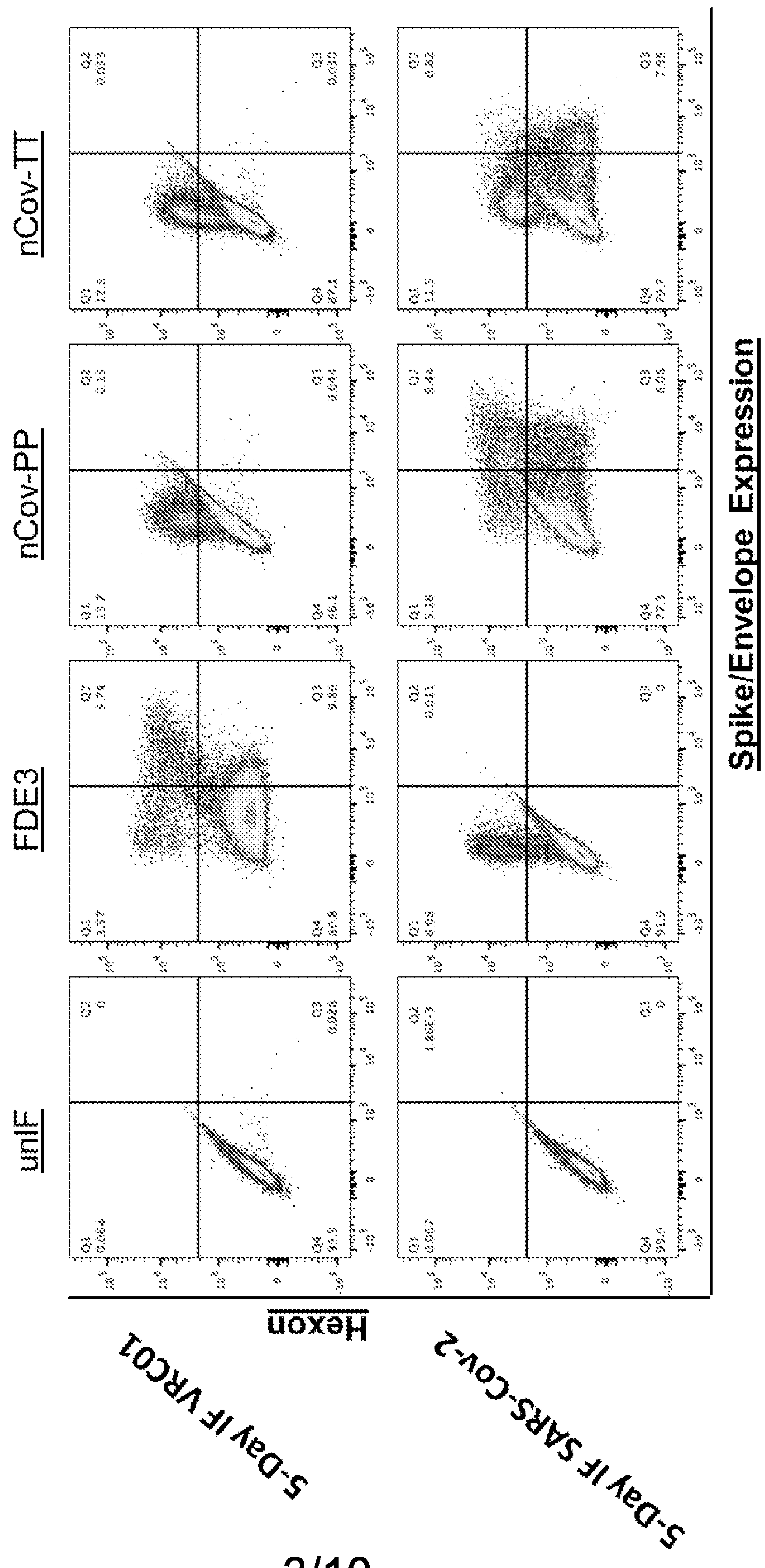

unIF
FDE3
nCov-PP
nCov-TT
Hexon
Spike/Envelope Expression
5-Day IF VRC01
5-Day IF SARS-Cov-2

FIG. 3

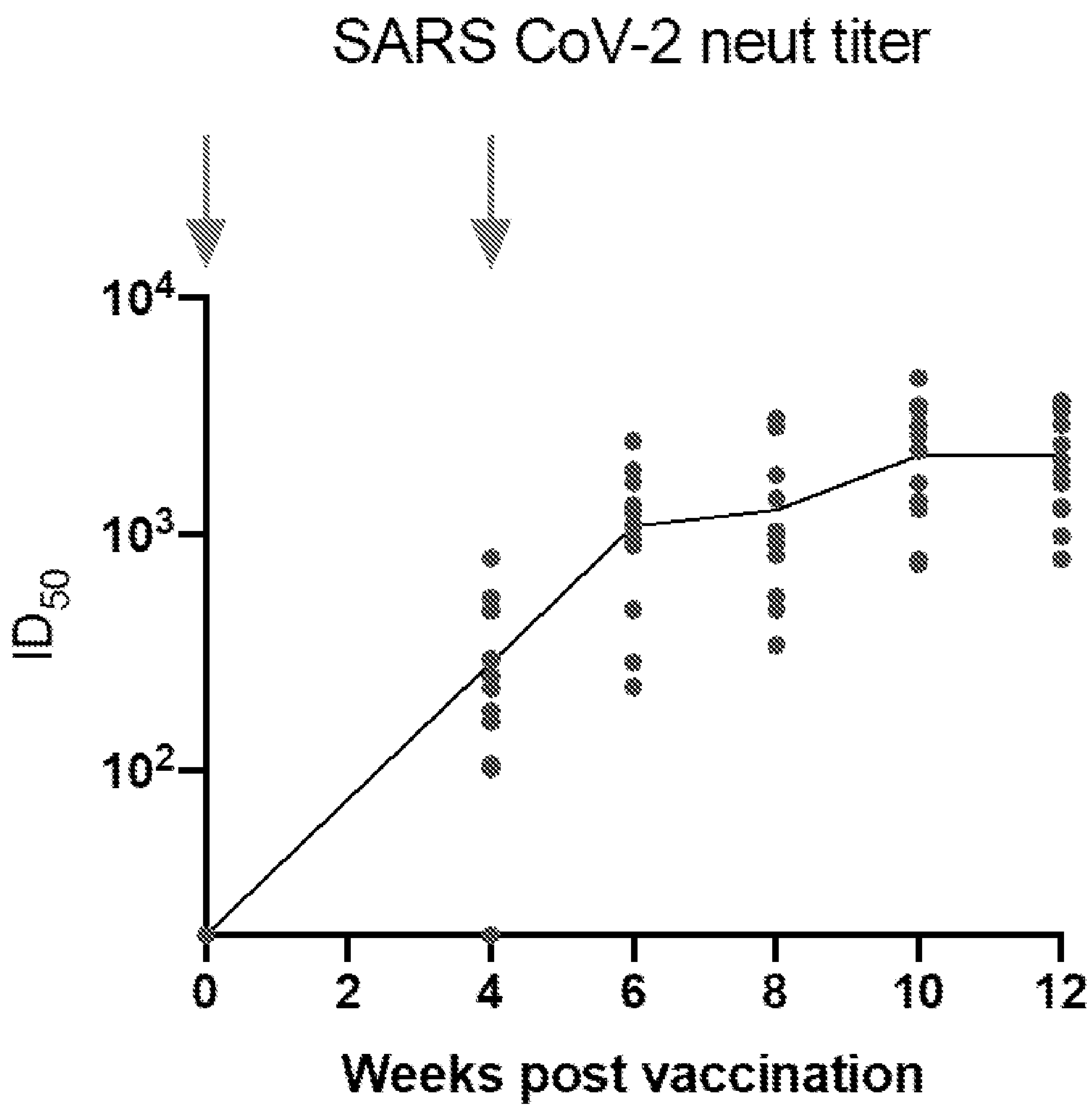

SARS CoV-2 neut titer
10^4
10^3
10^2
ID50
0
2
4
6
8
10
12
Weeks post vaccination

FIG. 4

```
nCoV-PP               QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN  960
nCoV-WT               QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN  960
nCoV-Tail-Truncation  QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN  960
nCoV-No-Endo          QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN  960
                      ************************************************************

nCoV-PP               TLVKQLSSNFGAISSVLNDILSRLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA  1020
nCoV-WT               TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA  1020
nCoV-Tail-Truncation  TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA  1020
nCoV-No-Endo          TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA  1020
                      *************************  *********************************

nCoV-PP               SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA  1080
nCoV-WT               SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA  1080
nCoV-Tail-Truncation  SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA  1080
nCoV-No-Endo          SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA  1080
                      ************************************************************

nCoV-PP               ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP  1140
nCoV-WT               ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP  1140
nCoV-Tail-Truncation  ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP  1140
nCoV-No-Endo          ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP  1140
                      ************************************************************

nCoV-PP               LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL  1200
nCoV-WT               LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL  1200
nCoV-Tail-Truncation  LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL  1200
nCoV-No-Endo          LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL  1200
                      ************************************************************

nCoV-PP               QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD  1260
nCoV-WT               QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD  1260
nCoV-Tail-Truncation  QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCS-----------  1249
nCoV-No-Endo          QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD  1260
                      *************************************************

nCoV-PP               SEPVLKGVKLHYT  1273  residues 901-1273 of SEQ ID NO: 3
nCoV-WT               SEPVLKGVKLHYT  1273  residues 901-1273 of SEQ ID NO: 2
nCoV-Tail-Truncation  -------------  1249  residues 901-1249 of SEQ ID NO: 4
nCoV-No-Endo          SEPVLKGV-----  1268  residues 901-1268 of SEQ ID NO: 5
```

FIG. 5B

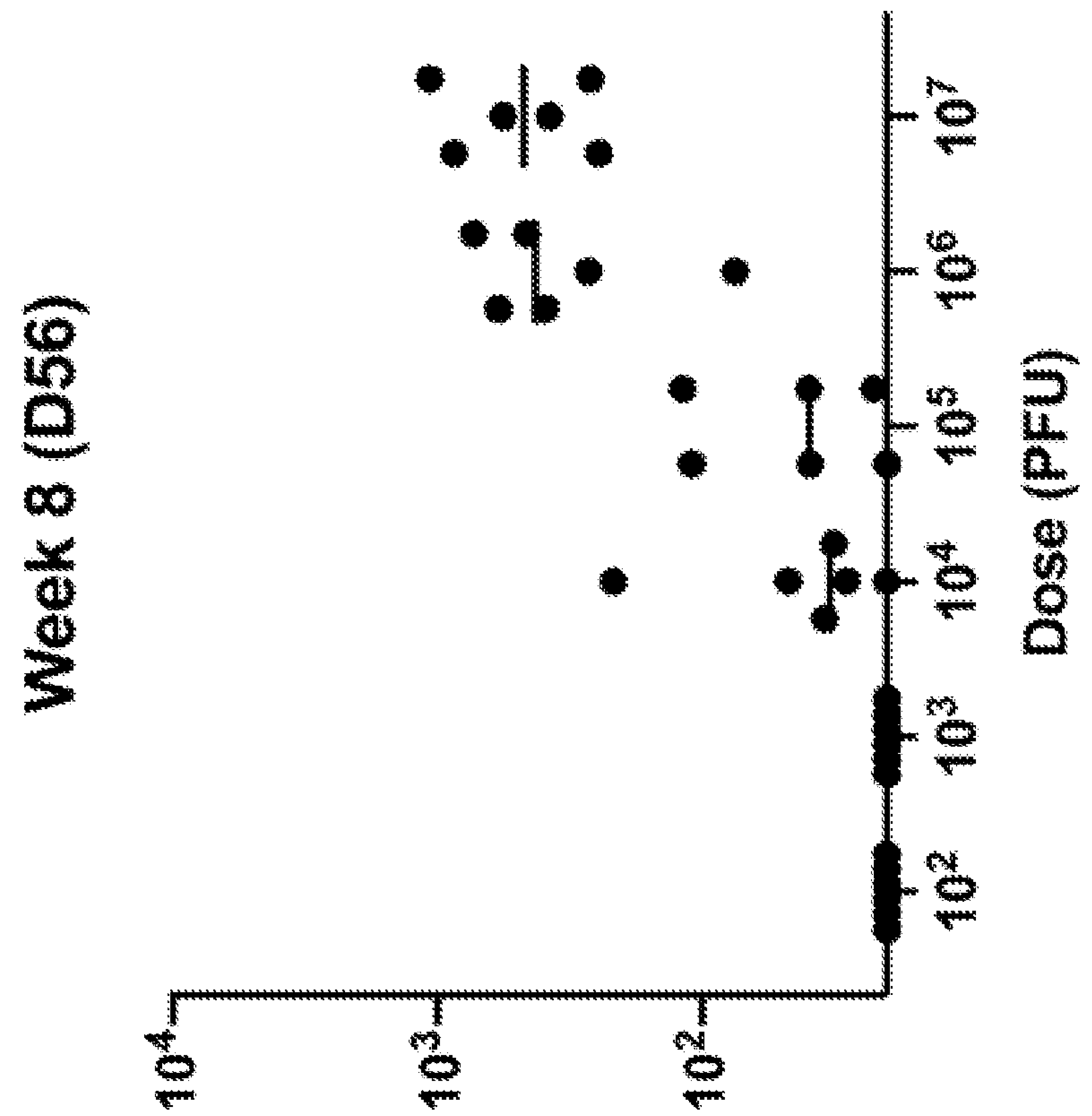

Week 8 (D56)
10^4
10^3
10^2
10^2
10^3
10^4
10^5
10^6
10^7
Dose (PFU)

FIG. 5A

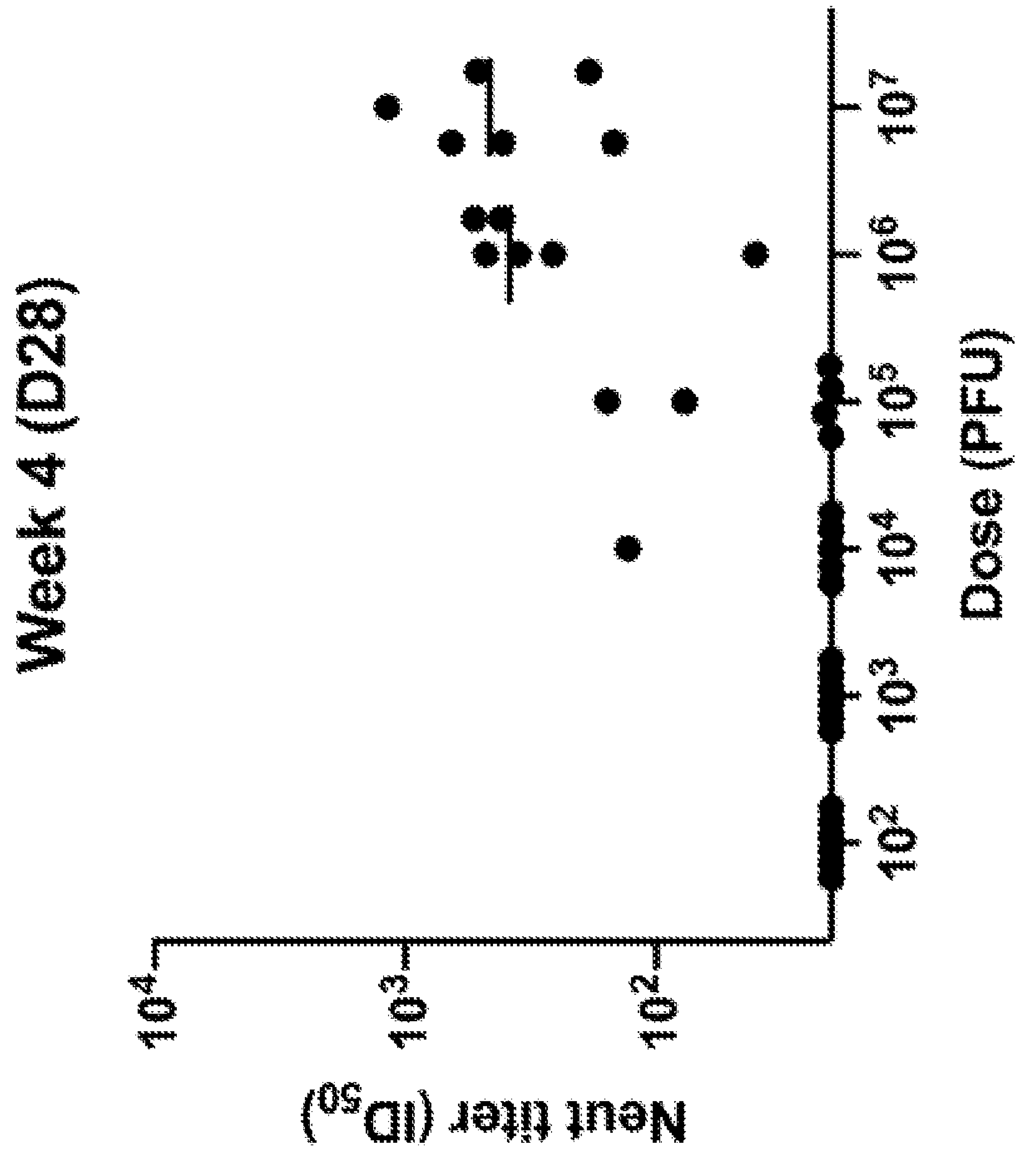

Week 4 (D28)
Neut titer (ID_{50})
10^4
10^3
10^2
10^2
10^3
10^4
10^5
10^6
10^7
Dose (PFU)

FIG. 6B

India (Delta)

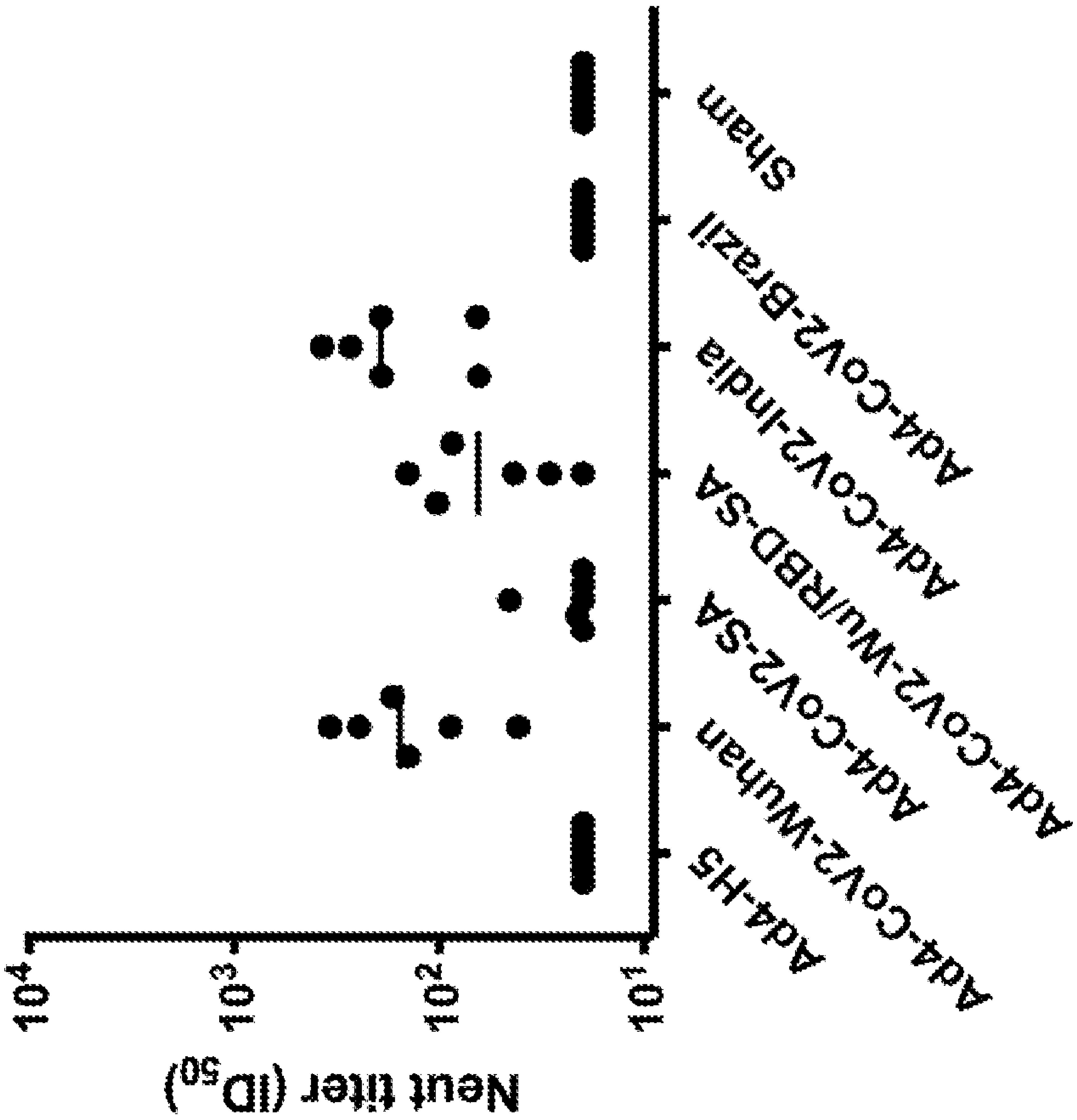


FIG. 6A

Wuhan (Alpha)

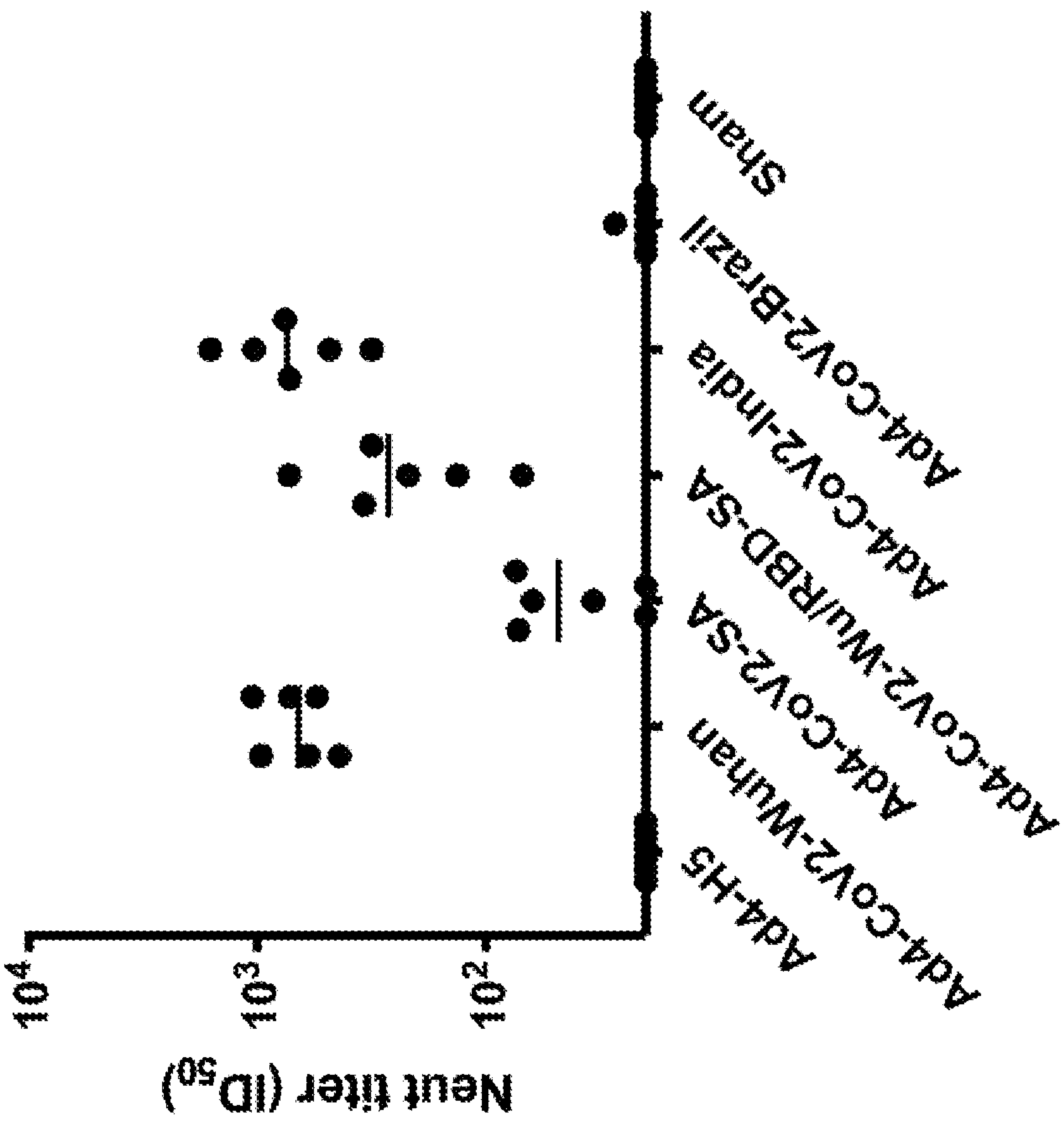


FIG. 6C

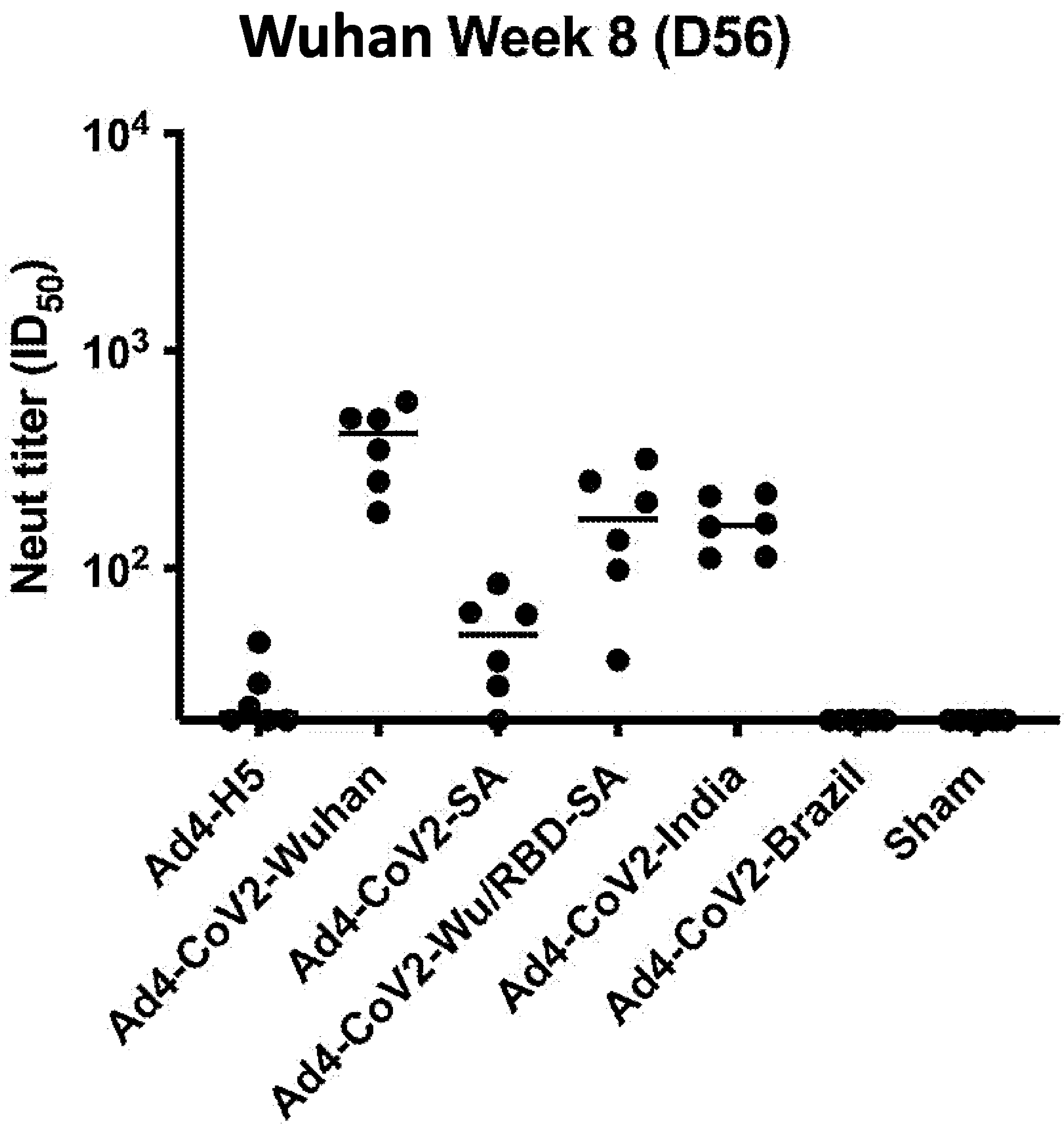

Wuhan Week 8 (D56)
10^4
10^3
10^2
Neut titer (ID_{50})
Ad4-H5
Ad4-CoV2-Wuhan
Ad4-CoV2-SA
Ad4-CoV2-Wu/RBD-SA
Ad4-CoV2-India
Ad4-CoV2-Brazil
Sham

FIG. 6D

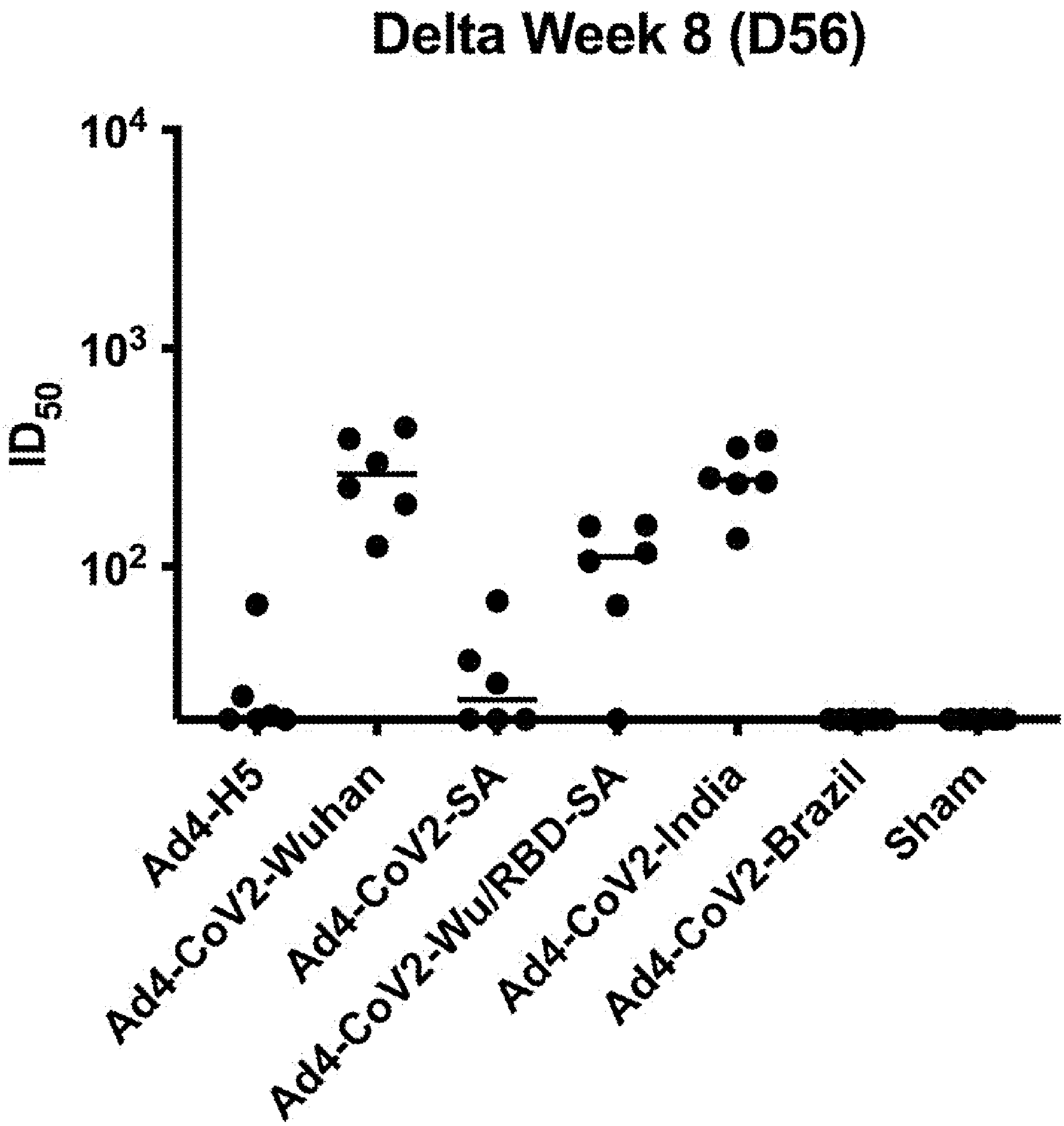
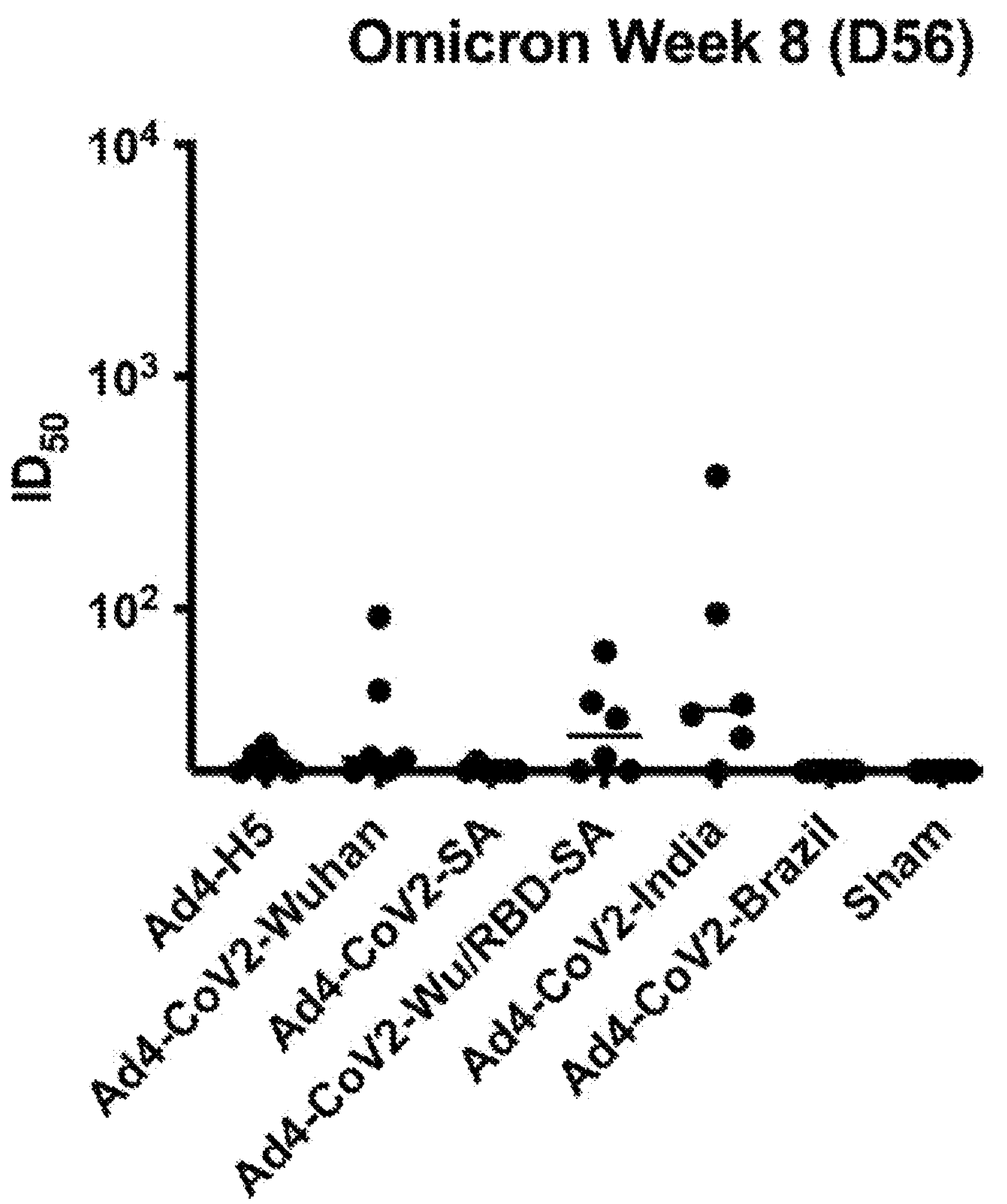

Delta Week 8 (D56)
10^4
10^3
10^2
ID_{50}
Ad4-H5
Ad4-CoV2-Wuhan
Ad4-CoV2-SA
Ad4-CoV2-Wu/RBD-SA
Ad4-CoV2-India
Ad4-CoV2-Brazil
Sham

FIG. 6E


Omicron Week 8 (D56)
10^4
10^3
10^2
ID_{50}
Ad4-H5
Ad4-CoV2-Wuhan
Ad4-CoV2-SA
Ad4-CoV2-Wu/RBD-SA
Ad4-CoV2-India
Ad4-CoV2-Brazil
Sham

SEQUENCE LISTING

<110> THE UNITED STATES OF AMERICA, AS REPRESENTED BY THE
SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES

<120> REPLICATION-COMPETENT ADENOVIRUS TYPE 4 SARS-COV-2 VACCINES AND
THEIR USE

<130> 4239-105641-02

<150> US 63/138,221
<151> 2021-01-15

<160> 19

<170> PatentIn version 3.5

<210> 1
<211> 40459
<212> DNA
<213> Artificial Sequence

<220>
<223> Recombinant vector

<400> 1
taaatttaaa tgaattccgt caagggcgac acaaaaggta ttctaaatgc ataataaata   60

ctgataacat cttatagttt gtattatatt ttgtattatc gttgacatgt ataattttga   120

tatcaaaaac tgattttccc tttattattt tcgagattta ttttcttaat tctctttaac   180

aaactagaaa tattgtatat acaaaaaatc ataaataata gatgaatagt ttaattatag   240

gtgttcatca atcgaaaaag caacgtatct tatttaaagt gcgttgcttt tttctcattt   300

ataaggttaa ataattctca tatatcaagc aaagtgacag gcgcccttaa atattctgac   360

aaatgctctt tccctaaact ccccccataa aaaaacccgc cgaagcgggt ttttacgtta   420

tttgcggatt aacgattact cgttatcaga accgcccagg atgcctggca gttccctact   480

ctcgccgctg cgctcggtcg ttcggctgcg ggacctcagc gctagcggag tgtatactgg   540

cttactatgt tggcactgat gagggtgtca gtgaagtgct tcatgtggca ggagaaaaaa   600

ggctgcaccg gtgcgtcagc agaatatgtg atacaggata tattccgctt cctcgctcac   660

tgactcgcta cgctcggtcg ttcgactgcg gcgagcggaa atggcttacg aacggggcgg   720 agatttcctg gaagatgcca ggaagatact taacagggaa gtgagagggc cgcggcaaag 780

ccgtttttcc ataggctccg cccccctgac aagcatcacg aaatctgacg ctcaaatcag 840

tggtggcgaa acccgacagg actataaaga taccaggcgt ttccccctgg cggctccctc 900

gtgcgctctc ctgttcctgc ctttcggttt accggtgtca ttccgctgtt atggccgcgt 960

ttgtctcatt ccacgcctga cactcagttc cgggtaggca gttcgctcca agctggactg 1020

tatgcacgaa ccccccgttc agtccgaccg ctgcgcctta tccggtaact atcgtcttga 1080

gtccaacccg gaaagacatg caaaagcacc actggcagca gccactggta attgatttag 1140

aggagttagt cttgaagtca tgcgccggtt aaggctaaac tgaaaggaca agttttggtg 1200

actgcgctcc tccaagccag ttacctcggt tcaaagagtt ggtagctcag agaaccttcg 1260

aaaaaccgcc ctgcaaggcg gttttttcgt tttcagagca agagattacg cgcagaccaa 1320

aacgatctca agaagatcat cttattaagc ttagaaaaac tcatcgagca tcaaatgaaa 1380

ttgcaattta ttcatatcag gattatcaat accatatttt tgaaaaagcc gtttctgtaa 1440

tgaaggagaa aactcaccga ggcagttcca taggatggca agatcctggt atcggtctgc 1500

gattccgact cgtccaacat caatacaacc tattaatttc ccctcgtcaa aaataaggtt 1560

atcaagtgag aaatcaccat gagtgacgac tgaatccggt gagaatggca aaagtttatg 1620

catttctttc cagacttgtt caacaggcca gccattacgc tcgtcatcaa aatcactcgc 1680

atcaaccaaa ccgttattca ttcgtgattg cgcctgagcg aggcgaaata cgcgatcgct 1740

gttaaaagga caattacaaa caggaatcga gtgcaaccgg cgcaggaaca ctgccagcgc 1800

atcaacaata ttttcacctg aatcaggata ttcttctaat acctggaacg ctgtttttcc 1860

ggggatcgca gtggtgagta accatgcatc atcaggagta cggataaaat gcttgatggt 1920

cggaagtggc ataaattccg tcagccagtt tagtctgacc atctcatctg taacatcatt 1980

ggcaacgcta cctttgccat gtttcagaaa caactctggc gcatcgggct tcccatacaa 2040

gcgatagatt gtcgcacctg attgcccgac attatcgcga gcccatttat acccatataa 2100

atcagcatcc atgttggaat ttaatcgcgg cctcgacgtt tcccgttgaa tatggctcat 2160

attcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca tgagcggata 2220 catatttgaa tgtatttaga aaaataaaca aataggggtc agtgttacaa ccaattaacc 2280

aattctgaac attatcgcga gcccatttat acctgaatat ggctcataac accccttgtt 2340

tgcctggcgg cagtagcgcg gtggtcccac ctgaccccat gccgaactca gaagtgaaac 2400

gccgtagcgc cgatggtagt gtggggactc cccatgcgag agtagggaac tgccaggcat 2460

caaataaaac gaaaggctca gtcgaaagac tgggcctttc gcccgggcta attagggggt 2520

gtcgccctta tcgctgagga tccatttaaa tttaattaac atcatcaata atataccta 2580

tttttttgt gtgagttaat atgcaaataa ggcgtgaaaa tttggggatg gggcgcgctg 2640

attggctgtg acagcggcgt tcgttagggg cggggcaggt gacgttttga tgacgcgact 2700

atgaggagga gttagtttgc aagttctggt ggggaaaagt gacgtcaaac gaggtgtggt 2760

ttaaacacgg aaatactcaa ttttcccacg ctgtctaaca ggaaatgagg tgtttttggg 2820

cggatgcaag tgaaaacgga ccattttcgc gcgaaaactg aatgaggaag tgaaatctga 2880

gtaatttagt gtttatgaca gggaggagta tttgccgagg gccgagtaga ctttgaccgt 2940

ttacgtgggg gtttcgatta ccgtgttttt cacctaaagt tccgcgtacg gtgtcaaagt 3000

ccggtgtttt tacgtaggtg tcagctgatc gtcagggtat ttaaacctgc gctctgcagt 3060

caagaggcca ctcttgagtg ccagcgagaa gagttttctc ctccgcgccg cgagtcagat 3120

ctacactttg aaatatgagg cacctaagag acctgcccga tgaggaaatt atcatcgctt 3180

ccgggagcga gattctggaa ctggtggtaa atgctatgat gggcgacgac catccggaac 3240

cccccacccc atttgagaca ccttcgctgc acgatttgta tgatctggag gtggatgtgc 3300

ccgaggacga ccccaacgag aaggcggtaa atgatttatt tagcgatgcc gcgctgctag 3360

ctgccgagga ggcttcaagc cctagctcag acagcgactc ttcactgcat accctagac 3420

acgacagagg tgagaaagag atccccgggc ttaaatggga aaagatggac ttgcgttgct 3480

atgaggaatg cctgccccca agcgatgatg aggacgagca ggcgattcag aacgcagcga 3540

gccatggagt gcaagccgtc agcgagagct ttgcactgga ctgcccgcct ttgcccggac 3600

acggctgtaa gtcttgtgaa tttcatcgca tcaatactgg agataaagct gtgttatgtg 3660

cactttgcta tatgagagcg tacaaccatt gtgtttacag taagtgtgat taagtgaact 3720 ttaaagggag gcaaagagta gggtgactgg gtgatgactg gtttatttat gtatatctgt 3780

tttttatata ggtcccgttt ctgacgcaga tgatgagacc cccactacag agtccacttt 3840

gtcacccccct gaaattggca cgtctccatc tgacaatatt gttagaccag ttcctgtaag 3900

agccactggg aggagagcag ctgtagaatg tttggatgat ttgcttcagg gtggagatga 3960

acctttggac ttgtgtaccc ggaaacgccc caggcattaa gtgccacaca tgtgtgttta 4020

cttgaggtga tgtcagtatt tatagggtgt ggagtgcaat aaaatatgtg ttgactttaa 4080

gtgcgtggtt tatgactcag ggaggggac tttgggtata taagcaggtg cagacctgtg 4140

tggttagctc agagcggtat ggagatttgg acggttttgg aagactttca caagactagg 4200

cagctgctag agaacgcctc gaacggagtc tcttacctgt ggagattctg cttcggcggt 4260

gacctagcta agctagtcta tagggccaaa caggattata gggaacaatt tgaggatatt 4320

ttgagagagt gtcctggtct tttgacgct cttaacttgg gccatcagtc tcactttaac 4380

cagagaattt caagagccct tgactttact actcctggca gaaccactgc agcagtagcc 4440

ttttttgctt ttatttttga caaatggagt caagaaaccc atttcagcag ggattaccag 4500

ctggatttct tagcagtagc tttgtggaga acatggaagt gccagcgcct gaatgcaatc 4560

tccggctact tgccggtaca gccgctagac actctgagga tcctgagtct ccagcagcag 4620

gaggatcaag aagagaatcc gagagccggc ctggaccctc cggcggagga gtagctgacc 4680

tgtttcctga actgcaccgg gtgctgacta ggtcttcgag tggtcgggag aggggtatta 4740

agcgggagag gcatgatgag actaatcaca gaattgaact gactgtgggt ctgatgagcc 4800

gcaagcgtcc agaaacagtg tggtggtatg aggtgcagtc aactggcaca gatgaggtgt 4860

cagtcatgca tgagagattt tccctagaac aagtcaagac ttgttggttg gagcctgagg 4920

atgattggga ggtagccatc aggaattatg ccaagctggc tctgaggcca gatagaaagt 4980

acaagattac taagctgata aatatcagaa atgcctgcta catctcaggg aatggggctg 5040

aagtggagat ctgtctccag gatagagtgg ctttcagatg ctgcatgatg aatatgtacc 5100

cgggagtggt ggacatggat ggggtcacct ttatgaacat gaggttcagg ggagatgggt 5160

ataatgggac ggtctttatg gccaatacca agctgacagt gcatggatgc tccttctttg 5220 ggtttaataa cacctgcatc gaggcttggg gtcaggtcgg tgttaagggg tgcagttttt 5280

cagccaactg gatgggggta gtgggcagga ccaagagtat gctgtctgtg aagaaatgct 5340

tgtttgagag gtgccacctg ggggtgatga gcgagggcga agccagaatc cgccactgtg 5400

cctctaccga gacgggctgt tttgtgctgt gcaagggcaa tgccaagatc aagcataata 5460

tgatctgtgg agcctcggac gagcgcggct accagatgct gacctgcgcc ggtgggaaca 5520

gtcatatgct ggccgccgtg catgtggctt cccattcccg caagccctgg cctgagttcg 5580

agcacaatgt catgaccagg tgcaatatgc atctgggggc tcgccgaggc atgtttatgc 5640

cctaccagtg caacctgaat tatgtaaagg tgctcctgga gcccgatgtc atgtccagag 5700

tgagcctgac gggggtgttt gacatgaatg tggaagtgtg gaagattcta agatatgatg 5760

aatacaagac caggtgtcga gcctgcgagt gcggagggaa gcatgccagg ttccagcccg 5820

tgtgtgtgga tgtgacggag gacctgcgac ccgatcattt ggtgttgtcc tgcaccggga 5880

cggagttcgg ctccagtggg gaagaatctg actagagtga gtagtgtttt ggggagggag 5940

aggacctgca taaggggcag aatgattaaa atctgtgctt ttctgtgtgt tgcagcagca 6000

tgagcggaaa cggctccttt gagggagggg tattcagccc ttatctgacg gggcgtctcc 6060

cctcctgggc gggagtgcgt caaaatgtga tgggatccac ggtggacggc cggcccgtac 6120

agcccgcgaa ctcttcaacc ctgacctatg caaccctgag ctcctcgtcg gtggacgcag 6180

ctgccgccgc agctgctgct tctgccgcca gcgccgtgcg cggaatggcc atgggcgccg 6240

gctattacgg cactctggtg gccaactcga gttccactaa taatcccgcc agcctgaacg 6300

aggagaagct gctgctgttg atggcccagc tcgaggcctt gacccagcgc ctgggcgagc 6360

tgacccagca ggtggctcag ctgcaggagc agacgcgggc cgcggttgcc acggtgaaat 6420

ccaaataaaa aatgaatcaa taaataaacg gagacggttg ttgattttaa aaatcagagt 6480

ctgaatcttt atttgatttt tcgcgcacgg taggccctgg accaccggcc tcgatcattg 6540

agcacccggt ggatcttttc caagacccgg tagaggtggg attggatatt gaggtacatg 6600

ggcatgagcc cgtcccgggg gtgaaggtag ctccattgca gggcctcgtg ctcgggggtg 6660

gtgttgtaaa tcacccagtc atagcaggga cgcagggcgt ggtgttgcac aatatctttg 6720 aggaggagac tgatggccac gggcagccct ttggtgtagg tgtttacaaa cctgttgagc 6780

tgggagggat gcatgcgggg ggagatgagg tgcatcttag cctggatctt cagattggcg 6840

atgttaccgc ccagatcccg cctgggattc atgttgtgca ggaccaccag cacggtgtat 6900

ccggtgcact tggggaattt gtcatgcaac ttggaaggga aggcatgaaa gaatttggag 6960

acgcccttgt ggccgcccag gttttccatg cactcatcca tgataatggc tatgggcccg 7020

tgggcggcgg cttgggcaaa gacgtttcgg gggtcggaca catcgtagtt gtggtcctgg 7080

gtgagatctt cataggccat tttaatgaat ttggggcgga gggtgcccga ttgggggacg 7140

aaggtaccct caatcccggg ggcgtagttt ccctcacaga tctgcatctc ccaggcctta 7200

agctccgagg gggggatcat gtccacctgc ggggcgataa agaaaacggt ttccggggcg 7260

ggggagatga gctgggcgga aagcaggttg cggagtagct gggacttacc gcagccggtg 7320

gggccgtaga taaccccaat gaccggctgc aggtggtagt tgagggagac acagctgccg 7380

tcctccctaa gaagggggc cacctcgttc atcatttggc gcacgtgcat gttctcgcgc 7440

accagttccg ccaggagtcg ctctccgccc agcgagagga gctcctggag cgaggcgaag 7500

tttttcagcg gcttgagccc gtcggccatg ggcattttgg aaagggtctg ttgcaggagt 7560

tccaagcggt cccagagctc ggtgatgtgc tctacggcat ctcgatccag cagacctcct 7620

cgtttcgcgg gttggggcga ctgcgggagt agggcgccag acgatgggcg tccagcgcgg 7680

ccagggtccg gtccttccag ggtcgcagcg tccgcgtcag ggtggtctcc gtcacggtaa 7740

aggggtgcgc gccgggctgg gcgcttgcga gggtgcgctt caggctcatc cggctggtcg 7800

agaaccgctc ccgatcggcg ccctgtgcgt cggccaggta gcaattgacc atgagttcgt 7860

agttgagcgc ctcggccgcg tggcctttgg cgcggagctt acctttggaa gtctgcccac 7920

aggcgggaca gaggagggac ttgagggcgt agagcttggg ggcgaggaag acggactcgg 7980

gggcgtaggc gtccgcgccg cagtgggcgc agacggtctc gcactccacg agccaggtga 8040

ggtcgggctg attgggatca aaaaccagtt ttccgccgtt ctttttgatg cgtttcttac 8100

ctctggtctc catgagctcg tgtccccgct gggtgacaaa gaggctgtcc gtgtccccgt 8160

aaaccgactt tatgggtcgg tcctcgagtg ggacgccgcg gtcctcgtcg tagaggaaac 8220

ccgaccactc tgagacgaag gcccgggtcc aagccagcac gaaggaggcc acgtgggagg 8280
gatagcggtc gttatccacc agcgggtcca ccttctccag tgtatgcaaa cacatgtccc 8340
cctcgtccac atccaggaag gtgattggct tgtaagtgta ggccacgtga ccgggggtcc 8400
cggccggggg ggtataaaag ggggcgggcc gctgctcgtc ttcactgtct tccggatcgc 8460
tgtccaggag cgccagctgt tggggtaggt attccctctc aaaggcgggc atgacctccg 8520
cactcaggtt gtcagtttct agaaacgagg aggatttgat attgacggtg ccggcggaga 8580
tgcctttcaa gagcccctcg tccatctggt cagaaaagac aatctttttg ttgtcgagtt 8640
tggtggcgaa ggagccgtag agggcgttgg agaggagctt ggcgatggag cgcatggtct 8700
ggttcttttc cttgtcggcg cgctccttgg cggcgatgtt gagctgcacg tactcgcgcg 8760
ccacgcactt ccattcgggg aagacggtgg ttagctcgtc tggcacgatt ctgacctgcc 8820
agccccggtt atgcagggtg atgaggtcaa cgctggtggc cacctcgccg cgcaggggct 8880
cgttggtcca gcagaggcgg ccgcccttgc gcgagcagaa gggggggcagg gggtccagca 8940
taagctcgtc gggggggtca gcatcgatgg tgaagatgcc tggcaggagg tcggggtcga 9000
agtagcttat gcaggtgccc agatcgtcca gagaagcttg ccattcgcgc acggccagcg 9060
cgcgctcgta gggactaagg ggcgtgcccc agggcatggg gtgggtgagc gcggaggcgt 9120
acatgccgca gatgtcgtag acgtagaggg gctcatcaag gatgccaatg taggtggggt 9180
agcagcggcc cccgcggatg ctggcgcgca cgtagtcata caactcgtgc gagggggcga 9240
ggagcccggc tccgagattg gcgcggctgg gtttttcggc gcggtagacg atctgacgga 9300
agatggcgtg ggagttggag gagatggtgg gtctttggaa gatgttgaag tgggcgtggg 9360
gcaggccgac cgagtcgcgg atgaagtggg cgtaggagtc ttgcagcttg gcgacaagct 9420
cggcggtgac gaggacgtcc agggcgcagt agtcaagggt ctcttggatg atgtcatact 9480
tgagctggcc cttttgtttc cacagctcgc ggttgagaag gaactcttcg cggtccttcc 9540
agtactcttc aagggggaac ccgtcctggt cggcacggta agagcctagc atgtagaact 9600
ggttaacggc cttgtaggcg cagcagccct tctccacggg gagggcatag gcctgggcgg 9660
ccttgcgcag ggaggtgtgc gtgagggcga aggtgtccct gaccatgacc tttaggaact 9720

ggtgcttgaa gtcgatatcg tcgcagcccc cctgctccca gagctggaag tccgtgcgct 9780

tcttgtaggc ggggttgggc aaagcgaaag taacatcgtt gaagaggatc ttgcccgcgc 9840

ggggcataaa gttgcgagtg atgcggaaag gctggggcac ctcggcccgg ttgttgatga 9900

cctgggcggc gagcacgatc tcgtcgaagc cgttaatgtt gtggcccaca atgtatagtt 9960

ccacgaaccg cgggcggccc ttgacgtggg gcagtttctt gagctcctcg taggtgagct 10020

cgtcggggtc gctgagcccg tgctgctcga gggcccagtc ggcgagatgg gggttggcgc 10080

ggaggaagga agtccagaga tccacggcca gggcggtttg cagacgatcc cggtactggc 10140

ggaactgctg acccacggcc atttttcgg gggtgacgca gtagaaggtg cgggggtcgc 10200

cgtgccaacg gtcccatttt agctggaggg cgagatcaag ggcgagctca acgagccggt 10260

cgtccccgga gagtttcatg accagcatga aggggacgag ctgcttgccg aaggacccca 10320

tccaggtgta ggtttccaca tcgtaggtga ggaagagcct ttcggtgcga ggatgcgagc 10380

cgatggggaa gaactggatc tcctgccacc agttggagga atggctgttg atgtgatgga 10440

agtagaaatg ccgacggcgc gccgaacatt cgtgcttgtg tttatacaag cggccacagt 10500

gctcgcaacg ctgcacggga tgcacgtgct gcacgagctg tacctgggtt cctttgacga 10560

ggaatttcag tgggaagtgg agtcgtggcg cctgcatctg gtgctgtact acgtcgtggt 10620

ggtcggcctg gccctcttct gcctcgatgg tggtcatgct gacgagcccg cgcgggaggc 10680

aggtccagac ctcggcgcga acgggtcgga gagcgaggac gagggcgcgc aggccggagc 10740

tgtccagggt cctgagacgc tgcggagtca ggtcagtggg cagcggcggc gcgcggttga 10800

cttgcaggag tttttcaagg gcgcgcggga ggtccagatg gtacttgatc tccaccgcgc 10860

cgttggtggc gacgtcgatg gcttgcagtg tcccgtgccc ctggggagtg accaccgtcc 10920

cccgtttctt cttggcgggc ggaagcggtt tggcttccat ggttaaaagc ggcggcgagg 10980

acgcgcgccg ggcggtaggg gcggctcggg acccggaggc agtggtggca ggggcacgtc 11040

ggcgccgcgc gcgggcaggt tctggtactg cgcccggaga agactggcgt gagcgacgac 11100

gcgacggttg acgtcctgga tctgacgcct ctgggtgaag gccacgggac ccgtgagttt 11160

gaacctgaaa gacagttcga cagaatcaat ctcggtatca ttgacggcgg cctgccgcag 11220 aatctcttgc acgtcgcccg agttgtcctg gtaggcaatc tcggtcatga actgctcgat 11280

ctcctcctcc tgaaggtctc cgcggccggc gcgctccacg gtggccgcga ggtcgttgga 11340

gatgcggccc atgagctgcg agaaggcgtt catgcccgcc tcgttccaga cgcggctgta 11400

aaccacggcg ccctcgggat cgcgggcgcg catgaccacc tgggcgaggt tgagctccac 11460

gtggcgcgca aaaaccgcgt agttgcagag gcgctggtag aggtagttga gcgtggtggc 11520

aatgtgctca gtgacaaaga agtacataat ccagcggcgg agcggcattt cgctgacgtc 11580

gcccagggct tccaagcgct ccatggcctc gtaaaagtcc acggcgaagt tgaaaaactg 11640

ggagttgcgt gcagatacgg tcaagtcctc ctccagaaga cggatgagct cggcgatggt 11700

ggcgcgcacc tcgcgctcga aggctcccgt gagttcctcc acttcctcct cttcatccac 11760

taacatctct tctacttcct cctcaggcgg tggtggcggg ggagggggcc tgcgtcgccg 11820

gcggcgcacg ggcagacggt cgatgaaacg ctcgatggtc tcgccgcgcc ggcgtcgcat 11880

ggtctcggtg acggcgcgcc cgtcctcgcg gggtcgcagc gtaaagacgc cgccgcgcat 11940

ctccaggtgg cccgggggt ccccgttggg cagggagagt gcgctgacga tgcatcttat 12000

caattgcccc gtagggactc cgcgcaagga cctaagcgtc tctagatcca cgggatctga 12060

aaaccgttga acgaaggctt cgagccagtc gcagtcgcaa ggtaggctga gcacggtttc 12120

ttctggcggc ggtggggtgt gggcgggggc gatgctgctg gtgatgaagt tgaaataggc 12180

ggttctgaga cggcggatgg tggcgaggag caccaggtct ttgggcccgg cttgctggat 12240

gcgcagacgg tcggccatgc cccaggcgtg gtcctgacac ctggccaggt ccttgtagta 12300

gtcctgcatg agccgctcca cgggcacctc ctcctcgccc gcgcggccgt gcatacgcgt 12360

gagcccaaac ccgcgctgcg gctggacgag cgccaggtca gcgacgacgc gctcggcgag 12420

gatggcctgc tggatctggg tgagggtggt ctggaagtcg tcaaagtcga cgaagcggtg 12480

gtaggctccg gtgttaatgg tgtaggagca gttggccatg acggaccagt tgacagtctg 12540

gtgaccgggc cgcgcgagct cgtggtactt gaggcgcgag taggcgcgcg agtcgaagat 12600

gtagtcgttg caggtgcgca ccaggtactg gtagccgatg aggaagtgcg gcggcggctg 12660

gcggtagagc ggccatcgct cggtggcggg ggcgccgggc gctaggtcct cgagcatggt 12720 gcggtggtag ccgtagatgt accttgacat ccaggtgatg ccggcggcgg tggtggaggc 12780

gcgagggaac tcgcggacgc ggttccagat gttgcgcagc ggcaggaagt agttcatggt 12840

gggcacggtc tggcccgtga ggcgcgcgca gtcgttgatg ctctagacat acgggcaaaa 12900

acgaaagcgg tcagcggctc gactccgtgg cctggaggct aagcgaacgg gttgggctgc 12960

gcgtgtaccc cggttcgaat ctcgaatcag gctggagccg cagctaacgt ggtactggca 13020

ctcccgtctc gacccaggcc tgcacaaaac ctccaggata cggaggcggg tcgttttgca 13080

aatttttggc ggtcgaaaaa agctagtaag cgcggaaagc ggccgaccgc aatggctcac 13140

tgccgtagat tggagaagaa tcgccagggt tgcgttgcgg tgtgccccgg ttcgagaccg 13200

ctcgggtcgg ccgaattccg cggctaacga gggcgtggct gccccgtcgt ttccaagacc 13260

ccataagcca gccgacttct ccagttacgg agcgagcccc tcttttgttt tgtttttgc 13320

cagatgcatc ccgtactgcg gcagatgcgc ccccaccctc caccgcaaca gcagccccct 13380

cctacgcaac agccggcgct tctgcctccg ccccagcagc agcaacttcc agccactacc 13440

gccgcggccg ccgtgagcgg ggccgggcag agtcagtatg acctggcttt ggaagagggc 13500

gaggggctgg cgcgcctggg ggcgtcgtcg ccggagcggc acccgcgcgt gcagatgaaa 13560

agggacgctc gcgaggccta cgtgcccaag cagaacctgt tcagagacag gagcggcgag 13620

gagcccgagg agatgcgcgc agcccgtttc cacgcggggc gggagctgcg gcgcggcctg 13680

gacagaaaga gggtgctgag ggacgaggat ttcgaggcgg acgagctgac ggggatcagc 13740

cctgcgcgcg cgcacgtggc cgcggccaac ctggtcacgg cgtacgagca gaccgtgaag 13800

gaggagagca acttccaaaa atccttcaac aaccacgtgc gcaccctgat cgcgcgcgag 13860

gaggtgaccc tgggcctgat gcacctgtgg gacctgctgg aggccattgt gcagaacccc 13920

accagcaaac cgctgacggc gcagctgttc ctggtggtgc agcacagtcg ggacaacgag 13980

acttttaggg aggcgctgct gaatatcacc gagcccgagg gccgctggct tctggacctg 14040

gtgaatattc tgcagagcat cgtggtgcag gagcgcgggc tgccgctgtc cgagaagctg 14100

gcggccatca acttttcggt gctgagtttg ggcaagtact acgctaggaa gatctacaag 14160

accccgtacg tgcccataga caaggaggtg aagatcgacg ggttttacat gcgcatgacc 14220 ctgaaagtgc tgaccctgag cgacgatctg ggggtgtacc gcaacgacag gatgcgccgc 14280

gcggtaagcg ccagcaggcg gcgcgagctg agcgatcagg agctgatgca cagcctgcag 14340

cgggccctga ccggggccgg gaccgagggg gagagctact ttgacatggg cgcggacctg 14400

cactggcagc ccagccgccg ggtcttggaa gccgcggcgg tcccttacgt agaagaggtg 14460

gacgatgagg atgagggcga gtacctggaa gactgatggc gcgaccgtat ttttgctaga 14520

tgcagcaaca gccacctcct gatcccgcaa tgcgggcggc gctgcagagc cagccgtccg 14580

gcattaactc ctcggacgat tggacccagg ccatgcaacg catcatggcg ctgacgaccc 14640

gcaaccccga agcctttaga cagcagcccc aggccaaccg gctctcggcc atcctggagg 14700

ccgtggtgcc ctcgcgctcc aaccccacgc acgagaaggt gctggccatc gtgaacgcgc 14760

tggttgagaa caaggccatt cgcggcgacg aggccgggct ggtgtacaac gcactgctgg 14820

agcgcgtggc ccgctacaac agcaccaacg tgcagaccaa cctggaccgc atggtgaccg 14880

acgtgcgcga agccgtggcc cagcgcgaac ggttccaccg cgagtccaac ctgggatcca 14940

tggtggcact gaacgccttc ctcagcacgc agcccgccaa cgtgccccgg ggccaggagg 15000

actacaccaa cttcattagc gccctgcggc taatggtgac cgaggtgccc cagagcgagg 15060

tgtaccagtc gggcccggac tacttcttcc agaccagtcg ccagggcttg cagaccgtga 15120

acctgagtca ggctttcaag aacttgcagg gactgtgggg cgtgcaggct ccggtcgggg 15180

accgcgcgac ggtgtcgagc ctgctgacgc cgaactcgcg cctgctgctg ctgctggtgg 15240

cgcccttcac ggacagcggt agtatcaacc gcaactcgta cctgggctac ctgattaacc 15300

tgtaccgcga ggccattggc caggcgcacg tggacgagca gacctaccag gagattaccc 15360

acgtgagccg cgcccttggc caggacgacc cgggcaatct ggaagccacc ctgaacttct 15420

tgctgaccaa ccggtcgcag aagatcccgc cccagtacgc gctgagcgcc gaggaggagc 15480

gtatattgag atacgtgcag caaagtgtgg gactgttcct gatgcaggag gggccaccc 15540

ccagcgccgc gctcgacatg accgcgcgca acatggagcc cagcatgtac gccagtaatc 15600

gcccgtttat taataagctg atggactacc tgcatcgggc ggccgccatg aactctgact 15660

atttcaccaa cgccatcctg aacccccact ggctcccgcc gccggggttc tacacgggcg 15720 agtacgacat gcccgacccc aatgacgggt ttctgtggga cgacgtggac agcagcgtgt 15780

tctccccccg accgggtgct aacgagcgcc ccttgtggaa gaaagagggc agcgaccggc 15840

gcccgtcctc ggcgctgtcc ggccgcacgg gtgctgccgc agcggtgccc gaggccgcca 15900

gtccctttcc gagcttgtca ctgaacagcg tccgcagtag cgagctgggc aggatcacgc 15960

gcccgcgctt gctgggcgag gaggagtact taaataactc gctgttgagg cccgagcggg 16020

agaagaactt ccccaataac gggatagaga gtctggtgga taagatgagc cgctggaaga 16080

cgtacgcgca tgagcacagg gacgatcccc gggcaacgca gggggccacc agccggggca 16140

gtgccgcccg taaacgccgc tggcacgaca ggcagcgggg actgatgtgg gacgatgagg 16200

attccgccga cgacagcagc gtgttggact tgggcgggag tggtggtggt aacccgttcg 16260

ctcacctgcg cccccgcgtc gggcgcctga tgtaaaaaga aaccaaaaat aaatggtact 16320

caccaaggcc atggcgacca gcgtgcgttc gtttcttctc tgttgtatct agtatgatga 16380

ggcgtgcgta cccggagggt cctcctccct cgtacgagag cgtgatgcag caggcaatgg 16440

cggcggcggc ggcgatgcag cccccgctgg aggctcctta cgtgccaccg cggtacctgg 16500

cgcctacgga ggggcgaaac agcattcgtt actcggagct ggcacccttg tacgatacca 16560

cccggttgta cctggtggac aacaagtcgg cggacatcgc ctcgctgaac taccagaacg 16620

accacagcaa ctttctgacc accgtggtgc agaacaacga tttcaccccc acggaggcca 16680

gcacccagac catcaacttt gacgagcgct cgcggtgggg cggtcagctg aaaaccatca 16740

tgcataccaa catgcccaac gtgaacgagt tcatgtacag caacaagttc aaggcgcggg 16800

tcatggtctc ccgcaagacc cccaacgggg tgacagtagg ggatgattat gatggtagtc 16860

aggatgagct gaaatacgag tgggtggagt ttgagctgcc cgaaggcaac ttctcggtga 16920

ccatgaccat tgacctgatg aacaacgcca tcatcgacaa ttacttggca gtggggcggc 16980

agaacggggt gctggagagc gacatcggcg tgaagttcga cacccggaac ttcaggctgg 17040

gttgggaccc cgtgaccgag ctggtcatgc ccggggtgta caccaacgag gccttccacc 17100

ccgacatcgt gctgttgccc ggctgcgggg tggactttac cgagagccgc ctcagtaata 17160

tgctgggcat ccgcaagagg cagcccttcc aggagggttt ccagatcatg tacgaggacc 17220

tggatggagg taacatcccc gcgctcttgg atgtcgaggc ctatgagaaa agcaaggagg 17280
agagcgtcgc cgcgtcaacc gcagccgtag ccaccgcctc taccgaggtc cggggcgata 17340
attttgctag cgccgcagca gtggcggcgg ccaaggctga tgaaaccgaa agtaagatag 17400
ttattcagcc ggtggagaag gatagcaagg ataggagcta caacgtgctc tcggacaaga 17460
aaaacaccgc ctaccgcagc tggtacctgg cctacaacta tggcgaccac gagaagggcg 17520
tgcgctcctg gacgctgctc accacctcgg acgtcacctg cggcgtggag caagtctact 17580
ggtcgctgcc cgacatgatg caagacccgg tcaccttccg ctccacgcgt caagttagca 17640
actacccggt ggtgggcgcc gagctcatgc ccgtctactc caagagcttc ttcaacgagc 17700
aggccgtcta ctcgcagcag ctgcgcgcct tcacctcgct cacgcacgtc ttcaaccgct 17760
tccctgagaa ccagatcctc gtccgcccgc ccgcgcccac cattaccacc gtcagtgaaa 17820
acgttcctgc tctcacagat cacgggaccc tgccgctgcg cagcagtatc cggggagtcc 17880
agcgcgtgac cgttactgac gccagacgcc gcacctgccc ctacgtctac aaggccctgg 17940
gcatagtcgc gccgcgcgtc ctctcgagcc gcaccttcta aaaaatgtcc attctcatct 18000
cgcccagtaa taacaccggt tggggtctgc gcgcgcccag caagatgtac ggaggcgctc 18060
gccaacgctc cacgcaacac cccgtgcgcg tgcgcgggca cttccgcgct ccctggggcg 18120
ccctcaaggg ccgcgtgcgg tcgcgcacca ccgtcgacga cgtgatcgac caggtggtgg 18180
ccgacgctcg caactacacc cccgccgccg cgcccgtctc caccgtggac gccgtcattg 18240
acagcgtggt gtccgacgcg cgccggtacg cccgcgccaa gagccggcgg cggcgcatcg 18300
cccggcggca ccgtagcacc accgccatgc gtgcggcgcg agccttgctg cgcagggcca 18360
ggcgcacggg acgcagggcc atgctcaggg cggccagacg cgcggcttca ggcgccagcg 18420
ccggcaggac tcggagacgc gcggccacgg cggcggcagc ggccatagcc agcatgtccc 18480
gcccgcggcg agggaacgtg tactgggtgc gcgacgccgc caccggtgtg cgcgtgcccg 18540
tgcgcacccg ccccccctcgc acttgaagat gttcacttcg cgatgttgat gtgtcccagc 18600
ggcgaggaga aggatgtcca agcgcaaatt caaggaagag atgctccagg tcatcgcgcc 18660
tgagatctac ggccccgcgg cggcggtgaa ggatgaaaga aatccccgca aaatcaagcg 18720

ggtcaaaaag gacaaaaagg aagaagatga tgtggacgat atggtagagt ttgtgcgcga 18780

gtttgccccc cggaggcgcg tgcagtggcg cgggcggaaa gtgcgtccgg tgctgagacc 18840

cggcaccacg gtggttttcg cgcctggcga gcggtccggc acgacatcca agcgctccta 18900

cgatgaggtg tacggggacg aggatattct cgagcaggcg gccgagcgcc tgggcgagtt 18960

tgcttacggc aagcgcaacc gccttgcgcc cctgaaggaa gaggtggtgt ccatcccgct 19020

ggaccacggc aaccccacgc cgagtcttaa gcccgtgacc ctgcagcagg tgctgccgag 19080

cgcggcgccg cgtcggggct tgaagcgcga gggcgaggat gtgtacccca ccatgcagct 19140

gatggtgccc aagcgccaga agctggaaga cgtgctggag accatgaagg tggacccgga 19200

cgtgcagccc gaggtcaagg tgaggcccat caagcaggtg gccccgggcc ttggcgtgca 19260

gaccgtggac atcaagatcc ccacggagcc catggaaacg cagaccgagg tcgtgaagcc 19320

catcaccagc accatggagg tgcagacgga tccttggatg ccggcggcgc cccgaaaacc 19380

ccggcgcaag tacggcgcgg ccagcctgct gatgcccaac tacgcgctgc atccttccat 19440

catccccacg ccgggctacc gcggcacgcg cttctaccac ggctataccg gctcccgccg 19500

ccgcaagacc accacccgcc gccgtcgtcg ccgcacagct gcaactcccg ctgccgccct 19560

ggtgcggaga gtgtaccgcc gcggccgcgc gcctctgacc ctgccgcggg cgcgctacca 19620

cccgagcatt accatttaac tttgccgtcg cctttgcaga tatggctctc acatgccgca 19680

ttcgcgtccc cattacgggc taccgaggaa gaaaaccgcg ccgtagaagg ctggcgggaa 19740

gcgggatgcg ccgccacccc caccggcggc ggcgcgccat cagcaagcgg ttggggggag 19800

gcttcctgcc cgcgctgatc cccatcatcg ccgcggcgat cggggcgatc cccggcattg 19860

cttccgtggc ggtgcaggcc tctcagcgcc actgagacac acacttggaa attgtaataa 19920

acccgaatgg actctgacgc tcctggtcct gtgatgtgtt tttgtagaca gatggaagac 19980

atcaattttt cgtccctggc tccgcgacac ggcacgcggc cgtttatggg cacctggagc 20040

gacatcggca ccagccaact gaacgggggc gccttcaatt ggagcagtct ctggagcggg 20100

cttaagaatt ttgggtccac gcttaaaacc tatggcagca aggcgtggaa cagcaccaca 20160

gggcaggcgc tgagagataa gctgaaagag cagaacttcc agcagaaggt agtcgatggc 20220 ctcgcctcag gcatcaacgg ggtggtggac ctggccaatc aggccgtgca gcggcagatc 20280

aacagccgcc tggacccggt tccccccgcc ggctccgtgg agatgccgca ggtggaggag 20340

gagctgcctc ccctggacaa gcggggcgac aagcgtcccc gtcccgacgc ggaggagacg 20400

ctgctgacgc acacggacga accgcccccg tacgaggagg cggtgaaact gggcctgccc 20460

accacgcgtc ccattgcgcc tctagctacc ggggtgctga aacccgagag tagtaagccc 20520

gcgaccttgg acttgcctcc tccgcccact ccccgcccct ccacagtggc taagcccctg 20580

ccgccggtgg ccgtggcccg cgcgcgaccg ggggctcgcc ctcaggcgaa ctggcagagc 20640

actctgaaca gcatcgtggg tctgggagtg cagagtgtga agcgccgccg ctgttattaa 20700

aaaacactgt agcgcttaac ttgcttgtct gtgtatatgt gtatgtccgc cgccgctgct 20760

gtccagaagg aggagtgaag agaaaggcgc gtcgtcgagt tgcaagatgg ccaccccatc 20820

gatgctgccc cagtgggcgt acatgcacat cgccggacag gacgcttcgg agtacctgag 20880

tccgggtctg gtgcagttcg cccgcgccac agacacctac ttcagtctgg ggaacaagtt 20940

taggaacccc acggtggcgc ctacccacga tgtgaccacc gaccgcagcc agcggctgac 21000

gctgcgcttt gtgcccgtgg accgggagga caacacctac tcgtacaaag tgcgctacac 21060

gctggccgtg ggcgacaacc gcgtgctgga catggccagc acctactttg acatccgcgg 21120

cgtgctggat cggggcccta gcttcaaacc ctactccggc actgcctaca acagcctggc 21180

tcccaaggga gcgcccaaca cctgccagtg gaaggattct gacagcaaaa tgcatacctt 21240

tggggcagct gccatgcccg gtgttactgg gaaaaagata gaagctgatg ggctgcctat 21300

tagaatagat tcaacttctg gaactgacac agtaatttat gctgataaaa ctttccaacc 21360

agaaccacaa gttggaaatg acagttgggt tgacaccaat ggtgcagagg aaaaatatgg 21420

aggcagagct ctaaaggaca ctacaaaaat gaaaccctgt tatggttcat tcgccaagcc 21480

taccaacaaa gaaggtggtc aggctaactt aaaagattca gaacccgccg ccaccactcc 21540

taactatgat atagacctgg ctttctttga cagcaaaact attgttgcta actacgatcc 21600

agatattgta atgtacacag aaaatgttga cttgcagact ccagatactc atattgtata 21660

caaacctgga acagaggaca ccagctctga atccaatttg ggtcagcagg ccatgcctaa 21720 cagacccaac tacattggct tcagagacaa ttttatcggg ctcatgtact acaacagcac 21780

tggcaatatg gggtgctgg ccggtcaggc ctctcagctg aatgctgtgg ttgacttgca 21840

agacagaaac actgaactgt cctaccagct cttgcttgac tctctgggtg acagaacccg 21900

gtatttcagt atgtggaatc aggcggtgga cagctatgat cctgatgtgc gcattattga 21960

aaaccatggt gtggaggatg aattgccaaa ctattgcttt ccgttgaatg gtgtgggatt 22020

gacagacact taccagggtg ttaaagttaa aacagatgca ggttctgaaa agtgggacaa 22080

agatgacacc acagttagta atgctaatga aatccatgta ggcaatcctt ttgccatgga 22140

aatcaacatc caagccaacc tgtggaggaa cttcctctat gccaatgttg ccctctattt 22200

gcctgataaa tacaaataca caccggccaa catcaccctg cccaccaaca ccaacaccta 22260

cgagtacatg aacggccggg tggtggcgcc ctcgctggtg gacgcctaca ttaacattgg 22320

ggcgcgctgg tcgctggacc ccatggacaa cgtaaatccc ttcaaccacc accgcaatgc 22380

gggcttgcgc taccgctcca tgctcctggg caacgggcgc tacgtgccat tccacatcca 22440

ggtgccccag aaatttttg ccattaagag cctcctgctc ctgcccgggt cctacaccta 22500

cgagtggaac ttccgcaagg acgtcaacat gatcctgcag agttcccttg gcaacgacct 22560

gcgcacagac ggggcctcca tcaccttcac cagcattaac ctctacgcca ccttcttccc 22620

catggcgcac aacaccgcct ccacgcttga ggccatgctg cgcaacgaca ccaatgacca 22680

atccttcaac gactacctct cggcggccaa catgctctat cccatcccgg ccaacgccac 22740

caacgtgccc atctccatcc cctcgcgcaa ctgggccgcc tttcgcggct ggtccttcac 22800

gcgtctcaag accaaagaga cgccctcgct gggctccggg ttcgacccct acttcgtcta 22860

ctcgggctcc atcccctacc tcgacggcac cttctacctc aaccacacct tcaagaaggt 22920

ctccatcacc ttcgactctt ccgtcagctg gcccggcaac gaccggctcc tgacgcccaa 22980

cgagttcgaa atcaagcgca ccgtcgacgg cgagggatac aacgtggccc agtgcaacat 23040

gaccaaggac tggttcctgg tccagatgct ggcccactac aacatcggct accagggctt 23100

ctacgtgccc gagggctaca aggaccgcat gtactccttc ttccgcaact tccagcccat 23160

gagccgccag gtggtggacg aggttaacta caaggactac caggccgtca ccctggccta 23220 ccaacacaac aactcgggct tcgttggata cctcgcgccc actatgcgcc agggccagcc 23280

ctaccccgcc aactacccct acccgctcat cggcaagagc gccgttacca gcgtcaccca 23340

gaaaaagttc atctgcgaca gggtcatgtg gcgcatcccc ttctccagca acttcatgtc 23400

catgggcgcg ctcaccgacc tcggccagaa catgctctat gctaactccg cccacgcgct 23460

agacatgaat ttcgaagtcg accccatgga tgagtccacc cttctctatg ttgtcttcga 23520

agtcttcgac gtcgtccgag tgcaccagcc ccaccgcggc gtcattgagg ccgtctacct 23580

gcgcaccccc ttctcagccg gtaacgccac cacataaatt cttgcttctt gcaagaagcc 23640

atggccgcgg gctccggcga gcaggagctc agggccatca tccgcgacct ggggtgcggg 23700

ccctacttcc tgggcacctt cgataagcga ttcccgggat tcatggcccc gcacaaggtg 23760

gcctgcgcca tcgtcaacac ggccggccgc gagaccgggg gcgagcattg gctggccttc 23820

gcctggaacc cgcgctcgaa cacctgctac ctcttcgacc ccttcgggtt ctcggaccag 23880

cgcctcaagc aaatctacca gttcgagtac gagggactgc tgcgccgcag cgccctggcc 23940

accaaggacc gctgcgttac cctggaaaag tccacccaga ccgtgcaggg tccgcgttcg 24000

gccgcctgcg ggcttttctg ctgcatgttc ctacacgcct tcgtgcactg gcccaaccgc 24060

cccatggaca aaaatcccac catgaacttg ctgacggggg tgcccaacgg catgctccag 24120

tcgccccagg tggaacctac cctgcgccgc aaccaggagg cactctaccg cttcctcaac 24180

tcccactctg catactttcg ctctcaccgc gcgcgcattg agaaggccac cgccttcgac 24240

cgcatgaatc aagacatgta acagtgtgtt ttaaaatatg tttaataaac agcacttttt 24300

atgtgacaca tgcatttgag ataattttat tcttaaaaat cgaaggggtt ctgccgggag 24360

gtttcggcat ggcccgcggg cagggacacg ttgcggaact ggtacttggc cagccacttg 24420

aactcgggga tcagcagttt cggcagcagg gtgtcgggga acgagtcggt ccacagcttc 24480

cgcgtcagtt gcagggcgcc cagcaggtcg ggcgcggaga tcttgaaatc gcagttggga 24540

cccgcgtttt gcgcgcgaga gttgcggtac acagggttgc agcactggaa caccatcagg 24600

gccggatgct tcacgctcgc cagcaccgta gcgtcggtga tcccgtccac gtcgaggtct 24660

tcggcgttgg ccatcccgaa gggggtcatc ttgcaggtct gccggcccat ggtgggcacg 24720 cagccgggct tgtggttgca atcgcagtgc agggggatca gcatcatctg ggcctggtcg 24780

gcgttcatcc ccgggtacat ggccttcatg aaagcctcca gctgcttaaa cgcctgctgg 24840

gccttggctc cctcggtgaa gaagaccccg caggacttgc tagaaaactg gttggtagcg 24900

cacccggcgt cgtgcacgca gcagcgcgcg tcgttgttgg ccagctgcac cacgctgcgc 24960

ccccagcggt tctgggtaat cttggcccgg tcggggttct cctttagcgc gcgttgcccg 25020

ttctcgcttg ccacatccat ctcgatcatg tgctccttct ggatcatggt ggtcccgtgc 25080

aggcaccgca gcttgccctc gacttcggta cagccgtgca gccacagcgc gcaccccgtg 25140

ctctcccagt tcttgtgggc gatctgggaa tgcgcatgca cgaacccctg caggaagcgg 25200

cccatcatgg tcgtcagggt cttgttactg gtaaaggtca gcggaatgcc gcggtgctcc 25260

tcgttgatgt acaggtggca gatgcggcga tacacctcgc cctgctcggg catcagttgg 25320

aagttggatt ttaggtcgct ttccacacgg tagcgctcca tcagcatatt catgatttcc 25380

atgcccttct cccaggccga tacaatgggc aggctcaggg ggttcgtcac cgccatctta 25440

gcgctagcag ccttcgtcag cgggtcgttc tcattgagag tctcaaagct ccgcttgccg 25500

tccttctcgg tgatccgcac ggggggggtag ctgaagccca cggccgccag ctcctcctcg 25560

gcctctcttt cgtcctcgct gtcctggctg acgtcctgca ggggcacatg cttcgttttg 25620

cggggtttct ttttgggcgg ctgctgcggc ggcggtggtt gttcctgagg cgagggggag 25680

cgcgagttct cgctcaccac tactatctct tcttcttggt ccgaggccac gcggcggtag 25740

gtatgtctct tcaggggcag aggcggaggc gacgggctct cgcggcccgg cgggtggctg 25800

gcagagcccc ttccgcgatc gggggtgcgc tcccggcggc gctctaactg acttcctccg 25860

cggccggcca ttgtgttctc ctagggaaca acaacaagca tggagactca gccatcgtcg 25920

ccaacctcgc catctgcccc caccgccgac aagaagcagc agcagaatga gagcttaacc 25980

gccccgccgc ccagccccgc cacctttgtc gcggccccag acatgcaaga gatggaggaa 26040

tccattcaga ttgacctggg ctatgtgacg cccgcggagc acgaggagga gcttgcagtg 26100

cgcttttcaa cccaggaaga gatacaccaa gaacagccag agcaggaagc aaagagcgag 26160

catgactacc tccaccagag cggggggggag gacgccctca tcaagcatct ggcccggcag 26220 gccatcatcg tcaaggacgc gctgcttgac cgcaccgagg tgcccctcag cgtggaggag 26280

ctcagccgcg cctacgagct caacctcttc tcgccgcgcg tgccccccaa gcgccagccc 26340

aacggcacct gcgagcccaa cccacgcctc aacttctacc cggtcttcgc ggtgcccgag 26400

gccctggcca cctaccacat ctttttcaag aaccaaagga tccctgtctc ctgtcgcgcc 26460

aaccgcaccc gcgccgactc ccttttcaac ctgggccccg gtgcccgcct acctgatatc 26520

gcctccttgg aagaggttcc caagatcttc gagggtctgg gcagcgacga gactcgggcc 26580

gcaaacgctc tgcaaggaga aggaggagat catgagcacc acagcgccct ggtggagttg 26640

gaaggcgaca acgcgcgtct ggcggtgctc aagcgcacga tcgagctgac ccatttcgcc 26700

tacccggcgc ttaacctgcc ccccaaagtc atgagcacgg ttatggatca ggtgctcatc 26760

aagcgcgcgt cgcccatctc caaggagatg caagaccccg agagctccga ggagggcaag 26820

cccgtggtca gcgacgagca gctggcgcgg tggctgggac cccaagctag tccccagagc 26880

ttggaagagc ggcgcaagct cataatggcc gtggtcctgg tgaccgcgga gctggagtgt 26940

ctgcgccgct tcttcgccga cgcagaaatt ctgcgcaagg tcgaggagaa cctgcactac 27000

atcttcaggc acgggttcgt acgccaggcc tgcaagatct ccaacgtgga gctgaccaac 27060

ctggtctcct acatgggcat cttgcacgag aaccgcctgg ggcagaacgt gctgcacacc 27120

accctgcgcg gggaggcccg ccgcgactac atccgcgact gcgtttacct ctacctctgc 27180

cacacctggc agacagccat gggcgtgtgg cagcagtgtc tggaggagca gaacctaaaa 27240

gagctctgca agctcctgca gaagaacctc aaggccctgt ggaccgggtt cgacgagcgc 27300

accaccgcct cggacctggc agacctcatt ttccccgagc gtctcaggct gacgctgcgc 27360

aacggtttgc ccgactttat gagtcaaagc atgttgcaaa actttcgctc tttcatcctc 27420

gaacgctccg ggatcctgcc ggccacctgc tccgcgctgc cctcggactt cgtgccgctg 27480

accttccgcg agtgcccccc gccgctgtgg agccactgct acctgctgcg cttggccaac 27540

tacctggcct accactcgga cgtgatcgag gacgtcagca gcgagggcct gctcgagtgc 27600

cactgccgct gcaacctctg cacgccgcac cgctccctgg cctgcaaccc ccagctgctg 27660

agcgagaccc agatcatcgg caccttcgag ttgcaagggc ccggcgatga gggttctgcc 27720 gccaaggggg gtctgaaact caccccgggg ctgtggacct cggcctactt gcgcaagttc 27780

gtgcccgagg actaccatcc cttcgagatc aggttctacg aggaccaatc ccagccgccc 27840

aaggccgagc tgtcggcctg cgtcatcacc cagggggcga tcctggccca attgcaagct 27900

atccagaaat cccgccaaga attcttgctg aaaaagggcc gcggggtcta ccttgatccc 27960

cagaccggtg aggagcttaa ccccggcttc ccccaggatg ccccgaggaa gcagcaagaa 28020

gctgaaagtg gagctgccgc ccgtggagga tttggaggaa gactgggaga gcagtcaggc 28080

agaggaggag gagatggaag actgggacag cactcaggca gaggacagcc tgcaagacag 28140

tctggaagac gaggaggagg cagaggaggt ggaagaagta gccgccgccg ccagaccgtc 28200

gtcctcggcg gagaaagcaa gcagcacgga taccatctcc gctccgggtc ggggtcccgc 28260

tcgaccccac agtagatggg acgagaccgg gcgattcccg aaccccacca cccagaccgg 28320

taagaaggag cggcagggat acaagtcctg gcgggggcac aaaaacgcca tcgtctcctg 28380

cttgcaagct tgcgggggca acatctcatt cacccggcgc tacctgctct ttcaccgcgg 28440

ggtgaacttc cccgcaaca tcttgcatta ctaccgtcac ctccacagcc cctactactt 28500

ccaagaagag gcagaaaaag acaaaaccag cagctagaaa atccacagcg gcggcggcgg 28560

caggtggact gaggatcgcg gcgaacgagc cggcgcagac ccgggaactg aggaaccgga 28620

tctttcccac cctctatgcc atcttccagc agagtcgggg gcaggagcag gaactgaaag 28680

tcaagaaccg ttctctgcgc tcgctcaccc gcagttgtct gtatcacaag agcgaagacc 28740

aacttcagcg cacgcttgag gacgccgagg ctctcttcaa caagtactgc gcactcactc 28800

ttaaagagta gcccgcgccc gcccacacac ggaaaaaggc gggaattacg tcacctgtgc 28860

acccccaccc agcaccgcta tgagcaaaga aattcccacg ccttacatgt ggagctacca 28920

gccccagatg ggcctggccg ccggcgccgc ccaggactac tccacccgca tgaattggct 28980

cagcgccggg cccgggatga tctcacgggt gaatgacatc cgcgcccacc gaaaccagat 29040

actcctagaa cagtcagcgc tcaccgccac gccccgcaat cacctcaatc cgcgtaattg 29100

gcccgccgcc ctagtgtacc aggaaattcc ccagcccacg accgtactac ttccgcgaga 29160

cgcccaggcc gaagtccagc tgactaactc aggtgtccag ctggcgggcg gcgccaccct 29220 gtgtcgtcac caccccgctc agggtataaa gcggctggtg atccggggca gaggcacaca 29280

gctcaacgac gaggtggtga gctcttcact gggtttgcga cctgacggag tcttccaact 29340

cgccggatcg ggaagatctt ttcggggcaa catctcattc acccggcgct acctgctctt 29400

tcaccgcggg gtgaacttcc cccgcaacat cttgcattac taccgtcacc tccacagccc 29460

ctactacttc caagaagagg cagaaaaaga caaaaccagc agctagaaaa tccacagcgg 29520

cggcggcggc aggtggactg aggatcgcgg cgaacgagcc ggcgcagacc cgggaactga 29580

ggaaccggat ctttcccacc ctctatgcca tcttccagca gagtcggggg caggagcagg 29640

aactgaaagt caagaaccgt tctctgcgct cgctcacccg cagttgtctg tatcacaaga 29700

gcgaagacca acttcagcgc acgcttgagg acgccgaggc tctcttcaac aagtactgcg 29760

cactcactct taaagagtag cccgcgcccg cccacacacg gaaaaaggcg ggaattacgt 29820

cacctgtgca cccccaccca gcaccgctat gagcaaagaa attcccacgc cttacatgtg 29880

gagctaccag ccccagatgg gcctggccgc cggcgccgcc caggactact ccacccgcat 29940

gaattggctc agcgccgggc ccgggatgat ctcacgggtg aatgacatcc gcgcccaccg 30000

aaaccagata ctcctagaac agtcagcgct caccgccacg ccccgcaatc acctcaatcc 30060

gcgtaattgg cccgccgccc tagtgtacca ggaaattccc cagcccacga ccgtactact 30120

tccgcgagac gcccaggccg aagtccagct gactaactca ggtgtccagc tggcgggcgg 30180

cgccaccctg tgtcgtcacc accccgctca gggtataaag cggctggtga tccggggcag 30240

aggcacacag ctcaacgacg aggtggtgag ctcttcactg ggtttgcgac ctgacggagt 30300

cttccaactc gccggatcgg gaagatcttc cttcacgcct cgtcaggccg tgctgacttt 30360

ggagagttct tcctcgcaac ctcgctcggg cggcatcggc actctccagt ttgtggagga 30420

gttcactccc tcggtctact tcaacccctt ctccggctcc cccggccact atccggacga 30480

gttcatcccg aacttcgatg ccatcagcga atcggtagac ggctacgatt gaatgtccca 30540

tggtggcgcg gctgacctag ctcggcttcg acacctggac cactgccgcc gctttcgctg 30600

cttcgctcgg gacctcgccg agtttaccta ctttgagctg tccgaggagc accctcaggg 30660

cccggcccac ggagtgcgga tcgtcgtcga agggggccta gactcccacc tgcttcgtat 30720 cttcagccag cgcccgatcc tggtccagcg ccaacagggc aacaccctcc tgaccettta 30780

ctgcatctgc aaccaccccg gcctgcacga aagtctttgt tgtctgctgt gtactgagta 30840

taataaaagc tgagatcagc gactactccg gactcgattg tgttccagca gtctggcgat 30900

accaagggtt gcatccactg ctcctgcgac tcccccgagt gcgttcacac cctcatcaag 30960

accctatgcg gcctccgcga cctcctcccc atgaactaat caactaaccc cttaccccat 31020

tacccatcca gtaaaaaaaa taaagattaa agagacgatg attttgaatt actagttatt 31080

aatagtaatc aattacgggg tcattagttc atagcccata tatggagttc cgcgttacat 31140

aacttacggt aaatggcccg cctggctgac cgcccaacga cccccgccca ttgacgtcaa 31200

taatgacgta tgttcccata gtaacgccaa tagggacttt ccattgacgt caatgggtgg 31260

agtatttacg gtaaactgcc cacttggcag tacatcaagt gtatcatatg ccaagtacgc 31320

cccctattga cgtcaatgac ggtaaatggc ccgcctggca ttatgcccag tacatgacct 31380

tatgggactt tcctacttgg cagtacatct acgtattagt catcgctatt accatggtga 31440

tgcggttttg gcagtacatc aatgggcgtg gatagcggtt tgactcacgg ggatttccaa 31500

gtctccaccc cattgacgtc aatgggagtt tgttttggca ccaaaatcaa cgggactttc 31560

caaaatgtcg taacaactcc gccccattga cgcaaatggg cggtaggcgt gtacggtggg 31620

aggtctatat aagcagagct cactgtcttc cggatcgctg tccaggagcg ccagctgttg 31680

ggctcgcggt tgagaaggaa ctcttcgcgg tccttccagt actcttcaag gggaacccg 31740

tcctggtcgg cacgggactc cgcgcaagga cctaagcgtc tccagatcca cgggatctga 31800

aaaccgttga acgaaggctt cgagccagtc gcagtcgcaa gtctagagcc accatgttcg 31860

tcttcctggt cctgctgccc ctggtctcat ctcagtgcgt gaatctgact acaagaactc 31920

agctgcctcc cgcctacacc aattccttca cccggggcgt gtactatcct gacaaggtgt 31980

ttagaagctc cgtgctgcac tctacacagg atctgtttct gccattcttt agcaacgtga 32040

cctggttcca cgccatccac gtgagcggca ccaatggcac aaagcggttc gacaatcccg 32100

tgctgccttt taacgatggc gtgtacttcg cctctaccga gaagagcaac atcatcagag 32160

gctggatctt tggcaccaca ctggactcca agacacagtc tctgctgatc gtgaacaatg 32220 ccaccaacgt ggtcatcaag gtgtgcgagt tccagttttg taatgatccc ttcctgggcg 32280

tgtactatca caagaacaat aagagctgga tggagtccga gtttagagtg tattctagcg 32340

ccaacaattg cacatttgag tacgtgtccc agcctttcct gatggacctg gagggcaagc 32400

agggcaattt caagaacctg agggagttcg tgtttaagaa tatcgatggc tacttcaaga 32460

tctactctaa gcacaccccc atcaacctgg tgcgcgacct gcctcagggc ttcagcgccc 32520

tggagccact ggtggatctg cctatcggca tcaacatcac ccggtttcag acactgctgg 32580

ccctgcacag aagctacctg acacccggcg actcctctag cggatggacc gcaggagcag 32640

cagcctacta tgtgggctat ctgcagccta ggaccttcct gctgaagtac aacgagaatg 32700

gcaccatcac agacgcagtg gattgcgccc tggacccccct gagcgagaca aagtgtacac 32760

tgaagtcctt taccgtggag aagggcatct atcagacatc caatttcagg gtgcagccaa 32820

ccgagtctat cgtgcgcttt cctaatatca caaacctgtg cccatttggc gaggtgttca 32880

acgcaaccag gttcgcaagc gtgtacgcat ggaataggaa gcgcatctct aactgcgtgg 32940

ccgactatag cgtgctgtac aactccgcct ctttcagcac ctttaagtgc tatggcgtgt 33000

cccccacaaa gctgaatgac ctgtgcttta ccaacgtgta cgccgattct ttcgtgatca 33060

ggggcgacga ggtgcgccag atcgcacctg gacagacagg caagatcgcc gactacaatt 33120

ataagctgcc agacgatttc accggctgcg tgatcgcctg gaacagcaac aatctggatt 33180

ccaaagtggg cggcaactac aattatctgt accggctgtt tagaaagagc aatctgaagc 33240

ccttcgagag ggacatctct acagagatct accaggccgg cagcacccct tgcaatggcg 33300

tggagggctt taactgttat ttcccactgc agtcctacgg cttccagccc acaaacggcg 33360

tgggctatca gccttaccgc gtggtggtgc tgagctttga gctgctgcac gcaccagcaa 33420

cagtgtgcgg acccaagaag tccaccaatc tggtgaagaa caagtgcgtg aacttcaact 33480

tcaacggcct gaccggaaca ggcgtgctga ccgagtccaa caagaagttc ctgccatttc 33540

agcagttcgg cagggacatc gcagatacca cagacgccgt gcgcgaccca cagaccctgg 33600

agatcctgga tatcacaccc tgctctttcg gcggcgtgag cgtgatcaca ccaggaacca 33660

atacaagcaa ccaggtggcc gtgctgtatc aggacgtgaa ttgtaccgag gtgcctgtgg 33720 ccatccacgc cgatcagctg accccaacat ggcgggtgta cagcaccggc tccaacgtgt 33780

tccagacaag agcaggatgc ctgatcggag cagagcacgt gaacaattcc tatgagtgcg 33840

acatcccaat cggcgccggc atctgtgcct cttaccagac ccagacaaac tctccaagga 33900

gagcacggag cgtggcatcc cagtctatca tcgcctatac catgtccctg ggcgccgaga 33960

attctgtggc ctactctaac aatagcatcg ccatcccaac caacttcaca atctctgtga 34020

ccacagagat cctgcccgtg tccatgacca agacatctgt ggactgcaca atgtatatct 34080

gtggcgattc taccgagtgc agcaacctgc tgctgcagta cggcagcttt tgtacccagc 34140

tgaatagagc cctgacaggc atcgccgtgg agcaggataa gaacacacag gaggtgttcg 34200

cccaggtgaa gcagatctac aagacccccc ctatcaagga ctttggcggc ttcaattttt 34260

cccagatcct gcctgatcca tccaagcctt ctaagcggag ctttatcgag gacctgctgt 34320

tcaacaaggt gaccctggcc gatgccggct tcatcaagca gtatggcgat tgcctgggcg 34380

acatcgcagc acgggacctg atctgtgccc agaagtttaa tggcctgacc gtgctgccac 34440

ccctgctgac agatgagatg atcgcacagt acacaagcgc cctgctggca ggaaccatca 34500

catccggatg gaccttcggc gcaggagccg ccctgcagat ccccttgcc atgcagatgg 34560

cctataggtt caacggcatc ggcgtgaccc agaatgtgct gtacgagaac cagaagctga 34620

tcgccaatca gtttaactcc gccatcggca agatccagga cagcctgtcc tctacagcct 34680

ccgccctggg caagctgcag gatgtggtga atcagaacgc ccaggccctg aataccctgg 34740

tgaagcagct gagctccaac ttcggcgcca tctctagcgt gctgaatgat atcctgagcc 34800

ggctggacaa ggtggaggca gaggtgcaga tcgaccggct gatcacaggc agactgcagt 34860

ctctgcagac ctatgtgaca cagcagctga tcagggcagc agagatcagg gcaagcgcca 34920

atctggcagc aaccaagatg tccgagtgcg tgctgggcca gtctaagaga gtggactttt 34980

gtggcaaggg ctatcacctg atgtccttcc ctcagtctgc cccacacggc gtggtgtttc 35040

tgcacgtgac ctacgtgccc gcccaggaga agaacttcac cacagcccct gccatctgcc 35100

acgatggcaa ggcccacttt ccaagggagg gcgtgttcgt gtccaacggc acccactggt 35160

ttgtgacaca gcgcaatttc tacgagcccc agatcatcac cacagacaat accttcgtga 35220 gcggcaactg tgacgtggtc atcggcatcg tgaacaatac cgtgtatgat ccactgcagc 35280

ccgagctgga cagctttaag gaggagctgg ataagtactt caagaatcac acctcccctg 35340

acgtggatct gggcgacatc agcggcatca atgcctccgt ggtgaacatc cagaaggaga 35400

tcgaccgcct gaacgaggtg gccaagaatc tgaacgagag cctgatcgat ctgcaggagc 35460

tgggcaagta tgagcagtac atcaagtggc catggtacat ctggctgggc ttcatcgccg 35520

gcctgatcgc catcgtgatg gtgaccatca tgctgtgctg tatgacatcc tgctgttctt 35580

gcctgaaggg ctgctgtagc tgtggctcct gctgtaagtt tgatgaggac gattccgaac 35640

ccgtgctgaa gggagtgaag ctgcattaca cctgaggatc cctcgagctg tgccttctag 35700

ttgccagcca tctgttgttt gcccctcccc cgtgccttcc ttgaccctgg aaggtgccac 35760

tcccactgtc ctttcctaat aaaatgagga aattgcatcg cattgtctga gtaggtgtca 35820

ttctattctg gggggtgggg tggggcagga cagcaagggg gaggattggg aagacaatag 35880

caggcatgct ggggatgcgg tgggctctat ggtgatcaat aaagaatcac ttacttgaaa 35940

tctgaaacca ggtctctgtc catgttttct gtcagcagca cttcgctccc ctcttcccag 36000

ctctggtact gcaggccccg gcgggctgca aacttcctcc acactctgaa ggggatgtca 36060

aattcctcct gtccctcaat cttcattttt tatttctatt agatgtccaa aaagcgcgcg 36120

cgggtggatg atggcttcga ccccgtgtat ccctacgatg cagacaacgc accgaccgtg 36180

cccttcatca accctccctt cgtctcttca gatggattcc aagaaaagcc cctgggggtg 36240

ttgtccctta ggctggccga ccctgtcacc accaagaatg gggaaattac cctcaagctg 36300

ggggaggggg tggaccttga cgactcggga aaactcattg caaacacagt aaacaaggcc 36360

attgcccctc tcagtttttc caacaacacc atttccctta acatggatac ccctttatac 36420

accaaagatg gaaaactatc cttacaagtt tctccaccat taagtatatt aaaatcaaca 36480

attttgaata cattagctct agcttttggc tcaggtttag gactcagtgg cagcgccctg 36540

gcagtacagt tagcctctcc acttacattt gatgataaag ggaatataaa gattacccta 36600

aacaggggat tgcatgttac aacaggagat gcaattgaaa gcaacatcag ttgggctaaa 36660

ggtataaaat ttgaagatgg tgccatagct acaaacattg gtaaggggct agagttcgga 36720 accagtagta cagaaacagg agttaataat gcttatccaa tccaagttaa acttggctct 36780

ggtctcagct ttgacagcac aggagccata atggctggca ataaagacta tgataaatta 36840

actttgtgga caacgcctga cccatcacca aactgtcaaa tacttgcaga aaatgatgca 36900

aaactaacac tttgcttaac taagtgtgac agtcaaatac tggccactgt atcagttttg 36960

gttgttagaa gtggaaactt aaacccaatt actggcacag taagcagtgc tcaagttttt 37020

ctacgttttg atgcaaatgg tgttctttta acagaacact ctacactaaa aaaatactgg 37080

ggctacaagc aaggagatag catagatggc actccataca ccaatgctgt tggttttatg 37140

ccaaattcaa cagcttatcc aaagacccaa agttctacta ctaaaaataa tatagtgggt 37200

caagtataca tgaatggaga tgtttcaaaa cccatgcttc ttactataac tcttaatggt 37260

actgatgaca ccaccagtgc atactcaatg tcattttcat acacctggac taacggaagc 37320

tatatcggag caacatttgg agctaactca tacaccttct cctacatagc ccaacaataa 37380

tcccaccctg catgccaacc caccttttcc ctctatttat aaatggaaac tgaaacaaaa 37440

ataaagttca agtgttttat tgattcaaca gtttttcaca ggattcgagt agttattttc 37500

cctccaccct cccatctcat ggaatacact atcctctccc cacgcacagc cttaaacatc 37560

tgaatgctat tggtaatgga catggttttg atctccacat tccacacagt ttcagagcga 37620

gacagtctcg ggtcggtcaa ggagatgaaa ccctccgggc actcctgcat ctgcacctca 37680

cagttcaaca gctgagggct gtcctcggtg attggaatca cagttatctg gaataagagc 37740

gatgagaatc ataatccgca aacgggatcg ggcggttgtg gcgcatcagg ccccgcagca 37800

gtcgctgtct gcgccgctcc gtcaagctgc tactcaaggg gtccgggtcc agggactccc 37860

tgcgcatgat gccaatggcc ctgagcatca gtcgcctggt acggcgggcg cagcagcgga 37920

tgcggatctc actcaggtcg gagcagtacg tgcagcacag caccaccaag ttgttcaaca 37980

gtccatagtt caacgtgctc cagccaaaac tcatttgtgg aactatgctg cccacatgtc 38040

catcgtacca gatcctgatg taaatcaggt ggcgtcccct ccagaacaca ctgcccatgt 38100

acatgatctc cttgggcatg tgcaggttca ccacctcccg gtaccacatc acccgctggt 38160

tgaacatgca gccctggata attctgcgga accagatggc aagtaccgtc ccgcccgcca 38220 tgcagcgcag ggaccccggg ttctggcaat ggcagtggat cacccaccgc tcgcgaccgt 38280

ggatcaactg ggaactaaac aagtctatgt tggcacagca caggcacacg ctcatgcatg 38340

tcttcagcac tctcaattcc tcgggggtca ggaccatatc ccagggcaca gggaactctt 38400

gcaggacagt gaacccggcc gaacagggca atcctcgcac ggaacttaca ttgtgcatgg 38460

acagggtatc gcaatcaggc agcaccggat gatcctccac cagagaagcg cggctctcgg 38520

tctcctcaca gcgaggtaag gtggccggcg gttggtacgg atgatggcga gataacgcta 38580

atcgtgttct ggatcgtgtc atgatggagc tgtttccgga cattttcgta tttcacaaag 38640

cagaacctgg tccgggcact gcacaccgct cgtcggcgac ggtctcggcg cttcgagcgc 38700

tcaatgttga agttatagaa cagccactcc ctcagaacgt gcagtatctc ctgagcctct 38760

tgggtgatga aaatcccatc cgccctgatg gctctgatta catcaaccac ggtggaatgg 38820

gccaaaccca gccagatgat gcaattttgt tgggtttcgg tgacggcggg ggagggaaga 38880

acaggaagaa ccatgattaa ctttattcca aacggtctcg gaacacttca aaatgcaggt 38940

cccggaggtg gcacctctcg cccccactgt gttggtggaa aataacagcc aggtcaaagg 39000

taacacggtt ctcgagatgt tccacggtgg cttccagcaa agcctccacg cgcacatcca 39060

gaaacaagag gacagcgaaa gcgggagcgt tttctaattc ctcaatcatc atattacact 39120

cctgcaccat gcctagataa ttttcatttt tccagccttg aatgattcgt attagttcct 39180

gaggtaaatc caagccagcc atgataaaaa gctcgcgcag agcgccctcc accggcattc 39240

ttaagcacac cctcataatt ccaacagatt ctgctcctgg ttcacctgta gtagattaac 39300

aagtggaata tcaattgctc tgccgcaatc cctaagctcc tcccttagca gtaactgtat 39360

gtactcattc atatcttctc cgaaattttt agccatagga ccaccaggaa caagagaagg 39420

gcaagccaca ttacagataa agcgaagtcc tccccagtga gcattgccaa atgtaagatt 39480

gaaataagca tgctggctag acccggtgat atcttccaga taactggaca gaaaatcagg 39540

caagcaattt ttaagaaaat taacaaaaga aaagtcgtct aggtgcacgt ttagagcctc 39600

aggaacaacg atggaataag tgcaaggagt acgttccagc atggttagtg tttttggtga 39660

tctgtagaac aaaaaataaa catgcaatat taaaccatgc tagcctggcg aacaggtgga 39720 taaatcactc tttccaacac caggcaggct acagggtctc cggcgcgacc attgtagaag 39780

ctgacattat gattaaaaag catcaccgac agaccttccc ggtggccggc atggatgatt 39840

cgagaagaag catacactcc gggaacattg gcgtccgtga gtgaaaaaaa gcgacctata 39900

aagccttgag gcactacaat gcttaatctt aattccagca aagcgacccc atgcggatga 39960

agcacaaaat tggcaggtgc gtaaaaaatg taattactcc ccttctgcac aggcagcaaa 40020

gcccccgctc cctccagaaa cacatacaaa acctgagcgt ccatagctta ccgagcacgg 40080

caggcgcaag agtcagagaa aaagctgagc tctaacctaa ctgcccgctt ctgtactcaa 40140

tatatagccc taacctcact gacgtaaagg ccaaggtcta aaaataccccg ccaacacgcc 40200

cagaaaccgg tgacacacta aaaaaatacg tgcacttcct caaacgccca aactggcgtc 40260

atttccggtt tcccacgcta cgtcacctct caacgacttt caaattccgt cgaccgttaa 40320

acacatcagt taccccgccc ctaacgaacg ccgctgtcac agccaatcag cgcgccccat 40380

ccccaaattt tcacgcctta tttgcatatt aactcacaca aaaaaaataa ggtatattat 40440

tgatgatgaa gcttttaat 40459

<210> 2
<211> 1273
<212> PRT
<213> SARS-CoV-2

<400> 2

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60

```
Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65                  70                  75                  80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
                85                  90                  95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
            100                 105                 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
        115                 120                 125

Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
    130                 135                 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145                 150                 155                 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu
                165                 170                 175

Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe
            180                 185                 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr
        195                 200                 205

Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu
    210                 215                 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr
225                 230                 235                 240

Leu Leu Ala Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser
                245                 250                 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro
            260                 265                 270
```

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala
        275   280   285

Val Asp Cys Ala Leu Asp Pro Leu Ser Glu Thr Lys Cys Thr Leu Lys
    290   295   300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val
305   310   315   320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys
                325   330   335

Pro Phe Gly Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala
            340   345   350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu
        355   360   365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro
    370   375   380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe
385   390   395   400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly
                405   410   415

Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys
            420   425   430

Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
        435   440   445

Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe
    450   455   460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
465   470   475   480

Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
                485   490   495

Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val
            500   505   510

Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys
        515   520   525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn
    530   535   540

Gly Leu Thr Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu
545   550   555   560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val
                565   570   575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe
            580   585   590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val
        595   600   605

Ala Val Leu Tyr Gln Asp Val Asn Cys Thr Glu Val Pro Val Ala Ile
    610   615   620

His Ala Asp Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser
625   630   635   640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val
                645   650   655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala
            660   665   670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala
        675   680   685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser
    690   695   700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile
705   710   715   720

Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val
                725   730   735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu
            740   745   750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr
        755   760   765

Gly Ile Ala Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln
    770   775   780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe
785   790   795   800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser
                805   810   815

Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly
            820   825   830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp
        835   840   845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
    850   855   860

Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly
865   870   875   880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile
                885   890   895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
            900   905   910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
        915   920   925

Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala
    930   935   940

Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn
945   950   955   960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val
                965   970   975

Leu Asn Asp Ile Leu Ser Arg Leu Asp Lys Val Glu Ala Glu Val Gln
            980   985   990

Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val
        995   1000   1005

Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn
    1010   1015   1020

Leu Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys
    1025   1030   1035

Arg Val Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro
    1040   1045   1050

Gln Ser Ala Pro His Gly Val Val Phe Leu His Val Thr Tyr Val
    1055   1060   1065

Pro Ala Gln Glu Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His
1070 1075 1080

Asp Gly Lys Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn
1085 1090 1095

Gly Thr His Trp Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln
1100 1105 1110

Ile Ile Thr Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val
1115 1120 1125

Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro
1130 1135 1140

Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn
1145 1150 1155

His Thr Ser Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn
1160 1165 1170

Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu
1175 1180 1185

Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu
1190 1195 1200

Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu
1205 1210 1215

Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met
1220 1225 1230

Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys
1235 1240 1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro
    1250   1255   1260

Val Leu Lys Gly Val Lys Leu His Tyr Thr
    1265   1270

<210> 3
<211> 1273
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic protein

<400> 3

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1   5   10   15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
            20   25   30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
        35   40   45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
    50   55   60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65   70   75   80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
                85   90   95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
            100   105   110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
        115   120   125

Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
    130   135   140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145   150   155   160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu
                165   170   175

Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe
            180   185   190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr
        195   200   205

Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu
    210   215   220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr
225   230   235   240

Leu Leu Ala Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser
                245   250   255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro
            260   265   270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala
        275   280   285

Val Asp Cys Ala Leu Asp Pro Leu Ser Glu Thr Lys Cys Thr Leu Lys
    290   295   300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val
305   310   315   320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys
                325   330   335

Pro Phe Gly Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala
            340   345   350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu
        355   360   365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro
    370   375   380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe
385   390   395   400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly
                405   410   415

Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys
            420   425   430

Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
        435   440   445

Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe
    450   455   460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
465   470   475   480

Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
                485   490   495

Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val
            500   505   510

Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys
        515   520   525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn
    530   535   540

Gly Leu Thr Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu
545   550   555   560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val
                565   570   575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe
            580   585   590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val
        595   600   605

Ala Val Leu Tyr Gln Asp Val Asn Cys Thr Glu Val Pro Val Ala Ile
    610   615   620

His Ala Asp Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser
625   630   635   640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val
                645   650   655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala
            660   665   670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala
        675   680   685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser
    690   695   700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile
705   710   715   720

Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val
                725   730   735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu
            740   745   750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr
        755   760   765

Gly Ile Ala Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln
    770   775   780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe
785   790   795   800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser
                805   810   815

Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly
            820   825   830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp
        835   840   845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
    850   855   860

Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly
865   870   875   880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile
                885   890   895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
            900   905   910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
        915   920   925

Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala
    930   935   940

Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn
945   950   955   960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val
                965   970   975

Leu Asn Asp Ile Leu Ser Arg Leu Asp Pro Pro Glu Ala Glu Val Gln
            980   985   990

Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val
        995   1000   1005

Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn
    1010   1015   1020

Leu Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys
    1025   1030   1035

Arg Val Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro
    1040   1045   1050

Gln Ser Ala Pro His Gly Val Val Phe Leu His Val Thr Tyr Val
    1055   1060   1065

Pro Ala Gln Glu Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His
    1070   1075   1080

Asp Gly Lys Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn
    1085   1090   1095

Gly Thr His Trp Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln
    1100   1105   1110

Ile Ile Thr Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val
    1115   1120   1125

Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro
    1130   1135   1140

Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn
    1145   1150   1155

His Thr Ser Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn
    1160   1165   1170

Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu
    1175   1180   1185

Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu
    1190   1195   1200

Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu
    1205   1210   1215

Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met
    1220   1225   1230

Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys
    1235   1240   1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro
    1250   1255   1260

Val Leu Lys Gly Val Lys Leu His Tyr Thr
    1265   1270

<210> 4
<211> 1249
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic protein

<400> 4

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1   5   10   15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
            20   25   30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
        35   40   45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
    50   55   60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65   70   75   80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
                85   90   95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
            100   105   110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
        115   120   125

Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
    130   135   140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145   150   155   160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu
                165   170   175

Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe
            180   185   190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr
        195   200   205

Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu
    210   215   220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr
225   230   235   240

Leu Leu Ala Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser
                245   250   255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro
            260   265   270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala
        275   280   285

Val Asp Cys Ala Leu Asp Pro Leu Ser Glu Thr Lys Cys Thr Leu Lys
    290   295   300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val
305   310   315   320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys
                325   330   335

Pro Phe Gly Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala
            340   345   350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu
        355   360   365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro
    370   375   380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe
385   390   395   400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly
                405   410   415

Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys
            420   425   430

Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
        435   440   445

Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe
    450   455   460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
465   470   475   480

Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
                485   490   495

Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val
            500   505   510

Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys
        515   520   525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn
    530   535   540

Gly Leu Thr Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu
545   550   555   560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val
                565   570   575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe
            580   585   590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val
        595   600   605

Ala Val Leu Tyr Gln Asp Val Asn Cys Thr Glu Val Pro Val Ala Ile
    610   615   620

His Ala Asp Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser
625   630   635   640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val
                645   650   655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala
            660   665   670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala
        675   680   685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser
    690   695   700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile
705   710   715   720

Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val
                725   730   735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu
            740   745   750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr
        755   760   765

Gly Ile Ala Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln
    770   775   780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe
785   790   795   800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser
                805   810   815

Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly
            820   825   830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp
        835   840   845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
    850   855   860

Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly
865   870   875   880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile
                885   890   895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
            900   905   910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
        915   920   925

Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala
    930   935   940

Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn
945   950   955   960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val
                965   970   975

Leu Asn Asp Ile Leu Ser Arg Leu Asp Lys Val Glu Ala Glu Val Gln
            980   985   990

Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val
995 1000 1005

Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn
1010 1015 1020

Leu Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys
1025 1030 1035

Arg Val Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro
1040 1045 1050

Gln Ser Ala Pro His Gly Val Val Phe Leu His Val Thr Tyr Val
1055 1060 1065

Pro Ala Gln Glu Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His
1070 1075 1080

Asp Gly Lys Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn
1085 1090 1095

Gly Thr His Trp Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln
1100 1105 1110

Ile Ile Thr Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val
1115 1120 1125

Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro
1130 1135 1140

Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn
1145 1150 1155

His Thr Ser Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn
1160 1165 1170

Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu
1175 1180 1185

Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu
    1190   1195   1200

Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu
    1205   1210   1215

Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met
    1220   1225   1230

Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys
    1235   1240   1245

Ser

<210> 5
<211> 1268
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic protein

<400> 5

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1   5   10   15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
            20   25   30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
        35   40   45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
    50   55   60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65   70   75   80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125

Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145 150 155 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu
165 170 175

Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe
180 185 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr
195 200 205

Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu
210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr
225 230 235 240

Leu Leu Ala Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser
245 250 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro
260 265 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala
        275   280   285

Val Asp Cys Ala Leu Asp Pro Leu Ser Glu Thr Lys Cys Thr Leu Lys
    290   295   300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val
305   310   315   320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys
                325   330   335

Pro Phe Gly Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala
            340   345   350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu
        355   360   365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro
    370   375   380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe
385   390   395   400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly
                405   410   415

Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys
            420   425   430

Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
        435   440   445

Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe
    450   455   460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
465   470   475   480

```
Asn Gly Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
                485                 490                 495

Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val
            500                 505                 510

Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys
        515                 520                 525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn
    530                 535                 540

Gly Leu Thr Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu
545                 550                 555                 560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val
                565                 570                 575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe
            580                 585                 590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val
        595                 600                 605

Ala Val Leu Tyr Gln Asp Val Asn Cys Thr Glu Val Pro Val Ala Ile
    610                 615                 620

His Ala Asp Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser
625                 630                 635                 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val
                645                 650                 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala
            660                 665                 670
```

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala
        675   680   685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser
    690   695   700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile
705   710   715   720

Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val
                725   730   735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu
            740   745   750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr
        755   760   765

Gly Ile Ala Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln
    770   775   780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe
785   790   795   800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser
                805   810   815

Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly
            820   825   830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp
        835   840   845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
    850   855   860

Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly
865   870   875   880

```
Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile
                885                 890                 895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
            900                 905                 910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
        915                 920                 925

Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala
    930                 935                 940

Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn
945                 950                 955                 960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val
                965                 970                 975

Leu Asn Asp Ile Leu Ser Arg Leu Asp Lys Val Glu Ala Glu Val Gln
            980                 985                 990

Ile Asp Arg Leu Ile Thr Gly Arg  Leu Gln Ser Leu Gln  Thr Tyr Val
        995                 1000                1005

Thr Gln  Gln Leu Ile Arg Ala  Ala Glu Ile Arg Ala  Ser Ala Asn
    1010                1015                1020

Leu Ala  Ala Thr Lys Met Ser  Glu Cys Val Leu Gly  Gln Ser Lys
    1025                1030                1035

Arg Val  Asp Phe Cys Gly Lys  Gly Tyr His Leu Met  Ser Phe Pro
    1040                1045                1050

Gln Ser  Ala Pro His Gly Val  Val Phe Leu His Val  Thr Tyr Val
    1055                1060                1065
```

Pro Ala Gln Glu Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His
1070 1075 1080

Asp Gly Lys Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn
1085 1090 1095

Gly Thr His Trp Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln
1100 1105 1110

Ile Ile Thr Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val
1115 1120 1125

Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro
1130 1135 1140

Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn
1145 1150 1155

His Thr Ser Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn
1160 1165 1170

Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu
1175 1180 1185

Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu
1190 1195 1200

Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu
1205 1210 1215

Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met
1220 1225 1230

Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys
1235 1240 1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro
1250 1255 1260

Val Leu Lys Gly Val
    1265

<210> 6
<211> 3822
<212> DNA
<213> SARS-CoV-2

<400> 6
atgtttgttt ttcttgtttt attgccacta gtctctagtc agtgtgttaa tcttacaacc   60

agaactcaat taccccctgc atacactaat tctttcacac gtggtgttta ttaccctgac   120

aaagttttca gatcctcagt tttacattca actcaggact tgttcttacc tttcttttcc   180

aatgttactt ggttccatgc tatacatgtc tctgggacca atggtactaa gaggtttgat   240

aaccctgtcc taccatttaa tgatggtgtt tattttgctt ccactgagaa gtctaacata   300

ataagaggct ggatttttgg tactacttta gattcgaaga cccagtccct acttattgtt   360

aataacgcta ctaatgttgt tattaaagtc tgtgaatttc aattttgtaa tgatccattt   420

ttgggtgttt attaccacaa aaacaacaaa agttggatgg aaagtgagtt cagagtttat   480

tctagtgcga ataattgcac ttttgaatat gtctctcagc cttttcttat ggaccttgaa   540

ggaaaacagg gtaatttcaa aaatcttagg gaatttgtgt ttaagaatat tgatggttat   600

tttaaaatat attctaagca cacgcctatt aatttagtgc gtgatctccc tcagggtttt   660

tcggctttag aaccattggt agatttgcca ataggtatta acatcactag gtttcaaact   720

ttacttgctt tacatagaag ttatttgact cctggtgatt cttcttcagg ttggacagct   780

ggtgctgcag cttattatgt gggttatctt caacctagga cttttctatt aaaatataat   840

gaaaatggaa ccattacaga tgctgtagac tgtgcacttg accctctctc agaaacaaag   900

tgtacgttga aatccttcac tgtagaaaaa ggaatctatc aaacttctaa ctttagagtc   960

caaccaacag aatctattgt tagatttcct aatattacaa acttgtgccc ttttggtgaa   1020

gtttttaacg ccaccagatt tgcatctgtt tatgcttgga acaggaagag aatcagcaac   1080

tgtgttgctg attattctgt cctatataat tccgcatcat tttccacttt taagtgttat   1140 ggagtgtctc ctactaaatt aaatgatctc tgctttacta atgtctatgc agattcattt 1200

gtaattagag gtgatgaagt cagacaaatc gctccagggc aaactggaaa gattgctgat 1260

tataattata aattaccaga tgattttaca ggctgcgtta tagcttggaa ttctaacaat 1320

cttgattcta aggttggtgg taattataat tacctgtata gattgtttag gaagtctaat 1380

ctcaaacctt ttgagagaga tatttcaact gaaatctatc aggccggtag cacaccttgt 1440

aatggtgttg aaggttttaa ttgttacttt cctttacaat catatggttt ccaacccact 1500

aatggtgttg gttaccaacc atacagagta gtagtacttt cttttgaact tctacatgca 1560

ccagcaactg tttgtggacc taaaaagtct actaatttgg ttaaaaacaa atgtgtcaat 1620

ttcaacttca atggtttaac aggcacaggt gttcttactg agtctaacaa aaagtttctg 1680

cctttccaac aatttggcag agacattgct gacactactg atgctgtccg tgatccacag 1740

acacttgaga ttcttgacat tacaccatgt tcttttggtg gtgtcagtgt tataacacca 1800

ggaacaaata cttctaacca ggttgctgtt ctttatcagg atgttaactg cacagaagtc 1860

cctgttgcta ttcatgcaga tcaacttact cctacttggc gtgtttattc tacaggttct 1920

aatgtttttc aaacacgtgc aggctgttta ataggggctg aacatgtcaa caactcatat 1980

gagtgtgaca tacccattgg tgcaggtata tgcgctagtt atcagactca gactaattct 2040

cctcggcggg cacgtagtgt agctagtcaa tccatcattg cctacactat gtcacttggt 2100

gcagaaaatt cagttgctta ctctaataac tctattgcca tacccacaaa ttttactatt 2160

agtgttacca cagaaattct accagtgtct atgaccaaga catcagtaga ttgtacaatg 2220

tacatttgtg gtgattcaac tgaatgcagc aatcttttgt tgcaatatgg cagtttttgt 2280

acacaattaa accgtgcttt aactggaata gctgttgaac aagacaaaaa cacccaagaa 2340

gtttttgcac aagtcaaaca aatttacaaa acaccaccaa ttaaagattt tggtggtttt 2400

aatttttcac aaatattacc agatccatca aaaccaagca agaggtcatt tattgaagat 2460

ctacttttca acaaagtgac acttgcagat gctggcttca tcaaacaata tggtgattgc 2520

cttggtgata ttgctgctag agacctcatt tgtgcacaaa agtttaacgg ccttactgtt 2580

ttgccacctt tgctcacaga tgaaatgatt gctcaataca cttctgcact gttagcgggt 2640 acaatcactt ctggttggac ctttggtgca ggtgctgcat tacaaatacc atttgctatg 2700

caaatggctt ataggtttaa tggtattgga gttacacaga atgttctcta tgagaaccaa 2760

aaattgattg ccaaccaatt taatagtgct attggcaaaa ttcaagactc actttcttcc 2820

acagcaagtg cacttggaaa acttcaagat gtggtcaacc aaaatgcaca agctttaaac 2880

acgcttgtta aacaacttag ctccaatttt ggtgcaattt caagtgtttt aaatgatatc 2940

ctttcacgtc ttgacaaagt tgaggctgaa gtgcaaattg ataggttgat cacaggcaga 3000

cttcaaagtt tgcagacata tgtgactcaa caattaatta gagctgcaga aatcagagct 3060

tctgctaatc ttgctgctac taaaatgtca gagtgtgtac ttggacaatc aaaaagagtt 3120

gatttttgtg gaaagggcta tcatcttatg tccttccctc agtcagcacc tcatggtgta 3180

gtcttcttgc atgtgactta tgtccctgca caagaaaaga acttcacaac tgctcctgcc 3240

atttgtcatg atggaaaagc acactttcct cgtgaaggtg tctttgtttc aaatggcaca 3300

cactggtttg taacacaaag gaatttttat gaaccacaaa tcattactac agacaacaca 3360

tttgtgtctg gtaactgtga tgttgtaata ggaattgtca acaacacagt ttatgatcct 3420

ttgcaacctg aattagactc attcaaggag gagttagata aatattttaa gaatcataca 3480

tcaccagatg ttgatttagg tgacatctct ggcattaatg cttcagttgt aaacattcaa 3540

aaagaaattg accgcctcaa tgaggttgcc aagaatttaa atgaatctct catcgatctc 3600

caagaacttg gaaagtatga gcagtatata aaatggccat ggtacatttg gctaggtttt 3660

atagctggct tgattgccat agtaatggtg acaattatgc tttgctgtat gaccagttgc 3720

tgtagttgtc tcaagggctg ttgttcttgt ggatcctgct gcaaatttga tgaagacgac 3780

tctgagccag tgctcaaagg agtcaaatta cattacacat aa 3822

<210> 7
<211> 1270
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic protein

<400> 7

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1   5   10   15

Asn Phe Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
            20   25   30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
        35   40   45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
    50   55   60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Ala
65   70   75   80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
                85   90   95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
            100   105   110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
        115   120   125

Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
    130   135   140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145   150   155   160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu
                165   170   175

Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe
            180   185   190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr

195 200 205

Pro Ile Asn Leu Val Arg Gly Leu Pro Gln Gly Phe Ser Ala Leu Glu
210 215 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr
225 230 235 240

Leu His Ile Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser Gly Trp Thr
245 250 255

Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro Arg Thr Phe
260 265 270

Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala Val Asp Cys
275 280 285

Ala Leu Asp Pro Leu Ser Glu Thr Lys Cys Thr Leu Lys Ser Phe Thr
290 295 300

Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val Gln Pro Thr
305 310 315 320

Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys Pro Phe Gly
325 330 335

Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala Trp Asn Arg
340 345 350

Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu Tyr Asn Ser
355 360 365

Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro Thr Lys Leu
370 375 380

Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe Val Ile Arg
385 390 395 400

Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly Asn Ile Ala
                405   410   415

Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys Val Ile Ala
            420   425   430

Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn Tyr
        435   440   445

Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe Glu Arg Asp
    450   455   460

Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys Asn Gly Val
465   470   475   480

Lys Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln Pro
                485   490   495

Thr Tyr Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val Leu Ser Phe
            500   505   510

Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys Lys Ser Thr
        515   520   525

Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn Gly Leu Thr
    530   535   540

Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu Pro Phe Gln
545   550   555   560

Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val Arg Asp Pro
                565   570   575

Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe Gly Gly Val
            580   585   590

Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val Ala Val Leu

595 600 605

Tyr Gln Gly Val Asn Cys Thr Glu Val Pro Val Ala Ile His Ala Asp
610 615 620

Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser Asn Val Phe
625 630 635 640

Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val Asn Asn Ser
645 650 655

Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala Ser Tyr Gln
660 665 670

Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala Ser Gln Ser
675 680 685

Ile Ile Ala Tyr Thr Met Ser Leu Gly Val Glu Asn Ser Val Ala Tyr
690 695 700

Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile Ser Val Thr
705 710 715 720

Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val Asp Cys Thr
725 730 735

Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu Leu Leu Gln
740 745 750

Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr Gly Ile Ala
755 760 765

Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln Val Lys Gln
770 775 780

Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe Asn Phe Ser
785 790 795 800

Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser Phe Ile Glu
                805   810   815

Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe Ile Lys
            820   825   830

Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp Leu Ile Cys
        835   840   845

Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu Leu Thr Asp
    850   855   860

Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly Thr Ile Thr
865   870   875   880

Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe Ala
                885   890   895

Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn Val
            900   905   910

Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn Ser Ala Ile
        915   920   925

Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala Leu Gly Lys
    930   935   940

Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu Val
945   950   955   960

Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn Asp
                965   970   975

Ile Leu Ser Arg Leu Asp Pro Pro Glu Ala Glu Val Gln Ile Asp Arg
            980   985   990

Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln Gln

```
        995                 1000                1005

Leu Ile  Arg Ala Ala Glu Ile  Arg Ala Ser Ala Asn  Leu Ala Ala
    1010                1015                1020

Thr Lys  Met Ser Glu Cys Val  Leu Gly Gln Ser Lys  Arg Val Asp
    1025                1030                1035

Phe Cys  Gly Lys Gly Tyr His  Leu Met Ser Phe Pro  Gln Ser Ala
    1040                1045                1050

Pro His  Gly Val Val Phe Leu  His Val Thr Tyr Val  Pro Ala Gln
    1055                1060                1065

Glu Lys  Asn Phe Thr Thr Ala  Pro Ala Ile Cys His  Asp Gly Lys
    1070                1075                1080

Ala His  Phe Pro Arg Glu Gly  Val Phe Val Ser Asn  Gly Thr His
    1085                1090                1095

Trp Phe  Val Thr Gln Arg Asn  Phe Tyr Glu Pro Gln  Ile Ile Thr
    1100                1105                1110

Thr Asp  Asn Thr Phe Val Ser  Gly Asn Cys Asp Val  Val Ile Gly
    1115                1120                1125

Ile Val  Asn Asn Thr Val Tyr  Asp Pro Leu Gln Pro  Glu Leu Asp
    1130                1135                1140

Ser Phe  Lys Glu Glu Leu Asp  Lys Tyr Phe Lys Asn  His Thr Ser
    1145                1150                1155

Pro Asp  Val Asp Leu Gly Asp  Ile Ser Gly Ile Asn  Ala Ser Val
    1160                1165                1170

Val Asn  Ile Gln Lys Glu Ile  Asp Arg Leu Asn Glu  Val Ala Lys
    1175                1180                1185
```

Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu Gly Lys Tyr
    1190   1195   1200

Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu Gly Phe Ile
    1205   1210   1215

Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met Leu Cys Cys
    1220   1225   1230

Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys Ser Cys Gly
    1235   1240   1245

Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val Leu Lys
    1250   1255   1260

Gly Val Lys Leu His Tyr Thr
    1265   1270

<210> 8
<211> 1273
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic protein

<400> 8

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1   5   10   15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
            20   25   30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
        35   40   45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
    50   55   60

```
Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65                  70                  75                  80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
                85                  90                  95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
            100                 105                 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
        115                 120                 125

Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Gly Val Tyr
    130                 135                 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145                 150                 155                 160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu
                165                 170                 175

Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe
            180                 185                 190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr
        195                 200                 205

Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu
    210                 215                 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr
225                 230                 235                 240

Leu Leu Ala Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser
                245                 250                 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro
```

```
            260                 265                 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala
        275                 280                 285

Val Asp Cys Ala Leu Asp Pro Leu Ser Glu Thr Lys Cys Thr Leu Lys
    290                 295                 300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val
305                 310                 315                 320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys
                325                 330                 335

Pro Phe Gly Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala
            340                 345                 350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu
        355                 360                 365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro
    370                 375                 380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe
385                 390                 395                 400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly
                405                 410                 415

Asn Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys
            420                 425                 430

Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
        435                 440                 445

Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe
    450                 455                 460
```

```
Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
465                 470                 475                 480

Asn Gly Val Lys Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
                485                 490                 495

Phe Gln Pro Thr Tyr Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val
            500                 505                 510

Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys
        515                 520                 525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn
    530                 535                 540

Gly Leu Thr Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu
545                 550                 555                 560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val
                565                 570                 575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe
            580                 585                 590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val
        595                 600                 605

Ala Val Leu Tyr Gln Gly Val Asn Cys Thr Glu Val Pro Val Ala Ile
    610                 615                 620

His Ala Asp Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser
625                 630                 635                 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val
                645                 650                 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala
```

660 665 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala
675 680 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser
690 695 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile
705 710 715 720

Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val
725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu
740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr
755 760 765

Gly Ile Ala Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln
770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe
785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser
805 810 815

Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly
820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp
835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
850 855 860

```
Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly
865                 870                 875                 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile
                885                 890                 895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
            900                 905                 910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
        915                 920                 925

Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala
    930                 935                 940

Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn
945                 950                 955                 960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val
                965                 970                 975

Leu Asn Asp Ile Leu Ser Arg Leu Asp Pro Pro Glu Ala Glu Val Gln
            980                 985                 990

Ile Asp Arg Leu Ile Thr Gly Arg  Leu Gln Ser Leu Gln  Thr Tyr Val
        995                 1000                1005

Thr Gln  Gln Leu Ile Arg Ala  Ala Glu Ile Arg Ala  Ser Ala Asn
    1010                1015                1020

Leu Ala  Ala Thr Lys Met Ser  Glu Cys Val Leu Gly  Gln Ser Lys
    1025                1030                1035

Arg Val  Asp Phe Cys Gly Lys  Gly Tyr His Leu Met  Ser Phe Pro
    1040                1045                1050

Gln Ser  Ala Pro His Gly Val  Val Phe Leu His Val  Thr Tyr Val
```

1055 1060 1065

Pro Ala Gln Glu Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His
1070 1075 1080

Asp Gly Lys Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn
1085 1090 1095

Gly Thr His Trp Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln
1100 1105 1110

Ile Ile Thr Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val
1115 1120 1125

Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro
1130 1135 1140

Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn
1145 1150 1155

His Thr Ser Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn
1160 1165 1170

Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu
1175 1180 1185

Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu
1190 1195 1200

Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu
1205 1210 1215

Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met
1220 1225 1230

Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys
1235 1240 1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro
1250 1255 1260

Val Leu Lys Gly Val Lys Leu His Tyr Thr
1265 1270

<210> 9
<211> 1273
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic protein

<400> 9

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15

Asn Leu Thr Thr Thr Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125

Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Asp Val Tyr
    130   135   140

Tyr His Lys Asn Asn Lys Ser Trp Met Lys Ser Glu Phe Arg Val Tyr
145   150   155   160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu
                165   170   175

Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe
            180   185   190

Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr
        195   200   205

Pro Ile Asn Leu Val Arg Asp Leu Pro His Gly Phe Ser Ala Leu Glu
    210   215   220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr
225   230   235   240

Leu Leu Ala Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser
                245   250   255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro
            260   265   270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala
        275   280   285

Val Asp Cys Ala Leu Asp Pro Leu Ser Glu Thr Lys Cys Thr Leu Lys
    290   295   300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val
305   310   315   320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys

325 330 335

Pro Phe Gly Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala
340 345 350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu
355 360 365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro
370 375 380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe
385 390 395 400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly
405 410 415

Lys Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys
420 425 430

Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
435 440 445

Tyr Asn Tyr Arg Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe
450 455 460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
465 470 475 480

Asn Gly Val Gln Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
485 490 495

Phe Gln Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val
500 505 510

Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys
515 520 525

```
Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn
    530                 535                 540

Gly Leu Thr Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu
545                 550                 555                 560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val
                565                 570                 575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe
            580                 585                 590

Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val
        595                 600                 605

Ala Val Leu Tyr Gln Gly Val Asn Cys Thr Glu Val Pro Val Ala Ile
    610                 615                 620

His Ala Asp Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser
625                 630                 635                 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val
                645                 650                 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala
            660                 665                 670

Ser Tyr Gln Thr Gln Thr Asn Ser Arg Arg Arg Ala Arg Ser Val Ala
        675                 680                 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser
    690                 695                 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile
705                 710                 715                 720

Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val
```

725 730 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu
740 745 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr
755 760 765

Gly Ile Ala Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln
770 775 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe
785 790 795 800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser
805 810 815

Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly
820 825 830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp
835 840 845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
850 855 860

Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly
865 870 875 880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile
885 890 895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
900 905 910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
915 920 925

Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala
    930   935   940

Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn
945   950   955   960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val
                965   970   975

Leu Asn Asp Ile Leu Ser Arg Leu Asp Pro Pro Glu Ala Glu Val Gln
            980   985   990

Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val
        995   1000   1005

Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn
    1010   1015   1020

Leu Ala Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys
    1025   1030   1035

Arg Val Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro
    1040   1045   1050

Gln Ser Ala Pro His Gly Val Val Phe Leu His Val Thr Tyr Val
    1055   1060   1065

Pro Ala Gln Glu Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His
    1070   1075   1080

Asp Gly Lys Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn
    1085   1090   1095

Gly Thr Asp Trp Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln
    1100   1105   1110

Ile Ile Thr Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val

1115 1120 1125

Val Ile Gly Ile Val Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro
1130 1135 1140

Glu Leu Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn
1145 1150 1155

His Thr Ser Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn
1160 1165 1170

Ala Ser Val Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu
1175 1180 1185

Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu
1190 1195 1200

Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu
1205 1210 1215

Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met
1220 1225 1230

Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys
1235 1240 1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro
1250 1255 1260

Val Leu Lys Gly Val Lys Leu His Tyr Thr
1265 1270

<210> 10
<211> 1273
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic protein

<400> 10

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1   5   10   15

Asn Phe Thr Asn Arg Thr Gln Leu Pro Ser Ala Tyr Thr Asn Ser Phe
            20   25   30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
        35   40   45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
    50   55   60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65   70   75   80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
                85   90   95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
            100   105   110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
        115   120   125

Lys Val Cys Glu Phe Gln Phe Cys Asn Tyr Pro Phe Leu Gly Val Tyr
    130   135   140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr
145   150   155   160

Ser Ser Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu
                165   170   175

Met Asp Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Ser Glu Phe
            180   185   190

```
Val Phe Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr
        195                 200                 205

Pro Ile Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu
    210                 215                 220

Pro Leu Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr
225                 230                 235                 240

Leu Leu Ala Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser
                245                 250                 255

Gly Trp Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro
            260                 265                 270

Arg Thr Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala
        275                 280                 285

Val Asp Cys Ala Leu Asp Pro Leu Ser Glu Thr Lys Cys Thr Leu Lys
    290                 295                 300

Ser Phe Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val
305                 310                 315                 320

Gln Pro Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys
                325                 330                 335

Pro Phe Gly Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala
            340                 345                 350

Trp Asn Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu
        355                 360                 365

Tyr Asn Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro
    370                 375                 380

Thr Lys Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe
```

385 390 395 400

Val Ile Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly
405 410 415

Thr Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys
420 425 430

Val Ile Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn
435 440 445

Tyr Asn Tyr Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe
450 455 460

Glu Arg Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Thr Pro Cys
465 470 475 480

Asn Gly Val Lys Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly
485 490 495

Phe Gln Pro Thr Tyr Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val
500 505 510

Leu Ser Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys
515 520 525

Lys Ser Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn
530 535 540

Gly Leu Thr Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu
545 550 555 560

Pro Phe Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val
565 570 575

Arg Asp Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe
580 585 590

```
Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val
        595                 600                 605

Ala Val Leu Tyr Gln Gly Val Asn Cys Thr Glu Val Pro Val Ala Ile
    610                 615                 620

His Ala Asp Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser
625                 630                 635                 640

Asn Val Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu Tyr Val
                645                 650                 655

Asn Asn Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala
            660                 665                 670

Ser Tyr Gln Thr Gln Thr Asn Ser Pro Arg Arg Ala Arg Ser Val Ala
        675                 680                 685

Ser Gln Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser
    690                 695                 700

Val Ala Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile
705                 710                 715                 720

Ser Val Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val
                725                 730                 735

Asp Cys Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu
            740                 745                 750

Leu Leu Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr
        755                 760                 765

Gly Ile Ala Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln
    770                 775                 780

Val Lys Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe
```

785   790   795   800

Asn Phe Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser
                805   810   815

Phe Ile Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly
            820   825   830

Phe Ile Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp
        835   840   845

Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu
    850   855   860

Leu Thr Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly
865   870   875   880

Thr Ile Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile
                885   890   895

Pro Phe Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr
            900   905   910

Gln Asn Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn
        915   920   925

Ser Ala Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala
    930   935   940

Leu Gly Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn
945   950   955   960

Thr Leu Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val
                965   970   975

Leu Asn Asp Ile Leu Ser Arg Leu Asp Pro Pro Glu Ala Glu Val Gln
            980   985   990

```
Ile Asp Arg Leu Ile Thr Gly Arg  Leu Gln Ser Leu Gln  Thr Tyr Val
        995                 1000                 1005

Thr Gln  Gln Leu Ile Arg Ala  Ala Glu Ile Arg Ala  Ser Ala Asn
    1010                 1015                 1020

Leu Ala  Ala Ile Lys Met Ser  Glu Cys Val Leu Gly  Gln Ser Lys
    1025                 1030                 1035

Arg Val  Asp Phe Cys Gly Lys  Gly Tyr His Leu Met  Ser Phe Pro
    1040                 1045                 1050

Gln Ser  Ala Pro His Gly Val  Val Phe Leu His Val  Thr Tyr Val
    1055                 1060                 1065

Pro Ala  Gln Glu Lys Asn Phe  Thr Thr Ala Pro Ala  Ile Cys His
    1070                 1075                 1080

Asp Gly  Lys Ala His Phe Pro  Arg Glu Gly Val Phe  Val Ser Asn
    1085                 1090                 1095

Gly Thr  His Trp Phe Val Thr  Gln Arg Asn Phe Tyr  Glu Pro Gln
    1100                 1105                 1110

Ile Ile  Thr Thr Asp Asn Thr  Phe Val Ser Gly Asn  Cys Asp Val
    1115                 1120                 1125

Val Ile  Gly Ile Val Asn Asn  Thr Val Tyr Asp Pro  Leu Gln Pro
    1130                 1135                 1140

Glu Leu  Asp Ser Phe Lys Glu  Glu Leu Asp Lys Tyr  Phe Lys Asn
    1145                 1150                 1155

His Thr  Ser Pro Asp Val Asp  Leu Gly Asp Ile Ser  Gly Ile Asn
    1160                 1165                 1170

Ala Ser  Phe Val Asn Ile Gln  Lys Glu Ile Asp Arg  Leu Asn Glu
```

    1175   1180   1185

Val Ala Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu
    1190   1195   1200

Gly Lys Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu
    1205   1210   1215

Gly Phe Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met
    1220   1225   1230

Leu Cys Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys
    1235   1240   1245

Ser Cys Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro
    1250   1255   1260

Val Leu Lys Gly Val Lys Leu His Tyr Thr
    1265   1270

<210> 11
<211> 1271
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic protein

<400> 11

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1   5   10   15

Asn Leu Arg Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
            20   25   30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
        35   40   45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp

50 55 60

Phe His Ala Ile His Val Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp
65 70 75 80

Asn Pro Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Thr Glu
85 90 95

Lys Ser Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser
100 105 110

Lys Thr Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile
115 120 125

Lys Val Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Asp Val Tyr
130 135 140

Tyr His Lys Asn Asn Lys Ser Trp Met Glu Ser Gly Val Tyr Ser Ser
145 150 155 160

Ala Asn Asn Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu Met Asp
165 170 175

Leu Glu Gly Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe Val Phe
180 185 190

Lys Asn Ile Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr Pro Ile
195 200 205

Asn Leu Val Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu Pro Leu
210 215 220

Val Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr Leu Leu
225 230 235 240

Ala Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser Gly Trp
245 250 255

Thr Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro Arg Thr
260 265 270

Phe Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala Val Asp
275 280 285

Cys Ala Leu Asp Pro Leu Ser Glu Thr Lys Cys Thr Leu Lys Ser Phe
290 295 300

Thr Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val Gln Pro
305 310 315 320

Thr Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys Pro Phe
325 330 335

Gly Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala Trp Asn
340 345 350

Arg Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu Tyr Asn
355 360 365

Ser Ala Ser Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Pro Thr Lys
370 375 380

Leu Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe Val Ile
385 390 395 400

Arg Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly Asn Ile
405 410 415

Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys Val Ile
420 425 430

Ala Trp Asn Ser Asn Asn Leu Asp Ser Lys Val Gly Gly Asn Tyr Asn
435 440 445

Tyr Arg Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe Glu Arg

450 455 460

Asp Ile Ser Thr Glu Ile Tyr Gln Ala Gly Ser Lys Pro Cys Asn Gly
465 470 475 480

Val Glu Gly Phe Asn Cys Tyr Phe Pro Leu Gln Ser Tyr Gly Phe Gln
485 490 495

Pro Thr Asn Gly Val Gly Tyr Gln Pro Tyr Arg Val Val Val Leu Ser
500 505 510

Phe Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys Lys Ser
515 520 525

Thr Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn Gly Leu
530 535 540

Thr Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu Pro Phe
545 550 555 560

Gln Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val Arg Asp
565 570 575

Pro Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe Gly Gly
580 585 590

Val Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val Ala Val
595 600 605

Leu Tyr Gln Gly Val Asn Cys Thr Glu Val Pro Val Ala Ile His Ala
610 615 620

Asp Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser Asn Val
625 630 635 640

Phe Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu His Val Asn Asn
645 650 655

Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala Ser Tyr
660 665 670

Gln Thr Gln Thr Asn Ser Arg Arg Arg Ala Arg Ser Val Ala Ser Gln
675 680 685

Ser Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser Val Ala
690 695 700

Tyr Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile Ser Val
705 710 715 720

Thr Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val Asp Cys
725 730 735

Thr Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu Leu Leu
740 745 750

Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Thr Gly Ile
755 760 765

Ala Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln Val Lys
770 775 780

Gln Ile Tyr Lys Thr Pro Pro Ile Lys Asp Phe Gly Gly Phe Asn Phe
785 790 795 800

Ser Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser Phe Ile
805 810 815

Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe Ile
820 825 830

Lys Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp Leu Ile
835 840 845

Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu Leu Thr

850 855 860

Asp Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly Thr Ile
865 870 875 880

Thr Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe
885 890 895

Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn
900 905 910

Val Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn Ser Ala
915 920 925

Ile Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala Leu Gly
930 935 940

Lys Leu Gln Asn Val Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu
945 950 955 960

Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn
965 970 975

Asp Ile Leu Ser Arg Leu Asp Pro Pro Glu Ala Glu Val Gln Ile Asp
980 985 990

Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln
995 1000 1005

Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu Ala
1010 1015 1020

Ala Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val
1025 1030 1035

Asp Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser
1040 1045 1050

Ala Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala
    1055   1060   1065

Gln Glu Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly
    1070   1075   1080

Lys Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr
    1085   1090   1095

His Trp Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile
    1100   1105   1110

Thr Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile
    1115   1120   1125

Gly Ile Val Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro Glu Leu
    1130   1135   1140

Asp Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn His Thr
    1145   1150   1155

Ser Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn Ala Ser
    1160   1165   1170

Val Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu Val Ala
    1175   1180   1185

Lys Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu Gly Lys
    1190   1195   1200

Tyr Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu Gly Phe
    1205   1210   1215

Ile Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met Leu Cys
    1220   1225   1230

Cys Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys Ser Cys

1235 1240 1245

Gly Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val Leu
1250 1255 1260

Lys Gly Val Lys Leu His Tyr Thr
1265 1270

<210> 12
<211> 1270
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic protein

<400> 12

Met Phe Val Phe Leu Val Leu Leu Pro Leu Val Ser Ser Gln Cys Val
1 5 10 15

Asn Leu Thr Thr Arg Thr Gln Leu Pro Pro Ala Tyr Thr Asn Ser Phe
20 25 30

Thr Arg Gly Val Tyr Tyr Pro Asp Lys Val Phe Arg Ser Ser Val Leu
35 40 45

His Ser Thr Gln Asp Leu Phe Leu Pro Phe Phe Ser Asn Val Thr Trp
50 55 60

Phe His Val Ile Ser Gly Thr Asn Gly Thr Lys Arg Phe Asp Asn Pro
65 70 75 80

Val Leu Pro Phe Asn Asp Gly Val Tyr Phe Ala Ser Ile Glu Lys Ser
85 90 95

Asn Ile Ile Arg Gly Trp Ile Phe Gly Thr Thr Leu Asp Ser Lys Thr
100 105 110

Gln Ser Leu Leu Ile Val Asn Asn Ala Thr Asn Val Val Ile Lys Val

        115   120   125

Cys Glu Phe Gln Phe Cys Asn Asp Pro Phe Leu Asp His Lys Asn Asn
    130   135   140

Lys Ser Trp Met Glu Ser Glu Phe Arg Val Tyr Ser Ser Ala Asn Asn
145   150   155   160

Cys Thr Phe Glu Tyr Val Ser Gln Pro Phe Leu Met Asp Leu Glu Gly
                165   170   175

Lys Gln Gly Asn Phe Lys Asn Leu Arg Glu Phe Val Phe Lys Asn Ile
            180   185   190

Asp Gly Tyr Phe Lys Ile Tyr Ser Lys His Thr Pro Ile Ile Val Glu
        195   200   205

Pro Glu Arg Asp Leu Pro Gln Gly Phe Ser Ala Leu Glu Pro Leu Val
    210   215   220

Asp Leu Pro Ile Gly Ile Asn Ile Thr Arg Phe Gln Thr Leu Leu Ala
225   230   235   240

Leu His Arg Ser Tyr Leu Thr Pro Gly Asp Ser Ser Ser Gly Trp Thr
                245   250   255

Ala Gly Ala Ala Ala Tyr Tyr Val Gly Tyr Leu Gln Pro Arg Thr Phe
            260   265   270

Leu Leu Lys Tyr Asn Glu Asn Gly Thr Ile Thr Asp Ala Val Asp Cys
        275   280   285

Ala Leu Asp Pro Leu Ser Glu Thr Lys Cys Thr Leu Lys Ser Phe Thr
    290   295   300

Val Glu Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val Gln Pro Thr
305   310   315   320

Glu Ser Ile Val Arg Phe Pro Asn Ile Thr Asn Leu Cys Pro Phe Asp
                325   330   335

Glu Val Phe Asn Ala Thr Arg Phe Ala Ser Val Tyr Ala Trp Asn Arg
            340   345   350

Lys Arg Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu Tyr Asn Leu
        355   360   365

Ala Pro Phe Phe Thr Phe Lys Cys Tyr Gly Val Ser Pro Thr Lys Leu
    370   375   380

Asn Asp Leu Cys Phe Thr Asn Val Tyr Ala Asp Ser Phe Val Ile Arg
385   390   395   400

Gly Asp Glu Val Arg Gln Ile Ala Pro Gly Gln Thr Gly Asn Ile Ala
                405   410   415

Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Thr Gly Cys Val Ile Ala
            420   425   430

Trp Asn Ser Asn Lys Leu Asp Ser Lys Val Ser Gly Asn Tyr Asn Tyr
        435   440   445

Leu Tyr Arg Leu Phe Arg Lys Ser Asn Leu Lys Pro Phe Glu Arg Asp
    450   455   460

Ile Ser Thr Glu Ile Tyr Gln Ala Gly Asn Lys Pro Cys Asn Gly Val
465   470   475   480

Ala Gly Phe Asn Cys Tyr Phe Pro Leu Arg Ser Tyr Ser Phe Arg Pro
                485   490   495

Thr Tyr Gly Val Gly His Gln Pro Tyr Arg Val Val Val Leu Ser Phe
            500   505   510

Glu Leu Leu His Ala Pro Ala Thr Val Cys Gly Pro Lys Lys Ser Thr

515 520 525

Asn Leu Val Lys Asn Lys Cys Val Asn Phe Asn Phe Asn Gly Leu Lys
530 535 540

Gly Thr Gly Val Leu Thr Glu Ser Asn Lys Lys Phe Leu Pro Phe Gln
545 550 555 560

Gln Phe Gly Arg Asp Ile Ala Asp Thr Thr Asp Ala Val Arg Asp Pro
565 570 575

Gln Thr Leu Glu Ile Leu Asp Ile Thr Pro Cys Ser Phe Gly Gly Val
580 585 590

Ser Val Ile Thr Pro Gly Thr Asn Thr Ser Asn Gln Val Ala Val Leu
595 600 605

Tyr Gln Gly Val Asn Cys Thr Glu Val Pro Val Ala Ile His Ala Asp
610 615 620

Gln Leu Thr Pro Thr Trp Arg Val Tyr Ser Thr Gly Ser Asn Val Phe
625 630 635 640

Gln Thr Arg Ala Gly Cys Leu Ile Gly Ala Glu Tyr Val Asn Asn Ser
645 650 655

Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys Ala Ser Tyr Gln
660 665 670

Thr Gln Thr Lys Ser His Arg Arg Ala Arg Ser Val Ala Ser Gln Ser
675 680 685

Ile Ile Ala Tyr Thr Met Ser Leu Gly Ala Glu Asn Ser Val Ala Tyr
690 695 700

Ser Asn Asn Ser Ile Ala Ile Pro Thr Asn Phe Thr Ile Ser Val Thr
705 710 715 720

Thr Glu Ile Leu Pro Val Ser Met Thr Lys Thr Ser Val Asp Cys Thr
                725   730   735

Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ser Asn Leu Leu Leu Gln
            740   745   750

Tyr Gly Ser Phe Cys Thr Gln Leu Lys Arg Ala Leu Thr Gly Ile Ala
        755   760   765

Val Glu Gln Asp Lys Asn Thr Gln Glu Val Phe Ala Gln Val Lys Gln
    770   775   780

Ile Tyr Lys Thr Pro Pro Ile Lys Tyr Phe Gly Gly Phe Asn Phe Ser
785   790   795   800

Gln Ile Leu Pro Asp Pro Ser Lys Pro Ser Lys Arg Ser Phe Ile Glu
                805   810   815

Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe Ile Lys
            820   825   830

Gln Tyr Gly Asp Cys Leu Gly Asp Ile Ala Ala Arg Asp Leu Ile Cys
        835   840   845

Ala Gln Lys Phe Lys Gly Leu Thr Val Leu Pro Pro Leu Leu Thr Asp
    850   855   860

Glu Met Ile Ala Gln Tyr Thr Ser Ala Leu Leu Ala Gly Thr Ile Thr
865   870   875   880

Ser Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe Ala
                885   890   895

Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn Val
            900   905   910

Leu Tyr Glu Asn Gln Lys Leu Ile Ala Asn Gln Phe Asn Ser Ala Ile

915 920 925

Gly Lys Ile Gln Asp Ser Leu Ser Ser Thr Ala Ser Ala Leu Gly Lys
930 935 940

Leu Gln Asp Val Val Asn His Asn Ala Gln Ala Leu Asn Thr Leu Val
945 950 955 960

Lys Gln Leu Ser Ser Lys Phe Gly Ala Ile Ser Ser Val Leu Asn Asp
965 970 975

Ile Phe Ser Arg Leu Asp Pro Pro Glu Ala Glu Val Gln Ile Asp Arg
980 985 990

Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln Gln
995 1000 1005

Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu Ala Ala
1010 1015 1020

Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Asp
1025 1030 1035

Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ser Ala
1040 1045 1050

Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ala Gln
1055 1060 1065

Glu Lys Asn Phe Thr Thr Ala Pro Ala Ile Cys His Asp Gly Lys
1070 1075 1080

Ala His Phe Pro Arg Glu Gly Val Phe Val Ser Asn Gly Thr His
1085 1090 1095

Trp Phe Val Thr Gln Arg Asn Phe Tyr Glu Pro Gln Ile Ile Thr
1100 1105 1110

Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly
    1115   1120   1125

Ile Val Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp
    1130   1135   1140

Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn His Thr Ser
    1145   1150   1155

Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn Ala Ser Val
    1160   1165   1170

Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu Val Ala Lys
    1175   1180   1185

Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu Gly Lys Tyr
    1190   1195   1200

Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Ile Trp Leu Gly Phe Ile
    1205   1210   1215

Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Met Leu Cys Cys
    1220   1225   1230

Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Cys Cys Ser Cys Gly
    1235   1240   1245

Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val Leu Lys
    1250   1255   1260

Gly Val Lys Leu His Tyr Thr
    1265   1270

<210> 13
<211> 3813
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic nucleic acid

<400> 13
atgttcgtgt ttctggtgct gctgcctctg gtgagctccc agtgcgtgaa cttcaccaca 60

agaacccagc tgccccctgc ctacaccaat tccttcacaa ggggcgtgta ctatcccgac 120

aaggtgtttc gctctagcgt gctgcactcc acacaggatc tgtttctgcc tttcttttct 180

aacgtgacct ggttccacgc catccacgtg agcggcacca atggcacaaa gcggttcgcc 240

aatccagtgc tgccctttaa cgacggcgtg tacttcgcct ccaccgagaa gtctaacatc 300

atcagaggct ggatctttgg caccacactg gatagcaaga cacagtccct gctgatcgtg 360

aacaatgcca ccaacgtggt catcaaggtg tgcgagttcc agttttgtaa tgacccattc 420

ctgggcgtgt actatcacaa gaacaataag tcttggatgg agagcgagtt tagggtgtac 480

tcctctgcca acaattgcac atttgagtac gtgagccagc ccttcctgat ggacctggag 540

ggcaagcagg gcaatttcaa gaacctgcgc gagttcgtgt ttaagaatat cgatggctac 600

ttcaagatct actccaagca caccccaatc aacctggtga ggggactgcc acagggcttc 660

tctgccctgg agccactggt ggacctgccc atcggcatca acatcacccg ctttcagaca 720

ctgcacatca gctacctgac accaggcgat agctcctctg gatggaccgc aggagcagca 780

gcctactatg tgggctacct gcagcccagg accttcctgc tgaagtataa cgagaatggc 840

accatcacag acgcagtgga ttgcgccctg gacccctgt ctgagaccaa gtgtacactg 900

aagagcttta ccgtggagaa gggcatctac cagacaagca atttccgggt gcagcctacc 960

gagtccatcg tgagatttcc caatatcaca aacctgtgcc cttttggcga ggtgttcaac 1020

gccacccgct tcgccagcgt gtatgcctgg aataggaagc gcatctccaa ctgcgtggcc 1080

gactattctg tgctgtacaa cagcgcctcc ttctctacct ttaagtgcta cggcgtgagc 1140

cccacaaagc tgaatgacct gtgctttacc aacgtgtatg ccgattcctt cgtgatcagg 1200

ggcgacgagg tgcgccagat cgcaccaggc cagacaggca atatcgccga ctacaactat 1260

aagctgcctg acgatttcac cggctgcgtg atcgcctgga acagcaacaa tctggatagc 1320

aaagtgggcg gcaactacaa ttatctgtac cggctgttta gaaagtctaa cctgaagcca 1380 ttcgagaggg acatctccac agagatctac caggccggct ctacccccctg caatggcgtg 1440

aagggcttta actgttattt ccctctgcag agctacggct tccagccaac ctacggcgtg 1500

ggctatcagc cctaccgcgt ggtggtgctg tcttttgagc tgctgcacgc acctgcaaca 1560

gtgtgcggcc caaagaagag caccaatctg gtgaagaaca agtgcgtgaa cttcaacttc 1620

aacggactga ccggcacagg cgtgctgacc gagtccaaca agaagttcct gccttttcag 1680

cagttcggcc gggacatcgc cgataccaca gacgccgtga gagaccctca gaccctggag 1740

atcctggata tcacaccatg ctccttcggc ggcgtgtctg tgatcacacc aggcaccaat 1800

acaagcaacc aggtggccgt gctgtaccag ggcgtgaatt gtaccgaggt gcccgtggca 1860

atccacgcag accagctgac ccctacatgg agggtgtatt ctaccggcag caacgtgttc 1920

cagacacgcg ccggatgcct gatcggagca gagcacgtga acaatagcta cgagtgcgat 1980

atccctatcg gcgccggcat ctgtgcctcc tatcagaccc agacaaactc cccacggaga 2040

gcccggtctg tggcaagcca gtccatcatc gcctacacca tgagcctggg cgtggagaac 2100

agcgtggcct attccaacaa ttctatcgcc atccctacca acttcacaat ctccgtgacc 2160

acagagatcc tgccagtgag catgaccaag acatccgtgg actgcacaat gtacatctgt 2220

ggcgattcca ccgagtgctc taacctgctg ctgcagtatg gctctttttg tacccagctg 2280

aatagagccc tgacaggcat cgccgtggag caggacaaga acacacagga ggtgttcgcc 2340

caggtgaagc agatctacaa gaccccaccc atcaaggact ttggcggctt caacttcagc 2400

cagatcctgc ccgatcctag caagccatcc aagcggtctt ttatcgagga cctgctgttc 2460

aacaaggtga ccctggccga tgccggcttc atcaagcagt acggcgattg cctgggcgac 2520

atcgcagcca gagacctgat ctgtgcccag aagtttaatg gcctgaccgt gctgcctcca 2580

ctgctgacag atgagatgat cgcccagtat acatctgccc tgctggcagg aaccatcaca 2640

agcggatgga ccttcggcgc aggagccgcc ctgcagatcc cctttgccat gcagatggcc 2700

tacaggttca acggcatcgg cgtgacccag aatgtgctgt atgagaacca gaagctgatc 2760

gccaatcagt ttaactccgc catcggcaag atccaggact ctctgagctc cacagcaagc 2820

gccctgggca agctgcagga tgtggtgaat cagaacgccc aggccctgaa taccctggtg 2880 aagcagctgt ctagcaactt cggcgccatc tcctctgtgc tgaatgatat cctgagccgg 2940

ctggaccctc ctgaggcaga ggtgcagatc gaccggctga tcacaggcag actgcagtcc 3000

ctgcagacct acgtgacaca gcagctgatc agggcagcag agatcagggc atctgccaat 3060

ctggccgcca ccaagatgag cgagtgcgtg ctgggccagt ccaagagagt ggacttttgt 3120

ggcaagggct accacctgat gagcttccca cagtccgccc ctcacggcgt ggtgtttctg 3180

cacgtgacct atgtgccagc ccaggagaag aacttcacca cagcaccagc catctgccac 3240

gatggcaagg cacactttcc tcgggagggc gtgttcgtga gcaacggcac ccactggttt 3300

gtgacacaga gaaatttcta cgagccacag atcatcacca cagacaatac cttcgtgagc 3360

ggcaactgtg acgtggtcat cggaatcgtg aacaataccg tgtacgatcc tctgcagcca 3420

gagctggact cttttaagga ggagctggat aagtatttca agaatcacac cagccccgac 3480

gtggatctgg gcgacatctc tggcatcaat gccagcgtgg tgaacatcca gaaggagatc 3540

gaccgcctga acgaggtggc caagaatctg aacgagtccc tgatcgatct gcaggagctg 3600

ggcaagtatg agcagtacat caagtggccc tggtacatct ggctgggctt catcgccggc 3660

ctgatcgcca tcgtgatggt gaccatcatg ctgtgctgta tgacaagctg ctgttcctgc 3720

ctgaagggct gctgttcttg tggcagctgc tgtaagtttg atgaggacga tagcgagcct 3780

gtgctgaagg gcgtgaagct gcactatacc tga 3813

<210> 14
<211> 3822
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic nucleic acid

<400> 14
atgttcgtgt ttctggtgct gctgcctctg gtgagctccc agtgcgtgaa cctgaccaca 60

aggacccagc tgccccctgc ctacaccaat tccttcacac ggggcgtgta ctatcccgac 120

aaggtgttta gatctagcgt gctgcactcc acacaggatc tgtttctgcc tttcttttct 180

aacgtgacct ggttccacgc catccacgtg agcggcacca atggcacaaa gcggttcgac 240 aatccagtgc tgccctttaa cgatggcgtg tacttcgcct ccaccgagaa gtctaacatc 300

atcagaggct ggatctttgg caccacactg gacagcaaga cacagtccct gctgatcgtg 360

aacaatgcca ccaacgtggt catcaaggtg tgcgagttcc agttttgtaa tgatccattc 420

ctgggcgtgt actatcacaa gaacaataag tcttggatgg agagcgagtt tcgcgtgtac 480

tcctctgcca acaattgcac atttgagtac gtgagccagc ccttcctgat ggacctggag 540

ggcaagcagg gcaatttcaa gaacctgagg gagttcgtgt ttaagaatat cgatggctac 600

ttcaagatct actccaagca caccccaatc aacctggtgc gcgacctgcc acagggcttc 660

tctgccctgg agccactggt ggatctgccc atcggcatca acatcacccg gtttcagaca 720

ctgctggccc tgcacagaag ctacctgaca ccaggcgaca gctcctctgg atggaccgca 780

ggagcagcag cctactatgt gggctacctg cagcccagga ccttcctgct gaagtataac 840

gagaatggca ccatcacaga cgcagtggat tgcgccctgg acccccctgtc tgagaccaag 900

tgtacactga agagctttac cgtggagaag ggcatctacc agacaagcaa tttcagggtg 960

cagcctaccg agtccatcgt gcgctttccc aatatcacaa acctgtgccc ttttggcgag 1020

gtgttcaacg ccacccgctt cgccagcgtg tatgcctgga ataggaagcg catctccaac 1080

tgcgtggccg actattctgt gctgtacaac agcgcctcct tctctacctt taagtgctac 1140

ggcgtgagcc ccacaaagct gaatgacctg tgctttacca acgtgtatgc cgattccttc 1200

gtgatcaggg gcgacgaggt gcgccagatc gcaccaggcc agacaggcaa tatcgccgac 1260

tacaactata agctgcctga cgatttcacc ggctgcgtga tcgcctggaa cagcaacaat 1320

ctggatagca aagtgggcgg caactacaat tatctgtacc ggctgtttag aaagtctaac 1380

ctgaagccat tcgagaggga catctccaca gagatctacc aggccggctc tacccccctgc 1440

aatggcgtga agggctttaa ctgttatttc cctctgcaga gctacggctt ccagccaacc 1500

tacggcgtgg gctatcagcc ctaccgcgtg gtggtgctgt cttttgagct gctgcacgca 1560

cctgcaacag tgtgcggccc aaagaagagc accaatctgg tgaagaacaa gtgcgtgaac 1620

ttcaacttca acggactgac cggcacaggc gtgctgaccg agtccaacaa gaagttcctg 1680

ccttttcagc agttcggcag ggacatcgca gataccacag acgccgtgcg cgaccctcag 1740 accctggaga tcctggatat cacaccatgc tccttcggcg gcgtgtctgt gatcacacca 1800

ggcaccaata caagcaacca ggtggccgtg ctgtaccagg gcgtgaattg taccgaggtg 1860

cccgtggcaa tccacgcaga ccagctgacc cctacatggc gggtgtattc taccggcagc 1920

aacgtgttcc agacaagagc cggatgcctg atcggagcag agcacgtgaa caatagctac 1980

gagtgcgata tccctatcgg cgccggcatc tgtgcctcct atcagaccca gacaaactcc 2040

ccacggagag cccggtctgt ggcaagccag tccatcatcg cctacaccat gagcctgggc 2100

gccgagaaca gcgtggccta ttccaacaat tctatcgcca tccctaccaa cttcacaatc 2160

tccgtgacca cagagatcct gccagtgagc atgaccaaga catccgtgga ctgcacaatg 2220

tacatctgtg gcgattccac cgagtgctct aacctgctgc tgcagtatgg ctctttttgt 2280

acccagctga atagagccct gacaggcatc gccgtggagc aggacaagaa cacacaggag 2340

gtgttcgccc aggtgaagca gatctacaag accccaccca tcaaggactt tggcggcttc 2400

aacttcagcc agatcctgcc cgatcctagc aagccatcca agcggtcttt tatcgaggac 2460

ctgctgttca acaaggtgac cctggccgat gccggcttca tcaagcagta cggcgattgc 2520

ctgggcgaca tcgcagccag agacctgatc tgtgcccaga agtttaatgg cctgaccgtg 2580

ctgcctccac tgctgacaga tgagatgatc gcccagtata catctgccct gctggcagga 2640

accatcacaa gcggatggac cttcggcgca ggagccgccc tgcagatccc ctttgccatg 2700

cagatggcct acagattcaa cggcatcggc gtgacccaga atgtgctgta tgagaaccag 2760

aagctgatcg ccaatcagtt taactccgcc atcggcaaga tccaggactc tctgagctcc 2820

acagcaagcg ccctgggcaa gctgcaggat gtggtgaatc agaacgccca ggccctgaat 2880

accctggtga agcagctgtc tagcaacttc ggcgccatct cctctgtgct gaatgatatc 2940

ctgagccggc tggacccacc agaggcagag gtgcagatcg accggctgat cacaggcaga 3000

ctgcagtccc tgcagaccta cgtgacacag cagctgatca gggcagcaga gatcagggca 3060

tctgccaatc tggccgccac caagatgagc gagtgcgtgc tgggccagtc caagagagtg 3120

gacttttgtg gcaagggcta ccacctgatg agcttcccac agtccgcccc tcacggcgtg 3180

gtgtttctgc acgtgaccta tgtgccagcc caggagaaga acttcaccac agcaccagcc 3240 atctgccacg atggcaaggc acactttccc cgggagggcg tgttcgtgag caacggaacc 3300

cactggtttg tgacacagcg caatttctac gagccacaga tcatcaccac agacaataca 3360

ttcgtgtccg gcaactgtga cgtggtcatc ggaatcgtga acaataccgt gtacgatcct 3420

ctgcagccag agctggactc ttttaaggag gagctggata agtatttcaa gaatcacacc 3480

agccccgacg tggatctggg cgacatctct ggcatcaatg ccagcgtggt gaacatccag 3540

aaggagatcg acaggctgaa cgaggtggcc aagaatctga acgagtccct gatcgatctg 3600

caggagctgg gcaagtatga gcagtacatc aagtggccct ggtacatctg gctgggcttc 3660

atcgccggcc tgatcgccat cgtgatggtg accatcatgc tgtgctgtat gacaagctgc 3720

tgttcctgcc tgaagggctg ctgttcttgt ggcagctgct gtaagtttga tgaggacgat 3780

agcgagcctg tgctgaaggg cgtgaagctg cactatacct ga 3822

<210> 15
<211> 3822
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic nucleic acid

<400> 15
atgttcgtgt ttctggtgct gctgcctctg gtgagctccc agtgcgtgaa cctgaccaca 60

accacacagc tgccccctgc ctataccaat tccttcacac gcggcgtgta ctatcctgac 120

aaggtgtttc ggtctagcgt gctgcactcc acacaggatc tgtttctgcc attcttttct 180

aacgtgacct ggttccacgc catccacgtg agcggcacca atggcacaaa gcggttcgac 240

aatccagtgc tgccctttaa cgatggcgtg tacttcgcct ccaccgagaa gtctaacatc 300

atccggggct ggatctttgg caccacactg gacagcaaga cacagtccct gctgatcgtg 360

aacaatgcca ccaacgtggt catcaaggtg tgcgagttcc agttttgtaa tgatcccttc 420

ctggacgtgt actatcacaa gaacaataag tcttggatga agagcgagtt tagagtgtat 480

tcctctgcca acaattgcac atttgagtac gtgtcccagc ctttcctgat ggacctggag 540

ggcaagcagg gcaatttcaa gaacctgaga gagttcgtgt ttaagaatat cgatggctac 600 ttcaagatct actccaagca caccccaatc aacctggtga gggacctgcc acacggcttc 660

tctgccctgg agccactggt ggatctgccc atcggcatca acatcaccag atttcagaca 720

ctgctggccc tgcacaggag ctacctgaca cccggcgaca gctcctctgg atggaccgcc 780

ggcgctgccg cctactatgt gggctatctg cagcctcgca ccttcctgct gaagtacaac 840

gagaatggca ccatcacaga cgcagtggat tgcgccctgg accccctgtc tgagaccaag 900

tgtacactga agagctttac cgtggagaag ggcatctatc agacaagcaa tttccgcgtg 960

cagccaaccg agtccatcgt gcggtttccc aatatcacaa acctgtgccc ttttggcgag 1020

gtgttcaacg caaccaggtt cgcaagcgtg tacgcatgga atcgcaagcg gatctccaac 1080

tgcgtggccg actattctgt gctgtacaac agcgcctcct tctctacctt taagtgctat 1140

ggcgtgagcc caacaaagct gaatgacctg tgctttacca acgtgtacgc cgattccttc 1200

gtgatccggg gcgacgaggt gcggcagatc gcaccaggac agacaggcaa gatcgcagac 1260

tacaattata agctgcctga cgatttcacc ggctgcgtga tcgcctggaa ctctaacaat 1320

ctggatagca aagtgggcgg caactacaat tatagataca ggctgtttag aaagtctaat 1380

ctgaagccat tcgagaggga catctccaca gagatctacc aggccggctc taccccctgc 1440

aatggcgtgc agggctttaa ctgttatttc cctctgcaga gctacggctt ccagccaacc 1500

aacggcgtgg gctatcagcc ctaccgggtg gtggtgctgt cttttgagct gctgcacgca 1560

cctgcaacag tgtgcggacc aaagaagagc accaatctgg tgaagaacaa gtgcgtgaac 1620

ttcaacttca acggactgac cggaacaggc gtgctgaccg agtccaacaa gaagttcctg 1680

ccatttcagc agttcggcag agacatcgcc gataccacag acgccgtgag ggaccctcag 1740

accctggaga tcctggatat cacaccatgc tccttcggcg gcgtgtctgt gatcacaccc 1800

ggcaccaata caagcaacca ggtggccgtg ctgtatcagg gcgtgaattg taccgaggtg 1860

ccagtggcaa tccacgcaga ccagctgacc cctacatggc gcgtgtactc taccggcagc 1920

aacgtgttcc agacaagggc aggatgcctg atcggagcag agcacgtgaa caatagctat 1980

gagtgcgata tccccatcgg cgccggcatc tgtgcctcct accagaccca gacaaactcc 2040

cggagaaggg ccagatctgt ggccagccag tccatcatcg cctataccat gagcctgggc 2100 gccgagaaca gcgtggccta ctccaacaat tctatcgcca tccctaccaa cttcacaatc 2160

tccgtgacca cagagatcct gccagtgagc atgaccaaga catccgtgga ctgcacaatg 2220

tatatctgtg gcgattccac cgagtgctct aacctgctgc tgcagtacgg ctctttttgt 2280

acccagctga atagggccct gacaggaatc gcagtggagc aggacaagaa cacacaggag 2340

gtgttcgccc aggtgaagca gatctacaag accccaccca tcaaggactt tggcggcttc 2400

aacttcagcc agatcctgcc cgatcctagc aagccctcca agcggagctt catcgaggac 2460

ctgctgttca acaaggtgac cctggccgat gccggcttca tcaagcagta tggcgattgc 2520

ctgggcgaca tcgcagcaag ggacctgatc tgtgcccaga agtttaatgg cctgaccgtg 2580

ctgcctccac tgctgacaga tgagatgatc gcccagtaca catctgccct gctggcagga 2640

accatcacaa gcggatggac cttcggcgca ggagccgccc tgcagatccc ttttgccatg 2700

cagatggcct atcgcttcaa cggcatcggc gtgacccaga atgtgctgta cgagaaccag 2760

aagctgatcg ccaatcagtt taactccgcc atcggcaaga tccaggactc tctgagctcc 2820

acagcaagcg ccctgggcaa gctgcaggat gtggtgaatc agaacgccca ggccctgaat 2880

accctggtga agcagctgtc tagcaacttc ggcgccatct cctctgtgct gaatgatatc 2940

ctgagcagac tggacccccc cgaggccgag gtgcagatcg acagactgat cacaggcagg 3000

ctgcagtccc tgcagaccta cgtgacacag cagctgatca gggccgccga gatcagggcc 3060

tctgccaatc tggccgccac caagatgagc gagtgcgtgc tgggccagtc caagagggtg 3120

gatttttgtg gcaagggcta tcacctgatg agcttcccac agtccgcccc tcacggagtg 3180

gtgtttctgc acgtgaccta cgtgccagcc caggagaaga acttcaccac agcaccagca 3240

atctgccacg acggcaaggc acactttcca agagagggcg tgttcgtgag caacggcacc 3300

gattggtttg tgacacagag gaatttctac gagccccaga tcatcaccac agacaataca 3360

ttcgtgtccg gcaactgtga cgtggtcatc ggcatcgtga acaataccgt gtatgatcct 3420

ctgcagccag agctggactc ttttaaggag gagctggata agtacttcaa gaatcacacc 3480

agccccgacg tggatctggg cgacatctct ggcatcaatg ccagcgtggt gaacatccag 3540

aaggagatcg accggctgaa cgaggtggcc aagaatctga acgagtccct gatcgatctg 3600 caggagctgg gcaagtatga gcagtacatc aagtggcctt ggtatatctg gctgggcttc 3660

atcgccggcc tgatcgccat cgtgatggtg accatcatgc tgtgctgtat gacaagctgc 3720

tgttcctgcc tgaagggctg ctgttcttgt ggcagctgct gtaagtttga tgaggacgat 3780

agcgagccag tgctgaaggg cgtgaagctg cactacacct ga 3822

<210> 16
<211> 3822
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic nucleic acid

<400> 16
atgttcgtgt ttctggtgct gctgcctctg gtgagctccc agtgcgtgaa tttcaccaac 60

agaacacagc tgccttctgc ctacaccaat agcttcacac ggggcgtgta ctatccagac 120

aaggtgttta gatctagcgt gctgcacagc acacaggatc tgtttctgcc attcttttcc 180

aacgtgacct ggttccacgc catccacgtg tccggcacca atggcacaaa gcggttcgac 240

aatcccgtgc tgccttttaa cgatggcgtg tacttcgcct ccaccgagaa gtctaacatc 300

atcagaggct ggatctttgg caccacactg gacagcaaga cacagtccct gctgatcgtg 360

aacaatgcca ccaacgtggt catcaaggtg tgcgagttcc agttttgtaa ttatcccttc 420

ctgggcgtgt actatcacaa gaacaataag tcttggatgg agagcgagtt tagggtgtac 480

tcctctgcca acaattgcac atttgagtat gtgagccagc ctttcctgat ggacctggag 540

ggcaagcagg gcaatttcaa gaacctgagc gagttcgtgt ttaagaatat cgatggctac 600

ttcaagatct actccaagca cacccccatc aacctggtgc gcgacctgcc tcagggcttc 660

tctgccctgg agcccctggt ggatctgcct atcggcatca acatcacccg gtttcagaca 720

ctgctggccc tgcacagaag ctacctgaca cccggcgaca gctcctctgg atggaccgcc 780

ggcgctgccg cctactatgt gggctacctg cagcctagga ccttcctgct gaagtataac 840

gagaatggca ccatcacaga cgcagtggat tgcgccctgg accccctgtc cgagaccaag 900

tgtacactga agtcttttac cgtggagaag ggcatctacc agacatctaa tttcagggtg 960 cagccaaccg agagcatcgt gcgctttcct aatatcacaa acctgtgccc atttggcgag 1020

gtgttcaacg ccacccgctt cgccagcgtg tatgcctgga ataggaagcg catcagcaac 1080

tgcgtggccg actattccgt gctgtacaac agcgcctcct tctctacctt taagtgttac 1140

ggcgtgtctc ctacaaagct gaatgacctg tgctttacca acgtgtatgc cgatagcttc 1200

gtgatcaggg gcgacgaggt gcgccagatc gcaccaggac agaccggaac aatcgcagac 1260

tacaattata agctgcctga cgatttcacc ggctgcgtga tcgcctggaa ctccaacaat 1320

ctggattcta aagtgggcgg caactacaat tatctgtacc ggctgtttag aaagtccaac 1380

ctgaagccat tcgagcggga catcagcaca gagatctacc aggcaggctc caccccatgc 1440

aatggagtga agggctttaa ctgttatttc ccactgcaga gctacggctt ccagcccaca 1500

tatggcgtgg gctatcagcc ttacagagtg gtggtgctgt cctttgagct gctgcacgca 1560

ccagcaacag tgtgcggacc caagaagtct accaatctgg tgaagaacaa gtgcgtgaac 1620

ttcaacttca acggactgac cggaacaggc gtgctgaccg agtccaacaa gaagttcctg 1680

ccatttcagc agttcggcag ggacatcgca gataccacag acgccgtgcg cgacccacag 1740

accctggaga tcctggatat cacaccctgc agcttcggcg gcgtgtccgt gatcacacca 1800

ggaaccaata caagcaacca ggtggccgtg ctgtaccagg gcgtgaattg taccgaggtg 1860

cctgtggcaa tccacgcaga ccagctgacc ccaacatggc gggtgtattc taccggcagc 1920

aacgtgttcc agacaagagc cggctgcctg atcggcgccg agtatgtgaa caattcttac 1980

gagtgcgata tccctatcgg cgccggcatc tgtgccagct accagaccca gacaaacagc 2040

ccacggagag cacggtccgt ggcaagccag tccatcatcg cctacaccat gtctctgggc 2100

gccgagaata gcgtggccta ttccaacaat tctatcgcca tcccaaccaa cttcacaatc 2160

tccgtgacca cagagatcct gcccgtgtct atgaccaaga caagcgtgga ctgcacaatg 2220

tacatctgtg gcgattccac cgagtgctct aacctgctgc tgcagtatgg cagcttttgt 2280

acccagctga atagagccct gacaggcatc gccgtggagc aggacaagaa cacacaggag 2340

gtgttcgccc aggtgaagca gatctacaag acccccccta tcaaggactt tggcggcttc 2400

aacttcagcc agatcctgcc tgatccaagc aagccatcca agaggtcttt tatcgaggac 2460 ctgctgttca acaaggtgac cctggccgat gccggcttca tcaagcagta cggcgattgc 2520

ctgggcgaca tcgcagcaag ggacctgatc tgtgcccaga agtttaatgg cctgaccgtg 2580

ctgccacccc tgctgacaga tgagatgatc gcccagtata catccgccct gctggccggc 2640

accatcacat ctggatggac cttcggcgca ggagccgccc tgcagatccc ctttgccatg 2700

cagatggcct acaggttcaa cggcatcggc gtgacccaga atgtgctgta tgagaaccag 2760

aagctgatcg ccaatcagtt taactccgcc atcggcaaga tccaggactc cctgagctcc 2820

acagcctctg ccctgggcaa gctgcaggat gtggtgaatc agaacgccca ggccctgaat 2880

accctggtga agcagctgtc tagcaacttc ggcgccatct cctctgtgct gaatgatatc 2940

ctgagccggc tggacccccc cgaggcagag gtgcagatcg accggctgat caccggcaga 3000

ctgcagagcc tgcagaccta cgtgacacag cagctgatca gggccgccga gatcagggca 3060

tccgccaatc tggccgccat caagatgtct gagtgcgtgc tgggccagag caagagagtg 3120

gacttttgtg gcaagggcta ccacctgatg agcttccctc agtccgcccc acacggagtg 3180

gtgtttctgc acgtgaccta tgtgcccgcc caggagaaga acttcaccac agcccctgcc 3240

atctgccacg atggcaaggc ccactttcca agggagggcg tgttcgtgtc caacggcacc 3300

cactggtttg tgacacagcg caatttctac gagccccaga tcatcaccac agacaatacc 3360

ttcgtgagcg gcaactgtga cgtggtcatc ggcatcgtga acaataccgt gtacgatcca 3420

ctgcagcccg agctggactc ctttaaggag gagctggata agtatttcaa gaatcacacc 3480

tctcccgacg tggatctggg cgacatctcc ggcatcaatg cctctttcgt gaacatccag 3540

aaggagatcg accgcctgaa cgaggtggcc aagaatctga acgagtccct gatcgatctg 3600

caggagctgg gcaagtatga gcagtacatc aagtggccct ggtacatctg gctgggcttc 3660

atcgccggcc tgatcgccat cgtgatggtg accatcatgc tgtgctgtat gacaagctgc 3720

tgttcctgcc tgaagggctg ctgttcttgt ggcagctgct gtaagtttga tgaggacgat 3780

agcgagcctg tgctgaaggg cgtgaagctg cactatacct ga 3822

<210> 17
<211> 3816
<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic nucleic acid

<400> 17
atgtttgtgt ttctggtgct gctgccactg gtgagtagcc agtgtgtgaa cctgagaacc   60

cgaacacagc tgcctcctgc ctataccaac agcttcacca gaggcgtgta ctaccctgac   120

aaggtgttcc gatctagcgt gctccatagc acccaggacc tgttcttgcc tttttttctct   180

aacgtgacat ggttccacgc cattcacgtg tctggcacca acggaacaaa aagattcgac   240

aaccctgtgc tgcccttcaa cgacggtgtc tattttgcca gcaccgagaa gagcaacatc   300

atcagaggct ggatcttcgg aaccaccctg gacagcaaga cccagagcct gctgatcgtc   360

aataacgcaa caaatgtggt gatcaaggtg tgcgagttcc aattttgcaa cgatcctttc   420

ctggatgtgt actaccacaa gaacaacaaa agctggatgg aaagtggagt ttatagcagc   480

gccaacaact gcaccttcga gtacgtgagc caacctttcc tgatggacct cgaagggaaa   540

cagggcaact tcaagaacct tagagagttc gtctttaaga acatcgacgg ctactttaaa   600

atctactcca agcacacccc catcaacctg gtgcgggacc tgcctcaggg ctttagcgcg   660

ctggaaccct tggttgacct gcccatcggc atcaacatca ctagattcca gacccttctg   720

gccctccacc ggtcttacct gacacctggc gacagtagtt ctggctggac agccggcgcc   780

gctgcctact acgtgggcta tctgcagcct agaaccttcc tgctgaagta caacgagaac   840

ggcaccatca ccgacgctgt ggattgcgcc ctggaccctc tgtccgaaac caagtgcaca   900

ctgaagtcct tcaccgtgga aaagggcatc taccagacct ctaacttccg ggtgcagcct   960

actgaaagca tcgtgcggtt cccaaacatt acaaacctgt gccctttcgg agaagttttc   1020

aacgccactc gcttcgcctc tgtctatgcc tggaacagaa agcggatcag caattgtgtg   1080

gccgattaca gcgtgctgta caacagcgcc agcttttcta cattcaagtg ctacggcgtg   1140

tctcccacca agctgaatga tctgtgcttc accaacgtgt acgccgactc gtttgtgatc   1200

cggggagacg aagtgcgcca gatcgcccct gggcagacag gaaacatcgc cgattacaat   1260

tacaaactgc ctgacgattt tacaggatgt gtgatagctt ggaactccaa caacctcgac   1320 agcaaagtgg gcggcaacta caattaccgg tacagactgt ttagaaagag caacctaaaa 1380

cccttcgaga gagatatctc taccgagatc taccaggccg gcagcaagcc ttgtaatggc 1440

gttgagggct tcaactgtta cttccctctg cagagctacg gcttccagcc caccaacggc 1500

gtcgggtacc agccttacag agttgtggtt ctgagcttcg agctgctcca cgctcctgcc 1560

accgtgtgtg gtcctaagaa aagcaccaac ctggtgaaga acaagtgcgt gaatttcaat 1620

ttcaacggcc tgacaggcac aggcgtgctg accgagagca acaaaaagtt cctgcccttc 1680

cagcagttcg gcagagatat tgccgatacc acagacgccg tgcgggaccc tcaaaccctg 1740

gaaatcttgg acatcacacc ttgcagcttc ggcggagtgt ctgtgatcac tcccgggacc 1800

aacaccagca accaggttgc cgtgctgtac cagggcgtca actgcaccga agtgccagtg 1860

gctatacacg ccgaccagct gacccctaca tggcgggtgt acagcaccgg cagcaacgtg 1920

ttccagacca gagccggctg cctgatcggc gcagagcacg tgaacaactc ttatgaatgc 1980

gacatcccca tcggagccgg catttgcgcc agctaccaga cacagaccaa tagcagaaga 2040

cgggctagaa gcgtggcctc gcagagcata atcgcataca caatgagcct gggagccgag 2100

aacagcgtgg cctacagcaa caatagtatc gccatcccca caaattttac catcagcgtg 2160

acaaccgaaa tcctgccagt gagcatgaca aagaccagcg tcgactgcac aatgtacata 2220

tgtggcgata gcacggagtg cagcaatctg ctgctccaat acggcagctt ctgcacccag 2280

ctgaatcggg cactgaccgg catcgccgtg gaacaggata aaaataccca ggaggtgttt 2340

gcccaggtga agcagatata taagacccct ccgatcaagg acttcggagg cttcaatttc 2400

agccagatcc tgcccgatcc aagcaagcct agcaagcggt ccttcatcga ggatctgctg 2460

ttcaataagg tgaccctggc cgacgccgga ttcatcaaac agtacggcga ctgcctgggc 2520

gacatcgccg ccagagatct gatctgtgct caaaagttca acggactgac agtcctgcca 2580

cctctgttga cagatgaaat gatcgctcag tacacctccg ccctcctggc cgggacgatc 2640

acctctggat ggaccttcgg cgccggcgct gcactgcaga tccctttcgc catgcagatg 2700

gcctacagat tcaacggcat cggagtgacc caaaacgtcc tgtacgagaa ccagaagctg 2760

atcgccaacc agttcaactc tgctatcggc aagatccagg acagcctcag cagcaccgcc 2820 agcgccctgg gcaaactcca gaacgtggtg aaccagaacg cacaggccct gaataccctg 2880

gtgaagcagc tgagcagcaa cttcggcgct atcagctctg tgctgaacga catcctgagc 2940

agactggacc ctcccgaggc cgaggtgcag attgacaggc tgatcacagg cagactgcag 3000

tcgctgcaaa cttacgtgac ccagcaactg atccgggccg ccgaaatcag ggccagcgcc 3060

aacctggctg ctacaaagat gtccgaatgc gtgttgggcc agtccaagag agtggacttc 3120

tgcggcaagg gataccacct gatgagcttc cctcagtccg ctccccacgg cgtcgtgttc 3180

ctgcatgtga catacgtgcc cgcccaggag aagaatttca ccaccgcccc tgccatctgc 3240

cacgacggca aggcccactt ccccagagag ggcgtgttcg tgtccaacgg cacccactgg 3300

ttcgtgaccc agcggaactt ctacgagcct cagatcatca ccaccgataa cacattcgtg 3360

tccggcaact gcgacgtggt tatcggcatc gtgaacaata ccgtgtacga ccctctgcag 3420

ccagaactgg attcttttaa ggaagagctg gacaaatact ttaagaacca cacatctcct 3480

gatgtggacc tgggcgacat cagcggcatc aacgcctccg tggtcaacat ccaaaaggag 3540

atcgatagac tgaacgaggt ggccaagaac ctcaacgagt ctctgattga cctgcaggag 3600

ctgggcaagt acgagcagta catcaagtgg ccttggtaca tctggctggg cttcatcgcc 3660

ggcctgatcg ctatcgtcat ggtgaccatc atgctgtgct gtatgacctc ctgctgcagc 3720

tgtctgaaag gctgctgttc ttgcggcagc tgttgcaagt ttgacgagga cgactccgag 3780

cccgtgctga agggggtgaa gctgcactac acgtga 3816

<210> 18
<211> 3813
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic nucleic acid

<400> 18
atgttcgtgt tcctggtgct gctgcccctg gtgtctagcc aatgtgtgaa cctgacaaca 60

aggacccagc ttcccccagc ttacaccaat tcatttacaa gaggcgtgta ttaccccgat 120

aaggtgttcc gaagcagcgt gctgcacagc acccaggatc tcttcctgcc tttttttcagc 180 aatgtgactt ggttccacgt gatcagcgga accaacggca ccaagcggtt tgacaatcct 240

gtgctgccct tcaacgacgg cgtgtacttc gccagcatcg agaagagcaa cattatccgg 300

ggctggatct tcggcaccac cctcgatagc aagacccaga gcttactgat cgtaaacaac 360

gccaccaatg tcgtaatcaa ggtctgtgaa tttcagttct gcaacgaccc ctttctggac 420

cacaagaaca acaagtcgtg gatggaaagc gagttcagag tgtacagctc cgctaacaat 480

tgtacattcg agtacgtgtc tcagcctttc ctgatggacc tggaaggcaa gcagggaaac 540

ttcaagaatc tgagggagtt cgtgttcaaa aacatcgacg gctacttcaa gatctacagc 600

aagcataccc ccatcatcgt tgaacctgag agagacctgc cacagggttt cagcgctctg 660

gagcctctgg ttgacctgcc catcggcatc aacatcaccc ggtttcagac actgttagcc 720

ctgcatagat cttacctgac cccaggcgat tcttcctctg gctggaccgc cggagccgca 780

gcctactacg tgggatatct gcagcccaga accttcctgc tgaaatacaa cgagaacgga 840

accatcaccg atgccgtgga ctgcgccctg gaccctctgt ctgaaaccaa gtgcaccctg 900

aagagcttca ccgtggaaaa gggcatctac cagaccagca actttcgggt gcagcccacc 960

gagagcatcg tgagatttcc aaacatcacc aacctgtgtc ctttcgacga ggtgtttaat 1020

gccacaagat tcgccagcgt gtacgcctgg aatagaaaaa gaatctccaa ctgcgtggct 1080

gattactcag tgctttacaa cctggcccca ttcttcacct tcaagtgcta cggcgttagc 1140

cctaccaagc tcaatgatct gtgcttcacg aacgtgtacg ccgacagctt cgtgatccgg 1200

ggcgacgaag tcagacagat cgcccctgga cagaccggta atatcgccga ctacaattac 1260

aagctgcctg atgatttcac aggttgcgtg atcgcctgga actccaacaa gctggacagc 1320

aaggtgtccg gcaactacaa ctacctgtat agacttttca gaaagtccaa cctgaagcca 1380

ttcgagcggg acatcagcac tgagatctac caggccggca acaaaccctg caacggagtt 1440

gccggattca actgctattt ccctctgaga tcttactcct tcagacctac atacggcgtg 1500

ggacaccagc cttacagagt agtggtgctc agcttcgagc ttctgcacgc tcctgccacc 1560

gtgtgcggcc ctaagaagag cacgaacctg gtgaagaaca aatgtgttaa ttttaacttc 1620

aacggcctga agggcacagg agtcctgacc gagagcaata aaaaattctt gcccttccag 1680 cagttcggaa gagacatcgc cgacaccaca gatgctgtga gagaccctca gaccctggaa 1740

atcctcgaca tcaccccttg cagcttcggc ggcgtcagcg tgatcacccc gggcaccaac 1800

acctctaacc aggtggccgt gctgtaccag ggcgtgaatt gcaccgaggt tcctgtggcc 1860

atccacgcgg accagctgac accaacatgg cgggtgtaca gcaccggctc caacgtgttt 1920

cagaccagag ccggctgtct gatcggcgcc gaatatgtga acaacagcta cgaatgcgac 1980

atcccaatcg gcgccggcat ttgcgccagc taccagacac agaccaaaag tcaccggaga 2040

gctcggagcg tggcctctca gagcattatc gcctatacca tgagcctggg ggccgagaac 2100

agcgtggcct attccaacaa cagcatcgcc atccctacca atttcaccat ctctgtgacc 2160

accgagatcc tgccagtgtc catgacaaag acaagcgtgg actgcaccat gtacatctgc 2220

ggcgactcta ccgagtgcag caacctgctg ctgcagtacg gcagcttttg cacacagctg 2280

aaacgggcgc tgacaggaat tgccgttgag caggacaaga acactcagga ggtgtttgcc 2340

caagtgaagc agatatataa gacccctcct atcaaatact tcggcggctt taacttcagc 2400

cagatcctcc ctgatccttc taagcctagc aagcgcagct tcatcgagga cctgctgttc 2460

aacaaggtaa ccctggctga cgccggcttc atcaagcagt acggtgattg cctgggcgac 2520

atcgcagccc gggacctgat ctgtgcccaa aaattcaagg gcctgactgt tctgcctcct 2580

ctgctgacag atgaaatgat cgcccagtac acctccgccc tgctggctgg cacaatcacc 2640

agcggctgga cattcggcgc cggcgccgcg ctgcagatcc ctttcgccat gcagatggcc 2700

tacagattca acggcatcgg agtgactcag aacgtgctgt acgaaaacca gaaactgatt 2760

gcaaatcagt ttaacagcgc aatcggcaag atccaggata gcctgtccag caccgcctcc 2820

gctctgggca agctgcaaga cgtggtgaac cacaatgccc aggctctgaa caccttggtg 2880

aagcagctga gcagcaagtt cggcgccatt tcttccgtgc tgaacgacat cttcagcaga 2940

ctcgatcctc ccgaggccga ggtgcagatc gacagactga tcacgggcag actgcagtct 3000

ctgcagacat acgtgacaca gcaactgatc agagccgctg aaatcagggc ctctgccaac 3060

ctggccgcca ccaagatgtc tgagtgcgtg ctcggccagt ctaaaagagt ggacttctgc 3120

ggcaaaggct accacctgat gagcttcccc cagagcgccc cccacggcgt ggtgttccta 3180 cacgttacct acgtgccggc tcaagaaaag aactttacca ccgcccctgc catctgccac 3240

gacggaaagg cccacttccc tcgggagggt gtgtttgtca gcaacggcac acactggttc 3300

gtgacacagc ggaacttcta cgagccccaa atcatcacaa cagataacac cttcgtcagc 3360

ggcaactgtg acgtggtgat cggcatcgtg aacaacaccg tgtatgaccc tctgcagcct 3420

gagctggaca gctttaagga agagctggac aagtacttca agaatcacac aagtcctgac 3480

gtggatctgg gcgatatcag tggcatcaac gcctctgtgg tgaacataca aaaggagatc 3540

gacagactga acgaggtggc aaagaacctg aatgaaagcc tgatcgacct gcaagaactg 3600

ggcaagtacg agcagtacat caagtggcct tggtacattt ggctgggatt tatcgcaggc 3660

ctcatcgcca tcgtgatggt gacaatcatg ctgtgttgca tgaccagctg ttgcagctgc 3720

ctgaaaggct gttgtagctg cggcagctgc tgcaagttcg atgaggacga cagcgagcct 3780

gtcctgaagg gggtgaagct gcactacaca tga 3813

<210> 19
<211> 3822
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic nucleic acid

<400> 19
atgttcgtct tcctggtcct gctgcccctg gtctcatctc agtgcgtgaa tctgactaca 60

agaactcagc tgcctcccgc ctacaccaat tccttcaccc ggggcgtgta ctatcctgac 120

aaggtgttta gaagctccgt gctgcactct acacaggatc tgtttctgcc attctttagc 180

aacgtgacct ggttccacgc catccacgtg agcggcacca atggcacaaa gcggttcgac 240

aatcccgtgc tgccttttaa cgatggcgtg tacttcgcct ctaccgagaa gagcaacatc 300

atcagaggct ggatctttgg caccacactg gactccaaga cacagtctct gctgatcgtg 360

aacaatgcca ccaacgtggt catcaaggtg tgcgagttcc agttttgtaa tgatcccttc 420

ctgggcgtgt actatcacaa gaacaataag agctggatgg agtccgagtt tagagtgtat 480

tctagcgcca acaattgcac atttgagtac gtgtcccagc ctttcctgat ggacctggag 540 ggcaagcagg gcaatttcaa gaacctgagg gagttcgtgt ttaagaatat cgatggctac 600

ttcaagatct actctaagca cacccccatc aacctggtgc gcgacctgcc tcagggcttc 660

agcgccctgg agccactggt ggatctgcct atcggcatca acatcacccg gtttcagaca 720

ctgctggccc tgcacagaag ctacctgaca cccggcgact cctctagcgg atggaccgca 780

ggagcagcag cctactatgt gggctatctg cagcctagga ccttcctgct gaagtacaac 840

gagaatggca ccatcacaga cgcagtggat tgcgccctgg accccctgag cgagacaaag 900

tgtacactga agtcctttac cgtggagaag ggcatctatc agacatccaa tttcagggtg 960

cagccaaccg agtctatcgt gcgctttcct aatatcacaa acctgtgccc atttggcgag 1020

gtgttcaacg caaccaggtt cgcaagcgtg tacgcatgga ataggaagcg catctctaac 1080

tgcgtggccg actatagcgt gctgtacaac tccgcctctt tcagcacctt taagtgctat 1140

ggcgtgtccc ccacaaagct gaatgacctg tgctttacca acgtgtacgc cgattctttc 1200

gtgatcaggg gcgacgaggt gcgccagatc gcacctggac agacaggcaa gatcgccgac 1260

tacaattata agctgccaga cgatttcacc ggctgcgtga tcgcctggaa cagcaacaat 1320

ctggattcca aagtgggcgg caactacaat tatctgtacc ggctgtttag aaagagcaat 1380

ctgaagccct tcgagaggga catctctaca gagatctacc aggccggcag caccccttgc 1440

aatggcgtgg agggctttaa ctgttatttc ccactgcagt cctacggctt ccagcccaca 1500

aacggcgtgg gctatcagcc ttaccgcgtg gtggtgctga gctttgagct gctgcacgca 1560

ccagcaacag tgtgcggacc caagaagtcc accaatctgg tgaagaacaa gtgcgtgaac 1620

ttcaacttca acggcctgac cggaacaggc gtgctgaccg agtccaacaa gaagttcctg 1680

ccatttcagc agttcggcag ggacatcgca gataccacag acgccgtgcg cgacccacag 1740

accctggaga tcctggatat cacaccctgc tctttcggcg gcgtgagcgt gatcacacca 1800

ggaaccaata caagcaacca ggtggccgtg ctgtatcagg acgtgaattg taccgaggtg 1860

cctgtggcca tccacgccga tcagctgacc ccaacatggc gggtgtacag caccggctcc 1920

aacgtgttcc agacaagagc aggatgcctg atcggagcag agcacgtgaa caattcctat 1980

gagtgcgaca tcccaatcgg cgccggcatc tgtgcctctt accagaccca gacaaactct 2040 ccaaggagag cacggagcgt ggcatcccag tctatcatcg cctataccat gtccctgggc 2100

gccgagaatt ctgtggccta ctctaacaat agcatcgcca tcccaaccaa cttcacaatc 2160

tctgtgacca cagagatcct gcccgtgtcc atgaccaaga catctgtgga ctgcacaatg 2220

tatatctgtg gcgattctac cgagtgcagc aacctgctgc tgcagtacgg cagcttttgt 2280

acccagctga atagagccct gacaggcatc gccgtggagc aggataagaa cacacaggag 2340

gtgttcgccc aggtgaagca gatctacaag acccccccta tcaaggactt tggcggcttc 2400

aatttttccc agatcctgcc tgatccatcc aagccttcta agcggagctt tatcgaggac 2460

ctgctgttca acaaggtgac cctggccgat gccggcttca tcaagcagta tggcgattgc 2520

ctgggcgaca tcgcagcacg ggacctgatc tgtgcccaga agtttaatgg cctgaccgtg 2580

ctgccacccc tgctgacaga tgagatgatc gcacagtaca caagcgccct gctggcagga 2640

accatcacat ccggatggac cttcggcgca ggagccgccc tgcagatccc ctttgccatg 2700

cagatggcct ataggttcaa cggcatcggc gtgacccaga atgtgctgta cgagaaccag 2760

aagctgatcg ccaatcagtt taactccgcc atcggcaaga tccaggacag cctgtcctct 2820

acagcctccg ccctgggcaa gctgcaggat gtggtgaatc agaacgccca ggccctgaat 2880

accctggtga agcagctgag ctccaacttc ggcgccatct ctagcgtgct gaatgatatc 2940

ctgagccggc tggacccccc cgaggcagag gtgcagatcg accggctgat cacaggcaga 3000

ctgcagtctc tgcagaccta tgtgacacag cagctgatca gggcagcaga gatcagggca 3060

agcgccaatc tggcagcaac caagatgtcc gagtgcgtgc tgggccagtc taagagagtg 3120

gacttttgtg gcaagggcta tcacctgatg tccttccctc agtctgcccc acacggcgtg 3180

gtgtttctgc acgtgaccta cgtgcccgcc caggagaaga acttcaccac agcccctgcc 3240

atctgccacg atggcaaggc ccactttcca agggagggcg tgttcgtgtc caacggcacc 3300

cactggtttg tgacacagcg caatttctac gagccccaga tcatcaccac agacaatacc 3360

ttcgtgagcg gcaactgtga cgtggtcatc ggcatcgtga acaataccgt gtatgatcca 3420

ctgcagcccg agctggacag ctttaaggag gagctggata agtacttcaa gaatcacacc 3480

tcccctgacg tggatctggg cgacatcagc ggcatcaatg cctccgtggt gaacatccag 3540 aaggagatcg accgcctgaa cgaggtggcc aagaatctga acgagagcct gatcgatctg 3600

caggagctgg gcaagtatga gcagtacatc aagtggccat ggtacatctg gctgggcttc 3660

atcgccggcc tgatcgccat cgtgatggtg accatcatgc tgtgctgtat gacatcctgc 3720

tgttcttgcc tgaagggctg ctgtagctgt ggctcctgct gtaagtttga tgaggacgat 3780

tccgaacccg tgctgaaggg agtgaagctg cattacacct ga 3822