

**CORONAVIRUS SPIKE GLYCOPROTEIN RECEPTOR BINDING DOMAINS
AND USES THEREOF**

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 63/465,978 filed May 12, 2023 and U.S. Provisional Patent Application No. 63/598,691 filed November 14, 2023, each of which is incorporated by reference herein in its entirety.

SEQUENCE LISTING

[0002] This application contains a Sequence Listing which has been submitted in .XML format via EFS-WEB and is hereby incorporated by reference in its entirety. Said .XML copy, created on May 8, 2023 is named 061291-509001WO_SeqList_ST26.xml and is 335 kilobytes in size.

BACKGROUND

[0003] The emergence of SARS-Cov-2, the virus that causes COVID-19, has heightened interest in vaccines for coronaviruses both for SARS-CoV-2 and also for other Coronaviruses. There is an unmet need for engineered Coronavirus polypeptides having improved properties and uses thereof.

SUMMARY

[0004] In one aspect, the disclosure provides a recombinant polypeptide, comprising antigenic fragment of a coronavirus Spike glycoprotein comprising a receptor-binding domain (RBD) and optionally, a coronavirus subdomain 1 (SD1). In some embodiments, the antigenic fragment of the polypeptides comprises or consist essentially of RDB-SD1.

[0005] In some embodiments, the SD1 comprises at least one substitution of a surface-exposed, non-polar amino acid residue with a polar amino acid residue. In some embodiments, the polypeptide comprises a substitution of one or more, two or more, or three amino acid at positions 560, 562, or 569 with a polar residue, wherein the substitutions are relative to a reference sequence according to SEQ ID NO: 1. In some embodiments, the polypeptide comprises a substitution of one or more, two or more, or three of the amino acid substitutions L560Q, F562T, F562Y, or I569S relative to a reference sequence according to SEQ ID NO: 1.

[0006] In some embodiments, the polypeptide comprises a substitution of one or more, two or more, or three or more amino acid at positions 544, 546, 560, 562, 564, 569, or 582 relative to a reference sequence according to SEQ ID NO: 1. In some embodiments, the polypeptide comprises a substitution of one or more, two or more, three or more, four or

more amino acid at positions N544L, N544M, N544Q, L546V, L560Q, F562T, F562Y, Q564C, Q564L, Q564N, Q564W, I569S, or L582S relative to a reference sequence according to SEQ ID NO: 1.

[0007] In some embodiments, the polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acids at positions 338, 358, 363, 365, 392, or 395 relative to a reference sequence according to SEQ ID NO: 1. In some embodiments, the polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acids at positions F338L, I358F, A363L, Y365F, Y365M, Y365W, F392W, or V395I relative to a reference sequence according to SEQ ID NO: 1. In some embodiments, the polypeptide comprises amino acid substitutions comprising F338L /Y365W, F392W, Y365F/V395I, Y365F/ F392W /V395I, F338L/A363L/Y365M, and/or I358F/Y365F/V395I.

[0008] In some embodiments, the polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acid at positions 329, 348, 350, 367, 375, 402, 407, 410, 418, 429, 433, 435, 452, 464, 510, 512, 514, 517, 518, 519, 520, 522, 527, or 528 relative to a reference sequence according to SEQ ID NO: 1. In some embodiments, the polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acid at positions F329K, F329R, F329Y, A348P, V350L, V367F, F375Y, I402V, V407L, I410F, I418V, F429W, V433I, A435I, A435V, L452R, F464Y, P512Q, V510I, V512F, V512I, S514T, L517A, L517D, L517S, L517T, L518N, L518Q, L518V, H519D, H519G, H519R, H519S, H519T, A520C, A520D, A520G, A520H, P521A, P521D, P521N, P521Q, P521S, A522G, A522I, P527N, K528I, K528Q, K528T, L518G, or L518S relative to a reference sequence according to SEQ ID NO: 1.

[0009] In some embodiments, the polypeptide comprises amino acid substitutions at positions 365, 395, 560, 562, and 569; and one or more, two or more, or three or more amino acid substitutions at positions 348, 402, 464, 514, 520, 526, and 527, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1. In some embodiments, the polypeptide comprises the amino acid substitutions comprising A348P/Y365F/V395I/L560Q/F562Y/I569S, Y365F/V395I/I402V/L560Q/F562Y/I569S, Y365F/V395I/S514T/L560Q/F562Y/I569S, Y365F/V395I/I402V/F464Y/L560Q/F562Y/I569S, and/or Y365F/V395I/G526S/P527N/L560Q/F562Y/I569S.

[0010] In some embodiments, the polypeptide comprises amino acid substitutions at positions 365, 395, 560, 562, and 569; and one or more, two or more, or three or more amino

acid substitutions at positions 329, 517, 519, 520, and 544, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1. In some embodiments, the polypeptide comprises the amino acid substitutions comprising Y365F/V395I/L560Q/F562Y/I569S, Y365F/V395I/F329Y/L560Q/F562Y/I569S, Y365F/V395I/L517T/H519S/L560Q/F562Y/I569S, Y365F/V395I/L517T/H519S/A520G/L560Q/F562Y/I569S, Y365F/V395I/L517T/H519S/N544L/L560Q/F562Y/I569S, and/or Y365F/V395I/F329Y/L517T/H519S/L560Q/F562Y/I569S.

[0011] In some embodiments, the polypeptide comprises amino acid substitutions at positions 365, 392, 395, 560, 562, and 569; and one or more, two or more, or three or more amino acid substitutions at positions 329, 517, 519, 520, and 544, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1. In some embodiments, the polypeptide comprises the amino acid substitutions comprising Y365F/F392W/V395I/L560Q/F562Y/I569S, Y365F/F392W/V395I/F329Y/L560Q/F562Y/I569S, Y365F/F392W/V395I/L517T/H519S/L560Q/F562Y/I569S, Y365F/F392W/V395I/L517T/H519S/A520G/L560Q/F562Y/I569S, Y365F/F392W/V395I/L517T/H519S/N544L/L560Q/F562Y/I569S, and/or Y365F/F392W/V395I/F329Y/L517T/H519S/L560Q/F562Y/I569S.

[0012] In some embodiments, the polypeptide comprises amino acid substitutions at positions 365, 395, 517, 519, 520, 560, 562, 564, and 569; and one or more, two or more, or three or more amino acid substitutions at positions 521, 544, and 546, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1. In some embodiments, the polypeptide comprises the amino acid substitutions comprising Y365F/V395I/L517T/H519G/A520C/Q564C/L560Q/F562Y/I569S, and/or Y365F/V395I/L517S/H519R/A520H/P521Q/N544L/L546V/Q564W/F562T/L560Q/F562Y/I569S.

[0013] In some embodiments, the polypeptide comprises amino acid substitutions at positions 365, 395, 517, 519, and 520; and one or more, or two or more amino acid substitutions at positions 518 and 392, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1. In some embodiments, the polypeptide comprises the amino acid substitutions comprising Y365F/V395I/L517T/H519S/A520G, Y365F/V395I/L517T/L518G/H519D/A520G, Y365F/V395I/L517T/L518S/H519S/A520G,

Y365F/F392W/V395I/L517T/H519S/A520G,
 Y365F/F392W/V395I/L517T/L518G/H519D/A520G, and/or
 Y365F/F392W/V395I/L517T/L518S/H519S/A520G.

[0014] In some embodiments, the polypeptide comprises amino acid substitutions at positions 329, 365, 395, 517, 519, 520, 560, 562, and 56, and one or more, two or more, or three or more amino acid substitutions at positions 392, 544, and 564, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1. In some embodiments, the polypeptide comprises the amino acid substitutions comprising Y365F/V395I/L517T/H519S/A520G/N544M/L560Q/F562Y/I569S/F329Y, Y365F/V395I/L517T/H519S/A520G/N544Q/L560Q/F562Y/I569S/F329Y, Y365F/V395I/L517T/H519S/A520G/Q564N/L560Q/F562Y/I569S/F329Y, Y365F/F392W/V395I/L517T/H519S/A520G/N544M/L560Q/F562Y/I569S/F329Y, Y365F/F392W/V395I/L517T/H519S/A520G/N544Q/L560Q/F562Y/I569S/F329Y and/or Y365F/F392W/V395I/L517T/H519S/A520G/Q564N/L560Q/F562Y/I569S/F329Y. In some embodiments, the polypeptide comprises the amino acid substitutions comprising Y365F/V395I/L517T/H519S/A520G/N544M/L560Q/F562Y/I569S, Y365F/V395I/L517T/H519S/A520G/N544Q/L560Q/F562Y/I569S, Y365F/V395I/L517T/H519S/A520G/Q564N/L560Q/F562Y/I569S, Y365F/F392W/V395I/L517T/H519S/A520G/N544M/L560Q/F562Y/I569S, Y365F/F392W/V395I/L517T/H519S/A520G/N544Q/L560Q/F562Y/I569S, and/or Y365F/F392W/V395I/L517T/H519S/A520G/Q564N/L560Q/F562Y/I569S.

[0015] In some embodiments, the polypeptide comprises amino acid substitutions at positions 365, 395, 517, 519, 520, 521, 504, 560, 562, 569, and 329; and one or more amino acid substitutions at positions 392, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1. In some embodiments, the polypeptide comprises the amino acid substitutions comprising Y365F/V395I/L517T/H519G/P521A/A520C/Q564C/L560Q/F562Y/I569S/F329Y, Y365F/V395I/L517T/H519G/P521N/A520C/Q564C/L560Q/F562Y/I569S/F329Y, Y365F/F392W/V395I/L517T/H519G/P521A/A520C/Q564C/L560Q/F562Y/I569S/F329Y, and/or Y365F/F392W/V395I/L517T/H519G/P521N/A520C/Q564C/L560Q/F562Y/I569S/F329Y.

[0016] In some embodiments, the polypeptide comprises a plurality of amino acid substitutions at positions 329, 365, 395, 526, 527, 528, 560, 562, and 569. In some

embodiments, the polypeptide comprises the amino acid substitutions comprising F329Y/Y365F/V395I/G526S/P527N/K528T/L560Q/F562Y/I569S, and/or F329Y/Y365F/V395I/G526S/P527N/K528I/L560Q/F562Y/I569S.

[0017] In some embodiments, the polypeptide comprises a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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 100% identical to any one of SEQ ID NOs: 11-19. In some embodiments, the polypeptide comprises a polypeptide sequence identical to any one of SEQ ID NOs: 11-19.

[0018] In some embodiments, the polypeptide expresses at a higher level than a reference polypeptide without the amino acid substitutions, optionally a reference polypeptide according to SEQ ID NOs: 2 or 10. In some embodiments, the expressed polypeptide is more thermally stable or more antigenic than SEQ ID NOs: 2 or 10.

[0019] In some embodiments, the polypeptide comprises a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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 100% identical to SEQ ID NOs: 2-10.

[0020] In another aspect, the disclosure provides a self-assembling protein nanostructure comprising a first component comprising a recombinant coronavirus polypeptide described herein and optionally, a second component comprising a second protein.

[0021] In some embodiments, the protein nanostructure of the recombinant coronavirus polypeptide comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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 100% identical to SEQ ID NOs: 11-19.

[0022] In some embodiments, the protein nanostructure of the recombinant coronavirus polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acids relative to a reference sequence according to SEQ ID NO: 1.

[0023] In some embodiments, the protein nanostructure component polypeptide and a recombinant coronavirus polypeptide are non-covalently coupled. In some embodiments, the protein nanostructure component polypeptide and the recombinant coronavirus polypeptide are covalently coupled.

[0024] In some embodiments, the protein nanostructure comprises a fusion protein which includes the recombinant coronavirus polypeptide, a linker, and the protein nanostructure component polypeptide. In some embodiments, the protein nanostructure comprises a fusion protein which includes, in N- to C-terminal order, the recombinant coronavirus polypeptide, a linker, and the protein nanostructure component polypeptide.

[0025] In some embodiments, the protein nanostructure of the recombinant coronavirus polypeptide comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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 100% identical to SEQ ID NOs: 11-19.

[0026] In some embodiments, the protein nanostructure of the fusion protein comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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 100% identical to SEQ ID NOs: 11-19.

[0027] In some embodiments, the first component of the protein nanostructure comprises I53-50A.

[0028] In some embodiments, the protein nanostructure is an I3-01/MI3 protein nanostructure. In some embodiments, the protein nanostructure is an I53-50AB protein nanostructure.

[0029] In some embodiments, the protein nanostructure is a ferritin protein nanostructure. In some embodiments, the protein nanostructure is an encapsulin protein nanostructure. In some embodiments, the protein nanostructure is a CP3 bacteriophage capsid protein nanostructure. In some embodiments, the protein nanostructure is a Q β bacteriophage capsid protein nanostructure. In some embodiments, the protein nanostructure is an AP205 bacteriophage capsid protein nanostructure.

[0030] In some embodiments, the recombinant polypeptide of the protein nanostructure comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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 100% identical to SEQ ID NOs: 11-19.

[0031] In some embodiments, the protein nanostructure component polypeptide comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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 100% identical to SEQ ID NOs: 20-21.

[0032] In some embodiments, the protein nanostructure comprises a second component polypeptide.

[0033] In some embodiments, the second component polypeptide is I53-50B.

[0034] In some embodiments, the second component polypeptide comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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 100% identical to SEQ ID NO: 22.

[0035] In some embodiments, the linker of the protein nanostructure comprises between 8 and 24 amino acid residues. In some embodiments, the linker of the protein nanostructure is selected from the group consisting of GGSGSGSGSGSGSGS, SGGSGSGSGSGSGSGS, EPEGSGSGSGSGSGSGS, YGGSGSGSGSGSGSGSGS, and GSGSGSGSGSGSGSGSGSGSGS.

[0036] In another aspect, the disclosure provides a polynucleotide encoding the polypeptide or the protein nanostructure disclosed herein. In some embodiments, the polynucleotide is mRNA.

[0037] In another aspect, the disclosure provides a pharmaceutical composition comprising the polypeptides, the protein nanostructures, or the polynucleotides disclosed above.

[0038] In another aspect, the disclosure provides vaccines comprising the polypeptides, the protein nanostructures, or the polynucleotides disclosed above. In some embodiments, the vaccine further comprises an adjuvant. In some embodiments, the adjuvant comprises squalene. In some embodiments, the vaccine is a bivalent vaccine. In some embodiments, the bivalent vaccine comprises the polypeptide or the protein nanostructure described herein.

[0039] In another aspect, the disclosure provides a method of preventing or treating a coronavirus related disease in a subject in need thereof, comprises administering to the subject the polypeptides, the protein nanostructures, the polynucleotides, the pharmaceutical component, or the vaccine disclosed above. In another aspect, the disclosure provides a

method of immunizing a subject in need thereof against infection by coronavirus, by administering to the subject the vaccine disclosed above. In another aspect, the disclosure provides kits, which may be used to prepare and administer polypeptides, protein nanostructures, polynucleotides, pharmaceutical compositions, or vaccines as described above.

BRIEF DESCRIPTION OF THE DRAWINGS

[0040] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

[0041] **FIGs. 1A-1C** show a structure of RBD and RBD-SD1 antigen fragments (PDB 7UB5). **FIG. 1A** shows Structure of RBD alone, from residues 328-519. The receptor binding motif (RBM) which binds ACE2 is shown in a box. **FIG. 1B** shows Structure of RBD-SD1 from residues 319-591. SD1 residues are colored black. **FIG. 1C** shows the same structure as **FIG. 1B**, with regions targeted for designed mutations according to the present disclosure are shaded, with regions comprising Rpk mutations, RBD-SD1 interface mutations, and RBM-proximal mutations are shaded and with density maps. The SD1resurf mutations are shown as shaded spheres and surrounded with a box.

[0042] **FIGs. 2A-2C** show analyses of supernatants for initial designs tested with ancestral RBD-SD1 fragment genetically fused to CompA. **FIG. 2A** shows Western blots of supernatants. **FIG. 2B** shows a first set of Bio-Layer Interferometry (BLI) data collected for ACE2-binding against select supernatant samples. Linker differences (in constructs RDBb1-5) were compared using WT RBD-SD1 antigens (top left) and stabilizing mutation differences were compared within constructs with 16-residue linkers. **FIG. 2C** shows a second set of BLI data collected for ACE2-binding against select supernatant samples. Data focused on designs containing Rpk8 (RBDb016 through 020), Rpk9 (RBDb021 through 025) or SD1resurf (RBDb036 through 040) mutations. RBD01 was used as a control for all studies. Maximum shift values for **FIGs. 2B-2C** are respectively shown in **Table 16** and **Table 17**.

[0043] **FIGs. 3A-3F** show BLI traces for purified CompAs from ancestral antigens. RBD01 was used as a control, which lacks stabilizing mutations and the SD1 domain. **FIG. 3A** shows ACE2 binding, which represents intactness of key epitopes that support neutralizing antibodies. Levels of binding were similar for all constructs. **FIG. 3B** shows ACE2 binding, which represents intactness of key epitopes that support neutralizing antibodies. Levels of binding were similar for all constructs. **FIG. 3C** and **FIG. 3D** shows CR3022-binding.

CR3022 is a non-neutralizing antibody that targets an epitope distal to the ACE2-binding site which isn't normally exposed in the Spike glycoprotein. **FIG. 3E** and **FIG. 3F** shows S309-binding. S309 is a broadly neutralizing antibody within sarbecoviruses that binds outside of the ACE2-binding site. Maximum shift values for **FIGs. 3A-F** are shown in **Table 18**.

[0044] **FIGs. 4A-4K** show thermal denaturation curves for ancestral CompAs using either intrinsic protein fluorescence (ITF) or SYPRO™ Orange Protein Gel Stain. **FIG. 4A** and **FIG. 4B**, respectively show ITF measurements for the RBD01 control and RBDb001 relative to RBDb002 to measure impact of adding SD1 domain or changing linkers. All showed similar melting temperatures. **FIG. 4C** to **FIG. 4F** show ITF measurements for RBD-SD1 CompAs containing different Rpk mutations compared to WT RBD-SD1 (RBDb002). Addition of Rpk9 mutations led to the highest melting temperatures. **FIG. 4G** ITF comparison of RBDb002 and RBDb037 (SD1resurf mutations), showing that addition of SD1 resurf mutations do not impact melting temperature. All plots in **FIG. 4A** to **FIG. 4G** show three replicate measurements (each labelled "rep1", "rep2", or "rep3"), and RBDb002 ITF measurements shown in each plot for **FIG. 4A** to **FIG. 4G** are identical to allow comparison across plots. Average melting temperatures for ITF measurements from RBDb002 and comparator samples are respectively shown in gray dotted lines or black dashed lines, and are also shown in **Table 19**. **FIG. 4H** to **FIG. 4K** shows SYPRO comparisons of RBD-SD1 CompAs containing different Rpk mutations compared to WT RBD-SD1 (RBDb002). Rpk9 mutations (RBDb022) showed combination of lower SYPRO baseline (suggesting improved foldedness) and a higher melting temperature. All plots in **FIG. 4H** to **FIG. 4K** show three replicate measurements (each labelled "rep1", "rep2", or "rep3"), and RBDb002 SYPRO measurements shown in each plot for **FIG. 4H** to **FIG. 4K** are identical to allow comparison across plots. Average melting temperatures for SYPRO measurements from RBDb002 and comparator samples are respectively shown in gray dotted lines or black dashed lines, and are also shown in **Table 20**.

[0045] **FIG. 5A** and **FIG. 5B** show assembly of ancestral CompAs with complementary CompB molecule, as measured by dynamic light scattering (DLS) (all shown as intensity distributions). **FIG. 5A** shows DLS measurements collected after assembly in the presence of buffer supplemented with 50mM Arginine and 4% sucrose, showing highly-efficient assembly of RBDb022 and RBDb037 based on the predicted size range of assembled protein nanostructures, whereas all other groups showed aggregation. **FIG. 5B** shows DLS measurements collected after assembly in the presence of buffer supplemented with 50mM

Arginine, showing highly-efficient assembly of RBDb022 and RBDb037 based on the predicted size range of assembled protein nanostructures, whereas all other groups showed aggregation. Labels for “MOPS” and “Tris/glycerol” indicate buffer backgrounds beyond the added excipients, which respectively refer to 50 mM MOPS, 0.75 % CHAPS, 150 mM NaCl, pH 7.4, or 20 mM Tris, 5 % Glycerol, 250 mM NaCl, pH 8. DLS results from **FIG. 5A** and **FIG. 5B** are also summarized in **Table 21**, along with more detailed buffer conditions for nanostructure assembly and analysis.

[0046] **FIG. 6** shows an analysis of assembly for RBDb092 (ancestral) and RBDb093 (BA.5), which contain the combination of Rpk9 and SD1resurf mutations. DLS was used for analysis (shown as an intensity distribution), with assembly performed in the presence of buffer supplemented with 100mM Arginine and 5% glycerol. RBDb092 successfully assembled with high efficiency, while RBDb093 showed aggregation. DLS results from **FIG. 6** are also summarized in **Table 22**.

[0047] **FIG. 7** shows an alignment of RBD-SD1 sequences (equivalent of residues 319-591 of SARS-CoV-2) of representative diverse sarbecoviruses. SD1 domains are marked by horizontal lines, and SD1 residues are italicized. Regions comprising Rpk8/Rpk9 mutations, the RBD-SD1 interface mutations, or the SD1resurf mutations are marked with symbols according to the legend. Gaps relative to other sequences are marked with dashes. This alignment shows that the regions comprising Rpk8/Rpk9 mutations and RBD-SD1-interface mutations are largely conserved, verifying that these mutations extend across diverse viral strains. Further, this alignment shows that the SD1resurf mutations L560Q and I569S are present in other sarbecoviruses, and therefore are applicable to diverse viral strains. I569V also occurs frequently in the alignment.

[0048] **FIG. 8** shows a Western blot analyses of BA.5 RBD-SD1 CompA designs in supernatants after small-scale expression from HEK293 cells. All double-digit construct numbers refer to the last two digits of the “RBDb0XX” naming scheme, in which the two digits replace the “XX”. With the exception of RBDb046, all designs contain Rpk8 and SD1resurf mutations, with constructs RBDb051-091 also containing further mutations. Constructs RBDb044-050 test the combination of Rpk8 and SD1resurf mutations in multiple Omicron strains, with RBDb047 using the BA.5 strain.

[0049] **FIG. 9A** to **FIG. 9C** show *in vitro* characterization of select purified designs for BA.5 RBD-SD1 CompAs. RBDb046 uses a WT RBD-SD1 antigen, while all other constructs further include stabilizing mutations. **FIG. 9A** shows melting (T_m) and onset (T_{onset}) temperatures measured using ITF. **FIG. 9B** shows binding rate measurements from

BLI against ACE2. Error bars show standard deviations between replicates. **FIG. 9C** shows max shifts from BLI against ACE2. Error bars show standard deviations between replicates. **[0050] FIG. 10** shows an assembly of select purified polypeptides for BA.5 RBD-SD1 CompAs as measured by DLS (shown as intensity distribution). Assembly was performed by buffer supplemented with 100mM Arginine. Various polypeptides assembled well without aggregation, exceptions being RBDb045, RBDb078, and RBDb081.

[0051] FIG. 11A and **FIG. 11B** show immunogenicity of select CompAs and protein nanostructures from ancestral and BA.5 antigens in mice. **FIG. 11A** shows ELISA binding titers against strain-matched ancestral or BA.5 Spike glycoprotein antigens. **FIG. 11B** shows pseudovirus neutralization against strain-matched ancestral or BA.5 pseudoviruses. These data show differences in immunogenicity between many CompAs, particularly when comparing RBDb046 to other BA.5 CompAs containing stabilizing mutations.

[0052] FIG.12A shows ACE2/anti-CompA binding ratio of selected wildtype and stabilized designs in various strains. **FIG.12B** shows total protein yields of selected wildtype and stabilized designs in various strains. **FIG.12C** shows melting temperature of selected wildtype and stabilized designs in various strains.

DETAILED DESCRIPTION

Definitions

[0053] Reference is made herein to particular features (including method steps) It is to be understood that the disclosure in this specification includes all possible combinations of such particular features. For example, where a particular feature is disclosed in the context of a particular aspect or embodiment, or a particular claim, that feature can also be used, to the extent possible, in combination with and/or in the context of other particular aspects and embodiments.

[0054] Where reference is made herein to a method comprising two or more defined steps, the defined steps can be carried out in any order or simultaneously (except where the context excludes that possibility), and the method can include one or more other steps which are carried out before any of the defined steps, between two of the defined steps, or after all the defined steps (except where the context excludes that possibility).

[0055] The practice of the technology will employ, unless indicated specifically to the contrary, conventional methods of chemistry, biochemistry, organic chemistry, molecular biology, microbiology, recombinant DNA techniques, genetics, immunology, and cell biology that are within the skill of the art, many of which are described below for the purpose of illustration. Such techniques are explained fully in the literature. It is to be understood

that this disclosure is not limited to the particular methodology, protocols, and reagents described, as these may vary, depending upon the context in which they are used by those of skill in the art.

[0056] All publications and patents mentioned herein are hereby incorporated by reference in their entirety as if each individual publication or patent was specifically and individually indicated to be incorporated by reference. In case of conflict, the present application, including any definitions herein, controls. However, mention of any reference, article, publication, patent, patent publication, and patent application cited herein is not, and should not be taken as an acknowledgment, or any form of suggestion, that they constitute valid prior art or form part of the common general knowledge in any country in the world.

[0057] The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described. Specifically, features described in one section may be combined with features in any other section of the description.

[0058] While illustrative embodiments are described and depicted, it will be appreciated that various changes can be made to these illustrative embodiments without departing from the spirit and scope of the invention.

[0059] Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. Various scientific dictionaries that include the terms included herein are well known and available to those in the art. Although any methods and materials similar or equivalent to those described herein find use in the practice or testing of the disclosure, some preferred methods and materials are described. Accordingly, the terms defined immediately below are more fully described by reference to the specification as a whole.

[0060] The singular forms “a”, “an”, and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise.

[0061] As used herein, the term “about” means a range of values including the specified value, which a person of ordinary skill in the art would consider reasonably similar to the specified value. In some embodiments, about means within a standard deviation using measurements generally acceptable in the art. In some embodiments, about means a range extending to +/- 10%, +/- 5%, +/- 3%, or +/- 1% of the specified value.

[0062] The term “at least” followed by a number is used herein to denote the start of a range beginning with that number (which may be a range having an upper limit or no upper limit, depending on the variable being defined). For example, “at least 1” means 1 or more than 1.

[0063] The term “at most” followed by a number is used herein to denote the end of a range ending with that number (which may be a range having 1 or 0 as its lower limit, or a range having no lower limit, depending upon the variable being defined). For example, “at most 4” means 4 or less than 4, and “at most 40%” means 40% or less than 40%. When, in this specification, a range is given as “(a first number) to (a second number)” or “(a first number)-(a second number)” this means a range whose lower limit is the first number and whose upper limit is the second number. For example, 25 to 100 mm means a range whose lower limit is 25 mm, and whose upper limit is 100 mm.

[0064] As used herein, “substantially” or “substantial” refers to the complete or nearly complete extent or degree of an action, characteristic, property, state, structure, item, or result. For example, an object that is “substantially” enclosed would mean that the object is either completely enclosed or nearly completely enclosed. The exact allowable degree of deviation from absolute completeness may in some cases depend on the specific context. However, generally speaking, the nearness of completion will be so as to have the same overall result as if absolute and total completion were obtained. The use of “substantially” is equally applicable when used in a negative connotation to refer to the complete or near complete lack of action, characteristic, property, state, structure, item, or result. For example, a composition that is “substantially free of” other active agents would either completely lack other active agents, or so nearly completely lack other active agents that the effect would be the same as if it completely lacked other active agents. In other words, a composition that is “substantially free of” an ingredient or element or another active agent may still contain such an item as long as there is no measurable effect thereof.

[0065] Throughout this specification, unless the context requires otherwise, the words “comprise”, “comprises” and “comprising” will be understood to imply the inclusion of a stated step or element or group of steps or elements but not the exclusion of any other step or element or group of steps or elements. For example, a composition “comprising” (or “which comprises”) components A, B, and C can consist of (*i.e.*, contain only) components A, B, and C, or can contain not only components A, B, and C but also one or more other components.

[0066] As used herein, the term “consisting of” refers to including, and limited to, whatever follows the phrase “consisting of.” Thus, the phrase “consisting of” indicates that the listed elements are required or mandatory, and that no other elements may be present. By “consisting essentially of” is meant including any elements listed after the phrase, and limited to other elements that do not interfere with or contribute to the activity or action

specified in the disclosure for the listed elements. Thus, the phrase “consisting essentially of” indicates that the listed elements are required or mandatory, but that no other elements are optional and may or may not be present depending upon whether or not they affect the activity or action of the listed elements.

[0067] Reference throughout this specification to, for example, “one embodiment”, “an embodiment”, “another embodiment”, “a particular embodiment”, “a related embodiment”, “in certain embodiments”, “in some embodiments”, “an additional embodiment”, or “a further embodiment” or combinations thereof means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present disclosure. Thus, the appearances of the foregoing phrases in various places throughout this specification are not necessarily all referring to the same embodiment. Furthermore, the particular features, structures, or characteristics may be combined in any suitable manner in one or more embodiments.

[0068] The following description includes information that may be useful in understanding the present invention. It is not an admission that any of the information provided herein is prior art or relevant to the presently claimed inventions, or that any publication specifically or implicitly referenced is prior art.

[0069] The term “identical” or percent “identity,” in the context of two or more nucleic acid or polypeptide sequences, refers to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence. Methods of alignment of sequences for comparison are well known in the art. Once aligned, the number of matches is determined by counting the number of positions where an identical nucleotide or amino acid residue is present in both sequences. The percent sequence identity is determined by dividing the number of matches in the alignment by the length of the reference sequence, followed by multiplying the resulting value by 100. For example, a peptide sequence that has 1166 matches when aligned with a test sequence having 1554 amino acids is 75.0 percent identical to the test sequence ($1166 \div 1554 * 100 = 75.0$). As the terms are used herein, gaps in the alignment do not decrease the percent sequence identity.

[0070] Unless otherwise specified, optimal alignment of sequences for comparison is conducted by the global alignment algorithm of Needleman and Wunsch, *Mol. Biol.* 48:443 (1970) as implemented by EMBOSS Needle (on the World Wide Web at ebi.ac.uk/Tools/psa/emboss_needle/) (Madeira et al. *Nucleic Acids Res.* 50(W1):W276-W279 (2022)). In embodiments, other alignment methods may be used, including without

limitation those described in Devereux, et al, *Nucleic Acids Res.* 12:387-95 (1984); Atschul et al. *J. Mo. Biol.* 215:403-10 (1990) (BLAST); Carrillo and Lipman *Siam J. Appl. Math.* 48(5) (1988); *Computational Molecular Biology* (Lesk, AM, ed., 1989); *Biocomputing Informatics and Genome Projects*, (Smith, DW, ed., 1993); *Computer Analysis of Sequence Data, Part I*, (Griffin and Griffin, eds., 1994); *Sequence Analysis in Molecular Biology* (von Heinje, 2012); *Sequence Analysis Primer* (Gribkov and Devereux, J., eds. 1993). In embodiments, sequence identity is calculated using the implementation of the Needleman-Wunsch algorithm provided by the National Library of Medicine (on the World Wide Web at

blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&BLAST_SPEC=GlobalAln).

[0071] For example, sequence identity can be determined by standard methods that are commonly used to compare the similarity of two polypeptide or two polynucleotide sequences. Using a computer program such as EMBOSS Needle or BLAST, two polypeptide or two polynucleotide sequences are aligned for optimal matching of their respective residues (either along the full length of one or both sequences, or along a pre-determined portion of one or both sequences). The programs provide a default opening penalty and a default gap penalty, and a scoring matrix such as PAM 250 (a standard scoring matrix; see Dayhoff et al., in *Atlas of Protein Sequence and Structure*, vol. 5, supp. 3 (1978)) that can be used in conjunction with the computer program.

[0072] An amino acid or nucleotide base “position” is denoted by a number that sequentially identifies each amino acid (or nucleotide base) in the reference sequence based on its position relative to the N-terminus (or 5'-end). Due to deletions, insertions, truncations, fusions, and the like that must be taken into account when determining an optimal alignment, in general the amino acid residue number in a test sequence determined by simply counting from the N-terminus will not necessarily be the same as the number of its corresponding position in the reference sequence. For example, in a case where a variant has a deletion relative to an aligned reference sequence, there will be no amino acid in the variant that corresponds to a position in the reference sequence at the site of deletion. Where there is an insertion in an aligned reference sequence, that insertion will not correspond to a numbered amino acid position in the reference sequence. In the case of truncations or fusions there can be stretches of amino acids in either the reference or aligned sequence that do not correspond to any amino acid in the corresponding sequence.

[0073] The terms “relative to” or “numbered with reference to” or “corresponding to,” when used in the context of the numbering of a given sequence, refers to the numbering of the

residues of the given sequence when aligned to a specified reference sequence, such that each position in the numbered sequence is associated a number corresponding to the number of the counterpart position in the reference sequence.

[0074] The term “coronavirus” refers to viruses of the subfamily Coronavirinae in the family Coronaviridae, in the order Nidovirales. Coronaviruses are enveloped viruses and have a positive –sense, single-stranded RNA genome with a nucleocapsid of helical symmetry. The term “coronavirus” encompasses all strains, genotypes, protectotypes, and sero types of infectious bronchitis virus. Examples of avian Coronaviruses are infectious bronchitis virus (IBV); guinea fowl Coronavirus (GfCo) and turkey Coronavirus (TCO turkey enteritis virus and bluecomb disease virus).

[0075] The term “SARS-CoV-2”, also known as Wuhan Coronavirus or 2019 novel Coronavirus, refers to a positive-sense, single stranded RNA virus of the genus betaCoronavirus that has emerged as a highly fatal cause of severe acute respiratory infection. 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) and zeta (P.2)).

[0076] B.1.1.529, also known as the omicron variant, is a variant of the original SARS-CoV-2. This variant has a total of 60 mutations compared to the original strain of SARS-CoV-2, specifically 50 nonsynonymous mutations, 8 synonymous mutations, 35 and 2 non-coding mutations. Thirty-two mutations affect the Spike glycoprotein (A67V, A69-70, T95I, G142D, A143-145, A211, L212I, ins214EPE, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, and L981F with reference to SEQ ID NO: 1), or which approximately half are located in the receptor binding domain (319-530).

[0077] The term “Spike glycoprotein” refers to a class I fusion glycoprotein initially synthesized as a precursor protein of approximately 1273 amino acids for SARS-CoV-2. 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. An exemplary SARS-CoV-2 Spike glycoprotein sequence is set forth herein as SEQ ID NO: 1 (NCBI Reference Sequence: YP_009724390.1 the sequence of which is incorporated by reference

herein in its entirety). As used herein, antigenic fragment is amino acids 319-591 in NCBI Reference Sequence: YP_009724390.1.

[0078] As used herein, the term “receptor-binding domain (RBD)” refers a part of the Spike glycoprotein. For example, the RBD is amino acids 319-541 in NCBI Reference Sequence: YP_009724390.1. As used herein, the RBD is amino acids 328-531 in NCBI Reference Sequence: YP_009724390.1. The RBD includes the receptor binding motif (RBM) which facilitates cellular entry by binding the ACE2 receptor.

[0079] As used herein, the term “subdomain-1 (SD1)” refers a conserved part of Spike glycoprotein adjacent to the receptor-binding domain. For example, the SD1 is amino acids 532-591 in NCBI Reference Sequence: YP_009724390.1.

[0080] The term “ancestral” refers to antigens or viruses based directly on the sequence of the original Wuhan-Hu-1 isolate of SARS-CoV-2 (Genbank identifier NC_045512.2)”

[0081] The term “recombinant” indicates that the material (e.g., a nucleic acid or protein) has been artificially or synthetically (non-naturally) altered by human intervention. The alteration can be performed on the material within, or removed from, its natural environment or state. Specifically, e.g., a Coronavirus is recombinant when it is produced by the expression of a recombinant nucleic acid. For example, a “recombinant nucleic acid” is one that is made by recombining nucleic acids, e.g., during cloning, DNA shuffling or other procedures, or by chemical or other mutagenesis; a “recombinant polypeptide” or “recombinant protein” is a polypeptide or protein which is produced by expression of a recombinant nucleic acid; and a “recombinant virus,” e.g., a recombinant Coronavirus, is produced by the expression of a recombinant nucleic acid. A “fusion protein” or “tetramers” described here are recombinant and do not occur in nature.

[0082] The terms “polypeptide”, “peptide” and “protein” are used interchangeably herein to refer to a polymer of amino acid residues, wherein the polymer may be conjugated to a moiety that does not consist of amino acids, such as a glycan or other post-translational modification, or a non-natural chemical moiety. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers. A “fusion protein” refers to a chimeric protein encoding two or more separate protein sequences that are recombinantly expressed as a single polymer molecule.

[0083] The term “recombinant polypeptide” refers a polypeptide that is not naturally occurring or has a sequence that is made by an artificial combination of two otherwise

separated segments of sequence. In several embodiments, a recombinant polypeptide is encoded by a heterologous (for example, recombinant) nucleic acid that has been introduced into a host cell, such as a bacterial or eukaryotic cell.

[0084] The term “antigen” or “immunogen” refers to its plain and ordinary meaning of a compound or composition that induces an immune response, cellular or humoral, e.g., cytotoxic T lymphocyte (CTL) response, a B cell response (for example, production of antibodies that specifically bind the epitope), an NK cell response or any combinations thereof, when administered to or expressed in an immunocompetent subject. Antigens can include polypeptides (including glycoproteins). In aspects, an antigen is a polypeptide or polypeptide complex including at least one component designed to elicit an immune response. For example, an antigen can include one or more immunogenic epitopes associated with a bacterial pathogen. The term antigen, as used herein, is not limited to the portion of the polypeptide or polypeptide complex that contains antigenic epitopes. An “epitope” or “antigenic determinant” refers to its plain and ordinary meaning as the part of an antigen that is recognized by the immune system, specifically by antibodies, B cells, or T cells.

[0085] The term “variant” refers to a polypeptide having one or more insertions, deletions, or amino acid substitutions relative to a reference polypeptide, but retains one or more properties of the reference protein.

[0086] The term “antigenic variant” refers to a variant that has one or more epitopes in common with a reference polypeptide and/or generates the same or similar immune response when administered to a subject as a reference polypeptide.

[0087] The term “antigenic fragment” refers to any fragment of a protein that generates an immune response (humoral or T cell response) to the protein in vivo. The antigenic fragment may be a linear epitope, discontinuous epitope, or a conformation epitope (e.g., a folded domain). The antigenic fragment may preserve the secondary, tertiary, and/or quaternary structure of the full-length protein.

[0088] The term “functional variant” refers to a variant that exhibits at least some of the activity as a reference polypeptide. For example, a functional variant of an assembly domain is able to promote multimerization and self-assembly to the same extent, or to similar extent, as a reference assembly domain and/or is able to multimerize and assemble with the same cognate assembly domains as a reference assembly domain.

[0089] The term “substitution” refers to replacing a single amino acid in a sequence with another amino acid residue. The standard form of abbreviations for amino acid substitution

are used. For example, V94R refers to substitution of valine (V) in a reference sequence with arginine (R). The abbreviation Arg94 refers to any sequence in which the 94th residue, relative to a reference sequence, is arginine (Arg). The term “combination of substitutions” refers to a plurality of amino acid substitutions which are included in a single polypeptide. As used herein, a slash “/” character identifies a combination of substitutions at each of the amino acid positions separated by the slash. As examples, “Y365F/V395I” is meant a Y-->F substitution at amino acid position 365 in combination with a V-->I substitution at amino acid position 395; “Y365F/ F392W /V395I” is meant is meant a Y-->F substitution at amino acid position 365 in combination with a F-->W substitution at amino acid position 392 and in combination with a V-->I substitution at amino acid position 395.

[0090] The following eight groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Glycine (G);
- 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);
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W);
- 7) Serine (S), Threonine (T); and
- 8) Cysteine (C), Methionine (M)

(see, e.g., Creighton, Proteins (1984)).

[0091] As used herein, “a surface-exposed amino acid residue” refers to an amino acid that is presented to the exterior of the polypeptide in which it is part. In some embodiments, a surface-exposed amino acid residue is available to interact with other polypeptides or with solvent. In other words, the amino acid residue is not buried within the polypeptide chain.

[0092] As used herein, “non-polar amino acid residue” refers to a hydrophobic amino acid or residue having a side chain that is uncharged at physiological pH and which has bonds in which the pair of electrons shared in common by two atoms is generally held equally by each of the two atoms (*i.e.*, the side chain is not polar). Genetically encoded non-polar amino acids include L-Gly (G), L-Leu (L), L-Val (V), L-Ile (I), L-Met (M) and L-Ala (A).

[0093] As used herein, “polar amino acid residue” refers to a hydrophilic amino acid or residue having a side chain that is uncharged at physiological pH, but which has at least one bond in which the pair of electrons shared in common by two atoms is held more closely by

one of the atoms. Genetically encoded polar amino acids include L-Asn (N), L-Gln (Q), L-Ser (S) and L-Thr (T).

[0094] An amino acid or nucleotide base “position” is denoted by a number that sequentially identifies each amino acid (or nucleotide base) in the reference sequence based on its position relative to the N-terminus (or 5'-end). Due to deletions, insertions, truncations, fusions, and the like that must be taken into account when determining an optimal alignment, in general the amino acid residue number in a test sequence determined by simply counting from the N-terminus will not necessarily be the same as the number of its corresponding position in the reference sequence. For example, in a case where a variant has a deletion relative to an aligned reference sequence, there will be no amino acid in the variant that corresponds to a position in the reference sequence at the site of deletion. Where there is an insertion in an aligned reference sequence, that insertion will not correspond to a numbered amino acid position in the reference sequence. In the case of truncations or fusions there can be stretches of amino acids in either the reference or aligned sequence that do not correspond to any amino acid in the corresponding sequence.

[0095] The term “expression” includes any step involved in the production of the polypeptide including, but not limited to, transcription, post-transcriptional modification, translation, post-translational modification, and secretion. In some aspects, expression can be detected using conventional techniques for detecting protein (*e.g.*, ELISA, Western blotting, flow cytometry, immunofluorescence, immunohistochemistry, *etc.*).

[0096] A “cell” as used herein, refers to a membrane-bound biological unit capable of carrying out metabolic or other function sufficient to preserve or replicate its genomic DNA. A cell can be identified by well-known methods in the art including, for example, presence of an intact membrane, staining by a particular dye, ability to produce progeny or, in the case of a gamete, ability to combine with a second gamete to produce a viable offspring. Cells may include prokaryotic and eukaryotic cells. Prokaryotic cells include but are not limited to bacteria. Eukaryotic cells include but are not limited to yeast cells and cells derived from plants and animals, for example mammalian, insect (*e.g.*, *spodoptera*) and human cells. Cells may be useful when they are naturally nonadherent or have been treated not to adhere to surfaces, for example by trypsinization.

[0097] The term “host cell” refers to any cell capable of use in expression of a recombinant polypeptide or polynucleotide. In some embodiments, the host cell is a bacterial cell. In some embodiments, the bacterial cell is an *E. coli* cell.

[0098] The terms “culturing” and “culture medium” refers to standard cell culture and recombinant protein expression techniques.

[0099] The term “manufacturing” refers to production of a recombinant polypeptide or protein nanostructure or polynucleotide at any scale, including at least 25-mL, 50-mL, 1-L, 1,000-L, 50,000-L, or greater scale.

[0100] The term “purify” refers to separating a molecule from other substances present in a composition. Polypeptides may be purified by affinity (*e.g.*, to an antibody or to a tag, *e.g.*, using a His-tag capture resin), by charge (*e.g.*, ion-exchange chromatography), by size (*e.g.*, preparative ultracentrifugation, size exclusion chromatography), or otherwise.

[0101] The term “stability” refers to the physical and chemical stability of formulations of polypeptides. Physical instability of a protein formulation may be caused by aggregation of the protein molecules to form higher order polymers or even precipitates. A “stable” formulation is one wherein the degree of aggregation of proteins therein is acceptably controlled, and does not increase unacceptably with time. Physical stability may be assessed by methods well-known in the art, including measurement of a sample's apparent attenuation of light (absorbance, or optical density). Such a measurement of light attenuation relates to the turbidity of a formulation. Turbidity is produced by aggregation or precipitation of proteins or complexes in the formulation.

[0102] The term “antigenicity” refers to the ability of an antigen to specifically bind to a T cell receptor or antibody and includes the reactivity of an antigen toward pre-existing antibodies in a subject. “Immunogenicity” refers to the ability of any antigen to induce an immune response and includes the intrinsic ability of an antigen to generate antibodies in a subject.

[0103] The term “neutralizing” (*e.g.*, “neutralizing antibody response”) refers to antibodies that prevent infection and/or reduce the level of infection by a pathogen. A neutralizing antibody response can be measured either in *in vitro* assays (*e.g.*, killing of target bacteria in the presence of the antibody) or in an *in vivo* assay (*e.g.*, by determining a protective dose of an antibody through administering the antibody to a subject prior to challenge with an infective dose of a pathogen).

[0104] An antibody “binds to” or is “specific to” or “specifically binds” (used interchangeably herein) to a target (*e.g.*, bacterial protein) are terms well understood in the art, and methods to determine such specific or preferential binding are also well known in the art. A molecule is said to exhibit “specific binding” or “preferential binding” if it reacts

or associates more frequently, more rapidly, with greater duration and/or with greater affinity with a particular cell or substance than it does with alternative cells or substances.

[0105] The term “protein nanostructure,” as used herein, refers to symmetric protein assemblies in which the subunits self-assemble in aqueous solution, without requiring lipids or macromolecules other than the protein nanostructure for assembly. Illustrative protein nanostructures are described in Hsia et al. *Nature* 35:136-9 (2016) and Bale et al. *Science* 353:389-394 (2016). In some embodiments, the protein nanostructure is a one-component protein nanostructure, in which a single polypeptide type provides the building blocks to self-assemble to form the protein nanostructure. In some embodiments, the protein nanostructure is a two-component protein nanostructure, in which two polypeptide types provide the building blocks to self-assemble to form the protein nanostructure. In some embodiments, the polypeptide types include an assembly domain, which causes the polypeptide to form symmetric dimeric, trimeric, tetrameric, hexameric components, or another multimeric component. In a two-component nanostructure, the two components differ in the selection of an assembly domain. In some embodiments, the assembly domain of the first polypeptide type causes the polypeptide to form a trimer; and the assembly domain of the second polypeptide type causes the polypeptide to form a pentamer.

[0106] In some embodiments of, or relating to, one-component nanostructures, two or more copies of the component further symmetrically self-assemble to form the nanostructure. In some embodiments of, or relating to, two-component nanostructures, two or more of each of the two, different components symmetrically self-assemble to form the nanostructure.

[0107] As used herein, the term “assembly domain” refers to the portion of a subunit of a component involved in forming the protein nanostructure through intra-component interactions and interactions with either other copies of the same component (in a one-component nanostructure) or with other components (e.g., in a two-component nanostructure).

[0108] The term “icosahedral particle” refers to protein nanostructures having a core with icosahedral symmetry. I53 refers to an icosahedral particle constructed from pentamers and trimers. I52 refers to an icosahedral particle constructed from pentamers and dimers. T33 refers to a tetrahedral particle constructed from two sets of trimers. T32 refers to a tetrahedral particle constructed from trimers and dimers. As used herein, the abbreviations I53-50A and I53-50B refer to the two components or polypeptide chains (A and B) of a particular I53-type particle (i.e., an I53-50 particle).

[0109] The potentially antigenic polypeptides may be attached to the core of the protein nanostructure either non-covalently or covalently, including as a fusion protein or by other means disclosed herein. Multimeric polypeptides may optionally be displayed along a symmetry axis of the protein nanostructure. Also provided are proteins and nucleic acid molecules encoding such proteins, formulations, and methods of use.

[0110] As used herein, the term “bound” when referring to two moieties means the two moieties are bonded, wherein the bond or bonds connecting the two moieties may be covalent or non-covalent. In some embodiments, the two moieties are covalently bonded to each other (e.g., directly or through a covalently bonded intermediary (a “linker”)). In some embodiments, the two moieties are non-covalently bonded (e.g., through ionic bond(s), Van der Waal’s bond(s)/interactions, hydrogen bond(s), polar bond(s), or combinations or mixtures thereof). For a fusion polypeptide that is N-terminally and C-terminally joined the linker is a peptide bond.

[0111] As used herein, the terms “bioconjugate” and “bioconjugate linker” refers to the resulting association between atoms or molecules of “bioconjugate reactive groups” or “bioconjugate reactive moieties”. The association can be direct or indirect. For example, a conjugate between a first bioconjugate reactive group (e.g., -NH_2 , -C(O)OH , $\text{-N-hydroxysuccinimide}$, or -maleimide) and a second bioconjugate reactive group (e.g., sulfhydryl, sulfur-containing amino acid, amine, amine sidechain containing amino acid, or carboxylate) provided herein can be direct, e.g., by covalent bond or linker (e.g. a first linker of second linker), or indirect, e.g., by non-covalent bond (e.g. electrostatic interactions (e.g. ionic bond, hydrogen bond, halogen bond), van der Waals interactions (e.g. dipole-dipole, dipole-induced dipole, London dispersion), ring stacking (pi effects), hydrophobic interactions and the like). In some embodiments, bioconjugates or bioconjugate linkers are formed using bioconjugate chemistry (i.e., the association of two bioconjugate reactive groups) including, but are not limited to nucleophilic substitutions (e.g., reactions of amines and alcohols with acyl halides, active esters), electrophilic substitutions (e.g., enamine reactions) and additions to carbon-carbon and carbon-heteroatom multiple bonds (e.g., Michael reaction, Diels-Alder addition). These and other useful reactions are discussed in, for example, March, *ADVANCED ORGANIC CHEMISTRY*, 3rd Ed., John Wiley & Sons, New York, 1985; Hermanson, *BIOCONJUGATE TECHNIQUES*, Academic Press, San Diego, 1996; and Feeney et al., *MODIFICATION OF PROTEINS*; *Advances in Chemistry Series*, Vol. 198, American Chemical Society, Washington, D.C., 1982. In some embodiments, the first bioconjugate reactive group (e.g., maleimide moiety) is covalently

attached to the second bioconjugate reactive group (e.g., a sulfhydryl). In some embodiments, the first bioconjugate reactive group (e.g., haloacetyl moiety) is covalently attached to the second bioconjugate reactive group (e.g., a sulfhydryl). In some embodiments, the first bioconjugate reactive group (e.g., pyridyl moiety) is covalently attached to the second bioconjugate reactive group (e.g., a sulfhydryl). In some embodiments, the first bioconjugate reactive group (e.g., -N-hydroxysuccinimide moiety) is covalently attached to the second bioconjugate reactive group (e.g., an amine). In some embodiments, the first bioconjugate reactive group (e.g., maleimide moiety) is covalently attached to the second bioconjugate reactive group (e.g., a sulfhydryl). In some embodiments, the first bioconjugate reactive group (e.g., -sulfo-N-hydroxysuccinimide moiety) is covalently attached to the second bioconjugate reactive group (e.g., an amine).

[0112] Useful bioconjugate reactive moieties used for bioconjugate chemistries herein include, for example:

(a) carboxyl groups and various derivatives thereof including, but not limited to, N-hydroxysuccinimide esters, N-hydroxybenztriazole esters, acid halides, acyl imidazoles, thioesters, p-nitrophenyl esters, alkyl, alkenyl, alkynyl and aromatic esters;

(b) hydroxyl groups which can be converted to esters, ethers, aldehydes, etc.

(c) haloalkyl groups wherein the halide can be later displaced with a nucleophilic group such as, for example, an amine, a carboxylate anion, thiol anion, carbanion, or an alkoxide ion, thereby resulting in the covalent attachment of a new group at the site of the halogen atom;

(d) dienophile groups which are capable of participating in Diels-Alder reactions such as, for example, maleimido or maleimide groups;

(e) aldehyde or ketone groups such that subsequent derivatization is possible via formation of carbonyl derivatives such as, for example, imines, hydrazones, semicarbazones or oximes, or via such mechanisms as Grignard addition or alkyllithium addition;

(f) sulfonyl halide groups for subsequent reaction with amines, for example, to form sulfonamides;

(g) thiol groups, which can be converted to disulfides, reacted with acyl halides, or bonded to metals such as gold, or react with maleimides;

(h) amine or sulfhydryl groups (e.g., present in cysteine), which can be, for example, acylated, alkylated or oxidized;

- (i) alkenes, which can undergo, for example, cycloadditions, acylation, Michael addition, etc;
- (j) epoxides, which can react with, for example, amines and hydroxyl compounds;
- (k) phosphoramidites and other standard functional groups useful in nucleic acid synthesis;
- (l) metal silicon oxide bonding; and
- (m) metal bonding to reactive phosphorus groups (e.g., phosphines) to form, for example, phosphate diester bonds.
- (n) azides coupled to alkynes using copper catalyzed cycloaddition click chemistry.
- (o) biotin conjugate can react with avidin or streptavidin to form an avidin-biotin complex or streptavidin-biotin complex.

[0113] The bioconjugate reactive groups can be chosen such that they do not participate in, or interfere with, the chemical stability of the conjugate described herein. Alternatively, a reactive functional group can be protected from participating in the crosslinking reaction by the presence of a protecting group. In some embodiments, the bioconjugate comprises a molecular entity derived from the reaction of an unsaturated bond, such as a maleimide, and a sulfhydryl group.

[0114] Another method of covalently linking two polypeptides is the SpyTag/SpyCatcher system. The peptide SpyTag (13 amino acids) spontaneously reacts with the protein SpyCatcher (12.3 kDa) to form an intermolecular isopeptide bond between the pair. Polynucleotide sequence encoding either SpyTag or SpyCatcher can be recombinantly introduced into the polynucleotide sequence encoding polypeptides of interest, forming a fusion protein. These fusion proteins can be covalently linked when mixed in a reaction through the SpyTag/SpyCatcher system.

[0115] Using the Tag/Catcher pair, bioconjugation can be achieved between two recombinant proteins that would otherwise be restrictive or impossible with traditional direct genetic fusion between the two proteins. For example, issues regarding protein folding, suboptimal expression host, and specialized post-translational modifications can be alleviated by separating the production of the proteins with the modularity of the Tag/Catcher system. The protein nanostructures may employ a variety of coupling techniques to attach an antigen, including but not limited to the SpyCatcher system

described in, *e.g.*, Escolano et al. *Nature* 570:468-473 (2019), He et al. *Sci Adv.* 7(12):eabf1591 (2021), and Tan et al. *Nat. Commun.* 12(1):542 (2021).

[0116] The term “domain” refers to any portion of a polypeptide that adopts a tertiary structure.

[0117] The terms “assembly domain” and “multimerize” refer to the ability of a polypeptide, or domain of a polypeptide, to form a tertiary structure with another polypeptide or domain of a polypeptide. In some embodiments, assembly domains can form dimers, trimers, tetramers, pentamers, or hexamers and/or to form heteromers with other assembly domains. In some embodiments, the assembly domain forms a protein nanostructure.

[0118] The term “trimerization domain” refers to an assembly domain that forms trimers.

[0119] The term “fragment” refers to a polypeptide having one or more N-terminal or C-terminal truncations compared to a reference polypeptide.

[0120] The term “functional fragment” refers to a fragment that retains at least one function of its reference polypeptide.

[0121] As used herein, the term “conjugated” when referring to two moieties means the two moieties are bonded, wherein the bond or bonds connecting the two moieties may be covalent or non-covalent. In some embodiments, the two moieties are covalently bonded to each other (*e.g.*, directly or through a covalently bonded intermediary). In some embodiments, the two moieties are non-covalently bonded (*e.g.*, electrostatic interactions (*e.g.*, ionic bond, hydrogen bond, halogen bond), Van der Waal’s bond(s)/interactions (*e.g.*, dipole-dipole, dipole-induced dipole, London dispersion), ring stacking (pi effects), hydrophobic interactions and the like), polar bond(s), or combinations or mixtures thereof.

[0122] The term “fusion protein” refers to a chimeric protein encoding two or more separate protein sequences that are recombinantly expressed as a polypeptide, thereby covalently coupling them. For a fusion polypeptide that is N-terminally and C-terminally joined the linker can be a peptide bond or peptide of any length.

[0123] The term “linker” refers to peptide bond or polypeptide that chemically joins the C terminus of a first polypeptide segment to the N terminus of a second polypeptide segment, thereby forming a fusion protein containing the two polypeptide segments.

[0124] The term “signal sequence” refers to a polypeptide sequence, typically at the N terminus of a polypeptide expressed in a host cell that directs the polypeptide to a particular cellular compartment. A signal sequence may be a secretion signal to cause the host cell to secrete the polypeptide into the media in which the host cell is cultured. The signal

sequence can be the “native” signal sequence, a signal sequence that occurs in nature as part of the polypeptide. The signal sequence can be a sequence that does not naturally occur with the polypeptide as found in nature. Various signal sequences are known, and it is within the skill of an ordinary artisan to select an appropriate signal peptide.

[0125] The term “secretes” refers to the ability of host cells to release expressed polypeptides into the media in which they are cultured.

[0126] As may be used herein, the terms “nucleic acid,” “nucleic acid molecule,” “nucleic acid oligomer,” “oligonucleotide,” “nucleic acid sequence,” “nucleic acid fragment” and “polynucleotide” are used interchangeably and are intended to include, but are not limited to, a polymeric form of nucleotides covalently linked together that may have various lengths, either deoxyribonucleotides or ribonucleotides, or analogs, derivatives, or modifications thereof. Different polynucleotides may have different three-dimensional structures, and may perform various functions, known or unknown. Non-limiting examples of polynucleotides include a gene, a gene fragment, an exon, an intron, intergenic DNA (including, without limitation, heterochromatic DNA), messenger RNA (mRNA), transfer RNA, ribosomal RNA, a ribozyme, cDNA, a recombinant polynucleotide, a branched polynucleotide, a plasmid, a vector, isolated DNA of a sequence, isolated RNA of a sequence, a nucleic acid probe, and a primer. Polynucleotides useful in the methods of the disclosure may comprise natural nucleic acid sequences and variants thereof, artificial nucleic acid sequences, or a combination of such sequences.

[0127] The term “messenger RNA” or “mRNA” refers to a polynucleotide that encodes at least one polypeptide. mRNA as used herein encompasses both modified and unmodified RNA. mRNA may contain one or more reading frames or regions.

[0128] The term “pharmaceutically acceptable excipients” and “pharmaceutically acceptable carrier” refer to a substance that aids the administration of an active agent to and absorption by a subject and can be included in the compositions of the present disclosure without causing a significant adverse toxicological effect on the patient and can mean excipients approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in animals, and more particularly in humans.. Non-limiting examples of pharmaceutically acceptable excipients include water, NaCl, normal saline solutions, lactated Ringer’s, normal sucrose, normal glucose, binders, fillers, disintegrants, lubricants, coatings, sweeteners, flavors, salt solutions (such as Ringer’s solution), alcohols, oils, gelatins, carbohydrates such as lactose, amylose or starch, fatty acid esters, hydroxymethylcellulose, polyvinyl pyrrolidone, and

colors, and the like. Such preparations can be sterilized and, if desired, mixed with auxiliary agents such as lubricants, preservatives, stabilizers, wetting agents, emulsifiers, salts for influencing osmotic pressure, buffers, coloring, and/or aromatic substances and the like that do not deleteriously react with the compounds of the disclosure. One of skill in the art will recognize that other pharmaceutical excipients are useful in the present disclosure.

[0129] The term “vaccine” refers to a composition that can provide active acquired immunity to a particular disease (*e.g.* COVID) or a pathogen (*e.g.* coronavirus). A vaccine typically contains one or more agents that can induce an immune response in a subject against a pathogen or disease, *i.e.*, a target pathogen or disease. The immunogenic agent stimulates the body’s immune system to recognize the agent as a threat or indication of the presence of the target pathogen or disease, thereby inducing immunological memory so that the immune system can more easily recognize and destroy any of the pathogen on subsequent exposure. Vaccines can be prophylactic (*e.g.*, preventing or ameliorating the effects of a future infection by any natural or pathogen, or of an anticipated occurrence of cancer in a predisposed subject) or therapeutic (*e.g.*, treating a disease, such as, but not limited to, treating meningitis in a subject in need thereof). The administration of vaccines is referred to vaccination. In some examples, a vaccine composition can provide antigenic molecules or a nucleic acid, *e.g.*, mRNA that encodes antigenic molecules (*e.g.*, peptides) to a subject. The antigenic molecule or nucleic acid that is delivered via the vaccine composition in the subject can be expressed into antigenic molecules and allow the subject to acquire immunity against the antigenic molecules. In the context of the vaccination against infectious disease, the vaccine composition can provide antigenic molecules or mRNA encoding antigenic molecules that are associated with a certain pathogen, *e.g.*, one or more peptides that are known to be expressed in the pathogen (*e.g.*, pathogenic bacterium or virus).

[0130] The term “bivalent vaccine” refers to a vaccine prepared from two different strains. Such vaccines may be combined with other immunogenic or antigenic components to provide a vaccine composition (*e.g.*, combined with one or more recombinant protein antigens).

[0131] The term “adjuvants” refers to a pharmaceutically acceptable substance that enhances the immune response to an antigen when co-administered with the antigen or administered before, during, or after administration of the antigen to a subject.

[0132] The term “subject” refers to a human or non-human animal to which a composition may be administered for vaccination, treatment, or other purpose. In some embodiments,

the non-human animal is a non-human primate, rabbit, hamster, gerbil, pig, cow, sheep, goat, guinea pig, rat, mouse, squirrel, wolf, fox, horse, zebra, giraffe, hyena, elephant, cat, dog, llama, or ferret.

[0133] The term “effective amount” refers to the amount of a composition that, when administered to a patient for treating a state, disorder or condition is sufficient to affect such treatment or when administered to a patient for generating an immune response is sufficient to generate such an immune response. The exact amounts will depend on the active ingredient, the state, disorder, or condition to be treated and its severity, and the age, weight, physical condition and responsiveness of the subject to be treated, and will be ascertainable by one skilled in the art using known techniques (see, e.g., Lieberman, *Pharmaceutical Dosage Forms* (vols. 1-3, 1992); Lloyd, *The Art, Science and Technology of Pharmaceutical Compounding* (1999); Pickar, *Dosage Calculations* (1999); and Remington: *The Science and Practice of Pharmacy*, 20th Edition, 2003, Gennaro, Ed., Lippincott, Williams & Wilkins).

[0134] As used herein, the term “delivering” means providing an entity to a destination. For example, delivering a therapeutic and/or prophylactic to a subject may involve administering a nanoparticle composition including the therapeutic and/or prophylactic to the subject (e.g., by an intravenous, intramuscular, intradermal, intraperitoneal, intratumoral, or subcutaneous route). Administration of a nanoparticle composition to a mammal or mammalian cell may involve contacting one or more cells with the nanoparticle composition.

[0135] As used herein, the term “prevent” refers to a decrease in the occurrence of disease symptoms in a patient. The prevention may be complete (no detectable symptoms) or partial, such that fewer symptoms are observed than would likely occur absent treatment.

[0136] “Treating” or “treatment” as used herein (and as well-understood in the art) broadly refers to any approach for obtaining beneficial or desired results in a subject’s condition, including clinical results. Beneficial or desired clinical results can include, but are not limited to, alleviation or amelioration of one or more symptoms or conditions, diminishment of the extent of a disease, stabilizing (*i.e.*, not worsening) the state of disease, prevention of a disease’s transmission or spread, delay or slowing of disease progression, amelioration or palliation of the disease state, diminishment of the reoccurrence of disease, and remission, whether partial or total and whether detectable or undetectable. In other words, “treatment” as used herein includes any cure, amelioration, or prevention of a disease. Treatment may prevent the disease from occurring; inhibit the disease’s spread; relieve the disease’s

symptoms, fully or partially remove the disease's underlying cause, shorten a disease's duration, or do a combination of these things.

[0137] The term “administering” refers to providing a composition to a subject in a manner that permits the composition to have its intended effect. Administration for vaccination or post-exposure prophylaxis may be performed by intramuscular injection, intravenous injection, intraperitoneal injection, or any other suitable route.

[0138] “Co-administer” means that a composition described herein is administered at the same time, just prior to, or just after the administration of one or more additional therapies. The compositions provided herein can be administered alone or can be co-administered to the subject. Co-administration is meant to include simultaneous or sequential administration of the compounds individually or in combination (more than one compound). Thus, the preparations can also be combined, when desired, with other active substances (e.g., to reduce metabolic degradation).

[0139] The terms “immunization” and “immunizing” refer to administering a composition to a subject in an amount sufficient to elicit, after one or more administering steps, a desired immune response. Immunization may comprise between one and ten, or more administrations (e.g., injections) of the composition, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more administrations. The first administration may elicit no detectable immune response as generally each subsequent administration will boost the immune response generated by prior administrations. The term “immunizing” as used herein includes post-exposure prophylaxis.

[0140] The term “protective immune response” refers to an immune response that prevents and/or reduces the severity of infection with a pathogen when the subject is later challenged with the pathogen, or to an immune response that generates a level of immune response that correlates with protection. For example, vaccination may generate a protective immune response if it results in production, in the plasma or serum, of the subject (e.g., human, pet, or agricultural animal), of neutralizing antibodies that protect the subject against subsequent infection and/or are present in a quantity observed to confer protection upon test subjects (e.g., New Zealand White (NZW) rabbits).

[0141] The term “antibody” refers to a polypeptide encoded by an immunoglobulin gene or functional fragments thereof that specifically binds and recognizes an antigen. The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon, and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma,

mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively.

[0142] The term “infection” or “infectious disease” refers to a disease or condition that can be caused by organisms such as a bacterium, virus, fungi or any other pathogenic microbial agents. In some embodiments, the infectious disease is caused by a pathogenic bacterium. In some embodiments, the infectious disease is a *Campylobacter jejuni*, *Enterococcus faecalis*, *Haemophilus influenzae*, *Helicobacter pylori*, *Klebsiella pneumoniae*, *Legionella pneumophila*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Staphylococcus aureus*, *Streptococcus pneumoniae*, or *Vibrio cholera* infection.

[0143] The term “viral infection” or “viral disease” refers to a disease or condition that is caused by a virus. Non-limiting examples of viral infections include hepatic viral diseases (e.g., hepatitis A, B, C, D, E), herpes virus infection (e.g., HSV-1, HSV-2, herpes zoster), flavivirus infection, Zika virus infection, cytomegalovirus infection, a respiratory viral infection (e.g., adenovirus infection, influenza, severe acute respiratory syndrome, coronavirus infection (e.g., SARS-CoV-1, SARS-CoV-2, MERS-CoV, COVID-19, MERS)), a gastrointestinal viral infection (e.g., norovirus infection, rotavirus infection, astrovirus infection), an exanthematous viral infection (e.g., measles, shingles, smallpox, rubella), viral hemorrhagic disease (e.g., Ebola, Lassa fever, dengue fever, yellow fever), a neurologic viral infection (e.g., West Nile viral infection, polio, viral meningitis, viral encephalitis, Japanese encephalitis, rabies), and human papilloma viral infection.

Recombinant coronavirus polypeptide

[0144] In one aspect, the disclosure provides a recombinant polypeptide, comprising an antigenic fragment of a coronavirus Spike glycoprotein comprising a receptor-binding domain (RBD) and optionally, a coronavirus subdomain 1 (SD1). The antigenic fragment comprises RBD-SD1. In some embodiments, the antigenic fragment comprises essentially of RBD-SD1. In some embodiment, the antigenic fragment comprises RBD-SD1 portion isolated from the remaining of the Spike glycoprotein. Some flanking sequence around the domain may be in use. In some embodiments, the antigenic fragment comprises RBD. The amino acid substitutions may be introduced into coronavirus Spike glycoproteins, or fragments thereof, from diverse Coronavirus strains (including, but not limited to MERS, SARS-CoV-1, and SARS-CoV-2).

[0145] An exemplary SARS-CoV-2 Spike glycoprotein sequence is set forth herein as SEQ ID NO: 1. In some embodiments, the polypeptide comprises a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at

least 87%, at least 88%, at least 89%, 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 100% identical to SEQ ID NO: 1.

```

1 MFVFLVLLPL VSSQCVNLTT RTQLPPAYTN SFTRGVYYPD KVFRSSVLHS
51 TQDLFLPFFS NVTWFHAIHV SGTNGTKRFD NPVLPFNDGV YFASTEKSNI
101 IRGWIFGTTL DSKTQSLIV NNATNVVIKV CEFQFCNDPF LGVYYHKNNK
151 SWMESEFRVY SSANNCTFEY VSQPFLMDLE GKQGNFKNLR EFVFKNIDGY
201 FKIYSKHTPI NLVRDLPQGF SALEPLVDLP IGINITRFQT LLALHRSYLT
251 PGDSSSGWTA GAAAYYVGYL QPRTFLLKYN ENGTITDAVD CALDPLSETK
301 CTLKSFTVEK GIYQTSNFRV QPTESIVRFP NITNLCPFGE VFNATRFASV
351 YAWNRRKRISN CVADYSVLYN SASFSTFKCY GVSPTKLNDL CFTNVYADSF
401 VIRGDEVROI APGQTGKIAD YNYKLDDFT GCVIAWNSNN LDSKVGNNYN
451 YLYRLFRKSN LKPFERDIST EIYQAGSTPC NGVEGFNCYF PLQSYGFQPT
501 NGVGYQPYRV VVLSFELLHA PATVCGPKKS TNLVKNKCVN FNFNGLTGTG
551 VLTESNKKFL PFQQFGRDIA DTTDAVRDPQ TLEILDITPC SFGGVSVITP
601 GTNTSNQVAV LYQDVNCTEV PVAIHADQLT PTWRVYSTGS NVFQTRAGCL
651 IGAEHVNNSY ECDIPIGAGI CASYQTQTN PRRARVASQ SIIAYTMSLG
701 AENSVAYSNN SIAIPTNFTI SVTTEILPVS MTKTSVDCTM YICGDSTECs
751 NLLQYGSFC TQLNRALTGI AVEQDKNTQE VFAQVKQIYK TPPIKDFGGF
801 NFSQILPDPS KPSKRSFIED LLFNKVTLD AGFIKQYGDC LGDIAARDLI
851 CAQKFENGLTV LPPLLTDEMI AQYTSALLAG TITSGWTFGA GAALQIPFAM
901 QMAYRFNGIG VTQNVLYENQ KLIANQFNSA IGKIQDSLSS TASALGKLQD
951 VVNQNAQALN TLVKQLSSNF GAISSVLNDI LSRLDKVEAE VQIDRLITGR
1001 LQSLQTYVTQ QLIRAAEIRA SANLAATKMS ECVLGQSKRV DFCGKGYHLM
1051 SFPQSAPHGV VFLHVTYVPA QEKNTTAPA ICHDGKAHFP REGVFSNGT
1101 HWFVTQRNFY EPQIITDNT FVSGNCDVVI GIVNNTVYDP LQPELDSFKE
1151 ELDKYFKNHT SPDVDLGDIS GINASVVNIQ KEIDRLNEVA KNLNESLIDL
1201 QELGKYEQYI KWPWYIWLGF IAGLIAVMV TIMLCMTSC CSCLKGCCSC
1251 GSCCKFDEDD SEPVLKGVKL HYT

```

(SEQ ID No: 1)

[0146] Illustrative antigenic fragment sequences from diverse strains of coronavirus are provided in **Table 1**. In some embodiments, the antigenic fragment sequence is at least 80%,

at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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 99%, or 100% identical to any polypeptide sequence listed in **Table 1**.

Table 1.

SEQ ID NO	Sequence	Strain
2	RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNR KRISNCVADYSVLYNSASFSTFKCYGVSP TKLNDLCF TNVYADSFVIRGDEV RQIAPGQTGKIADYNYKLPDD FTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPF ERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTN GVGYQP YRVVLSFELLHAPATVCGPKKSTNLVKNK CVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTD AVRDPQTLEILDITPCS	Ancestral
3	RVQPTESIVRFPNITNLCPFDEVFNATRFASVYAWNR KRISNCVADYSVLYNFAPFFAFKCYGVSP TKLNDLCF TNVYADSFVIRGNEVSQIAPGQTGNIADYNYKLPDDF TGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPF RDISTEIIYQAGNKPCNGVAGVNCYFPLQSYGFRPTYG VGHQP YRVVLSFELLHAPATVCGPKKSTNLVKNKC VNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDA VRDPQTLEILDITPCS	BA.5
4	RVQPTESIVRFPNITNLCPFDEVFNATRFASVYAWNR KRISNCVADYSVLYNLAPFFTFKCYGVSP TKLNDLCF TNVYADSFVIRGDEV RQIAPGQTGNIADYNYKLPDD FTGCVIAWNSNKLDSKVSGNYNYLYRLFRKSNLKPF ERDISTEIIYQAGNKPCNGVAGFNCYFPLRSYSFRPTY GVGHQP YRVVLSFELLHAPATVCGPKKSTNLVKNK CVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTD AVRDPQTLEILDITPCS	BA.1

5	RVQPTESIVRFPNITNLCPFHEVFNATTFASVYAWNR KRISNCVADYSVIYNFAPFFAFKCYGVSPTKLNDLCF TNVYADSFVIRGNEVSQIAPGQTGNIADYNYKLPDDF TGCVIAWNSNKLDKPSGNYNLYRLFRKSKLPFE RDISTEIQAGNKPCNGVAGSNCYSPLQSYGFRPTYG VGHQPYRVVVLSELLHAPATVCGPKKSTNLVKNKC VNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDA VRDPQTLEILDITPCS	XBB
6	RVVPSGDVVRFPNITNLCPFGEVFNATKFPSVYAWER KKISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCF SNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDD FMGCVLAWNTRNIDATSTGNYNYKYRYLRHGKLRP FERDISNVPFSPDGKPCPPALNCYWPLNDYGFTTT GIGYQPYRVVVLSELLNAPATVCGPKLSTDLIKNQC VNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDS VRDPKTSEILDISPCS	SARS-CoV-1
7	RVQPTESIVRFPNITNLCPFDEVFNATRFASVYAWNR KRISNCVADYSVLYNFAPFFAFKCYGVSPTKLNDLCF TNVYADSFVIRGNEVSQIAPGQTGNIADYNYKLPDDF TGCVIAWNSNKLDKVGGNYNLYRLFRKSNLKPF RDISTEIQAGNKPCNGVAGFNCYFPLRSYGFRPTYG VGHQPYRVVVLSELLHAPATVCGPKKSTNLVKNKC VNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDA VRDPQTLEILDITPCS	BA.2
8	RVQPTESIVRFPNITNLCPFDEVFNATRFASVYAWNR KRISNCVADYSVLYNFAPFFAFKCYGVSPTKLNDLCF TNVYADSFVIRGNEVSQIAPGQTGNIADYNYKLPDDF TGCVIAWNSNKLDKVGGNYNLYRLFRKSNLKPF RDISTEIQAGNKPCNGVAGFNCYFPLRSYGFRPTYG VGHQPYRVVVLSELLHAPATVCGPKKSTNLVKNKC VNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDA VRDPQTLEILDITPCS	BA.2.12

9	RVQPTESIVRFPNITNLCPFDEVFNATRFASVYAWNR KRISNCVADYSVLYNFAPFFAFKCYGVSP TKLNDLCF TNVYADSFVIRGNEVSQIAPGQTGNIADYNYKLPDDF TGCVIAWNSNKLD SKVGGNYNYQYRLFRKSNLKPF RDISTEIQAGNKPCNGVAGFNCYFPLRSYGFRPTYG VGHQPYRVVVL SFELLHAPATVCGPKKSTNLVKNKC VNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDA VRDPQTLEILDITPCS	BA.2.75
10	RVQPTESIVRFPNITNLCPFHEVFNATTFASVYAWNR KRISNCVADYSVIYNFAPFFAFKCYGVSP TKLNDLCF TNVYADSFVIRGNEVSQIAPGQTGNIADYNYKLPDDF TGCVIAWNSNKLD SKPSGNYNYLYRLFRKSKLKPF RDISTEIQAGNKPCNGVAGPNCYSPLQSYGFRPTYG VGHQPYRVVVL SFELLHAPATVCGPKKSTNLVKNKC VNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDA VRDPQTLEILDITPCS	XBB.1.5
245	RVQPTESIVRFPNITNLCPFHEVFNATTFASVYAWNR KRISNCVADYSVIYNFAPFFAFKCYGVSP TKLNDLCF TNVYADSFVIRGNEVSQIAPGQTGNIADYNYKLPDDF TGCVIAWNSNKLD SKPSGNYNYLYRLLRKSCLKPF RDISTEIQAGNKPCNGVAGPNCYSPLQSYGFRPTYG VGHQPYRVVVL SFELLHAPATVCGPKKSTNLVKNKC VNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDA VRDPQTLEILDITPCS	EG.5.1
246	RVQPTESIVRFPNVTNLCPFHEVFNATRFASVYAWNR TRISNCVADYSVLYNFAPFFAFKCYGVSP TKLNDLCF TNVYADSFVIKGNEVSQIAPGQTGNIADYNYKLPDDF TGCVIAWNSNKLD SKHSGNYDYWYRLFRKSKLKPF ERDISTEIQAGNKPCKGKGPNCYFPLQSYGFRPTYG VGHQPYRVVVL SFELLHAPATVCGPKKSTNLVKNKC VNFNFNGLTGTGVLTKSNKKFLPFQQFGRDIVDTTD AVRDPQTLEILDITPCS	BA.2.86

247	RVQPTESIVRFPNVTNLCPFHEVFNATRFASVYAWNR TRISNCVADYSVLYNFAPFFAFKCYGVSPTKLNDLCF TNVYADSFVIKGNEVSQIAPGQTGNIADYNYKLPDDF TGCVIAWNSNKLDSKHSGNYDYWYRSFRKSKLPFE RDISTEIQAGNKPCGKGPNCYFPLQSYGFRPTYGV GHQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCV NFNFNGLTGTGVLTKSNKKFLPFQQFGRDIVDTDA VRDPQTLEILDITPCS	JN.1
248	RVQPTDSIVRFPNITNLCPFGEVFNATTFASVYAWNR KRISNCVADYSVLYNSTSFSTFKCYGVSPTKLNDLCF TNVYADSFVITGDEVQRQIAPGQTGKIADYNYKLPDDF TGCVIAWNSKHIDAKEGGNFNYLYRLFRKANLKPFE RDISTEIQAGSKPCNGQTGLNCYYPLYRYGFYPTDG VGHQPYRVVLSFELLNAPATVCGPKKSTNLVKNKC VNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDA VRDPQTLEILDITPCS	RaTG13
249	RVQPTISIVRFPNITNLCPFGEVFNASKFASVYAWNR KRISNCVADYSVLYNSTSFSTFKCYGVSPTKLNDLCF TNVYADSFVVKGDEVQRQIAPGQTGVIADYNYKLPDD FTGCVIAWNSVKQDALTTGGNYGYLYRLFRKSKLPF ERDISTEIQAGSTPCNGQVGLNCYYPLERYGFHPTT GVNYQPFRVVLSFELLNGPATVCGPKLSTTLVKDK CVNFNFNGLTGTGVLTTSKKQFLPFQQFGRDISDTTD AVRDPQTLEILDITPCS	Pangolin17
272	RVAPSKEVVRFNITNLCPFGEVFNATTFPSVYAWER KRISNCVADYSVLYNSTSFSTFKCYGVSATKLNDLCF SNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDD FTGCVLAWNTRNIDATQTGNYNKYRSLRHGKLRPF ERDISNVPFSPDGKCTPPAFNCYWPLNDYGFYITNGI GYQPYRVVLSFELLNAPATVCGPKLSTDLIKNQCV NFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDSV RDPKTSEILDISPCS	WIV1

273	RVAPSKVVRFNPITNLCPFGEVFNATTFPSVYAWER KRISNCVADYSVLYNSTSFSTFKCYGVSATKLNDLCF SNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDD FLGCVLAWNTNSKDSSTSGNLYLYRWVRRSKLNP YERDLSNDIYSPGGQSCSAVGPNPCYNPLRPYGFFTTA GVGHQPYPYRVVLSFELLNAPATVCGPKLSTDLIKQ CVNFNENGLTGTGVLTPSSKRFQPFQFGRDVSDFTD SVRDPKTSEILDISPCS	SHC014
274	RILPSTEVRFPNITNFCPFDKVFNATRFPNVYAWQR TKISDCIADYTVLYNSTSFSTFKCYGVSPSKLIDLCFTS VYADTFLIRFSEVRQIAPGETGVIADYNYKLPDDFTG CVLAWNTAQDIGSYFYRSHRAVKLKPFERDLSSDE NGVRTLSTYDFNPVPLDYQATR VVLSFELLNAPA TVCGPKLSTQLVKNRCVNFNENGLRGTGVLTDSDKR FQSFQFGRDSADFTDSVRDPQTLQILDISPCS	RmYN02
275	RVSPSTEVRFPNITNLCPFQVFNASNFPVYAWER LRISDCVADYAVLYNSSSFSTFKCYGVSPTKLNDLC FSSVYADYFVVKGDDVRQIAPAQTGVIADYNYKLPD DFTGCVLAWNTNSVDSKSGNNFYRLFRHGKIKPYE RDISNVLYNSAGGTCSSISQLGCYEPLKSYGFTPTVG VGYPYRVVLSFELLNAPATVCGPKKSTELVKNKC VNFNENGLTGTGVLTSSTKKFQPFQFGRDVSDFTD VRDPKTFEILDISPCS	BTKY72
238	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVY YDPKVFRSSVLHSTQDLFLPFFSNVTWFHAISGTNGT KRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKT QSLIVN NATNVVIKVCEFQFCNDPFLGVYHKNNKS WMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNF KNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALE PLVDLPIGINITRFQTLLALHRSYLT PGDSSSGWTAGA AAYYVGYLQPRTFLLKYNENGITITDAVDCALDPLSE TKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLC PFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSA	SARS-CoV-2 Alpha (B.1.1.7)

	SFSTFKCYGVSP TKLNDLCFTNVYADSFVIRGDEV RQ IAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKV GGNYNYLYRLFRKSNLKPFERDISTEIQAGSTPCNG VEGFNCYFPLQSYGFQPTYGVGYQP YRVVLSFELL HAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTE SNKKFLPFQQFGRDIDDTDAVRDPQTLEILDITPCSF GGVSVITPGTNTSNQVAVLYQGVNCTEVPVAIHADQ LTPTWRVYSTG SNVFQTRAGCLIGAEHVNSYEC DIP IGAGICASYQTQTNSHRRARSVASQSIIAYTMSLGAE NSVAYSNN SIAIPINF TISVTTEILPVSMKTSVDCTMY ICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNT QEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRS FIEDLLFNKVT LADAGFIKQYGDCLGDIAARDLICAQ KFNGLTVLP LLTDEMIAQYTSALLAGTITSGWTFGA GAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQ FNSAIGKIQDSL SASTASALGKLQDVVNQNAQALNTLV KQLSSNF GAISSVLNDILARLDKVEAEVQIDRLITGRL QSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQS KRVDFCGKGYHLM SFPQSAPHGVVFLHVTYVPAQE KNFTTAPAICH DGKAHFPREGVFVSNGTHWFVTQRN FYEPQIITHTNTFVSGNCDVVIGIVNNTVYDPLQPELD SFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEID RLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWL GFI AGLIAIVMVTIMLCCMTSCC SCLKGCCSCGSCCKFDE DDSEPVLKGVKLHYT	
239	MFVFLVLLPLVSSQCVNL TTRTQLPPAYTNSFTRGVY Y PDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTN G TKRFANPVL PFNDGVYFASTEKSNIIRGWIFGTTLDS K TQSL LIVN NATNVVIKVCE FQFCNDPFLGVYYHKN N KSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQ G NFKNLREFVFKNIDGYFKIYSKHTPINLVRGLPQGFS A LEPLVDLPIGINITRFQTLHRSYLT PGDSSSGWTAGA A AYYYVGYLQPRTFLLKYNENG TITDAVDCALDPLSE	SARS-CoV-2 Beta (B.1.351)

	TKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLC PFGEVFNATRFASVYAWNKRISNCVADYSVLYNSA SFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQR IAPGQTGNIADYNYKLPDDFTGCVIAWNSNNLDSKV GGNYNYLYRLFRKSNLKPFERDISTEIQAGSTPCNG VKGFNICYFPLQSYGFQPTYGVGYQPYRVVLSFELL HAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTE SNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSF GGVSVITPGTNTSNQVAVLYQGVNCTEVPVAIHADQ LTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSYECDIP IGAGICASYQTQTNSPRRARSVASQSIAYTMSLGVEN SVAYSNNISIAIPTNFTISVTTEILPVSMTKTSVDCTMYI CGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNT QEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRS FIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQ KFNGLTVLPPLLDEMIQYTSALLAGTITSGWTFGA GAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQ FNSAIGKIQDSLSSASALGKLQDVVNQNAQALNTLV KQLSSNFGAISSVLNDILSRDKVEAEVQIDRLITGRL QSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQS KRVDFCGKGYHLSFQSPHGVVFLHVTYVPAQE KNFTTAPAICHGDKAHFPREGVVFVSNGTHWFVTQRN FYEPQIITDNTFVSGNCDVVIGIVNNTVYDPLQPELD SFKEELDKYFKNHTSPDVLGDISGINASVVNIQKEID RLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFI AGLIAIVMVTIMLCCMTSCCCLKGCCSCGSCCKFDE DDSEPVLKGVKLHYT	
240	MFVFLVLLPLVSSQCVNFTNRTQLPSAYTNSFTRGVY YPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTN GTRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDS KTQSLIVN NATNVVIK VCEFQFCNYPFLGVYYHKN NKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQ GNFKNLSEFVFKNIDGYFKIYSKHTPINLVRLDPQGFS	SARS-CoV-2 Gamma (P.1)

	ALEPLVDLPIGINITRFQTLALHRSYLT PGDSSSGWT AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDP LSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITN LCPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYN SASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGTIADYNYKLPDDFTGCVIAWNSNNLDS KVGGNYNLYRLFRKSNLKPFERDISTEIQAGSTPC NGVKGFNCYFPLQSYGFQPTYGVGYQPYRVVLSFE LLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVL TESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPC SFGGVSVITPGTNTSNQVAVLYQGVNCTEVPVAIHA DQLTPTWRVYSTGNSNVFQTRAGCLIGA EYVNNSEYEC DIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLG AENSVAYSNNIAIPTNFTISVTTEILPVSMTKTSVDCT MYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK NTQEVFAQVKQIYKTPIKDFGGFNFSQILPDPSKPSK RSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLIC AQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTF GAGAAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIA NQFNSAIGKIQDSLSTASALGKLQDVVNQNAQALN TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLIT GRLQSLQTYVTQQLIRAAEIRASANLAAIKMSECVLG QSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPA QEKNFTTAPAICHDKGAHFPREGVFVSNGTHWFVTQ RNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPE LDSFKEELDKYFKNHTSPDVDLGDISGINASFVNIQK EIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIW LGFIAGLIAIVMTIMLCCMTSCCCLKGCCSCGSCC KFDEDDSEPVLKGVKLHYT	
241	MFVFLVLLPLVSIQCVNLTTRTQLPPAYTNSFTRGVY YPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTN GTRKFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDS KTQSLIVNNA TNVVIKVCEFCNDPFLGVYYHKN	SARS-CoV-2 Epsilon (B.1.427)

	NKSCMESEFRVYSSANNCTFEYVSQPFLMDLEGKQG NFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFS LEPLVDLPIGINITRFQTLALHRSYLTPGDSSSGWTA GAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPL SETKCTLSFTVEKGIYQTSNFRVQPTESIVRFPNITNL CPFGEVFNATRFASVYAWNKRISNCVADYSVLYNS ASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNYKLPPDFTGCVIAWNSNNLDSK VGGNYNYRYRLFRKSNLKPFERDISTEIQAGSTPCN GVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFEL LHAPATVCGPKKSTNLVKNKCVNFNENGLTGTGVLT ESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCS FGGVSVITPGTNTSNQVAVLYQGVNCTEVPVAIHAD QLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSECD IPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGA ENSVAYSNNIAIPTNFTISVTTEILPVSMTKTSVDCT MYICGDSTECNLLLQYGSFCTQLNRALTGIAVEQDK NTQEVFAQVKQIYKTPIKDFGGFNFSQILPDPSKPSK RSFIEDLLFNKVTADAGFIKQYGDCLGDIAARDLIC AQKFENGLTVLPPLLTDEMIAQYTSALLAGTITSGWTF GAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIA NQFNSAIGKIQDSLSTASALGKLQDVVNQNAQALN TLVKQLSSNFGAISSVLNDILSRDKVEAEVQIDRLIT GRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVL GQSKRVDFCGKGYHLSFPQSAPHGVVFLHVTYVP AQEKNFTTAPAICHGDKAHFPREGVFVSNGTHWFVT QRNFYEPQIITDNTFVSGNCDVVIGIVNNTVYDPLQP ELDSFKEELDKYFKNHTSPDVLGDISGINASVVNIQ KEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYI WLGFIAGLIAIVMVTIMLCCMTSCCCLKGCCSCGSC CKFDEDDSEPVLKGVKLHYT	
242	MFVFLVLLPLVSSQCVNLRTRTQLPPAYTNSFTRGVY YPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSNTN	SARS-CoV-2 Delta

GTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDS KTQSLIVNNAATNVVIKVCEFQFCNDPFLDVYYHKN NKSWMESGVYSSANNCTFEYVSQPFLMDLEGKQGN FKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSAL EPLVDLPIGINITRFQTLALHRSYLT PGDSSSGWTAG AAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLS ETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNL CPFGEVFNATRFASVYAWNKRISNCVADYSVLYNS ASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEV QIAPGQTGKIADYNYKL PDDFTGCVIAWNSNNLDSK VGGNYNYRYRLFRKSNLKPFERDISTEIYQAGSKPCN GVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVSFEL LHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLT ESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCS FGGVSVITPGTNTSNQVAVLYQGVNCTEVPVAIHAD QLTPTWRVYSTGSNVFQTRAGCLIGAEHVNSYECD IPIGAGICASYQTQTSRARRSVASQSIIAYTMSLGA ENSVAYSNNIAIPTNFTISVTTEILPVSMTKTSVDCT MYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDK NTQEVFAQVKQIYKTPIKDFGGFNFSQILPDPSPSK RSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLIC AQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTF GAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIA NQFNSAIGKIQDSLSTASALGKLQNVVNQNAQALN TLVKQLSSNFGAISSVLNDILSRDLKVEAEVQIDRLIT GRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVL GQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVP AQEKNFTTAPAICHGDKAHFPREGVFVSNGTHWFVT QRNFYEPQIITDNTFVSGNCDVVIGIVNNTVYDPLQP ELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQ KEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYI WLGFIAGLIAIVMTIMLCCMTSCCCLKGCCSCGSC CKFDEDDSEPVLKGVKLHYT	(B.1.617.2) complete spike
--	---------------------------------------

243	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVY YDPKVFRRSSVLHSTQDLFLPFFSNVTWFHVISGTNGT KRFDNPVLPFNDGVYFASIEKSNIRGWIFGTTLDSKT QSLIVNNAATNVVIKVC EFQFCNDPFLDHKNNKSWM ESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNL REFVFKNIDGYFKIYSKHTPIIVREPEDLPQGFSALEPL VDLPIGINITRFQTLALHRSYLT PGDSSSGWTAGAA AYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSET KCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPF DEVFNATRFASVYAWNRRKISNCVADYSVLYNLAPF FTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQA PGQTGNIAADYNKLPDDFTGCVIAWNSNKLDSKVSG NYYLYRLFRKSNLKPFERDISTEIQAGNKPCNGVA GFNCYFPLRSYSFRPTYGVGHQPYRVVVL SFELLHAP ATVCGPKKSTNLVKNKCVNFNFNGLKGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGG VSVITPGTNTSNQVAVLYQGVNCTEVPVAIHADQLTP TWRVYSTGSNVFQTRAGCLIGAEYVNNSYECDIPIGA GICASYQTQTKSHRRARSVASQSIIAYTMSLGAENSV AYSNNIAIPTNFTISVTTEILPVSMTKTSVDCTMYICG DSTECNLLLQYGSFCTQLKRALTGIAVEQDKNTQEV FAQVKQIYKTPPIKYFGGFNFSQILPDPSKPSKRSFIED LLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFK GLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAA LQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNS AIGKIQDSLSTASALGKLQDVVNHNAQALNTLVKQ LSSKFGAISSVLNDIFSRLDKVEAEVQIDRLITGRLQSL QTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRV DFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFT TAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEP QIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKE ELDKYFKNHTSPDVLGDISGINASVVNIQKEIDRLN EVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIA	SARS-CoV-2 Omicron (BA1, previously B.1.1.529) complete spike
-----	---	---

	LIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD SEPVCLKGVKLHYT	
244	MFVFLVLLPLVSSQCVNLITRTQSYTNSFTRGVYYPD KVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTK RFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQ SLIVNNATNVVIKVCEFQFCNDPFLDVYYHKNNKS WMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNF KNLREFVFKNIDGYFKIYSKHTPINLGRDLPQGFSALE PLVDLPIGINITRFQTLLALHRSYLTGDSGWTAGA AAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSE TKCTLSFTVEKGIYQTSNFRVQPTESIVRFPNITNLC PFDEVFNATRFASVYAWNRKRISNCVADYSVLYNFA PFFAFKCYGVSPTKLNDLCFTNVYADSFVIRGNEVSQ IAPGQTGNIADYNYKLPDDFTGCVIAWNSNKLDISKV GGNYNYLYRLFRKSNLKPFERDISTEIYQAGNKPCNG VAGFNCYFPLRSYGFRPTYGVGHQPYRVVLSFELL HAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTE SNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSF GGVSVITPGTNTSNQVAVLYQGVNCTEVPVAIHADQ LTPTWRVYSTGSNVFQTRAGCLIGAEYVNNSYECDIP IGAGICASYQTQTKSHRRARSVASQSIIAYTMSLGAE NSVAYSNNIAIPTNFTISVTTEILPVSMKTTSVDCTM YICGDSTECNLLLQYGSFCTQLKRALTGIAVEQDKN TQEVFAQVKQIYKTPPIKYFGGFNFSQILPDPSKPSKR SFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICA QKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFG AGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIAN QFNSAIGKIQDSLSTASALGKLQDVVNHNAQALNTL VKQLSSKFGAISSVLNDILSRDKVEAEVQIDRLITGR LQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQS KRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQE KNFTTAPAICHGDKAHFPREGVFVSNGTHWFVTQRN FYEPQIITDNTFVSGNCDVVIGIVNNTVYDPLQPELD	SARS-CoV-2 complete spike

	SFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEID RLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFI AGLIAIVMVTIMLCCMTSCCSCCLKGCCSCGSCCKFDE DDSEPVLKGVKLHYT	
254	MFVFLVLLPLVSSQCVNLITRTQSYTNSFTRGVYYPD KVFRSSVLHSTQDLFLPFFSNVTWFHAIHVS GTNGTK RFDNPALPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQ SLLIVNNATNVVIKVCEFQFCNDPFLDVYQKNNKSW MESEFRVYSSANNCTFEYVSQPFLMDLEGKEGNFKN LREFVFKNIDGYFKIYSKHTPINLERDLPQGFSALEPL VDLPIGINITRFQTLALHRSYLT PVDSSSGW TAGAA A Y Y V G Y L Q P R T F L L K Y N E N G T I T D A V D C A L D P L S E T K C T L K S F T V E K G I Y Q T S N F R V Q P T E S I V R F P N I T N L C P F H E V F N A T T F A S V Y A W N R K R I S N C V A D Y S V I Y N F A P F F A F K C Y G V S P T K L N D L C F T N V Y A D S F V I R G N E V S Q I A P G Q T G N I A D Y N Y K L P D D F T G C V I A W N S N K L D S K P S G N Y N Y L Y R L F R K S K L K P F E R D I S T E I Y Q A G N K P C N G V A G P N C Y S P L Q S Y G F R P T Y G V G H Q P Y R V V L S F E L L H A P A T V C G P K K S T N L V K N K C V N F N F N G L T G T G V L T E S N K K F L P F Q Q F G R D I A D T T D A V R D P Q T L E I L D I T P C S F G G V S V I T P G T N T S N Q V A V L Y Q G V N C T E V P V A I H A D Q L T P T W R V Y S T G S N V F Q T R A G C L I G A E Y V N N S Y E C D I P I G A G I C A S Y Q T Q T K S H R R A R S V A S Q S I I A Y T M S L G A E N S V A Y S N N S I A I P T N F T I S V T T E I L P V S M T K T S V D C T M Y I C G D S T E C S N L L L Q Y G S F C T Q L K R A L T G I A V E Q D K N T Q E V F A Q V K Q I Y K T P P I K Y F G G F N F S Q I L P D P S K P S K R S F I E D L L F N K V T L A D A G F I K Q Y G D C L G D I A A R D L I C A Q K F N G L T V L P P L L T D E M I A Q Y T S A L L A G T I T S G W T F G A G A A L Q I P F A M Q M A Y R F N G I G V T Q N V L Y E N Q K L I A N Q F N S A I G K I Q D S L S S T A S A L G K L Q D V V N H N A Q A L N T L V K Q L S S K F G A I S S V L N D I L S R L D K V E A E V Q I D R L I T G R L Q S L Q T Y V T Q Q L I R A A E I R A S A N L A A T K M S E C V L G Q S K R V D F C G K G Y H L M S F P Q S A P H G V V F L H V T Y V P A Q E K N	XBB.1.5 complete spike

	FTTAPAICHGDKAHFPREGVFSNGTHWFVTQRNFY EPQIITDNTFVSGNCDVVIGIVNNTVYDPLQPELDSF KEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDR LNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFI AGLIAIVMVTIMLCCMTSCCSCCLKGCCSCGSCCKFDE DDSEPVLKGVKLHYT	
255	MFVFLVLLPLVSSQCVNLITRTQSYTNSFTRGVYYPD KVFRSSVLHSTHDLFLPFFSNVTWFHAIHVSGTNGTK RFDNPALPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQ SLLIVN NATNVVIK VCEFQFCNDPFLDVYQKNNKSW MESEFRVYSSANNCTFEYVSQPFLMDLEGKEGNFKN LREFVFKNIDGYFKIYSKHTPINLERDLPQGFSALEPL VDLPIGINITRFQTLALHRSYLTVPDSSSGWTAGAA AYYVGYLQPRTFLLKYNENG TITDAVDCALDPLSET KCTLSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPF HEVFNATTFASVYAWNRRKRISNCVADYSVIYNFAPF FAFKCYGVSPTKLNDLCFTNVYADSFVIRGNEVSQIA PGQTGNIA DYNKLPDDFTGCVIAWNSNKLDSKPSG NYNYLYRLLRKSCLKPFERDISTEIQAGNKPCNGVA GPNCYSPLQSYGFRPTYGVGHQPYRVVLSFELLHA PATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTL EILDITPCSFGG VSVITPGTNTSNQVAVLYQGVNCTEVPVAIHADQLTP TWRVYSTGSNVFQTRAGCLIGA EYVNNSYECDIPIGA GICASYQTQTKSHRRARSVASQSIIAYTMSLGAENSV AYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICG DSTECSNLLLQYGSFCTQLKRALTGIAVEQDKNTQEV FAQVKQIYKTPPIKYFGGFNFSQILPDPSKPSKRSFIED LLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFN GLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAA LQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNS AIGKIQDSLSSSTASALGKLQDVVNHNAQALNTLVKQ LSSKFGAISSVLNDILSRDKVEAEVQIDRLITGRLQS	EG.5.1 complete spike

	LQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKR VDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEK FTTAPAICHGDKAHFPREGVFVSNGTHWFVTQRNFY EPQIITDNTFVSGNCDVVIGIVNNTVYDPLQPELDSF KEELDKYFKNHTSPDVLGDISGINASVVNIQKEIDR LNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFI AGLIAIVMVTIMLCCMTSCCSCCLKGCCSCGSCCKFDE DDSEPVLKGVKLHYT	
256	MFVFLVLLPLVSSQCVMPFLNLITTTQSYTNSFTRGV YYPDKVFRSSVLHLTQDLFLPFFSNVTWFHAISGTNG TKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSK TQSLNATNVFIKVCEFQFCNDPFLDVYHKNNKS WMESESGVYSSANNCTFEYVSQPFLMDLEGKQGNF KNLREFVFKNIDGYFKIYSKHTPIGRDFPQGFSALEP LVDLPIGINITRFQTLALNRSYLTGDSGWTAGAA DYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSET KCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNVTNLCP FHEVFNATRFASVYAWNRTNISNCVADYSVLYNFAP FFAFKCYGVSPKLNLCFTNVYADSFVIKGNESQI APGQTGNIADYNYKLDDFTGCVIAWNSNKLDSKHS GNYDYWYRLFRKSKLPFERDISTEIQAGNKPCKG KGPNCYFPLQSYGFRPTYGVGHQPYRVVLSFELLH APATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTKS NKKFLPFQQFGRDIVDTTDAVRDPQTLEILDITPCSFG GVSVITPGTNTSNQVAVLYQGVNCTEVSVAIHADQL TPTWRVYSTGSNVFQTRAGCLIGAEYVNNSECDIPI GAGICASYQTQTKSRRRARSVASQSIIAYTMSLGAEN SVAYSNNIAIPTNFTISVTTEILPVSMTKTSVDCTMYI CGDSTECNLLLQYGSFCTQLKRALTGIAVEQDKNT QEVFAQVKQIYKTPPIKYFGGFNFSQILPDPSKPSKRS FIEDLLFNKVTADAGFIKQYGDCLGDIAARDLICAQ KFNGLTVLPPLTDEMIAQYTSALLAGTITSGWTFGA GAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQ	BA.2.86 complete spike

	FNSAIGKIQDSLSTASALGKLQDVVNHNAQALNTLV KQLSSKFGAISSVLNDILSRDKVEAEVQIDRLITGRL QSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQS KRVDFCGKGYHLMSEFPQSAPHGVVFLHVTYVPAQE KNFTTAPAICHGDKAHFPREGVVFVSNNGTHWFTQRN FYEPQIITDNTFVSGNCDVVIGIVNNTVYDPLQLELD SFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEID RLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGF AGLIAIVMVTIMLCCMTSCCCLKGCCSCGSCCKFDE DDSEPVKGVKLHYT	
257	MFVFLVLLPLVSSQCVNLITTTQSYTNSFTRGVYYPD KVFRSSVLHLTQDLFLPFFSNVTWFWHAISGTNGTKRF DNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSL LIVNNATNVFIKVCEFQFCNDPFLDVYHKNNKSWME SESGVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLR EFVFKNIDGYFKIYSKHTPIIGRDFPQGFSALEPLVDLP IGINITRFQTLALNRSYLT PGDSSSGWTAGAADYYV GYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLK SFTVEKGIYQTSNFRVQPTESIVRFPNVTNLCPFHEVF NATRFASVYAWNRTNISNCVADYSVLYNFAPFFAFK CYGVSPTKLNDLCFTNVYADSFVIKGNEVSQIAPGQT GNIADYNYKLPDDFTGCVIAWNSNKLDSKHSGNYD YWYRSFRKSKLKPFERDISTEIQAGNKPCGKGPNC YFPLQSYGFRPTYGVGHQPYRVVVLSELLHAPATV CGPKKSTNLVKNKCVNFNFNGLTGTGVLTksnkkfl PFQQFGRDIVDTTDAVRDPQTLEILDITPCSFGGVSVI TPGTNTSNQVAVLYQGVNCTEVSVAIHADQLTPTWR VYSTGSNVFQTRAGCLIGAEYVNNSYECDIPIGAGIC ASYQTQTKSRRRARSVASQSIIAYTMSLGAENSVAYS NNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDST ECSNLLLQYGSFCTQLKRALTGIAVEQDKNTQEVEFA QVKQIYKTPPIKYFGGFNFSQILPDPSKPSKRSFIEDLL FNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGL	JN.1 complete spike

	TVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQ IPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAI GKIQDSLFSTASALGKLQDVVNHNAAQALNTLVKQLS SKFGAISSVLNDILSRDKVEAEVQIDRLITGRLQSLQ TYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVD FCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTT APAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQ IITDNTFVSGNCDVVIGIVNNTVYDPLQLELDSFKEE LDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNE VAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLI AIVMVTIMLCCMTSCCCLKGCCSCGSCCKFDEDDS EPVLKGVKLHYT	
258	MFIFLLFLTSTSGSDLDRCTTFDDVQAPNYTQHTSSM RGVYYPDEIFRSDTLYLTQDLFLPFYSNVTGFH TINHTFGNPVIPFKDGIYFAATEKSNVVRGWVFGSTM NNKSQSVIINNSTNVVIRACNFELCDNPFFAV SKPMGTQTHTMIFDNAFNCTFEYISDAFSLDVSEKSG NFKHLREFVFKNKDGFLYVYKGYQPIDVVRDLP SGFNTLKPIFKLPLGINITNFRAILTAFSPAQDIWGTS AAYFVGYLKPTTFMLKYDENGITITDAVDCSQ NPLAELKCSVKSFEIDKGIYQTSNFRVVPBGDVVRFP NITNLCPFGEVFNATKFPSVYAWERKKISNCVA DYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSF VVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCV LAWNTRNIDATSTGNYNKYRYLRHGKLRPFERDIS NVPFSPDGKPCTPPALNCYWPLNDYGFYTTTGIG YQPYRVVLSFELLNAPATVCGPKLSTDLIKNQCVNF NFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTD SVRDPKTSEILDISPCSFSGVSVITPGTNASSEVAVLY QDVNCTDVSTAIHADQLTPAWRIYSTGNNVFQ TQAGCLIGAEHVDTSYECDIPIGAGICASYHTVSLIRS TSQKSIVAYTMSLGADSSIAYSNNTIAIPTNF	SARS-CoV-1 complete spike

	SISITTEVMPVSMAKTSVDCNMYICGDSTECANLLLQ YGSFCTQLNRALSGIAAEQDRNTREVFAQVKQM YKTPTLKYFGGFNFSQILPDPLKPTKRSFIEDLLFNKV TLADAGFMKQYGECLGDINARDLICAQKFNGL TVLPPLTDDMIAAYTAALVSGTATAGWTFGAGAAL QIPFAMQMAYRFNGIGVTQNVLYENQKQIANQFN KAISQIQESLTTTSTALGKLQDVVNQNAQALNTLVK QLSSNFGAISSVLNDILSRDKVEAEVQIDRLIT GRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVL GQSKRVDFCGKGYHLMSFPQAAPHGVVFLHVTYV PSQERNFTTAPAICHEGKAYFPREGVVFVNGTSWFIT QRNFFSPQIITDNTFVSGNCDVVIGIINNTVY DPLQPELDSFKEELDKYFKNHTSPDVLGDISGINAS VVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQ YIKWPWYVWLGFIAGLIAIVMTILLCCMTSCCCLK GACSCGSCCKFDEDDSEPVKGVKLHYT	
259	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSSTRGVY YPDKVFRSSVLHLTQDLFLPFFSNVTWFHAIHVSGTN GIKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDS KTQSLIVNNAATNVVIKVCEFQFCNDPFLGVYYHKN NKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQ GNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPPGFS ALEPLVDLPIGINITRFQTLALHRSYLT PGDSSSGWT AGAAAYYVGYLQPRTFLLKYNENGITDAVDCALDP LSETKCTLKSFTVEKGIYQTSNFRVQPTDSIVRFPNIT NLCPFGEVFNATTFASVYAWNRRKRISNCVADYSVLY NSTSFSTFKCYGVSPTKLNDLCFTNVYADSFVITGDE VRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSKHID AKEGGNFNYLYRLFRKANLKPFERDISTEIYQAGSKP CNGQTGLNCYYPYRYGFYPTDGVGHQPYPYRVVLS FELLNAPATVCGPKKSTNLVKNKCVNFNFNGLTGTG VLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDIT PCSFGGVSVITPGTNASNQVAVLYQDVNCTEVPVAIH	RaTG13 complete spike

	ADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSE CDIPIGAGICASYQTQTNRSVASQSIAYTMSLGAEN SVAYSNNIAIPTNFTISVTTEILPVSMKTSVDCTMYI CGDSTECNLLLQYGSFCTQLNRALTGIAVEQDKNT QEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRS FIEDLLFNKVTADAGFIKQYGDCLGDIAARDLCAQ KFNGLTVPPLLDEMIQYTSALLAGTITSGWTFGA GAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQ FNSAIGKIQDSLSTASALGKLQDVVNQNAQALNTLV KQLSSNFGAISSVLNDILSRDKVEAEVQIDRLITGRL QSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQS KRVDFCGKGYHLSFPQSAPHGVVFLHVTYVPAQE KNFTTAPAICHGDKAHFPREGVVFVSNHWFVTQRN FYEPQIITDNTFVSGSCDVVIGIVNNTVYDPLQPELD SFKEELDKYFKNHTSPDVLGDISGINASVVNIQKEID RLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGI AGLIAIIMVTIMLCCMTSCCCLKGCCSCGSCCKFDE DDSEPVLKGVKLHYT	
260	MFVFLFVLPLVSSQCVNLTRTGIPPGYTNSSTRGVY YDPKVFRSSILHLTQDLFLPFFSNVTWENTINYQGGF KKFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDAR TQSLIVNNAATNVVIKVEFQFCTDPFLGVYYHNNN KTWVENEFRVYSSANNCTFEYISQPFLMDLEGKQGN FKNLREFVFKNVDGYFKIYSKHTPIDLVRDLPRGFAA LEPLVDLPIGINITRFQTLALHRSYLTGKLESGWTT GAAAYYVGYLQQRFTLLSYNQNGTITDAVDCSLDPL SETKCTLKSLTVEKGIYQTSNFRVQPTISIVRFPNITNL CPFGEVFNASKFASVYAWNRKRISNCVADYSVLYNS TSFSTFKCYGVSPTKLNDLCFTNVYADSFVVKGDEV RQIAPGQTGVIADYNYKLPDDFTGCVIAWNSVKQDA LTGGNYGYLYRLFRKSKLKPFERDISTEIQAGSTPC NGQVGLNCYYPLERYGFHPTTG VNYQPFRVVVLSFE LLNGPATVCGPKLSTTLVKDKCVNFNFNGLTGTGVL	Pangolin17 complete spike

	TTSKKQFLPFQQFGRDISDTTDAVRDPQTLEILDITPC SFGGVSVITPGTNTSNQVAVLYQDVNCTEVPMAIHA EQLTPAWRVYSAGANVFQTRAGCLVGAEHVNNSE CDIPVGAGICASYHSMSSFRSVNQRSIIAYTMSLGAE NSVAYSNNIAIPTNFTISVTTEILPVSMKTSTVDCTM YICGDSIECSNLLLQYGSFCTQLNRALTGIAVEQDKN TQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKR SFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICA QKFNGLTVLPLLTDEMAQYTSALLAGTITSGWTFG AGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIAN QFNSAIGKIQDLSSTASALGKLQDVVNQNAQALNTL VKQLSSNFGAISSVLNDILSRDLKVEAEVQIDRLITGR LQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQS KRVDFCGKGYHLSFPQSAPHGVVFLHVTYVPAQE KNFTTAPAICHEGKAHFPREGVFVSNGTHWFITQRNF YEPQIITDNTFVSGSCDVVIGIVNNTVYDPLQPELDS FKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDR LNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGF AGLIAIIMVTIMLCCMTSCCCLKGCCSCGSCCKFDE DDSEPVLKGVKLHYT	
261	MKLLVLVFATLVSSYTIEKCLDFDDRTPPANTQFLSS HRGVYYPDIFRSNVLHLVQDHFLPFDSNVTRFITFG LNFDNPIIPFKDGIYFAATEKSNVIRGWVFGSTMNNK SQSVIIMNNSTNLVIRACNFELCDNPFFVVLKSNNTQI PSYIFNNAFNCTFEYVSKDFNLDLGEKPGNFKDLREF VFRNKDGLHVVSGYQPISAASGLPTGFNALKPIFKL PLGINITNFRLLTAFPPRPDYWGTSAAAYFVGYLKP TTFMLKYDENGITDAVDCSQNPLAELKCSVKSFEID KGIYQTSNFRVAPSKEVVRFPNITNLCPFGEVFNATTF PSVYAWERKRISNCVADYSVLYNSTSFSTFKCYGVS ATKLNLCFSNVYADSFVVKGDDVRQIAPGQTGVIA DYNKLPDDFTGCVLAWNTRNIDATQTGNYNKYR SLRHGKL RPFERDISNVPFSPDGKPCTPPAFNCYWPL	WIV1 complete spike

	NDYGFYITNGIGYQPYRVVLSFELLNAPATVCGPKL STDLIKNQCVNFNFNGLTGTGVLTTPSSKRFQPFQQFG RDVSDFTDSVRDPKTSEILDISPCSFGGVSVITPGTNTS SEVAVLYQDVNCTDVPVAIHADQLTPSWRVHSTGN NVFQTQAGCLIGAEHVDTSYECDIPGAGICASYHTV SSLRSTSQKSIVAYTMSLGADSSIAYSNNTIAIPTNFSI SITTEVMPVSMAKTSVDCNMYICGDSTECANLLLQY GSFCTQLNRALSGIAVEQDRNTREVFAQVKQMYKTP TLKDFGGFNFSQILPDPLKPTKRSFIEDLLFNKVTLAD AGFMKQYGECLGDINARDLICAQKFNGLTVLPPLT DDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQ MAYRFNGIGVTQNVLYENQKQIANQFNKAISQIQESL TTTSTALGKLQDVVNQNAQALNTLVKQLSSNFGAIS SVLNDILSRDKVEAEVQIDRLITGRLQSLQTYVTQQ LIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGY HLMSFPQAAPHGVVFLHVTYVPSQERNFTTAPAICHE GKAYFPREGVVFVNGTSWFITQRNFFSPQIITDNTFV SGSCDVVIGIINNTVYDPLQPELDSFKEELDKYFKNH TSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESL IDLQELGKYEQYIKWPWYVWLGFIAGLIAIVMVTILL CCMTSCCCLKGACSCGSCCKFDEDDSEPVLKGVKL HYT	
262	MKLLVLVFATLVSSYTIEKCLDFDDRTPPANTQFLSS HRGVYYPDDIFRSNVLHLVQDHFLPFDSNVTRFITFG LNFDNPIIPFRDGIYFAATEKSNVIRGWVFGSTMNNK SQSVIIMNNSTNLVIRACNFELCDNPFFVVLKSNNTQI PSYIFNNAFNCTFEYVSKDFNLDLGEKPGNFKDLREF VFRNKDGFHLHVYSGYQPISAASGLPTGFNALKPIFKL PLGINITNFRLLTAFPPRPDYWGTSAAAYFVGYLKP TTFMLKYDENGITITDAVDCSQNPLAELKCSVKSFEID KGIYQTSNFRVAPSKEVVRFPNITNLCPFGEVFNATTF PSVYAWERKRISNCVADYSVLYNSTSFSTFKCYGVS ATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIA	SHC014 complete spike

	DYNKLPDDFLGCVLAWNTNSKDSSTSGNYNYLYR WVRRSKLNPYERDLSNDIYSPGGQSCSAVGPNPCYNP LRPYGFFTTAGVGHQPYRVVVLSELLNAPATVCGP KLSTDLIKNCVNFNFENGLTGTGVLTSSKRFQPFQQ FGRDVSDFDTSVRDPKTSEILDSPCSFGGVSVITPGT NTSSEVAVLYQDVNCTDVPVAIHADQLTPSWRVYST GNNVFQTAQAGCLIGAEHVDTSYECDIPGAGICASYH TVSSLRSTSQKSIVAYTMSLGADSSIAYSNNTIAIPTNF SISITTEVMPVSMAKTSVDCNMYICGDSTECANLLLQ YGSFCTQLNRALSGIAVEQDRNTREVFAQVKQMYKT PTLKDFGGFNFSQILPDPLKPTKRSFIEDLLFNKVTLA DAGFMKQYGECLGDINARDLICAQKFENGLTVLPPLL TDDMIAAYTAALVSGTATAGWTFGAGAALQIPFAM QMAYRFNGIGVTQNVLYENQKQIANQFNKAISQIQES LTTTSTALGKLQDVVNQNAQALNTLVKQLSSNFGAI SSVLNDILSRDKVEAEVQIDRLITGRLQSLQTYVTQ QLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKG YHLSFPQAAPHGVVFLHVTYVPSQERNFTTAPAIC HEGKAYFPREGVVFVNGTSWFITQRNFFSPQIITDNT FVSGSCDVVIGIINNTVYDPLQPELDSFKEELDKYFKN HTSPDVLGDISGINASVVNIQKEIDRLNEVAKNLNE SLIDLQELGKYEQYIKWPWYVWLGFIAGLIAIVMTI LLCCMTSCCCLKGACSCGSCCKFDEDDSEPVLKGV KLHYT	
263	MFILLLIGYTAATTCVTGPTIENKQNVSSLMRGVYYP DDIYRSNVNVLFTVPFLRFNSTLTWYNFWNQAYTSR VMEFGDGIYFSTVDKSNVARGWIFGTTLDNTTQSAL LINNGSLVTIQVCYFQFCANPAFLVAGGQQTSAAVYI SSHNCTYSEVLSHISLDFTEDTGSFKRLREFVFKNND GFLHIYGAYQPHVIDIGATPSLPSTFKPLIPLWKLPLG LNITNFKVVLTFRSNSQPLQANFAVGSCLKLTTIMLSF DVNGTIDKAVDCSSDPLSELKCTLKSFNVSKGIYPTS NFRILPSTEVRFPNITNFCPFDKVFNATRFPNVYAW	RmYN02 complete spike

	QRTKISDCIADYTVLYNSTSFSTFKCYGVSPSKLIDLC FTSVYADTFLIRFSEVRQIAPGETGVIADYNYKLPDDF TGCVLAWNTAQQDIGSYFYRSHRAVKLKPFERDLSS DENGVRTLSTYDFNPNVPLDYQATRVVLSFELLNA PATVCGPKLSTQLVKNRCVNFNFNGLRGTGVLTDSD KRFQSFQQFGRDSADFTDSVRDPQTLQILDISPCSFGG VSVITPGTQTSSKVAVLYQDVNCTDVPTALGLDQISA AWRVYAIGNDVFQTQAGCLVGAEHTNISYECDPIGA GVCASYNSPAARVG TNSIIAYAMSIGAESSIAYSNN SI AIPTNFSIQVTTEVLPVSVSKTSVDCTMYICGDSQECN KLLLQYGSFCAQLNRALSGVAVEQDKNTEAVFAQV KQIYKAPVIKDLGGFNFSQLLPDPTKPSQRSFIEDLLF DKVTLS DAGFIKQYGDCLGDIAARDLICAQKFNGLT VLPPLLTDEMIAAYTSALVSGTATAGWTFGIGAAALQ VPFAMQMAYRFNGIGVTQNVLYENQKLIANQFN SAI GKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLS SNFGAISSVLNDILSRDKVEAEVQIDRLITGRLQSLQ TYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVD FCGKGYHLMSFPQSAPHGVVFLHVTVPAQEKNFTT APAICYDGKAHFPREGVFVSNGTHWFVTQRNFYEPQ IITDNTFVSGSCDVVIGIVNNTVYDPLQPELDSFKEE LDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNE VAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLI AIIMVTIMLCCMTSCCCLKGCCSCGSCCKFDEDDSE PVLKGVKLHYT	
264	MKFFILLSLLSFTTAQEGCGILSNKSNPALTQYFSSRR GFYYFDDTFRSSVRVLT TGYFLPFNSNL TGYSSRNSV TGRLIQFDNPNI PFKDGLYFAATERSNVIRGWIFGSTL DN TTQSAVLFNNGTHIVINVCNFYFCQDPMLAVANG SHFKSWVFLNATNCTYNRVHGFEIDPSPNTGSFIHLR EHVFRNV DGF LYVYHNYERVDVYDNFPQGFSVLKPI FKLPFGLNITQFKVIMTLFSPTTSSFNADASVYFVGHL KPLTMLAEFDENG TITDAVDCSQDPLSELKCTTKSLT	BTKY72 complete spike

VEKGIYQTSNFRVSPSTEVVRFNPITNLCPFGQVFNAS NFPSVYAWERLRISDCVADYAVLYNSSSSSFSTFKCYG VSPTKLNLCFSSVYADYFVVKGDDVRQIAPAQTGV IADYNYKLPDDFTGCVLAWNTNSVDSKSGNNFYR LFRHGKIKPYERDISNVLYNSAGGTCSSISQLGCYEPL KSYGFTPTVGVGYQPYRVVLSFELLNAPATVCGPK KSTELVKNKCVNFNFNGLTGTGVLTSSTKKFQPFQQ FGRDVSDFDTSVRDPKTFEILDISPCSYGGVSVITPGT NTSKAVAVLYQDVNCTDVPTMIHVEQVSSDWRVYA FNSYGNMFQTQAGCLVGAIYENTTYECDIPGAGICA KFGSDKIRMGQESIVAYTMSIGEDQSIAYSNNIIAIPTN FSISVTTEVLVPVSMKTSVDCNMYICGDSTECNLLL QYGSFCTQLNRALSGIAVEQDRNTRDVFAQTKSIYKT PNIKDFGGFNFSQILPDPKKLSYRSFIEDLLYNKVTLS DPGFMKQYGDCLGGINARDLICAQKFNGLTVLPPLL TDDMIAAYTAALISGTATAGYTFGAGAALQIPFAMQ MAYRFNGIGVTQNVLYENQKQIANQFNNAISKIQDSL TTTSAALGKLQDVINQNAVALNTLVKQLSSNFGAISS VLNDILSRDKVEAEVQIDRLITGRLQSLQTYVTQQLI RAAEIRASANLAATKMSECVLGQSKRVDFCGKGYH LMSFPQAAPHGVVFLHVTYVPSQQQNFTTAPAICHN GKAYFPREGVFMNGTHWFITQRNFYSPQVITTDNT FESGSCDVVIGIINNTVYDPLQPELESFKQELDKYFKN HTSPDVDFGDISGINASVVDIKKEIAHLNEIAKNLNE LIDLQELGKYEQYVKWPWYVWLGFVAGIVAILMSVI MLCCMTNCCSCFKGMCSCGSCCKFDEDHSEPVS GVKLHYT	
---	--

[0147] In some embodiments, the SD1 comprises at least one substitution of a surface-exposed, non-polar amino acid residue with a polar amino acid residue. In some embodiments, the polypeptide comprises a substitution of one or more, two or more, or three amino acid at positions 560, 562, or 569 with a polar residue, wherein the substitutions are relative to a reference sequence according to SEQ ID NO: 1. In some embodiments, the polypeptide comprises a substitution of one or more, two or more, or three of the amino acid

substitutions L560Q, F562T, F562Y, or I569S relative to a reference sequence according to SEQ ID NO: 1. In some embodiments, the polypeptide comprises a substitution of one or more, two or more, or three or more amino acid at positions 544, 546, 560, 562, 564, 569, or 582 relative to a reference sequence according to SEQ ID NO: 1. In some embodiments, the polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acid at positions N544L, N544M, N544Q, L546V, L560Q, F562T, F562Y, Q564C, Q564L, Q564N, Q564W, I569S, or L582S relative to a reference sequence according to SEQ ID NO: 1

[0148] In some embodiments, the polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acids at positions 338, 358, 363, 365, 392, or 395 relative to a reference sequence according to SEQ ID NO: 1. In some embodiments, the polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acids at positions F338L, I358F, A363L, Y365F, Y365M, Y365W, F392W, or V395I relative to a reference sequence according to SEQ ID NO: 1. In some embodiments, the polypeptide comprises amino acid substitutions comprising as shown in **Table 2**.

Table 2.

F338L /Y365W
F392W
Y365F/V395I
Y365F/ F392W /V395I
F338L/A363L/Y365M
I358F/Y365F/V395I

[0149] In some embodiments, the polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acid at positions 329, 348, 350, 367, 375, 402, 407, 410, 418, 429, 433, 435, 452, 464, 510, 512, 514, 517, 518, 519, 520, 522, 527, or 528 relative to a reference sequence according to SEQ ID NO: 1.

[0150] In some embodiments, the polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acid at positions F329K, F329R, F329Y, A348P, V350L, V367F, F375Y, I402V, V407L, I410F, I418V, F429W, V433I, A435I, A435V, L452R, F464Y, P512Q, V510I, V512F, V512I, S514T, L517A, L517D, L517S, L517T, L518N, L518Q, L518V, H519D, H519G, H519R, H519S, H519T, A520C, A520D, A520G, A520H, P521A, P521D, P521N, P521Q, P521S, A522G, A522I, P527N, K528I, K528Q, K528T, L518G, or L518S relative to a reference sequence according to SEQ ID NO: 1.

[0151] In some embodiments, the polypeptide comprises amino acid substitutions at positions 365, 395, 560, 562, and 569; and one or more, two or more, or three or more amino acid substitutions at positions 348, 402, 464, 514, 520, 526, and 527, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1.

[0152] In some embodiments, the polypeptide comprises the amino acid substitutions comprising as shown in **Table 3**.

Table 3.

A348P/Y365F/V395I/L560Q/F562Y/I569S
Y365F/V395I/I402V/L560Q/F562Y/I569S
Y365F/V395I/S514T/L560Q/F562Y/I569S
Y365F/V395I/I402V/F464Y/L560Q/F562Y/I569S
Y365F/V395I/G526S/P527N/L560Q/F562Y/I569S

[0153] In some embodiments, the polypeptide comprises amino acid substitutions at positions 365, 395, 560, 562, and 569; and one or more, two or more, or three or more amino acid substitutions at positions 348, 402, 464, 514, 520, 526, and 527, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1.

[0154] In some embodiments, the polypeptide comprises amino acid substitutions at positions 365, 395, 560, 562, and 569; and one or more, two or more, or three or more amino acid substitutions at positions 329, 517, 519, 520, and 544, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1.

[0155] In some embodiments, the polypeptide comprises the amino acid substitutions comprising as shown in **Table 4**.

Table 4.

Y365F/V395I/L560Q/F562Y/I569S
Y365F/V395I/F329Y/L560Q/F562Y/I569S
Y365F/V395I/L517T/H519S/L560Q/F562Y/I569S
Y365F/V395I/L517T/H519S/A520G/L560Q/F562Y/I569S
Y365F/V395I/L517T/H519S/N544L/L560Q/F562Y/I569S
Y365F/V395I/F329Y/L517T/H519S/L560Q/F562Y/I569S

[0156] In some embodiments, the polypeptide comprises amino acid substitutions at positions 365, 392, 395, 560, 562, and 569; and one or more, two or more, or three or more

amino acid substitutions at positions 329, 517, 519, 520, and 544, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1.

[0157] In some embodiments, the polypeptide comprises the amino acid substitutions comprising as shown in **Table 5**.

Table 5.

Y365F/VF392W/395I/L560Q/F562Y/I569S
Y365F/VF392W/V395I/F329Y/L560Q/F562Y/I569S
Y365F/VF392W/V395I/L517T/H519S/L560Q/F562Y/I569S
Y365F/VF392W/V395I/L517T/H519S/A520G/L560Q/F562Y/I569S
Y365F/VF392W/V395I/L517T/H519S/N544L/L560Q/F562Y/I569S
Y365F/VF392W/V395I/ F329Y/L517T/H519S/L560Q/F562Y/I569S

[0158] In some embodiments, the polypeptide comprises amino acid substitutions at positions 365, 395, 517, 519, 520, 560, 562, 564, and 569; and one or more, two or more, or three or more amino acid substitutions at positions 521, 544, and 546, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1

[0159] In some embodiments, the polypeptide comprises the amino acid substitutions comprising as shown in **Table 6**.

Table 6.

Y365F/V395I/L517T/H519G/A520C/Q564C/L560Q/F562Y/I569S
Y365F/V395I/L517S/H519R/A520H/P521Q/N544L/L546V/Q564W/F562T/L560Q/F562Y/I569S

[0160] In some embodiments, the polypeptide comprises amino acid substitutions at positions 365, 395, 517, 519, and 520; and one or more, or two or more amino acid substitutions at positions 518 and 392, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1.

[0161] In some embodiments, the polypeptide comprises the amino acid substitutions comprising as shown in **Table 7**.

Table 7.

Y365F/V395I/L517T/H519S/A520G
Y365F/V395I/L517T/L518G/H519D/A520G
Y365F/V395I/L517T/L518S/H519S/A520G

Y365F/F392W/V395I/L517T/H519S/A520G
Y365F/F392W/V395I/L517T/L518G/H519D/A520G
Y365F/F392W/V395I/L517T/L518S/H519S/A520G

[0162] In some embodiments, the polypeptide comprises amino acid substitutions at positions 329, 365, 395, 517, 519, 520, 560, 562, and 569; and one or more, two or more, or three or more amino acid substitutions at positions 392, 544, and 564, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1

[0163] In some embodiments, the polypeptide comprises the amino acid substitutions comprising as shown in **Table 8**.

Table 8.

Y365F/V395I/L517T/H519S/A520G/N544M/L560Q/F562Y/I569S/F329Y
Y365F/V395I/L517T/H519S/A520G/N544Q/L560Q/F562Y/I569S/F329Y
Y365F/V395I/L517T/H519S/A520G/Q564N/L560Q/F562Y/I569S/F329Y
Y365F/F392W/V395I/L517T/H519S/A520G/N544M/L560Q/F562Y/I569S/F329Y
Y365F/F392W/V395I/L517T/H519S/A520G/N544Q/L560Q/F562Y/I569S/F329Y
Y365F/F392W/V395I/L517T/H519S/A520G/Q564N/L560Q/F562Y/I569S/F329Y
Y365F/V395I/L517T/H519S/A520G/N544M/L560Q/F562Y/I569S
Y365F/V395I/L517T/H519S/A520G/N544Q/L560Q/F562Y/I569S
Y365F/V395I/L517T/H519S/A520G/Q564N/L560Q/F562Y/I569S
Y365F/F392W/V395I/L517T/H519S/A520G/N544M/L560Q/F562Y/I569S
Y365F/F392W/V395I/L517T/H519S/A520G/N544Q/L560Q/F562Y/I569S
Y365F/F392W/V395I/L517T/H519S/A520G/Q564N/L560Q/F562Y/I569S

[0164] In some embodiments, the polypeptide comprises amino acid substitutions at positions 365, 395, 517, 519, 520, 521, 504, 560, 562, 569, and 329; and one or more amino acid substitutions at positions 392, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1

[0165] In some embodiments, the polypeptide comprises the amino acid substitutions comprising as shown in **Table 9**.

Table 9.

Y365F/V395I/L517T/H519G/P521A/A520C/Q564C/L560Q/F562Y/I569S/F329Y
Y365F/V395I/L517T/H519G/P521N/A520C/Q564C/L560Q/F562Y/I569S/F329Y

Y365F/F392W/V395I/L517T/H519G/P521A/A520C/Q564C/L560Q/F562Y/I569S/F329Y
Y365F/F392W/V395I/L517T/H519G/P521N/A520C/Q564C/L560Q/F562Y/I569S/F329Y

[0166] In some embodiments, the polypeptide comprises a plurality of amino acid substitutions at positions 329, 365, 395, 526, 527, 528, 560, 562, and 569.

[0167] In some embodiments, the polypeptide comprises the amino acid substitutions comprising as shown in **Table 10**.

Table 10.

F329Y/Y365F/V395I/G526S/P527N/K528T/L560Q/F562Y/I569S
F329Y/Y365F/V395I/G526S/P527N/K528I/L560Q/F562Y/I569S

[0168] In some embodiments, the polypeptide comprises a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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 100% identical to SEQ ID NOs: 2-10.

[0169] In some embodiments, the polypeptide comprises a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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 100% identical to any one of SEQ ID NOs: 11-19.

[0170] In some embodiments, the polypeptide comprises a polypeptide sequence identical to any one of SEQ ID NOs: 11-19.

[0171] In some embodiments, the polypeptide is expressed at a higher level than a reference polypeptide without the amino acid substitutions, optionally a reference polypeptide according to SEQ ID NOs: 2 or 10.

[0172] In some embodiments, expression can be detected using conventional techniques for detecting protein (*e.g.*, ELISA, Western blotting, flow cytometry, immunofluorescence, immunohistochemistry, *etc.*).

[0173] In some embodiments, the expressed polypeptide is more thermally stable or more antigenic than SEQ ID NOs: 2 or 10.

[0174] In some embodiments, the polypeptide comprises a plurality of amino acid substitutions at positions 329, 367, 383, 547, 568, 570, and 619. In some embodiments, the polypeptide comprises the amino acid substitutions at position F329Y, F329S, V367F, S383C, T547C, D568N, A570C, and E619Q.

[0175] In some embodiments, an antigenicity is measured via Bio-Layer Interferometry (BLI) using immobilized ACE2-Fc, S309, CR3022 antibodies.

[0176] As used herein, the term “thermal stability” or “thermostability” of a protein is herein understood as follows as the protein stability in terms of denaturation due to temperature alteration and can be determined by the kinetics of protein denaturation during heating. Thermal stability can be evaluated by measuring the melting temperature (T_m) or onset temperature (T_{onset}) of a composition of the invention using any suitable technique. The melting temperature is the temperature at which the first derivative of the melting curve reaches a local maximum. The onset temperature is the temperature at which a linear regression of the baseline intersects with a linear regression of the melting transition.

[0177] In some embodiments, a nano differential scanning fluorimetry (nanoDSF) can be used to determine the thermal stability by monitoring intrinsic tryptophan fluorescence. In some embodiments, Static light scattering (SLS) can be used to reveal an aggregation temperature.

[0178] In some embodiments, thermal stability can be evaluated by analytical spectroscopy. An exemplary analytical spectroscopy method is Circular Dichroism (CD) spectroscopy. CD spectrometry measures the optical activity of a composition as a function of increasing temperature. Circular dichroism (CD) spectroscopy measures differences in the absorption of left-handed polarized light versus right-handed polarized light which arise due to structural asymmetry. A disordered or unfolded structure results in a CD spectrum very different from that of an ordered or folded structure. The CD spectrum reflects the sensitivity of the proteins to the denaturing effects of increasing temperature and is therefore indicative of a protein's thermal stability (see, e.g., van Mierlo and Steemsma, J. Biotechnol., 79(3):281-98, 2000).

[0179] Another exemplary analytical spectroscopy method for measuring thermal stability is Fluorescence Emission Spectroscopy. Fluorescence-based methods to evaluate thermal stability monitor changes in the fluorescence of intrinsic fluorophores (e.g., tryptophan and tyrosine amino acids) or extrinsic fluorophores (e.g., ANS or SYPRO Orange) upon thermal unfolding, (see, e.g., Niesen F. PL, Berglund H. and Vedadi M. : The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nature

Protocols 2007, 2:2212-21.) Yet another exemplary analytical spectroscopy method for measuring antibody, light or heavy chain thermal stability is Nuclear Magnetic Resonance (NMR) spectroscopy (see, e.g., van Mierlo and Steemsma, J. Biotechnol., 79(3):281 -98, 2000).

[0180] In some embodiments, the melting temperature (T_m) of an expressed polypeptide as described herein is about 30°C. In some embodiments, the melting temperature (T_m) is about 35°C. In some embodiments, the melting temperature (T_m) is about 40°C. In some embodiments, the melting temperature (T_m) is about 45°C. In some embodiments, the melting temperature (T_m) is about 50°C. In some embodiments, the melting temperature (T_m) is about 55°C. In some embodiments, the melting temperature (T_m) is about 60°C. In some embodiments, the melting temperature (T_m) is about 65°C. In some embodiments, the melting temperature (T_m) is about 70°C. In some embodiments, the melting temperature (T_m) is about 75°C.

[0181] In some embodiments, the melting temperature (T_m) of an expressed polypeptide as described herein is at least about 30°C. In some embodiments, the melting temperature (T_m) is at least about 35°C. In some embodiments, the melting temperature (T_m) is at least about 40°C. In some embodiments, the melting temperature (T_m) is at least about 45°C. In some embodiments, the melting temperature (T_m) is at least about 50°C. In some embodiments, the melting temperature (T_m) is at least about 55°C. In some embodiments, the melting temperature (T_m) is at least about 60°C. In some embodiments, the melting temperature (T_m) is at least about 65°C. In some embodiments, the melting temperature (T_m) is at least about 70°C. In some embodiments, the melting temperature (T_m) is at least about 75°C.

[0182] In some embodiments, the melting temperature (T_m) is from 30°C to 75°C. In some embodiments, the melting temperature (T_m) is from 35°C to 75°C. In some embodiments, the melting temperature (T_m) is from 40°C to 75°C. In some embodiments, the melting temperature (T_m) is from 45°C to 75°C. In some embodiments, the melting temperature (T_m) is from 50°C to 75°C. In some embodiments, the melting temperature (T_m) is from 55°C to 75°C. In some embodiments, the melting temperature (T_m) is from 60°C to 75°C. In some embodiments, the melting temperature (T_m) is from 65°C to 75°C. In some embodiments, the melting temperature (T_m) is from 70°C to 75°C. In some embodiments, the melting temperature (T_m) is from 30°C to 70°C. In some embodiments, the melting temperature (T_m) is from 35°C to 70°C. In some embodiments, the melting temperature (T_m) is from 40°C to 70°C. In some embodiments, the melting temperature (T_m) is from 45°C to 70°C. In some embodiments, the melting temperature (T_m) is from 50°C to 70°C.

In some embodiments, the melting temperature (T_m) is from 55°C to 70°C. In some embodiments, the melting temperature (T_m) is from 60°C to 70°C. In some embodiments, the melting temperature (T_m) is from 65°C to 70°C. In some embodiments, the melting temperature (T_m) is from 30°C to 65°C. In some embodiments, the melting temperature (T_m) is from 35°C to 65°C. In some embodiments, the melting temperature (T_m) is from 40°C to 65°C. In some embodiments, the melting temperature (T_m) is from 45°C to 65°C. In some embodiments, the melting temperature (T_m) is from 50°C to 65°C. In some embodiments, the melting temperature (T_m) is from 55°C to 65°C. In some embodiments, the melting temperature (T_m) is from 60°C to 65°C. In some embodiments, the melting temperature (T_m) is from 35°C to 60°C. In some embodiments, the melting temperature (T_m) is from 40°C to 60°C. In some embodiments, the melting temperature (T_m) is from 45°C to 60°C. In some embodiments, the melting temperature (T_m) is from 50°C to 60°C. In some embodiments, the melting temperature (T_m) is from 55°C to 60°C.

[0183] In some embodiments, the melting temperature (T_m) of an expressed polypeptide as described herein is about 0.1°C increased. In some embodiments, the melting temperature (T_m) of an expressed polypeptide as described herein is about 0.2°C increased. In some embodiments, the melting temperature (T_m) of an expressed polypeptide as described herein is about 0.3°C increased. In some embodiments, the melting temperature (T_m) of an expressed polypeptide as described herein is about 0.4°C increased. In some embodiments, the melting temperature (T_m) of an expressed polypeptide as described herein is about 0.5°C increased. In some embodiments, the melting temperature (T_m) of an expressed polypeptide as described herein is about 0.6°C increased. In some embodiments, the melting temperature (T_m) of an expressed polypeptide as described herein is about 0.7°C increased. In some embodiments, the melting temperature (T_m) of an expressed polypeptide as described herein is about 0.8°C increased. In some embodiments, the melting temperature (T_m) of an expressed polypeptide as described herein is about 0.9°C increased. In some embodiments, the melting temperature (T_m) of an expressed polypeptide as described herein is about 1.0°C increased.

Linkers

[0184] In some embodiments, the recombinant polypeptide and a protein nanostructure may be genetically fused such that they are both present in a single polypeptide, termed a “fusion protein.” The linkage between the polypeptide and the protein nanostructure allows the recombinant polypeptide to be displayed on the exterior of the self-assembling protein nanostructure.

[0185] A wide variety of polypeptide sequences can be used to link the proteins, or antigenic fragments thereof and the protein nanostructure. In some cases, the linker comprises a polypeptide sequence that can be included in the encoding polynucleotide sequence. Any suitable linker polypeptide can be used. In some embodiments, the linker imposes a rigid relative orientation of the antigenic protein (*e.g.*, coronavirus polypeptide comprising a receptor-binding domain (RBD) and an RBD-adjacent subdomain 1 (SD1) from the Spike glycoprotein) or antigenic fragment thereof to the protein nanostructure. In some embodiments, the linker flexibly links the antigenic protein (*e.g.*, Coronavirus polypeptide comprising a receptor-binding domain (RBD) and an RBD-adjacent subdomain 1 (SD1) from the Spike glycoprotein) or antigenic fragment thereof to the protein nanostructure. In some embodiments, the encoded polypeptides can include a linker between regions. In some embodiments, the polypeptide is a fusion protein which includes the recombinant coronavirus polypeptide, a linker, and the protein nanostructure component polypeptide. In some embodiments, the polypeptide is a fusion protein, which includes, in N- to C-terminal order, the recombinant coronavirus polypeptide, a linker, and the protein nanostructure component polypeptide. The linker can be a polypeptide. A wide variety of polypeptide sequences can be used and are well known in the art. In some embodiments, the linker may comprise a Gly-Ser linker (*i.e.*, a linker consisting of glycine and serine residues) of any suitable length. In some embodiments, the Gly-Ser linker may be 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more amino acid residues in length. Non-limiting examples of Gly-Ser linkers are presented below:

[0186] Illustrative polypeptide linkers are provided in **Table 11**. In some embodiments, the polypeptide linker is any sequence in **Table 11**.

[0187] In some embodiments, the linker comprises between 3 and 30 amino acid residues. In some embodiments, the linker comprises between 4 and 24 amino acid residues. In some embodiments, the linker comprises between 8 and 24 amino acid residues. In some embodiments, the linker comprises between 10 and 24 amino acid residues. In some embodiments, the linker comprises between 12 and 24 amino acid residues. In some embodiments, the linker comprises between 16 and 24 amino acid residues. In some embodiments, the linker comprises between 18 and 24 amino acid residues. In some embodiments, the linker comprises between 20 and 24 amino acid residues. In some embodiments, the linker comprises between 4 and 20 amino acid residues. In some embodiments, the linker comprises between 8 and 20 amino acid residues. In some embodiments, the linker comprises between 10 and 20 amino acid residues. In some

embodiments, the linker comprises between 12 and 20 amino acid residues. In some embodiments, the linker comprises between 16 and 20 amino acid residues. In some embodiments, the linker comprises between 8 and 18 amino acid residues. In some embodiments, the linker comprises between 12 and 16 amino acid residues.

[0188] In some embodiments, the linker comprises 3 amino acid residues. In some embodiments, the linker comprises 4 amino acid residues. In some embodiments, the linker comprises 7 amino acid residues. In some embodiments, the linker comprises 8 amino acid residues. In some embodiments, the linker comprises 10 amino acid residues. In some embodiments, the linker comprises 12 amino acid residues. In some embodiments, the linker comprises 16 amino acid residues. In some embodiments, the linker comprises 18 amino acid residues. In some embodiments, the linker comprises 19 amino acid residues. In some embodiments, the linker comprises 20 amino acid residues. In some embodiments, the linker comprises 24 amino acid residues. In some embodiments, the linker comprises 30 amino acid residues.

[0189] In some embodiments, the encoded polypeptides can include a linker between regions. In some embodiments, the polypeptide is a fusion protein which includes the recombinant SARS-CoV-2 polypeptide, a linker, a N-terminal extension linker, and the protein nanostructure component polypeptide. In some embodiments, the polypeptide is a fusion protein, which includes, in N- to C-terminal order, the recombinant SARS-CoV-2 polypeptide, a linker, a N-terminal extension linker, and the protein nanostructure component polypeptide. In some embodiments, the N-terminal extension linker is I53-50A helical extension. In some embodiments, polypeptide sequence of N-terminal extension linker is EKAAKAEEAARK (SEQ ID NO: 237).

[0190] In some embodiments, the linker is selected from the group shown in **Table 11**.

Table 11.

Sequence	Length	SEQ ID NO:
GSS	3	85
GS GS	4	86
GGSGEKP	7	87
GGSGQKP	7	88
GGSGGSGS	8	89
GGSGGSGEKP	10	90
GGSGGSGQKP	10	91

GGSGGSGGSGGS	12	92
GSGGSGSGSGGS	12	93
GGGGGSGGGSGGGGS	15	94
GGGGSGGGGSGGGGS	15	95
GGSGGSGSGGSGGSGS	16	96
GGGGSGGGGSGGGGSGG	17	97
SGGGSGGSGSGGSGGSGS	18	98
EPEGSGGSGSGGSGGSGS	19	99
YGGSGGSGGSGSGGSGGSGS	20	100
GGSGGSGSGGSGGSGSGGSGGS	24	101
GSGGSGGSGGSGGSGSGGSGGSGS	24	102
KSDELLGSGGSGSGSGGSEKAAKAEAAARK	30	103

Self-assembling protein nanostructures

[0191] In another aspect, the disclosure provides a self-assembling protein nanostructure including a first component comprising a recombinant polypeptide that includes an antigenic fragment of a coronavirus Spike glycoprotein as described above. Optionally, the self-assembling protein nanostructure can include a second component comprising a second protein. The first component, as a trimeric assembly, comprises a protein-protein interface that induces a region of the first component to self-associate to form trimeric building blocks. In protein nanostructures that have two or more components, each copy of the first assembly domain further comprises a surface-exposed interface that interacts with a complementary surface-exposed interface on the second assembly domain. Similarly stated, the second component, as a pentameric assembly, is adapted to multimerize with first component. As described in King et al. (Nature 510, 103-108, 2014), Bale et al. (Science 353, 389-394, 2016), and patent publications WO2014124301 A1 and US20160122392 A1, the complementary protein-protein interface between a first component and a second component drives the assembly of multiple copies of the trimeric assembly domain and second assembly domain into a target protein nanostructure. In some embodiments, each of the trimeric assembly domains of the protein nanostructure bears an antigenic protein, or antigenic fragment thereof, linked thereto (*e.g.*, as a genetic fusion); these protein nanostructures display the proteins at full valency. In other embodiments, the protein nanostructures of the disclosure comprise one or more first assembly domains bearing

antigenic proteins, or antigenic fragments thereof (*e.g.*, as genetic fusions) as well as one or more first assembly domains that do not bear antigenic proteins; these protein nanostructures display the proteins at partial valency. The first assembly domains can be any polypeptide sequence that forms a trimer and interacts with a second assembly domains to drive assembly to a target protein nanostructure. In some embodiments, the protein nanostructure comprises a first polypeptide and a second polypeptide selected from those disclosed in US 20130274441 A1, US 20150356240 A1, US 20160122392 A1, WO 2018187325 A1, each of which is incorporated by reference herein in its entirety.

[0192] In some embodiments, a single component self-assembles into the protein nanostructure. In some embodiments, one or more purified samples of first and second components for use in forming a protein nanostructure are mixed in an approximately equimolar molar ratio in aqueous conditions (*e.g.*, an I53-50A/B icosahedral protein nanostructure). The first and second components (through the assembly domains) interact with one another to drive assembly of the target protein nanostructure. Successful assembly of the target protein nanostructure can be confirmed by analyzing the *in vitro* assembly reaction by common biochemical or biophysical methods used to assess the physical size of proteins or protein assemblies, including but not limited to size exclusion chromatography, native (non-denaturing) gel electrophoresis, dynamic light scattering, multi-angle light scattering, analytical ultracentrifugation, negative stain electron microscopy, cryo-electron microscopy, or X-ray crystallography. If necessary, the assembled protein nanostructure can be purified from other species or molecules present in the *in vitro* assembly reaction using preparative techniques commonly used to isolate proteins by their physical size, including but not limited to size exclusion chromatography, preparative ultracentrifugation, tangential flow filtration, or preparative gel electrophoresis. The presence of the antigenic protein in the protein nanostructure can be assessed by techniques commonly used to determine the identity of protein molecules in aqueous solutions, including but not limited to SDS-PAGE, mass spectrometry, protein sequencing, ELISA, surface plasmon resonance, biolayer interferometry, or amino acid analysis. The accessibility of the protein on the exterior of the protein nanostructure, as well as its conformation or antigenicity, can be assessed by techniques commonly used to detect the presence and conformation of an antigen, including but not limited to binding by monoclonal antibodies, conformation-specific monoclonal antibodies, surface plasmon resonance, biolayer interferometry, or antisera specific to the antigen.

[0193] In various embodiments, the protein nanostructures of the disclosure comprise two or more distinct first polypeptides bearing different antigenic proteins as genetic fusions; these protein nanostructures co-display multiple different proteins (*e.g.*, different variants of coronavirus polypeptide that includes a receptor-binding domain (RBD) and an RBD-adjacent subdomain 1 (SD1) from the Spike glycoprotein) on the same protein nanostructure. These multi-antigen protein nanostructures are produced by performing *in vitro* assembly with mixtures of two or more antigens each comprising an assembly domain. The fraction of each antigen in the mixture determines the average valency of each antigenic protein in the resulting protein nanostructures. The presence and average valency of each antigen in a given sample can be assessed by quantitative analysis using the techniques described above for evaluating the presence of antigenic proteins in full-valency protein nanostructures.

[0194] The protein nanostructures of the present invention may comprise multimeric protein assemblies adapted for display of coronavirus polypeptides that include a receptor-binding domain (RBD) and an RBD-adjacent subdomain 1 (SD1) from the Spike glycoprotein, or antigenic variants or fragments thereof. The protein nanostructures of the present invention comprise at least a first component and, optionally, a second component. The first component may include substitution of at least one amino acid residue or by addition at the N- or C-terminus of one or more amino acid residues. In some embodiments the fusion protein comprises a protein sequence determined by computational methods. This first component may form the entire core of the protein nanostructure; or the core of the protein nanostructure may include a second component or third, fourth, fifth component and so on. In some embodiments the first component is a trimeric component in which the assembly domains form trimers related by 3-fold rotational symmetry, and/or the second component is a pentameric component, in which the assembly domains form pentamers related by 5-fold rotational symmetry. In some embodiments, the combination of the two components form an “icosahedral particle” having I_{53} symmetry. Together these components may be arranged such that the members of each component are related to one another by symmetry operators. A general computational method for designing self-assembling protein materials, involving symmetrical docking of protein building blocks in a target symmetric architecture, is disclosed in Patent Pub. No. US 2015/0356240 A1.

[0195] The “core” of the protein nanostructure is used herein to describe the central portion of the protein nanostructure that links together the several copies of polypeptides that include an antigenic fragment of the coronavirus Spike glycoprotein, displayed by the

protein nanostructure. In some embodiments, the fusion protein comprises a first polypeptide comprising an antigenic fragment of the Spike glycoprotein of a coronavirus a linker, and a first assembly domain. In some embodiments, the antigenic fragment of the Spike glycoprotein of a coronavirus, is non-covalently or covalently linked to the assembly domain. For example, an antibody or antigenic fragment thereof may be fused to the first component and configured to bind a portion of the first component, or a chemical tag on the first component. For example, a streptavidin-biotin (or neuravidin-biotin) linker can be employed. Or various bioconjugate linkers may be used. In some embodiments of the present disclosure, the antigen comprises further polypeptide sequences in addition to the coronavirus polypeptides.

[0196] A non-limiting example of an embodiment is an antigenic fragment of the Spike glycoprotein of a coronavirus genetically fused to a component of the protein nanostructure, which optionally is expressed recombinantly in a host cell (*e.g.*, *E. coli* cells); along with a pentameric protein assembly as a second component, which is optionally expressed recombinantly in the same or a different host cell. These two components self-assemble into a protein nanostructure displaying 60 antigenic fragment monomers around an icosahedral core. In some embodiments, an antigenic fragment of the Spike glycoprotein of a coronavirus mixed with another antigen protein in the same protein nanostructure, such as two different variants of the Spike glycoprotein of a coronavirus. In some embodiments, the protein nanostructure comprises, in addition to one or more antigenic fragment of the Spike glycoprotein of a coronavirus including the receptor-binding domain (RBD) and, optionally an RBD-adjacent subdomain 1 (SD1), antigens of other pathogenic organisms, and thus may be used as a combination vaccine. In some embodiments, the protein nanostructure is further linked to polypeptides or other agents capable of acting as an adjuvant. In some embodiments, the first component and/or the second component comprise one or more T cell epitopes, optionally a T cell epitope of heterologous origin.

[0197] Other potential arrangements of components of the present disclosure include the protein nanostructure adapted for display of up to 12, 24, or 60 monomers. In some embodiments a component may comprise polypeptide linked to diverse antigens, such that the protein nanostructure displays different antigens on the same nanostructure. In some embodiments, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more different antigens are displayed. Non-limiting illustrative protein nanostructure are provided in Bale et al. *Science* 353:389-94 (2016); Heinze et al. *J.*

Phys. Chem B. 120:5945-5952 (2016); King et al. *Nature* 510:103-108 (2014); and King et al. *Science* 336:1171-71 (2012).

Attachment Modalities

[0198] The protein nanostructures of the present disclosure display antigenic proteins in various ways including as gene fusions or by other means disclosed herein. As used herein, “linked to” or “attached to” denotes any means known in the art for causing two polypeptides to associate. The association may be direct or indirect, reversible, or irreversible, weak, or strong, covalent or non-covalent, and selective or nonselective.

[0199] In some embodiments, attachment is achieved by genetic engineering to create an N- or C-terminal fusion of potentially antigenic polypeptides of the protein nanostructure.

[0200] In some embodiments, attachment is achieved by post-translational covalent attachment of one or more pluralities of antigenic protein. In some embodiments chemical cross-linking is used to non-specifically attach the antigen to a protein nanostructure. In some embodiments chemical cross-linking is used to specifically attach the antigenic protein to a protein nanostructure (*e.g.*, to the first polypeptide or the second polypeptide). Various specific and non-specific cross-linking chemistries are known in the art, such as Click chemistry and other methods. In general, any cross-linking chemistry/bioconjugate used to link two proteins may be adapted for use in the presently disclosed protein nanostructures. In particular, chemistries used in creation of immunoconjugates or antibody drug conjugates may be used. In some embodiments a protein nanostructure is created using a cleavable or non-cleavable linker. Processes and methods for conjugation of antigens to carriers are provided by, *e.g.*, Patent Pub. No. US 2008/0145373 A1.

[0201] In some embodiments, attachment is achieved by non-covalent attachment between a component and the antigen. In some embodiments the antigenic protein is engineered to be negatively charged on at least one surface and the core polypeptide is engineered to be positively charged on at least one surface, or positively and negatively charged, respectively. This can promote intermolecular association between the antigenic protein and the component core polypeptide by electrostatic force. In some embodiments shape complementarity is employed to cause linkage of antigen protein to component core. Shape complementarity can be pre-existing or rationally designed. In some embodiments computational designed of protein-protein interfaces is used to achieve attachment. In some embodiments, the antigen is biotin-labeled and the polypeptide comprises a streptavidin, or vice versa. In some embodiments, streptavidin is displayed by gene fusion or otherwise as a tetramer on a 4-fold axis of the component core and the biotin-labeled antigen is

monomeric, dimeric, or tetrameric, permitting association to the component core in a configuration appropriate for native multimerization of the antigen. In some embodiments a protein-based adaptor is used to capture the antigenic protein. In some embodiments the polypeptide is fused to a protein capable of binding a complementary protein, which is fused to the antigenic protein.

[0202] Immunogenicity of an antigenic protein may be controlled by altering the orientation of the antigenic protein relative to the component core. Depending on how the antigenic protein is attached to the component core of the protein nanostructure, the antigenic protein may be displayed in various orientations. In some embodiments, the antigenic protein is displayed so that one or more epitopes are oriented at or towards the distal end of the antigenic protein, such that these epitope(s) are preferentially accessible to the immune system. In some embodiments the orientation will recapitulate the orientation of the Spike glycoprotein protein with respect to the virus. The choice of orientation may direct the immune system to one or the other epitope.

[0203] In some embodiments, epitope preference is controlled by other means, such as positioning of glycans on the protein nanostructure by addition or subtraction of the N-linked glycan sequence motif N-X-[T/S] at predetermined positions in the amino acid sequence of a component including in the amino acid sequence of the antigenic protein.

[0204] In some embodiments the epitopes found at intermediate distances from the proximal to the distal end will be the preferred over epitopes more distally located depending on various considerations including but not limited to the overall geometry of the protein nanostructure, surface hydrophobicity, surface charge, and competitive binding of proteins endogenously present in the subject or proteins exogenously provided in the vaccine composition. The present disclosure encompasses all known methods of rational design of protein structure and the foregoing is not intended to be limiting.

Polypeptide Sequences

[0205] Patent Pub No. US 2015/0356240 A1 describes various methods for designing protein assemblies. As described in US Patent Pub No. US 2016/0122392 A1 and in International Patent Pub. No. WO 2014/124301 A1, the isolated polypeptides shown in **Table 12** were designed for their ability to self-assemble in pairs to form protein nanostructures, such as icosahedral particles. The design involved design of suitable interface residues for each member of the polypeptide pair that can be assembled to form the protein nanostructure. The protein nanostructure so formed include symmetrically repeated, non-natural, non-covalent polypeptide-polypeptide interfaces that orient a first

assembly and a second assembly into a protein nanostructure, such as one with an icosahedral symmetry. Thus, in one embodiment a first repeat element and second repeat element of the component are selected from the group consisting of SEQ ID NOs: 23-73. In each case, an N-terminal methionine residue present in the full-length protein is included but may be removed to make a fusion that is not included in the sequence. The identified residues in **Table 12** are numbered beginning with an N-terminal methionine (not shown). In various embodiments, one or more additional residues are deleted from the N-terminus and/or additional residues are added to the N-terminus (*e.g.*, to form a helical extension).

Table 12

Name	Component Multimer	Amino Acid Sequence	Identified interface residues
I53-34A SEQ ID NO:23	trimer	EGMDPLAVLAESRLLPLLTVRGGEDLAGLAT VLELMGVGALEITLRTEKGLEALKALRKSGLL LGAGTVRSPKEAEEAEEAGAAFLVSPGLLEEV AALAQARGVPYLPGLVLTPTTEVERALALGLSA LKFFPAEPFQGVRLRAYAEVFPEVRFLPTGGI KEEHLPHYAALPNLLAVGGSWLLQGDAAV MKKVKAALKALLSPQAPG	I53-34A: 28,32,36,3 7,186,188, 191,192,19 5
I53-34B SEQ ID NO:24	pentamer	TKKVGIVDTTFARVDMAEAAIRTLKALSPNIKI IRKTVPGIKDLPVACKKLLEEEGCDIVMALGM PGKAEKDKVCAHEASLGLMLAQLMTNKHIIIE VFVHEDEAKDDDEL DILALVRAIEHAANVYY LLFKPEYLTRMAGKGLRQGREDA GPARE	I53-34B: 19,20,23,2 4,27,109,1 13,116,117 ,120,124,1 48
I53-40A SEQ ID NO:25	pentamer	TKKVGIVDTTFARVDMASAAITLKMESPNIKI IRKTVPGIKDLPVACKKLLEEEGCDIVMALGM MPGKAEKDKVCAHEASLGLMLAQLMTNKHII EVFVHEDEAKDDAELKILAAARRAIEHALNVY YLLFKPEYLTRMAGKGLRQGFEDAGPARE	I53-40A: 20,23,24,2 7,28,109,1 12,113,116 ,120,124
I53-40B SEQ ID NO:26	trimer	STINNQLKALKVIPVIAIDNAEDIPLGKVLAE NGLPAAEITFRSSAAVKAIMLLRSAQPEMLIGA GTILNGVQALAAKEAGATFVVSPGFNPNTVR ACQIIGIDIVPGVNNPSTVEAALEMGLTTLKFF PAEASGGISMVKSLSVGPYGDIRLMPTGGITPSN IDNYLAIPQVLACGGTWMVDKKLVTNGEWD EIARLTREIVEQVNP	I53-40B: 47,51,54,5 8,74,102
I53-47A SEQ ID NO:27	trimer	PIFTLNTNIKATDVPSDFLSLTSLVGLILSKPG SYVAVHINTDQQLSFGGSTNPAAFGTLM SIGG IEPSKNRDHSAVLFDHLNAMLGIPKNRMYIHF VNLNGDDVGWNGTTF	I53-47A: 22,25,29,7 2,79,86,87
I53-47B	pentamer	NQSHSKDYETVRIAVVRARWHADIVDACVE AFEIAMAAIGGDRFAVDVFDVPGAYEIP L HAR	I53-47B: 28,31,35,3

Name	Component Multimer	Amino Acid Sequence	Identified interface residues
SEQ ID NO:28		TLAETGRYGAVLGTAFFVNVNGGIYRHEFVASA VIDGMMNVQLSTGVPVLSAVLTPHRYRDSAE HHRFFAAHFAVKGVEAARACIEILAAREKIAA	6,39,131,132,135,139,146
I53-50A SEQ ID NO:29	trimer	EELFKKHKIVAVLRANSVEEAIEKAVAVFAGG VHLIEITFTVPDADTVIKALSVLKEKGAIIGAG TVTSVEQCRKAVESGAEFIVSPHLDEEISQFCK EKGVFYMPGVMTPTLVKAMKLGHTILKLFP GEVVGPQFVKAMKGPFPNVKFPVPTGGVNLDN VCEWFKAGVLA VGVGSALVKGTPDEVREKA KAFVEKIRGCTE	I53-50A: 25,29,33,54,57
I53-50B SEQ ID NO:30	pentamer	NQSHSKDYETVRIAVVRARWHAIEVDACVSA FEAAMADIGGDRFAVDVFDVPGAYEIPHAR TLAETGRYGAVLGTAFFVNVNGGIYRHEFVASA VIDGMMNVQLSTGVPVLSAVLTPHRYRDSAE HTLLFLALFAVKGMEAAARACVEILAAREKIAA	I53-50B: 24,28,36,124,125,127,128,129,131,132,133,135,139
I53-51A SEQ ID NO:31	trimer	FTKSGDDGNTNVINKRVGKDSPLVNFGLDLD ELNSFIGFAISKIPWEDMKDLERVQVELFEIG EDLSTQSSKKKIDESYVLWLLAATAIYRIESGP VKLFVIPGGSEEASVLHVTRSVARRVERNAV YTKELPEINRMIVYLNRLSSLLFAMALVANK RRNQSEKIYEIGKSW	I53-51A: 80,83,86,87,88,90,91,94,166,172,176
I53-51B SEQ ID NO:32	pentamer	NQSHSKDYETVRIAVVRARWHADIVDQCVR AFEEAMADAGGDRFAVDVFDVPGAYEIPHAR TLAETGRYGAVLGTAFFVNVNGGIYRHEFVAS AVIDGMMNVQLSTGVPVLSAVLTPHRYRSSR EHHEFFREHFVMVKGVEAAAACITILAAREKIAA	I53-51B: 31,35,36,40,122,124,128,131,135,139,143,146,147
I52-03A SEQ ID NO:33	pentamer	GHTKGPTPQQHDGSALRIGIVHARWNKTIIMP LLIGTIAKLLECGVKASNIVVQSVPGSWELPIA VQRLYSASQLQTPSSGPSLSAGDLLGSSTTDLT ALPTTTASSTGPFDAIAIGVLIKGETMHFEYIA DSVSHGLMRVQLDTGVPVIFGVLTVLTDQQA KARAGVIEGSHNHGEDWGLAAVEMGVRRRD WAAGKTE	I52-03A: 28,32,36,39,44,49
I52-03B SEQ ID NO:34	dimer	YEVDHADVYDLFYLGKGDYAAEASDIADLV RSRTPEASSLLDVACGTGTHLEHFTKEFGDTA GLELSEDMLTTHARKRLPDATLHQGDMRDFQL GRKFSAVVSVMFSSVGYLKTVAELGA AVASFA EHLEPGGVVVVEPWVFPETFADGWVSADV RRDGRTVARVSHSVREGNATRMEVHFTVADP GKGVRHFSDVHLITLFHQREYEA AFMAAGLR VEYLEGGPSGRGLFVG VPA	I52-03B: 94,115,116,206,213
I52-32A SEQ ID NO:35	dimer	GMKEKFVLIITHGDFGKGLLSGAEVIGKQEN VHTVGLNLGDNIEKVAKEVMRIIAKLAEDKE IIIVVDLFGGSPFNIALEMMKTFDVKVITGINM	I52-32A: 47,49,53,55

Name	Component Multimer	Amino Acid Sequence	Identified interface residues
		PMLVELLTSINVYDTTELLENISKIGKDGKIVIEKSSSLKM	4,57,58,61,83,87,88
I52-32B SEQ ID NO:36	pentamer	KYDGSKLRIGILHARWNLEHAAALVAGAIKRLQEFGVKAENIIETVPGSFELPYGSKLFVEKQKRLGKPLDAIPIGVLIKSTMHFEYICDSTTHQLMKLNFELGIPVIFGVLTCLTDEQAEARAGLIEGKMHNHGEDWGAAAVEMATKFN	I52-32B: 19,20,23,30,40
I52-33A SEQ ID NO:37	pentamer	AVKGLGEVDQKYDGSKLRIGILHARWNRKIILALVAGAVLRLLEFGVKAENIIETVPGSFELPYGSKLFVEKQKRLGKPLDAIPIGVLIKSTMHFEYICDSTTHQLMKLNFELGIPVIFGVLTCLTDEQAEARAGLIEGKMHNHGEDWGAAAVEMATKFN	I52-33A: 33,41,44,50
I52-33B SEQ ID NO:38	dimer	GANWYLDNESSRLSFTSTKNADIAEVHRFLVLHGKVDPKGLAEVEVETESISTGIPLRDMLLRVLVFQVSKFPVAQINAQLDMRPINN LAPGAQLELRLPLTVSLRGKSHSYNAELLATRLDERRFQVVTLEPLVIHAQDFDMVRAFNALRLVAGLSAVSLSVPVGAVLIFTAR	I52-33B: 61,63,66,67,72,147,148,154,155
I32-06A SEQ ID NO:39	dimer	TDYIRDGSAIKALSFAIILAEADLRHIPQDLQRLAVRVIHACGMVDVANDLAFSEGAGKAGRNALLAGAPILCDARMVAEGITRSLPADNRVIYTLSDPSVPELAKKIGNTRSAAALDLWLPHIEGSI VAIGNAPTALFRLFELLDAGAPKPALIIGMPVGVGAAESKDELAANSRGVPYVIVRGRRGGSAMTAAAVNALASERE	I32-06A: 9,12,13,14,20,30,33,34
I32-06B SEQ ID NO:40	trimer	ITVFGLKSKLAPRREKLAEVIYSSLHLGLDIPKGKHAIRFLCLEKEDFYYPFDRSDDYTVIEINLMAGRSEETKMLLIFLLFIALERKLGIRAH DVEITIKEQPAHCWGFRGRTGDSARDLDYDIYV	I32-06B: 24,71,73,76,77,80,81,84,85,88,114,118
I32-19A SEQ ID NO:41	trimer	GSDLQKLQRFSTCDISDGLLN VYNIPTGGYFPNLTAISPPQNSSIVGTAYTVLFAPIDDRPAVN YIDSVPPNSILVLALEPHLQSQFHPFIKITQAMYGGLMSTRAQYLKSN GTVVFGRI RDVDEHRTL NHPVFAYGVGSCAPKAVVKAVGTNVQLKILTS DGVTQTICPGDYIAGDNNGIVRIPVQETDISKLVTYIEKSIEVDRLVSEAIKNGLPAAQAQTAR RMVLKDYI	I32-19A: 208,213,218,222,225,226,229,233
I32-19B SEQ ID NO:42	dimer	SGMRVYLGADHAGYELKQAIIFLKMTGHEPIDCGALRYDADDYPAFCIAAATRTVADPGSLGIVLGGSGNGEQIAANKVPGARCALAWSVQT AALAREHNNAQLIGIGGRMHTLEEALRIVKAFVTPWSKAQRHQRRIDILAEYERTHEAPPVPGAPA	I32-19B: 20,23,24,27,117,118,122,125

Name	Component Multimer	Amino Acid Sequence	Identified interface residues
I32-28A SEQ ID NO:43	trimer	GDDARIAAIGDVELNSQIGVLLAEPLPDDVR AALSAIQHDLFDLGGELCIPGHAAITEDHLLRL ALWL VHYNGQLPPLEEFILPGGARGAALAHV CRTVCRAERSIKALGASEPLNIAPAAVYNLL SDLLFVLARVLNRAAGGADVLWDRTRAH	I32-28A: 60,61,64,6 7,68,71,11 0,120,123, 124,128
I32-28B SEQ ID NO:44	dimer	ILSAEQSFTLRHPHGQAAALAFVREPAALAG VQRLRGLSDGGEQVWGELLVRVPLLGEVDLP FRSEIVRTPQGAELRPLTLTGERAWVAVSGQA TAAEGGEMAFQFQQAHLATPEAEGEGGA EVMVQAAAGVTLLL VAMALPQGLAAGLPPA	I32-28B: 35,36,54,1 22,129,137 ,140,141,1 44,148
I53-40A.1 SEQ ID NO:45	pentamer	TKKVGIVDTTFARVDMASAILTLKMESPNIK IIRKTVPGIKDLPVACKKLEEEGCDIVMALG MPGKKEKDKVCAHEASLGLMLAQLMTNKHII EVFVHEDEAKDDAELKILARRAIEHALNVY YLLFKPEYLTRMAGKGLRQGFEDAGPARE	I53-40A: 20,23,24,2 7,28,109,1 12,113,116 ,120,124
I53-40B.1 SEQ ID NO:46	trimer	DDINNQLKRLKVIPVIAIDNAEDIPLGKVLAE NGLPAAEITFRSSAAVKAIMLLRSAQPEMLIG AGTILNGVQALAAKEAGADFVSPGFNPNTV RACQIIGIDIVPGVNNPSTVEQALEMGLTTLKF FPAEASGGISMVKS LVGPYGDIRLMPTGGITPD NIDNYLAIPQVLACGGTWMVDKKLVRNGEW DEIARLTREIVEQVNP	I53-40B: 47,51,54,5 8,74,102
I53-47A.1 SEQ ID NO:47	trimer	PIFTLNTNIKADDVPSDFLSLTSRLVGLILSKPG SYVAVHINTDQQLSFGGSTNPAAFGLMSIGG IEPDKNRDHSAVLF DHLNAMLGIPKNRMYIHF VNLNGDDVGWNGTTF	I53-47A: 22,25,29,7 2,79,86,87
I53-47A.1N egT2 SEQ ID NO:48	trimer	PIFTLNTNIKADDVPSDFLSLTSRLVGLILSEPG SYVAVHINTDQQLSFGGSTNPAAFGLMSIGG IEPDKNEDHSAVLF DHLNAMLGIPKNRMYIHF VDLDGDDVGWNGTTF	I53-47A: 22,25,29,7 2,79,86,87
I53-47B.1 SEQ ID NO:49	pentamer	NQSHSKDHETVRIAVVRARWHADIVDACVE AFEIAMAAIGGDRFAVDVFDVPGAYEIPHAR TLAETGRYGAVLGTAFVVNGGIYRHEFVASA VIDGMMNVQLDTGVPVLSAVLTPHRYRDSDE HHRFFAAHFAVKGVEAARACIEILNAREKIAA	I53-47B: 28,31,35,3 6,39,131,1 32,135,139 ,146
I53-47B.1N egT2 SEQ ID NO:50	pentamer	NQSHSKDHETVRIAVVRARWHADIVDACVE AFEIAMAAIGGDRFAVDVFDVPGAYEIPHAR TLAETGRYGAVLGTAFVVDDGGIYDHEFVASA VIDGMMNVQLDTGVPVLSAVLTPHEYEDSDE DHEFFAAHFAVKGVEAARACIEILNAREKIAA	I53-47B: 28,31,35,3 6,39,131,1 32,135,139 ,146

Name	Component Multimer	Amino Acid Sequence	Identified interface residues
I53-50A.1 SEQ ID NO:51	trimer	EELFKKHKIVAVLRANSVEEAIEKAVAVFAGG VHLIEITFTVPDADTVIKALSVLKEKGAIIGAG TVTSVEQCRKAVESGAEFIVSPHLDEEISQFCK EKGVFYMPGVMTPTTELVKAMKLGHDILKLFP GEVVGPPQFVKAMKGPFPNVKFVPTGGVNLDN VCEWFKAGVLAVGVGDALVKGDPDEVREKA KKFVEKIRGCTE	I53-50A: 25,29,33,5 4,57
I53-50A.1N egT2 SEQ ID NO:52	trimer	EELFKKHKIVAVLRANSVEEAIEKAVAVFAGG VHLIEITFTVPDADTVIKALSVLKEKGAIIGAG TVTSVEQCRKAVESGAEFIVSPHLDEEISQFCK EKGVFYMPGVMTPTTELVKAMKLGHDILKLFP GEVVGPEFVEAMKGPFPNVKFVPTGGVDLDD VCEWFDAGVLAVGVGDALVEGDPDEVREDA KEFVEEIRGCTE	I53-50A: 25,29,33,5 4,57
I53-50A.1P osT1 SEQ ID NO:53	trimer	EELFKKHKIVAVLRANSVEEAIEKAVAVFAGG VHLIEITFTVPDADTVIKALSVLKEKGAIIGAG TVTSVEQCRKAVESGAEFIVSPHLDEEISQFCK EKGVFYMPGVMTPTTELVKAMKLGHDILKLFP GEVVGPPQFVKAMKGPFPNVKFVPTGGVNLDN VCKWFKAGVLAVGVGKALVKGKPDEVREKA KKFVKKIRGCTE	I53-50A: 25,29,33,5 4,57
I53-50B.1 SEQ ID NO:54	pentamer	NQSHKDHETVRIAVVRARWHAIEIVDACVSA FEAAMRDIGGDRFAVDVFDVPGAYEIP LHART LAETGRYGAVLGTAFFVNGGIYRHEFVASAVI DGMMNVQLDTGVPVLSAVLTPHRYRDSDAH TLLFLALFAVKGMEAAARACVEILAAREKIAA	I53-50B: 24,28,36,1 24,125,127 ,128,129,1 31,132,133 ,135,139
I53-50B.1N egT2 SEQ ID NO:55	pentamer	NQSHKDHETVRIAVVRARWHAIEIVDACVSA FEAAMRDIGGDRFAVDVFDVPGAYEIP LHART LAETGRYGAVLGTAFFVDGGIYDHEFVASAV IDGMMNVQLDTGVPVLSAVLTPHEYEDSDAD TLLFLALFAVKGMEAAARACVEILAAREKIAA	I53-50B: 24,28,36,1 24,125,127 ,128,129,1 31,132,133 ,135,139
I53-50B.4Po sT1 SEQ ID NO:56	trimer	NQSHKDHETVRIAVVRARWHAIEIVDACVSA FEAAMRDIGGDRFAVDVFDVPGAYEIP LHART LAETGRYGAVLGTAFFVNGGIYRHEFVASAVI NGMMNVQLNTGVPVLSAVLTPHNYDKSKAH TLLFLALFAVKGMEAAARACVEILAAREKIAA	I53-50B: 24,28,36,1 24,125,127 ,128,129,1 31,132,133 ,135,139
I53-40A genus SEQ ID NO:57	pentamer	TKKVGIVDTTFARVDMASAAITLKMESPNIK IIRKTVPGIKDLPVACKLLEEEGCDIVMALG MPGK(A/K)EKDKVCAHEASLGLMLAQLMTN KHIIIEVFVHEDEAKDDAELKILAARRAIEHAL NVYYLLFKPEYLTRMAGKGLRQGFEDAGPAR E	

Name	Component Multimer	Amino Acid Sequence	Identified interface residues
I53-40B genus SEQ ID NO:58	trimer	(S/D)(T/D)INNQLK(A/R)LKVIPVIAIDNAEDIPL GKVLAEENGLPAAEITFRSSAAVKAIMLLRSAQ PEMLIGAGTILNGVQALAAKEAGA(T/D)FVVS PGFNPNTVRACQIIGIDIVPGVNNPSTVE(A/Q)A LEMGLTTLKFFPAEASGGISMVKS LVGPYGD RLMPTGGITP(S/D)NIDNYLAIPQVLACGGTW MVDK KLV(T/R)NGEWDEIARLTREIVEQVNP	
I53-47A genus SEQ ID NO:59	trimer	PIFTLNTNIKA(T/D)DVPSDFLSLTSRLVGLILS(K/E)PGSYVAVHINTDQQLSFGGSTNPAAFGTL MSIGGIEP(S/D)KN(R/E)DHSAVLFDHLNAMLG IPKNRMYIHFV(N/D)L(N/D)GDDVGWNGTTF	
I53-47B genus SEQ ID NO:60	pentamer	NQHSHKD(Y/H)ETVRIAVVRARWHADIVDAC VEAFEIAMAAGGDRFAVDVFDVPGAYEIPHL ARTLAETGRYGAVLGTAFVV(N/D)GGIY(R/D) HEFVASAVIDGMMNVQL(S/D)TGVPLSAVL PH(R/E)Y(R/E)DS(A/D)E(H/D)H(R/E)FFAAHFA VKGVEAARACIEIL(A/N)AREKIAA	
I53-50A genus SEQ ID NO:61	trimer	EELFKKHKIVAVLRANSVEEAIEKAVAVFAGG VHLIEITFTVPDADTVIKALSVLKEKGAIIGAG TVTSVEQCRKAVESGAEFIVSPHLDEEISQFCK EKGVFYMPGVMTPTLVKAMKLGH(T/D)ILK LFPGEVVGP(Q/E)FV(K/E)AMKGPFPNVKFVPT GGV(N/D)LD(N/D)VC(E/K)WF(K/D)AGVLAVG VG(S/K/D)ALV(K/E)G(T/D/K)PDEVRE(K/D)AK(A/E/K)FV(E/K)(K/E)IRGCTE	
I53-50B genus SEQ ID NO:62	pentamer	NQHSHKD(Y/H)ETVRIAVVRARWHAEIVDAC VSAFEAAM(A/R)DIGGDRFAVDVFDVPGAYEI PLHARTLAETGRYGAVLGTAFVV(N/D)GGIY(R/D)HEFVASAVI(D/N)GMMNVQL(S/D/N)TGV PLSAVLTPH(R/E/N)Y(R/D/E)(D/K)S(D/K)A(H/ D)TLLFLALFAVKGMEARACVEILAAREKIA A	
T32- 28A SEQ ID NO:63	dimer	GEVPIGDPKELNGMEIAAVYLQPIEMEPRGIDL AASLADIHLEADIIHALKNNPNGFPEGFWMPYL TIAYALANADTGAIKTGTLMMPVADDGPHYG ANIAMEKDKKGGFGVGTYALTFLISNPEKQGF GRHVDEETGVGKWFEPFVVITYFFKYTGTPK	
T32- 28B SEQ ID NO:64	trimer	SQAIGILELTSIAKGMELGDAMLKSANVDLLV SKTISPGKFLLMLGGDIGAIQQA IETGTSQAGE MLVDSLVLANIHP SVLP AISGLNSVDKRQAVG IVETWSVAACISAADLAVKGSNVT LVRVHMA FGIGGKCYMVVAGDVLDVAAAVATASLAAG AKGLLVYASII PRPHEAMWRQMVEG	
T33- 09A	trimer	EEVVLITVPSALVAVKIAHALVEERLAACVNI VPGLTSIYRWQGSVSDHELLLVKTTTHAFP	

Name	Component Multimer	Amino Acid Sequence	Identified interface residues
SEQ ID NO:65		KLKERVKALHPYTVPEIVALPIAEGNREYLDWLRENTG	
T33-09B SEQ ID NO:66	trimer	VRGIRGAITVEEDTPAAILAATIELLLKMLEANGIQSYEELAAVIFTVTEDLTSAFPAAEARLIGMHRVPLLSAREVPVPGSLPRVIRVLALWNTDTPQDRVRHVYLNEAVRLRPDLESAQ	
T33-15A SEQ ID NO:67	trimer	SKAKIGIVTVSDRASAGITADISGKAAILALNLYLTSEWEPIYQVIPDEQDVIIETLIKMADEQDCCLIVTTGGTGPAKRDTVPEATEAVCDRMMMPGF GELMRAESLKEVPTAILSRQTAGLRGDSLIVNLPGDPASISDCLLAVFPAIPYCIDLMEGPYLECNEAMIKPFRPKAK	
T33-15B SEQ ID NO:68	trimer	VRGIRGAITVNSDTPSTIIATILLEKMLEANGIQSYEELAAVIFTVTEDLTSAFPAAEARQIGMHRVPLLSAREVPVPGSLPRVIRVLALWNTDTPQDRVRHVYLSEAVRLRPDLESAQ	
T33-21A SEQ ID NO:69	trimer	RITTKVGDKGSTRFLFGGEEVWKDSPIIEANGTLDELTSFIGEAKHYVDEEMKGILEEIQNDIYKIMGEIGSKGKIEGISEERIAWLLKLILRYMEMVNLKSFVLPGGTLES AKLDVCRTIARRALRKVLTVTREFGIGAEAAAYLLALS DLLFLLARVIEIEKNKLKEVRS	
T33-21B SEQ ID NO:70	trimer	PHLVIEATANLRLETSPGELLEQANKALFASGQFGEADIKSRFVTLEAYRQGTAAVERAYLHACL SILDGRDIATRTLLGASLC AVLAEAVAGGGEEGVQVSVEVREMERLSYAKRVVARQR	
T33-28A SEQ ID NO:71	trimer	ESVNTSFLSPSLVTIRDFDNGQFAVLRIGRTGFPADKGDIDLCLDKMIGVRAAQIFLGDDTEDGFKGPHIRIRCVDIDDKHTYNAMVYVDLIVGTGASEVERETAEEEEAKLALRVALQVDIADEHSCVTQFEMKLREELLSSDSFHPDKDEYYKDFL	
T33-28B SEQ ID NO:72	trimer	PVIQTFVSTPLDHHKRLLLAIIRIVTRVVLGKPEDLVMMTFHDSTPMHFFGSTDPVACVRVEALGGYGPSEPEKVT SIVTAAITAVCGIVADRIFVLYFSPLHCGWNGTNF	
T33-31A SEQ ID NO:73	trimer	EEVVLITVPSALVAVKIAHALVEERLAACVNIVPGLTSIYREEGSVVSDHELLLLVKTTTDAFPKLKERVKELHPYEVPEIVALPIAEGNREYLDWLRENTG	
I53-50A ΔCys SEQ ID NO: 74	trimer	EELFKKHKIVAVLRANSVEEAIEKAVAVFAGGVHLIEITFTVPDADTVIKALS VLKEKGAIIGAGTVTSVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVNLDN	

Name	Component Multimer	Amino Acid Sequence	Identified interface residues
		VAEWFKAGVLAVGVGSALVKGTPDEVREKA KAFVEKIRGATE	
T33_dn 2A SEQ ID NO: 75		NLAEKMYKAGNAMYRKGGYTTIAHAYTLALL KDPNNAEAWYNLGNAAAYKKGEYDEAIEAYQ KALEDPNNAEAWYNLGNAYYKQGDYDEAI EYYKKALRLDPRNVDAIENLIEAEEKQG	
T33_dn 2B SEQ ID NO: 76		EEAELAYLLGELAYKLGEYRIAIRAYRIALKR DPNNAEAWYNLGNAYYKQGDYREAIRYYLR ALKLDPENAEAWYNLGNALYKQGGYDLAIIA YQAAL EEDPNNAEAKQNLGNAKQKQG	
T33_dn 5A SEQ ID NO: 77		NSAEAMYKMGNAAYKQGDYILAHAYLLALE KDPNNAEAWYNLGNAAAYKQGDYDEAIEYYQ KALEDPNNAEAWYNLGNAYYKQGDYDEAI EYYEKALEDPNNAEALKNLLEAIAEQD	
T33_dn 5A SEQ ID NO: 78		TDPLAVILYIAILKAEKSIARAKAAEALGKIGD ERAVEPLIKALKDEDALVRAAADALGQIGD ERAVEPLIKALKDEEGLVRASAAIALGQIGDE RAVQPLIKALTDERDLVRVAAVALGRIGDE KAVRPLIIVLKDEEGEVREAAIALGSIGGERV RAAMEKLAERGTGFARKVAVNYLETHK	
T33_dn 10A SEQ ID NO: 79		EEAELAYLLGELAYKLGEYRIAIRAYRIALKR DPNNAEAWYNLGNAYYKQGDYDEAIEYYQK ALEDPNNAEAWYNLGNAYYKQGDYDEAIE YYEKALEDPENLEALQNLLNAMDKQG	
T33_dn 10B SEQ ID NO: 80		IEEVVAEMIDILAESSKKIEELARAADNKTTE KAVAEAEIEIARLATAAIQLIEALAKNLASEEF MARASIAELAKKAIEAIYRLADNHTTDTFM ARAI AAIANLAVTAILAIAALASNHTTEEFMA RAISIAELAKKAIEAIYRLADNHTTDFMAA AIEAIALLATLAILAIALASNHTTEKFMAIRAI MAIAILAKAIEAIYRLADNHTSPTYIEKAIEAI EKIARKAIAIEMLAKNITTEYKEKAKKIIDI RKLAKMAIKKLEDNRT	
I53_dn5 A SEQ ID NO: 81	pentamer	KYDGSKLRIGILHARWNAEIIALVLGALKRL QFEGVKRENIITVPGSFELPYGSKLFVEKQK RLGKPLDAIPIGVLIKSTMHFEYICDSTTHQL MKLNFELGIPVIFGVLTCLTDEQAEARAGLIEG KMHNHGEDWGAAAVEMATKFN	
I53_dn5 B SEQ ID NO: 82	trimer	EEAELAYLLGELAYKLGEYRIAIRAYRIALKR DPNNAEAWYNLGNAYYKQGRYREAIEYYQK ALEDPNNAEAWYNLGNAYYERGEREYEEAIEY YRKALRLDPNNADAMQNLLNAKMREE	
I53_dn5 A.1	pentamer	KYDGSKLRIGILHARGNAEIIALVLGALKRLQ EFGVKRENIITVPGSFELPYGSKLFVEKQKR LGKPLDAIPIGVLI RGSTPHFDYIADSTTHQLM	

Name	Component Multimer	Amino Acid Sequence	Identified interface residues
SEQ ID NO: 83		KLNFELGIPVIFGVITADTDEQAEARAGLIEGK MHNHGEDWGAAAVEMATKFN	
I53_dn5 A.2 SEQ ID NO: 84	pentamer	KYDGSKLRIgilHARGNAEIIELVLGALKRLQ EFGVKRENIHETVPGSFELPYGSKLFVEKQKR LGKPLDAIPIGVLRGSAHFDYIADSTTHQL MKLNFELGIPVIFGVLTTESEDEQAEERAGTKA GNHGEDWGAAAVEMATKFN	
I3-01 SEQ ID NO: 209		MKMEELFKKHKIVAVLRANSVEEAKKKALA VFLGGVHLIEITFTVPDADTVIKELSFLKEMGA IIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEIS QFCKEKGVFYMPGVMTPTTELVKAMKLGHTIL KLFPGEVVGPPQFVKAMKGPFPPNVKFVPTGGV LDNVCEWFKAGVLA VGVGSALVKGTPVEV AEKAKAFVEKIRGCTE	none
I3-01 (M31) SEQ ID NO: 210		MKIEELFKKHKIVAVLRANSVEEAKKKALAV FLGGVHLIEITFTVPDADTVIKELSFLKEMGAII GAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQ FCKEKGVFYMPGVMTPTTELVKAMKLGHTILK LFPGEVVGPPQFVKAMKGPFPPNVKFVPTGGVN LDNVCEWFKAGVLA VGVGSALVKGTPVEVA EKAKAFVEKIRGCTE	
1WA3- ref SEQ ID NO: 211		MKMEELFKKHKIVAVLRANSVEEAKKALAV FEGGVHLIEITFTVPDADTVIKELSFLKEKGAI GAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQ FCKEKGVFYMPGVMTPTTELVKAMKLGHTILK LFPGEVVGPPQFVKAMKGPFPPNVKFVPTGGVN LDNVCEWFKAGVLA VGVGSALVKGTPDEV EKAKAFVEKIRGCTE	
1WA3-1 SEQ ID NO: 212		(MK)MEELFKKHKIVAVLRANSVEEAKKKAL AVFLGGVHLIEITFTVPDADTVIKELSFLKEMG AIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEE ISQFCKEKGVFYMPGVMTPTTELVKAMKLGHT ILKLFPGEVVGPPQFVKAMKGPFPPNVKFVPTGG VNLNVCEWFKAGVLA VGVGSALVKGTPVE VAEKAKAFVEKIRGCTE	none
1WA3-2 SEQ ID NO: 213		(MK)IEELFKKHKIVAVLRANSVEEAKKKALA VFLGGVHLIEITFTVPDADTVIKELSFLKEMGA IIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEIS QFCKEKGVFYMPGVMTPTTELVKAMKLGHTIL KLFPGEVVGPPQFVKAMKGPFPPNVKFVPTGGV LDNVCEWFKAGVLA VGVGSALVKGTPVEV AEKAKAFVEKIRGCTE	E33L, K61M, D187V, R190A
1WA3-3 SEQ ID NO: 214		(MK)MEELFKKHKIVAVLRANSVEEAKKKAL AVFLGGVHLIEITFTVPDADTVIKELSFLKEMG AIIGAGTVTSVEQARKAVESGAEFIVSPHLDEE	E33L, K61M,

Name	Component Multimer	Amino Acid Sequence	Identified interface residues
		ISQFAKEKGVFYMPGVMTPTTELVKAMKLGHT ILKLFPGEVVGPPQFVKAMKGPPPNVKFVPTGG VNLDNVCEWFKAGVLA VGVGSALVKGTPVE VAEKAKAFVEKIRGCTE	D187V, R190A
1WA3-4 SEQ ID NO: 215		(MK)MEELFKKHKIVAVLRANSVEEAKMKAL AVFVGGVHLIEITFTVPDADTVIKELSFLKELG AIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEE ISQFCKEKGVFYMPGVMTPTTELVKAMKLGHT ILKLFPGEVVGPPQFVKAMKGPPPNVKFVPTGG VNLDNVCEWFKAGVLA VGVGSALVKGTTAE VAAKAAAFVEKIRGCTE	E33M, K61L, D187A, R190A
1WA3-5 SEQ ID NO: 216		(MK)MEELFKKHKIVAVLRANSVEEAKKKAL AVFLGGVDLIEITFTVPDADTVIKELSFLKEMG AIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEE ISQFCKEKGVFYMPGVMTPTTELVKAMKLGHT ILKLFPGEVVGPPQFVKAMKGPPPNVKFVPTGG VNLDNVCEWFKAGVQAVGVGEALNKGTPVE VAEKAKAFVEKIRGCTE	E33L, K61M, D187V, R190A
1WA3-6 SEQ ID NO: 217		(MK)MEELFKKHKIVAVLRANSVEEAKKKAL AVFMGGVDLIEITFTVPDADTVIKELSFLKELG AIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEE ISQFCKEKGVFYMPGVMTPTTELVKAMKLGHT ILKLFPGEVVGPPQFVKAMKGPPPNVKFVPTGG VNLDNVCEWFKAGVQAVGVGEALNKGTPAE VAEKAKAFVEKIRGCTE	E33M, K61L, D187A, R190A
I3-01 H35D SEQ ID NO: 218		(METDTLLLWVLLLWVPGSTGDYKDEK)MEE LFKKHKIVAVLRANSVEEAKKKALAVFLGGV DLIEITFTVPDADTVIKELSFLKEMGAIIGAGTV TSVEQCRKAVESGAEFIVSPHLDEEISQFCKEK GVFYMPGVMTPTTELVKAMKLGHTILKLFPGE VVGPPQFVKAMKGPPPNVKFVPTGGVNLDNV CEWFKAGVLA VGVGSALVKGTPVEVAEKAK AFVEKIRGCTE(QKLISEEDLHHHHHH)	
I3-01 K25D SEQ ID NO: 219		(METDTLLLWVLLLWVPGSTGDYKDEK)MEE LFKKHKIVAVLRANSVEEAKKDALAVFLGGV HLIEITFTVPDADTVIKELSFLKEMGAIIGAGTV TSVEQCRKAVESGAEFIVSPHLDEEISQFCKEK GVFYMPGVMTPTTELVKAMKLGHTILKLFPGE VVGPPQFVKAMKGPPPNVKFVPTGGVNLDNV CEWFKAGVLA VGVGSALVKGTPVEVAEKAK AFVEKIRGCTE(QKLISEEDLHHHHHH)	
I3-01 K25N SEQ ID NO: 220		(METDTLLLWVLLLWVPGSTGDYKDEK)MEE LFKKHKIVAVLRANSVEEAKKNALAVFLGGV HLIEITFTVPDADTVIKELSFLKEMGAIIGAGTV TSVEQCRKAVESGAEFIVSPHLDEEISQFCKEK GVFYMPGVMTPTTELVKAMKLGHTILKLFPGE	

Name	Component Multimer	Amino Acid Sequence	Identified interface residues
		VVGPQFVKAMKGPPNVK FVPTGGVNLDNV CEWFKAGVLA VGVGSALVKGTPVEVAEKAK AFVEKIRGCTE(QKLISEEDLHHHHHH)	
I3-01 L171Q SEQ ID NO: 221		(METDTLLLWVLLLWVPGSTGDYKDEK)MEE LFKKHKIVAVLRANSVEEAKKKALAVFLGGV HLIEITFTVPDADTVIKELSFLKEMGAIIGAGTV TSVEQCRKAVESGAEFIVSPHLDEEISQFCKEK GVFYMPGVMTPTEL VKAMKLGHTILKLPGE VVGPQFVKAMKGPPNVK FVPTGGVNLDNV CEWFKAGVQAVGVGSALVKGTPVEVAEKAK AFVEKIRGCTE(QKLISEEDLHHHHHH)	
I3-01 L171Q/ S177E/ V180N SEQ ID NO: 222		(METDTLLLWVLLLWVPGSTGDYKDEK)MEE LFKKHKIVAVLRANSVEEAKKKALAVFLGGV HLIEITFTVPDADTVIKELSFLKEMGAIIGAGTV TSVEQCRKAVESGAEFIVSPHLDEEISQFCKEK GVFYMPGVMTPTEL VKAMKLGHTILKLPGE VVGPQFVKAMKGPPNVK FVPTGGVNLDNV CEWFKAGVQAVGVGEALNKGTPVEVAEKAK AFVEKIRGCTE(QKLISEEDLHHHHHH)	
I3-01 'secretion mutations' (H35D/ L171Q/ S177E/ V180N) SEQ ID NO: 223		(METDTLLLWVLLLWVPGSTGDYKDEK)MEE LFKKHKIVAVLRANSVEEAKKKALAVFLGGV DLIEITFTVPDADTVIKELSFLKEMGAIIGAGTV TSVEQCRKAVESGAEFIVSPHLDEEISQFCKEK GVFYMPGVMTPTEL VKAMKLGHTILKLPGE VVGPQFVKAMKGPPNVK FVPTGGVNLDNV CEWFKAGVQAVGVGEALNKGTPVEVAEKAK AFVEKIRGCTE(QKLISEEDLHHHHHH)	
I3-01 'negative interior' SEQ ID NO: 224		(METDTLLLWVLLLWVPGSTGD)MEELFKEHK IVAVLRANSVEEAKKKALAVFLGGVHLIEITF TVPDADTVIKELSFLKEMGAIIGAGTVTSVEQ AREAVESGAEFIVSPHLDEEISQFAKEEGVFY MPGVMTPTEL VKAMKLGHTILKLPGEVVGP QFVEAMKGPPNVK FVPTGGVNLCNVAEWF AGVLA VGVGSALVEGTPVEVAEKAKAFVEKI EGATE(QKLISEEDLHHHHHH)	
I3-01 'negative interior with secretion'		(METDTLLLWVLLLWVPGSTGD)MEELFKEHK IVAVLRANSVEEAKKKALAVFLGGVDLIEITF TVPDADTVIKELSFLKEMGAIIGAGTVTSVEQ AREAVESGAEFIVSPHLDEEISQFAKEEGVFY MPGVMTPTEL VKAMKLGHTILKLPGEVVGP QFVEAMKGPPNVK FVPTGGVNLDNVAEWF AGVQAVGVGEALNEGTPVEVAEKAKAFVEKI EGATE(QKLISEEDLHHHHHH)	

Name	Component Multimer	Amino Acid Sequence	Identified interface residues
mutations' SEQ ID NO: 225			

[0206] **Table 12** provides the amino acid sequence of a first repeat element and second repeat element of embodiments of the present disclosure. In each case, the pairs of sequences together form an I53 multimer with icosahedral symmetry. The right-hand column in **Table 12** identifies the residue numbers in each illustrative polypeptide that were identified as present at the interface of resulting assembled protein nanostructures (*i.e.*: “identified interface residues”). As can be seen, the number of interface residues for the illustrative polypeptides of SEQ ID NOs: 23-56 range from 4-13. In various embodiments, a first repeat element and second repeat element comprise an amino acid sequence that is at least 75%, at least 80%, at least 85%, 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 100% identical over its length, and identical at least at 1, at least at 2, at least at 3, at least at 4, at least at 5, at least at 6, at least at 7, at least at 8, at least at 9, at least at 10, at least at 11, at least at 12, or at least at 13 identified interface positions (depending on the number of interface residues for a given polypeptide), to the amino acid sequence of a polypeptide selected from the group consisting of SEQ ID NOs: 23-56. SEQ ID NOs: 57-84 represent other amino acid sequences of a first repeat element and second repeat element from embodiments of the present disclosure. In other embodiments, a first repeat element and/or second repeat element comprise an amino acid sequence that is at least 75%, at least 80%, at least 85%, 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 100% identical over its length, and identical at least at 20%, at least 25%, at least 33%, at least 40%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 90%, or 100% of the identified interface positions, to the amino acid sequence of a polypeptide selected from the group consisting of SEQ ID NOs: 23-84, 209-225.

[0207] As is the case with proteins in general, the polypeptides are expected to tolerate some variation in the designed sequences without disrupting subsequent assembly into protein nanostructures: particularly when such variation comprises conservative amino acid

substitutions. As used here, “conservative amino acid substitution” means that: hydrophobic amino acids (Ala, Cys, Gly, Pro, Met, Val, Ile, Leu) are substituted with other hydrophobic amino acids; hydrophobic amino acids with bulky side chains (Phe, Tyr, Trp) are substituted with other hydrophobic amino acids with bulky side chains; amino acids with positively charged side chains (Arg, His, Lys) are substituted with other amino acids with positively charged side chains; amino acids with negatively charged side chains (Asp, Glu) are substituted with other amino acids with negatively charged side chains; and amino acids with polar uncharged side chains (Ser, Thr, Asn, Gln) are substituted with other amino acids with polar uncharged side chains.

[0208] In various embodiments of the protein nanostructures of the invention, a first repeat element and second repeat element, or the vice versa, comprise polypeptides with the amino acid sequence selected from the following pairs, or modified versions thereof (*i.e.*, permissible modifications as disclosed for the polypeptides of the invention: isolated polypeptides comprising an amino acid sequence that is at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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 100% over its length, and/or identical at least at one identified interface position, to the amino acid sequence indicated by the SEQ ID NO):

SEQ ID NO: 23 and SEQ ID NO: 24 (I53-34A and I53-34B);
 SEQ ID NO: 25 and SEQ ID NO: 26 (I53-40A and I53-40B);
 SEQ ID NO: 25 and SEQ ID NO: 46 (I53-40A and I53-40B.1);
 SEQ ID NO: 45 and SEQ ID NO: 26 (I53-40A.1 and I53-40B);
 SEQ ID NO: 57 and SEQ ID NO: 58 (I53-40A genus and I53-40B genus);
 SEQ ID NO: 27 and SEQ ID NO: 28 (I53-47A and I53-47B);
 SEQ ID NO: 27 and SEQ ID NO: 49 (I53-47A and I53-47B.1);
 SEQ ID NO: 27 and SEQ ID NO: 50 (I53-47A and I53-47B.1NegT2);
 SEQ ID NO: 47 and SEQ ID NO: 28 (I53-47A.1 and I53-47B);
 SEQ ID NO: 47 and SEQ ID NO: 49 (I53-47A.1 and I53-47B.1);
 SEQ ID NO: 47 and SEQ ID NO: 50 (I53-47A.1 and I53-47B.1NegT2);
 SEQ ID NO: 48 and SEQ ID NO: 28 (I53-47A.1NegT2 and I53-47B);
 SEQ ID NO: 48 and SEQ ID NO: 49 (I53-47A.1NegT2 and I53-47B.1);
 SEQ ID NO: 48 and SEQ ID NO: 50 (I53-47A.1NegT2 and I53-47B.1NegT2);
 SEQ ID NO: 59 and SEQ ID NO: 60 (I53-47A genus and I53-47B genus);

SEQ ID NO: 29 and SEQ ID NO: 30 (I53-50A and I53-50B);
 SEQ ID NO: 29 and SEQ ID NO: 54 (I53-50A and I53-50B.1);
 SEQ ID NO: 29 and SEQ ID NO: 55 (I53-50A and I53-50B.1NegT2);
 SEQ ID NO: 29 and SEQ ID NO: 56 (I53-50A and I53-50B.4PosT1);
 SEQ ID NO: 51 and SEQ ID NO: 30 (I53-50A.1 and I53-50B);
 SEQ ID NO: 51 and SEQ ID NO: 54 (I53-50A.1 and I53-50B.1);
 SEQ ID NO: 51 and SEQ ID NO: 55 (I53-50A.1 and I53-50B.1NegT2);
 SEQ ID NO: 51 and SEQ ID NO: 56 (I53-50A.1 and I53-50B.4PosT1);
 SEQ ID NO: 52 and SEQ ID NO: 30 (I53-50A.1NegT2 and I53-50B);
 SEQ ID NO: 52 and SEQ ID NO: 54 (I53-50A.1NegT2 and I53-50B.1);
 SEQ ID NO: 52 and SEQ ID NO: 55 (I53-50A.1NegT2 and I53-50B.1NegT2);
 SEQ ID NO: 52 and SEQ ID NO: 56 (I53-50A.1NegT2 and I53-50B.4PosT1);
 SEQ ID NO: 53 and SEQ ID NO: 30 (I53-50A.1PosT1 and I53-50B);
 SEQ ID NO: 53 and SEQ ID NO: 54 (I53-50A.1PosT1 and I53-50B.1);
 SEQ ID NO: 53 and SEQ ID NO: 55 (I53-50A.1PosT1 and I53-50B.1NegT2);
 SEQ ID NO: 53 and SEQ ID NO: 56 (I53-50A.1PosT1 and I53-50B.4PosT1);
 SEQ ID NO: 61 and SEQ ID NO: 62 (I53-50A genus and I53-50B genus);
 SEQ ID NO: 31 and SEQ ID NO: 32 (I53-51A and I53-51B);
 SEQ ID NO: 33 and SEQ ID NO: 34 (I52-03A and I52-03B);
 SEQ ID NO: 35 and SEQ ID NO: 36 (I52-32A and I52-32B);
 SEQ ID NO: 37 and SEQ ID NO: 38 (I52-33A and I52-33B);
 SEQ ID NO: 39 and SEQ ID NO: 40 (I32-06A and I32-06B);
 SEQ ID NO: 41 and SEQ ID NO: 42 (I32-19A and I32-19B);
 SEQ ID NO: 43 and SEQ ID NO: 44 (I32-28A and I32-28B);
 SEQ ID NO: 45 and SEQ ID NO: 46 (I53-40A.1 and I53-40B.1);
 SEQ ID NO: 63 and SEQ ID NO: 64 (T32-28A and T32-28B);
 SEQ ID NO: 65 and SEQ ID NO: 66 (T33-09A and T33-09B);
 SEQ ID NO: 67 and SEQ ID NO: 68 (T33-15A and T33-15B);
 SEQ ID NO: 69 and SEQ ID NO: 70 (T33-21A and T33-21B);
 SEQ ID NO: 71 and SEQ ID NO: 72 (T33-28A and T32-28B); and
 SEQ ID NO: 73 and SEQ ID NO: 66 (T33-31A and T33-09B (also referred to as T33-31B)).

[0209] In some embodiments, the assembly domains are I53-50A (trimer, optionally linked to an antigenic fragment of a coronavirus Spike glycoprotein as described above) and

I53-50B (pentamer). I53_dn5 nanostructures are described in US 2022/0072120 A1, the contents of which are incorporated by reference. I53_dn5 variants may include one or more amino acid substitutions, such as C94A, C119A, W18G, K84R, M88P, E91D, L117I, or L120D (together “I53_dn5A.1”; Ueda et al. eLife 9:e57659 (2020) or A25E, M88A, C119T, L120E, A127E, L131T, I132K, E133A, or a deletion of positions 135-137 (“I53_dn5A.2”; Wang et al. bioRxiv 2022.08.04.502842).

[0210] In some embodiments, the one or more antigenic fragments of coronavirus Spike glycoprotein, or antigenic fragments thereof, are expressed as a fusion protein with a first assembly domain. In some embodiments, the first assembly domain and the antigenic fragment of a coronavirus Spike glycoprotein are joined by a linker sequence.

[0211] Non-limiting examples of designed protein complexes useful in protein nanostructures of the present disclosure include those disclosed in U.S. Patent No. 9,630,994; Int'l Pat. Pub No. WO2018187325A1; U.S. Pat. Pub. No. 2018/0137234 A1; U.S. Pat. Pub. No. 2019/0155988A2, each of which is incorporated herein in its entirety.

[0212] In various embodiments of the protein nanostructures of the disclosure, the assembly domains are polypeptides with the amino acid sequence selected from the following pairs, or modified versions thereof (*i.e.*, permissible modifications as disclosed for the polypeptides of the invention: isolated polypeptides comprising an amino acid sequence that is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% over its length, and/or identical at least at one identified interface position, to the amino acid sequence indicated by the SEQ ID NO):

SEQ ID NO: 75 and SEQ ID NO: 76 (T33_dn2A and T33_dn2B);

SEQ ID NO: 77 and SEQ ID NO: 78 (T33_dn5A and T33_dn5B);

SEQ ID NO: 79 and SEQ ID NO: 80 (T33_dn10A and T33_dn10B); or

SEQ ID NO: 81 and SEQ ID NO: 82 (I53_dn5A and I53_dn5B).

[0213] In another aspect, the disclosure provides a self-assembling protein nanostructure comprising a first component comprising a recombinant coronavirus polypeptide described herein and optionally, a second component comprising a second protein.

[0214] In some embodiments, the protein nanostructure of the recombinant coronavirus polypeptide comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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 100% identical to SEQ ID NOs: 11-19.

[0215] In some embodiments, the protein nanostructure of the recombinant coronavirus polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acids relative to a reference sequence according to SEQ ID NO: 1.

[0216] In some embodiments, the protein nanostructure component polypeptide and the recombinant coronavirus polypeptide are non-covalently coupled.

[0217] In some embodiments, the protein nanostructure component polypeptide and the recombinant coronavirus polypeptide are covalently coupled.

[0218] In some embodiments, the protein nanostructure comprises a fusion protein which includes the recombinant coronavirus polypeptide, a linker, and the protein nanostructure component polypeptide.

[0219] In some embodiments, the protein nanostructure comprises a fusion protein which includes, in N- to C-terminal order, the recombinant coronavirus polypeptide, a linker, and the protein nanostructure component polypeptide.

[0220] In some embodiments, the protein nanostructure of the recombinant coronavirus polypeptide comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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 100% identical to SEQ ID NOs: 11-19.

[0221] In some embodiments, the protein nanostructure of the fusion protein comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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 100% identical to SEQ ID NOs: 11-19.

[0222] In some embodiments, the first component of the protein nanostructure is I53-50A. In some embodiments, the first component of the protein nanostructure is I53-34A. In some embodiments, the first component of the protein nanostructure is I53-40B. In some embodiments, the first component of the protein nanostructure is I53-47A. In some embodiments, the first component of the protein nanostructure is I53-51A.

[0223] In some embodiments, the protein nanostructure is an I3-01/MI3 protein nanostructure. In some embodiments, the protein nanostructure is an I53-50AB protein nanostructure.

[0224] In some embodiments, the protein nanostructure is a ferritin protein nanostructure. In some embodiments, the protein nanostructure is an encapsulin protein nanostructure. In some embodiments, the protein nanostructure is a CP3 bacteriophage capsid protein nanostructure. In some embodiments, the protein nanostructure is a Q β bacteriophage capsid protein nanostructure. In some embodiments, the protein nanostructure is an AP205 bacteriophage capsid protein nanostructure. In some embodiments, the protein nanostructure comprises a first polypeptide and a second polypeptide selected from those disclosed in WO 2021252688 which is incorporated by reference herein in its entirety.

[0225] In some embodiments, the recombinant polypeptide of the protein nanostructure comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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 100% identical to SEQ ID NOs: 11-19.

[0226] In some embodiments, the protein nanostructure component polypeptide comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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 100% identical to SEQ ID NOs: 20-21.

[0227] In some embodiments, the protein nanostructure comprises a second component polypeptide.

[0228] In some embodiments, the second component polypeptide is I53-50B. In some embodiments, the second component polypeptide is I53-34B. In some embodiments, the second component polypeptide is I53-40B. In some embodiments, the second component polypeptide is I53-47B. In some embodiments, the second component polypeptide is I53-51B. . In some embodiments, the second component of the protein nanostructure is I53-50B.4PosT1.

[0229] In some embodiments, the second component polypeptide comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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 100% identical to SEQ ID NO: 22.

Ferritin-based nanostructures

[0230] In some embodiments, the assembly domain is a ferritin polypeptide. In some embodiments, the assembly domain of a ferritin protein nanostructure comprises a polypeptide sequence at least 70%, at least 80%, 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 100% identical to any one of the following sequences:

MLSKDIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLFDHAAE
EYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYEHEQHISESIN
NIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFDKILDKVELIGNENHGLY
LADQYVKGIAKSRS. (SEQ ID NO: 204)

MLKPEMIEKLNEQMNELEYSSLLYQQMSAWCSYHTFEGAAFLRRHAQE
EMTHMQRLFDYLTDTGNLPRINTVESPFAYSSLDLQETQKHEQLITQK
INELAHAAMTNQDYPTFNFLQWYVSEQHEEEKLFKSIIDKLSLAGKSGEG
LYFIDKELSTLDAQN. (SEQ ID NO: 205)

NFHQDCEAGLNRTVNLKFHSSYVYLSMASYFNRDDVALSNFAKFFRERS
EEEKEHAIEYQNRGRVFLQSVKPERDDWANGLEALQTALKLQK
SVNQALLDLHAVAADKSDPHMTDFLESPYLSVESVETIKKLGDHITSLKKL
WSSHPGMAEYLFNKHTLG. (SEQ ID NO: 206)

QFSKDIEKLLNEQVNKEMQSSNLYMSMSSWCYTHSLDGAGLFLFDHAAE
EYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYEHEQHISESIN
NIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFDKILDKIELIGNENHGLYL
ADQYVKGIAKSRSKSGS. (SEQ ID NO: 207)

SGESQVRQNFKPEMEEKLNEQMNELEYSSLLYQQMSAWCSYHTFEGAA
AFLRRHAQEEMTHMQRLFDYLTDTGNLPRINTVESPFAYSSLDLQET
YKHEQLITQKINELAHAAMTNQDYPTFNFLQWYVSEQHEEEKLFKSIIDK
LSLAGKSGEGLYFIDKELSTLDGS. (SEQ ID NO: 208)

Other nanostructures or nanoparticles

[0231] In some embodiments, the ectodomains described herein are displayed on any nanostructure or nanoparticle known in the art. Illustrative nanostructures and nanoparticles include, but are not limited to Human papillomavirus (HPV) virus-like particles (VLPs), Chikungunya VLPs, AP205 capsid protein VLPs, phage VLPs (e.g., bacteriophage), an encapsulin protein nanostructure, a CP3 bacteriophage capsid protein nanostructure, a Q β bacteriophage capsid protein nanostructure. Display on these and other platforms may be performed by creating a fusion protein of the ectodomain to a relevant protein of the system,

by bioconjugate chemistry (e.g., SpyCatcher), or other means known in the art. The protein nanostructure may be a lumazine synthase nanoparticle as described, e.g., in Geng et al. PLoS Pathog. 17(9):e1009897 (2021). The protein nanostructure may be a ferritin nanoparticle as described, e.g., in Joyce et al. bioRxiv 2021.05.09.443331 and in U.S. Pat. Pub. No. US 2019/0330279 A1.

Polynucleotides

[0232] Provided herein are polynucleotides encoding the polypeptides or the protein nanostructures described herein. The polynucleotide sequence may comprise RNA or DNA. Isolated polynucleotides refer to those that have been removed from their normal surrounding nucleic acid sequences in the genome or in cDNA sequences. Such polynucleotide sequences may comprise additional sequences useful for promoting expression and/or purification of the encoded protein, including but not limited to polyA sequences, modified Kozak sequences, and sequences encoding epitope tags, export signals, and secretory signals, nuclear localization signals, and plasma membrane localization signals. It will be apparent to those of skill in the art, based on the teachings herein, what nucleic acid sequences will encode the proteins of the disclosure.

[0233] In a further aspect, the present disclosure provides recombinant expression vectors comprising the isolated nucleic acid of any embodiment or combination of embodiments of the disclosure operatively linked a suitable control sequence. “Recombinant expression vector” includes vectors that operatively link a nucleic acid coding region or gene to any control sequences capable of effecting expression of the gene product. “Control sequences” operably linked to the nucleic acid sequences of the disclosure are nucleic acid sequences capable of effecting the expression of the nucleic acid molecules. The control sequences need not be contiguous with the nucleic acid sequences, so long as they function to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between a promoter sequence and the nucleic acid sequences and the promoter sequence can still be considered “operably linked” to the coding sequence. Other such control sequences include, but are not limited to, polyadenylation signals, termination signals, and ribosome binding sites. Such expression vectors can be of any type known in the art, including but not limited to plasmid and viral-based expression vectors. The control sequence used to drive expression of the disclosed nucleic acid sequences in a mammalian system may be constitutive (driven by any of a variety of promoters, including but not limited to, CMV, SV40, RSV, actin, EF) or inducible (driven by any of a number of inducible promoters including, but not limited to, tetracycline, ecdysone, steroid

responsive). The construction of expression vectors for use in transfecting prokaryotic cells is also well known in the art, and thus can be accomplished via standard techniques. (See, for example, Sambrook, Fritsch, and Maniatis, in: *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1989; *Gene Transfer and Expression Protocols*, pp. 109-128, ed. E.J. Murray, The Humana Press Inc., Clifton, N.J.), and the Ambion 1998 Catalog (Ambion, Austin, TX). The expression vector can be replicable in the host organisms either as an episome or by integration into host chromosomal DNA. Suitable expression vectors and hosts are well known in the art. In some embodiments, the expression vector comprises a plasmid. However, the disclosure is intended to include other expression vectors that serve equivalent functions, such as viral vectors.

[0234] In another aspect, the present disclosure provides host cells that have been transfected or transduced with the recombinant expression vectors disclosed herein, wherein the host cells can be either prokaryotic or eukaryotic. The cells can be transiently or stably transfected or transduced. Such transfection or transduction of expression vectors into prokaryotic and eukaryotic cells can be accomplished via any technique known in the art, including but not limited to standard bacterial transformations, calcium phosphate coprecipitation, electroporation, or liposome mediated-, DEAE dextran mediated-, polycationic mediated-, or viral mediated transfection. (See, for example, *Molecular Cloning: A Laboratory Manual* (Sambrook, et al., 1989, Cold Spring Harbor Laboratory Press; *Culture of Animal Cells: A Manual of Basic Technique*, 2nd Ed. (R.I. Freshney. 1987. Liss, Inc. New York, NY).

[0235] In some embodiments, the polynucleotide is a messenger RNA (mRNA). Methods of generating a polynucleotide either by chemical synthesis or by *in vitro* transcription (IVT) (for mRNA) are well known in the art.

[0236] The basic structure of an encoding mRNA can resemble “mature” eukaryotic mRNA and can include some or all of the following features including (i) a protein-encoding open reading frame (ORF), flanked by (ii) 5′ and 3′ untranslated regions (UTRs), and at the end sides (iii) a 7-methyl guanosine 5′ cap structure and (iv) a 3′ poly(A) tail. The non-coding structural features can be individually optimized to modulate the mRNA stability, translation efficiency, and immunogenicity. By incorporating modified nucleosides, mRNA transcripts referred to as “nucleoside-modified mRNA” can be produced with reduced immunostimulatory activity, and therefore an improved safety profile can be obtained. In addition, modified nucleosides allow the design of mRNA vaccines with strongly enhanced stability and translation capacity, as they can avoid the direct antiviral pathways that are

induced by type IFNs and are programmed to degrade and inhibit invading mRNA. For instance, the replacement of uridine with pseudouridine reduces the activity of 2'-5'-oligoadenylate synthetase, which regulates the mRNA cleavage by RNase L. In addition, lower activities are measured for protein kinase R, an enzyme that is associated with the inhibition of the mRNA translation process.

[0237] Polynucleotides of the disclosure may include one or more modified (*e.g.*, altered or alternative) nucleobases, nucleosides, nucleotides, or combinations thereof. The polynucleotides can include any useful modification or alteration, such as to the nucleobase, the sugar, or the internucleoside linkage (*e.g.*, to a linking phosphate/to a phosphodiester linkage/to the phosphodiester backbone). In some embodiments, alterations (*e.g.*, one or more alterations) are present in each of the nucleobase, the sugar, and the internucleoside linkage. Alterations according to the present disclosure may be alterations of polynucleotides, *e.g.*, the substitution of the 2'—OH of the ribofuranosyl of an RNA ring to 2'—H, threose nucleic acids (TNAs), glycol nucleic acids (GNAs), peptide nucleic acids (PNAs), locked nucleic acids (LNAs), or hybrids thereof.

[0238] The polynucleotide may include a 5'-cap structure. The 5'-cap structure of a polynucleotide is involved in nuclear export and increasing polynucleotide stability and binds the mRNA Cap Binding Protein (CBP). A 5'-UTR may be provided as a flanking region to the mRNA. In some embodiments the polynucleotide sequence is codon-optimized. The polynucleotide may include a “polyA sequence” or “polyadenylation signal,” terms used interchangeably.

[0239] In some embodiments, the mRNA comprises between 100 bases and 5 kilobases (kb).

[0240] In some embodiments, the mRNA comprises at least 100 bases, at least 200 bases, at least 400 bases, at least 600 bases, or at least 800 bases.

[0241] In some embodiments, the mRNA comprises at least 1 kb, at least 2 kb, at least 3 kb, at least 4 kb, or at least 5 kb.

[0242] In some embodiments, the mRNA comprises between 100 bases and 5 kb, between 100 bases and 3 kb, between 100 bases and 2 kb, between 100 bases and 1 kb.

[0243] In some embodiments, the mRNA comprises between 1 kb and 5 kb, between 1 kb and 4 kb, or between 1 kb and 3 kb.

Signal sequence

[0244] In some embodiments, the encoded polypeptide includes a peptide region that is a signal peptide. Signal sequences are well known in the art. The signal peptide may be the

native signal sequence or can be replaced with another signal sequence. Signal sequences function to prompt a cell to translocate the protein, usually to the cellular membrane. The core of the signal sequence often contains a long stretch of hydrophobic amino acids (about 5–16 residues long) that has a tendency to form a single alpha-helix and is also referred to as the “h-region”. In addition, many signal sequences begin with a short positively charged stretch of amino acids, which may help to enforce proper topology of the polypeptide during translocation by what is known as the positive-inside rule. Because of its close location to the N-terminus it is called the “n-region”. At the end of the signal sequence there is typically a stretch of amino acids that is recognized and cleaved by signal peptidase.

[0245] In some embodiments, the polypeptide does not include a signal peptide.

[0246] Nonlimiting examples of signal peptides are provided in **Table 13**

Table 13.

Sequence	Length	SEQ ID NO:
MELLILKANAITTILTAVTFCFASG	25	104
MSWKVMIISLLITPQHGL	19	105
MPISILLIITTMIMASHC	18	106
MKAILVVLLYTFTTANA	17	107
MFVFLVLLPLVSSQC	15	108
MGILPSPGMPALLSLVLLSVLLMGCVAETGS	32	109
MGILPSPGMPALLSLVLLSVLLMGCVAETGT	32	110

Trimerization Domains

[0247] In some embodiments, the polypeptide may comprise a trimerization domain, such as FoldOn or a GCN4 trimerization. In some embodiments, the linker sequence comprises a FoldOn, wherein the FoldOn sequence is GYIPEAPRDGQAYVRKDGEWVLLSTFL (SEQ ID NO: 226).

[0248] In some embodiments, the polypeptide may comprise a trimerization domain, wherein the trimerization domain sequence is DKIEEILSKIYHIENEIARIKKLIGE (GEN) (SEQ ID NO: 227). In some embodiments, the polypeptide may comprise a trimerization domain, wherein the trimerization domain sequence is EKFHQIEKEFSEVEGRIQDLEK (HA) (SEQ ID NO: 228).

[0249] In some embodiments, the polypeptide may comprise a trimerization domain, wherein the trimerization domain sequence is EDKIEEILSKIYHIENEIARIKKLIGEA (coiled-coil isoleucine zipper) (SEQ ID NO: 229).

[0250] In some embodiments, the polypeptide may comprise a trimerization domain, wherein the trimerization domain sequence is GSGYIPEAPRDGQAYVRKDGEWVLLSTFL (bacteriophage T4 fibrin) (SEQ ID NO: 230).

[0251] In some embodiments, a trimerization sequence is RMKQIEDKIEEILSKIYHIENEIARIKKLIGEA (GCN4) (SEQ ID NO: 231). In some embodiments, a trimerization domain is a GCN4 variant. In some embodiments, the GCN4 variant sequence is RMKQIEDKIEEILSKIYHIENEIARIKKLIGERGGR (SEQ ID NO: 232), RMKQIEDKIEEILSKIYHIENEIARIKKLIGNRTGGR (SEQ ID NO: 233), RMKQIEDKIENTSKIYHIENEIARIKKLIGNRTGGR (SEQ ID NO: 234), RMKQIEDKIEEILSKIYNITNEIARIKKLIGNRTGGR (SEQ ID NO: 235), or RMKQIEDKIENTSKIYNITNEIARIKKLIGNRTGGR (SEQ ID NO: 236).

Pharmaceutical Compositions

[0252] In another aspect, the disclosure provides a pharmaceutical composition comprising the polypeptides, the protein nanostructures, or the polynucleotides disclosed above.

[0253] The pharmaceutical compositions described herein can be administered in combination with one or more additional therapeutic agents. The additional therapeutic agents may include, but are not limited to antibiotics or antibacterial agents, antiemetic agents, antifungal agents, anti-inflammatory agents, antiviral agents, immunomodulatory agents, cytokines, antidepressants, hormones, alkylating agents, antimetabolites, antitumour antibiotics, antimetabolic agents, topoisomerase inhibitors, cytostatic agents, anti-invasion agents, antiangiogenic agents, inhibitors of growth factor function inhibitors of viral replication, viral enzyme inhibitors, anticancer agents, α -interferons, β -interferon, ribavirin, hormones, and other toll-like receptor modulators, immunoglobulins (Igs), and antibodies modulating Ig function (such as anti-IgE (omalizumab)).

[0254] A pharmaceutical composition may be sterile and/or pyrogen free. Pharmaceutical compositions may be isotonic with respect to humans.

[0255] Pharmaceutical compositions may include one or more antimicrobial agents, particularly when packaged in multiple dose format. Antimicrobial agents include, but are not limited to, thiomersal and 2-phenoxyethanol. In some embodiments, it may be desirable to use either a mercury-free preservative or no preservative at all.

[0256] In some embodiments, the pharmaceutical composition is buffered, including, but not limited to a Tris buffer, a histidine buffer, a phosphate buffer, a citrate buffer or an acetate buffer. The pharmaceutical composition may also include a lyoprotectant, e.g., sucrose, sorbitol or trehalose. In some embodiments, the composition includes a preservative e.g., benzalkonium chloride, benzethonium, chlorohexidine, phenol, m-cresol, benzyl alcohol, methylparaben, propylparaben, chlorobutanol, o-cresol, p-cresol, chlorocresol, phenylmercuric nitrate, thimerosal, benzoic acid, and various mixtures thereof. In other embodiments, the pharmaceutical composition includes a bulking agent, like glycine. In yet other embodiments, the pharmaceutical composition includes a surfactant e.g., polysorbate-20, polysorbate-40, polysorbate- 60, polysorbate-65, polysorbate-80 polysorbate-85, poloxamer-188, sorbitan monolaurate, sorbitan monopalmitate, sorbitan monostearate, sorbitan monooleate, sorbitan trilaurate, sorbitan tristearate, sorbitan trioleate, or a combination thereof. The pharmaceutical composition may also include a tonicity adjusting agent, e.g., a compound that renders the formulation substantially isotonic or isoosmotic with human blood. Illustrative tonicity adjusting agents include sucrose, sorbitol, glycine, methionine, mannitol, dextrose, inositol, sodium chloride, arginine and arginine hydrochloride. In other embodiments, the pharmaceutical composition additionally includes a stabilizer, e.g., a molecule which substantially prevents or reduces chemical and/or physical instability of the protein nanostructure, in lyophilized or liquid form. Illustrative stabilizers include sucrose, sorbitol, glycine, inositol, sodium chloride, methionine, arginine, and arginine hydrochloride.

[0257] The pH of the composition is usually between about 4.5 to about 11, such as between about 5 to about 11, between about 5.5 to about 11, between about 6 to about 11, between about 5 to about 10.5, between about 5.5 to about 10.5, between about 6 to about 10.5, between about 5 to about 10, between about 5.5 to about 10, between about 6 to about 10, between about 5 to about 9.5, between about 5.5 to about 9.5, between about 6 to about 9.5, between about 5 to about 9, between about 5.5 to about 9, between about 6 to about 9, between about 5 to about 8.5, between about 5.5 to about 8.5, between about 6 to about 8.5, between about 5 to about 8, between about 5.5 to about 8, between about 6 to about 8, about 4.5, about 5, about 6.5, about 6, about 6.5, about 7, about 7.5, about 8, about 8.5, about 9, about 9.5, about 10, about 10.5, about 11, etc. Stable pH may be maintained by the use of a buffer e.g., a Tris buffer, a citrate buffer, phosphate buffer, or a histidine buffer. Thus a composition will generally include a buffer.

[0258] Provided herein are compositions for a medicament, e.g., for use in inducing or enhancing an immune response in a subject in need thereof, such as a mammal.

[0259] Provided herein are compositions for the manufacture of a medicament for inducing or enhancing an immune response in a subject in need thereof, such as a mammal.

Vaccines

[0260] In another aspect, the disclosure provides vaccines comprising the polypeptides, the protein nanostructures, or the polynucleotides disclosed above. Such compositions can be used to raise antibodies in a mammal (*e.g.*, a human). The vaccines compositions of the disclosure typically include a pharmaceutically acceptable carrier, and a thorough discussion of such carriers is available in Remington: The Science and Practice of Pharmacy.

[0261] Antigens could alternatively be used for vaccines that do not contain protein nanostructure with smaller oligomerization methods using FoldOn tag, wherein the FoldOn tag sequence is EKAAKAEEAARK (SEQ ID NO: 85).

[0262] Vaccine compositions may include an immunological adjuvant. Illustrative adjuvants include the following: 1. mineral-containing compositions; 2. oil emulsions; 3. saponin formulations; 4. virosomes and virus-like particles; 5. bacterial or microbial derivatives; 6. bioadhesives and mucoadhesives; 7. liposomes; 8. polyoxyethylene ether and polyoxyethylene ester formulations; 9. polyphosphazene (pcpp); 10. muramyl peptides; 11. imidazoquinolone compounds; 12. thiosemicarbazone compounds; 13. tryptanthrin compounds; 14. human immunomodulators; 15. lipopeptides; 16. benzonaphthyridines; 17. microparticles; 18. immunostimulatory polynucleotide (such as RNA or DNA; *e.g.*, cpg-containing oligonucleotides).

[0263] For example, the composition may include an aluminum salt adjuvant, an oil in water emulsion (*e.g.*, an oil-in-water emulsion comprising squalene, such as MF59 or AS03), a TLR7 agonist (such as imidazoquinoline or imiquimod), or a combination thereof. Suitable aluminum salts include hydroxides (*e.g.*, oxyhydroxides), phosphates (*e.g.*, hydroxyphosphates, orthophosphates), (*e.g.*, see chapters 8 & 9 of Vaccine Design. (1995) eds. Powell & Newman. ISBN: 030644867X. Plenum), or mixtures thereof. The salts can take any suitable form (*e.g.*, gel, crystalline, amorphous, etc.), with adsorption of antigen to the salt being an example. The concentration of Al⁺⁺⁺ in a composition for administration to a patient may be less than 5mg/ml *e.g.*, <4 mg/ml, <3 mg/ml, <2 mg/ml, <1 mg/ml, etc. A preferred range is between 0.3 and 1 mg/ml. A maximum of 0.85mg/dose is preferred.

Aluminum hydroxide and aluminum phosphate adjuvants are suitable for use with the disclosure.

[0264] Exemplary adjuvants that may be used in a pharmaceutical composition provided herein include, but are not limited to, 3M-052, Adju-PhosTM, AlhydrogelTM, AdjumerTM, albumin-heparin microparticles, Algal Glucan, Algammulin, Alum, Antigen Formulation, AS-2 adjuvant, ASO1, ASO3, autologous dendritic cells, autologous PBMC, AvridineTM, B7-2, BAK, BAY R1005, BECC TLR-4 agonists, Bupivacaine, Bupivacaine-HCl, BWZL, Calcitriol, Calcium Phosphate Gel, CCR5 peptides, CFA, Cholera holotoxin (CT) and Cholera toxin B subunit (CTB), Cholera toxin A1-subunit-Protein A D-fragment fusion protein, CpG, CPG-1018, CRL1005, Cytokine-containing Liposomes, D-Murapalmitine, DDA, DHEA, Diphtheria toxoid, DL-PGL, DMPC, DMPG, DOC/Alum Complex, Fowlpox, Freund's Complete Adjuvant, Gamma Inulin, Gerbu Adjuvant, GM-CSF, GMDP, hGM-CSF, hIL-12 (N222L), hTNF-alpha, IFA, IFN-gamma in pcDNA3, IL-12 DNA, IL-12 plasmid, IL-12/GMCSF plasmid (Sykes), IL-2 in pcDNA3, IL-2/Ig plasmid, IL-2/Ig protein, IL-4, IL-4 in pcDNA3, ImiquimodTM, ImmTherTM, Immunoliposomes Containing Antibodies to Costimulatory Molecules, Interferon-gamma, Interleukin-1 beta, Interleukin-12, Interleukin-2, Interleukin-7, ISCOM(s)TM, Iscoprep 7.0.3TM, Keyhole Limpet Hemocyanin, Lipid-based Adjuvant, Liposomes, Loxoribine, LT(R192G), LT-OA or LT Oral Adjuvant, LT-R192G, LTK63, LTK72, Matrix-MTM adjuvant, MF59, MONTANIDE ISA 51, MONTANIDE ISA 720, MPLTM., MPL-SE, MTP-PE, MTP-PE Liposomes, Murametide, Murapalmitine, NAGO, nCT native Cholera Toxin, Non-Ionic Surfactant Vesicles, non-toxic mutant E112K of Cholera Toxin mCT-E112K, p-Hydroxybenzoique acid methyl ester, pCIL-10, pCIL12, pCMVmCAT1, pCMVN, Peptomer-NP, Pleuran, PLG, PLGA, PGA, and PLA, Pluronic L121, PMMA, PODDSTM, Poly rA: Poly rU, Polysorbate 80, Protein Cochleates, QS-21, Quadri A saponin, Quil-A, Rehydragel HPA, Rehydragel LV, RIBI, Ribi like adjuvant system (MPL, TMD, CWS), S-28463, SAF-1, Sclavo peptide, Sendai Proteoliposomes, Sendai-containing Lipid Matrices, Span 85, Specol, Squalane 1, Squalene 2, Stearyl Tyrosine, SWE, Tetanus toxoid (TT), TheramideTM, Threonyl muramyl dipeptide (TMDP), Ty Particles, and Walter Reed Liposomes.

[0265] In some embodiments, the adjuvant is an aluminum hydroxide gel (e.g., AlhydrogelTM). In preferred embodiments, the adjuvant is SWE. In preferred embodiments, the adjuvant is MF59. MF59 is an oil-in-water emulsion containing squalene (4.3%) in citric acid buffer with stabilizing nonionic surfactants Tween 80 (0.5%) and Span 85 (0.5%). MF59 has been shown to be well-tolerated in humans and is used in vaccines against

seasonal influenza (see Ko and Kang, *Hum Vaccin Immunother.* 2018; 14(12): 3041–3045; U.S. Patent No. 6,299,884).

[0266] For example, the composition may include an aluminum salt adjuvant, an oil in water emulsion (e.g., an oil-in-water emulsion comprising squalene, such as MF59, SWE, or AS03), a TLR4 agonist, a TLR9 agonist (such as CpG oligodeoxynucleotides), a TLR7 agonist (such as imidazoquinoline or imiquimod), or a combination thereof. In some embodiments, the adjuvant is a combination of an aluminum salt and CPG1018. Suitable aluminum salts include hydroxides (e.g., oxyhydroxides), phosphates (e.g., hydroxyphosphates, orthophosphates), (e.g., see chapters 8 & 9 of *Vaccine Design*. (1995) eds. Powell & Newman. ISBN: 030644867X. Plenum), or mixtures thereof. The salts can take any suitable form (e.g., gel, crystalline, amorphous, etc.), with adsorption of antigen to the salt being an example. The concentration of Al⁺⁺⁺ in a composition for administration to a patient may be less than 5mg/ml e.g., <4 mg/ml, <3 mg/ml, <2 mg/ml, <1 mg/ml, etc. A preferred range is between 0.3 and 1 mg/ml. A maximum of 0.85mg/dose is preferred. Aluminum hydroxide and aluminum phosphate adjuvants are suitable for use with the disclosure. In some embodiments, a pharmaceutical composition provided herein comprises aluminum hydroxide as an adjuvant.

[0267] In some embodiments, the adjuvant is a squalene emulsion. In some embodiments, the adjuvant is a TLR4 immunostimulant (e.g., SLA, GLA), e.g., as described in Van Hoven et al. *PLoS One*. 11(2):e0149610 (2016). In some embodiments, the adjuvant is a TLR7/8 immunostimulant (e.g., R848, IMQ, 3M-052), e.g., as described in Dowling D. *ImmunoHorizons* (6):185-197 (2018). In some embodiments, the adjuvant is a TLR9 immunostimulant (CpG), e.g., as described in Bode et al. *Expert Rev Vaccines*. 10(4):499–511 (2011). In some embodiments, the adjuvant is saponin (QS21), e.g., as described in Zhu et al. *Nat Prod Chem Res*. 3(4):e113 (2016).

[0268] In some embodiments, the vaccine comprises a combination of two or more adjuvants (e.g., squalene emulsion and alum or a TLR4 immunostimulant).

[0269] A vaccine composition comprises an immunologically effective amount of its antigen(s). An “immunologically effective amount” or “effective amount” is an amount which, when administered to a subject, is effective for eliciting an antibody response against the antigen. This amount can vary depending upon the health and physical condition of the individual to be treated, their age, the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is

expected that the amount will fall in a relatively broad range that can be determined through routine trials. The antigen content of compositions of the disclosure will generally be expressed in terms of the mass of protein per dose. A dose of 10-500 μ g (e.g., 50 μ g) per antigen can be useful.

[0270] In some embodiments, the vaccine (immunogenic composition) is a stable emulsion.

[0271] In some embodiments, the vaccine is a bivalent vaccine. In some embodiments, the bivalent vaccine comprises the polypeptide or the protein nanostructure described herein. The bivalent vaccine can comprise polypeptides, protein nanostructures, or polynucleotides for two different strains of a Coronavirus, in any combination. For example, the bivalent vaccine could either contain a cocktail of two different immunogens that separately display each antigen, or “mosaic” immunogens that co-display both antigens. For example, the bivalent vaccine can include a polypeptide of the original virus strain to provide broad protection against COVID-19 and a component of the omicron variant to provide better protection against COVID-19 caused by the omicron variant.

Methods of Use

[0272] In another aspect, the disclosure provides a method of preventing or treating a coronavirus related disease in a subject in need thereof, comprises administering to the subject the polypeptides, the protein nanostructures, the polynucleotides, the pharmaceutical component, or the vaccine disclosed above.

[0273] In another aspect, the disclosure provides a method of immunizing a subject in need thereof against infection by coronavirus, by administering to the subject the vaccine disclosed above.

[0274] In some embodiments, such a method protects against the development of coronaviral-associated disease (e.g., severe disease), for example, COVID-19. The subject may be any suitable mammalian subject, including but not limited to a human subject. In some embodiments, a subject is a human child, e.g., a child of less than 12 months of age. In some embodiments, a subject is a human toddler, e.g., of about 1 to about 3 years of age or of about 1 to about 5 years of age. In some embodiments, the subject is a human adult of more than 60 years of age. In some embodiments, the subject is a human adult of more than 65 years of age. In particular embodiments, the subject is dependent on the help of others or with serious health concerns or risks (e.g., a frail elderly person). In some embodiments, the subject is a healthy adult of 18-60 years of age. In some embodiments, the subject is a healthy adult of 18-45 years of age. In another embodiment, the subject is a pregnant woman. In some embodiments the subject is an immunocompromised human adult. In some

embodiments, the subject is a human adult suffering from chronic underlying heart and/or lung disease or from functional disability. In some embodiments, the subject is at risk of severe viral disease (*e.g.*, LRTI or pneumonia).

[0275] The immunogenic compositions provided herein may be used to vaccinate an unborn child. The administration of certain inactivated vaccines is recommended during pregnancy to induce immunity in the unborn child, for example, the tetanus toxoid, reduced diphtheria toxoid and acellular pertussis (Tdap) vaccine and the influenza vaccine. Thus, in some embodiments, provided herein is a method of generating an immune response in an unborn child, the method comprising administering an effective amount of the immunogenic composition provided herein to the mother of said unborn child. The immunogenic composition may be administered at any suitable time point in the pregnancy, *e.g.*, in the last trimester of the pregnancy.

[0276] A composition provided herein may be co-administered with other treatments, such as other vaccines. Thus, in some embodiments, a subject treated in accordance with a method provided herein may also be administered one or more seasonal or pandemic vaccines such as an influenza vaccine. In some embodiments, a subject treated in accordance with the methods provided herein may also be administered a pneumococcal, Recombinant Zoster (Shingles), or Tdap vaccine. One, two, or more vaccines may be co-administered with an immunogenic composition provided herein. “Co-administration” includes both concurrently as well as subsequent administration. For example, the one, two, or more vaccines and an immunogenic composition provided herein may be administered on the same day. In some embodiments, the one, two, or more vaccines and an immunogenic composition provided herein are administered within 1 hour, within 2 hours, within 3 hours, within 4 hours, within 5 hours, within 6 hours, within 8 hours, within 10 hours, or within 12 hours.

[0277] In another aspect, provided herein is a method of treating a subject suffering from viral infection. As used herein, “treat” or “treating” includes, but is not limited to, accomplishing one or more of the following: (a) reducing viral titer in the subject; (b) limiting any increase of viral titer in the subject; (c) reducing the severity of viral infection symptoms; (d) limiting or preventing development of symptoms after viral infection; (e) inhibiting worsening of symptoms of viral infection; (f) limiting or preventing recurrence of symptoms of viral infection in subjects that were previously symptomatic for viral infection; and/or (g) increasing survival. In some embodiments, a method of vaccinating decreases the subject’s risk of becoming infected with a virus. In some embodiments, a

method of vaccinating limits the development of a viral infection. In some embodiments, a method of vaccinating decreases the severity of the symptoms of a viral infection. Symptoms of a Coronavirus infection include, but are not limited to, fever or chills, cough, shortness of breath or difficulty breathing, fatigue, muscle or body aches, headache, new loss of taste or smell, sore throat, congestion or runny nose, nausea or vomiting, and diarrhea.

[0278] In some embodiments, the methods provided herein may be used to prevent a viral infection or illness (*e.g.*, acute respiratory disease) in a subject. As used herein, “prevent” or “preventing” includes, but is not limited to accomplishing one or more of the following: (a) generating an immune response (antibody and/or cell-based, *e.g.*, CD4 T cells, memory B cells, and/or CD8 T cells) to a virus in the subject expected to confer protection against lower respiratory tract infection (LRTI) caused by or associated with a virus in the subject; (b) generating neutralizing antibodies against a virus in the subject expected to reduce the severity of LRTI caused by or associated with a Coronavirus in the subject; (c) preventing LRTI caused by or associated with a virus in a subject, detected as an increase in the titer of the virus in the subject or by an increase in one or more symptoms of viral infection; (d) preventing severe LRTI caused by or associated with Coronavirus in a subject detected as an increase in the titer of the virus of the subject or by an increase in one or more severe symptoms of viral infection; (e) reducing the risk of LRTI or severe LRTI caused by or associated with Coronavirus within a population of subjects; or (f) causing a seroresponse (or seroconversion) of a subject, such as generating neutralizing antibodies against a virus at least 4-fold higher than a baseline antibody level in the subject. Prevention may be assessed by comparing immune responses, especially correlates of protection, in subjects administered a vaccine to the same subjects before administration (termed baseline), subjects administered a placebo, or subjects administered a comparator vaccine.

[0279] As used herein, “limiting” the development a viral infection, refers to accomplishing one or more of the following: (a) generating an immune response (antibody and/or cell-based, *e.g.*, CD4 T cells, memory B cells, and/or CD8 T cells) to a virus in the subject expected to limit an increase in viral titer or symptoms in the subject; (b) generating neutralizing antibodies against a Coronavirus in the subject at a level expected to limit an increase in viral titer or symptoms in the subject; (c) causing reduced viral titers in the subject after exposure to a Coronavirus compared to subjects not administered the protein complex; and (d) caused reduced incidence or severity of symptoms after viral infection.

[0280] Clinical efficacy of a vaccine for Coronavirus infection can be assessed by various means known in the art, including but not limited to placebo-controlled clinical efficacy studies to measure viral load or symptoms of viral disease in vaccinated versus control subjects. Correlates of protection may also be defined, such as neutralizing antibody titers (typically expressed as geometric mean titers), fold increases above baselines (typically expressed as geometric fold rise), and seroresponse rate (a percentage of subjects that achieve a fold rise in neutralizing antibody titers above a predetermined threshold).

[0281] In some embodiments, the methods described herein generate an immune response in a subject not known to be infected with a virus, wherein the immune response serves to limit development of infection and symptoms of a viral infection. In some embodiments, the immune response comprises generation of neutralizing antibodies and/or cell-based responses against a virus. In some embodiments, the immune response comprises generation of Spike glycoprotein protein-specific responses with a mean geometric titer of at least 1×10^3 , at least 1×10^4 , at least 1×10^5 , at least 1×10^6 , at least 1×10^7 , at least 1×10^8 , or at least 1×10^9 . In a further embodiment, the immune response comprises generation of antibodies against multiple antigenic epitopes on the Spike glycoprotein protein-specific protein trimer.

[0282] In one aspect, the methods provided herein may results in an increase in antibody titers in a subject, *e.g.*, in an increase in virus-specific neutralizing antibodies or virus-specific binding antibodies. Antibody titers may be determined using any suitable assays known in the art or described herein including, without limitation, binding enzyme-linked immunosorbent assays (ELISA), enzyme-linked immune absorbent spot (ELISpot), competition ELISAs, immunoprecipitation, immunoblotting, and agglutination assays.

[0283] In some embodiments, a neutralization or microneutralization (MN) assay may be used to measure increases in neutralizing antibodies in a subject after administration of a composition as described herein. Microneutralization refers to a neutralizing performed in a miniaturized format, such as a 96-well plate. A (micro)neutralization assay is used to test for the inhibition of a virus by antibodies (*e.g.*, purified antibodies, serum, or plasma). The assay measures the level of antibodies present in a sample that are able to neutralize a virus *in vitro*. Generally, microneutralization assays for clinical samples are performed with a serial dilution of serum mixed with a fixed concentration of virus. Methods for performing (micro)neutralization assays are well known. Illustrative microneutralization assays are described, see, *e.g.*, van Baalen et al. Vaccine 35 (2017) 46–52.

[0284] In some embodiments, the methods provided herein may result in an increase in immune cells in a subject (*e.g.*, an increase in virus-specific memory B cells and/or virus-specific T cells). The number of immune cells in a subject may be determined using any suitable assay known in the art or described herein, including, without limitation, FACS and flow cytometry.

[0285] In some embodiments, the composition is administered via any suitable route, including intranasally, sublingually, orally, parentally, by inhalation spray, rectally, or topically in dosage unit formulations containing conventional pharmaceutically acceptable carriers, adjuvants, and vehicles. The term parenteral as used herein includes, subcutaneous, intravenous, intra-arterial, intramuscular, intrasternal, intratendinous, intraspinal, intracranial, intrathoracic, infusion techniques or intraperitoneally.

Kits

[0286] In another aspect, the disclosure provides kits, which may be used to prepare and administer polypeptides, protein nanostructures, polynucleotides, pharmaceutical compositions, or vaccines as described above. In some embodiments, a kit provided herein comprises polypeptides, protein nanostructures, polynucleotides, pharmaceutical compositions, or vaccines as disclosed herein, and instructions for use in a methods as described above. In some embodiments, a kit comprises one or more unit doses as disclosed herein, and instructions for use in a method of the disclosure. In some embodiments, the kit comprises a vial comprising a single dose of a pharmaceutical composition or vaccine provided herein. In some embodiments, a kit comprises a vial comprising multiple doses provided herein. In some embodiments, a kit further comprises instructions for use of the pharmaceutical composition or vaccine. In some embodiments, a kit further comprises a diluent for preparing dilutions of the pharmaceutical composition or vaccine prior to administration. In some embodiments, the vaccine comprises an adjuvant. In some embodiments, a kit comprises a vaccine and an adjuvant which must be mixed prior to administration.

EXAMPLES

Materials and Methods

[0287] **Structural modeling.** PyMol version 2.5.2 or ChimeraX version 1.3 was used to analyze structural models and generate images. Structure-based prediction using Colabfold (Ref 11) was performed with default settings against the RDB-SD1 region of each designed sequence.

[0288] HEK293 expression. One day prior to transfections, Expi293F cells (Thermo Fisher Scientific) were counted and seeded at 2.5×10^6 cells/mL and subsequently incubated overnight at 36 °C with shaking at 120rpm. The next day the cells were counted and diluted to 3×10^6 cells/mL. Cells were then transiently transfected with the desired plasmid according to manufacturer directions (Thermo Fisher Scientific), with the exception that Transporter 5 (Polysciences) was used as the transfection reagent at a ratio of 3.0 ug of Transporter 5 per 1 ug of plasmid DNA. Transfected cells were then incubated at 36 °C for 4 days while shaking at 120rpm.

[0289] Western blots. Western Blots were performed against HEK293 cell supernatants after small-scale expression using Anti-His HRP conjugated antibody. Supernatant was diluted 2-fold in 2x Laemmli loading buffer (Bio-Rad) with 5% (v/v) β -mercaptoethanol, then heated to 95°C for 10 minutes. 10 μ L was loaded onto a NuPAGE 4-12% Bis-Tris protein gel along with 7 μ L of His-tagged protein standard/ladder (Invitrogen, #LC5606), and ran for 30 minutes at 200V. Proteins were transferred onto an Immuno-Blot PDVF membrane (Bio-Rad). Following transfer, the blot was blocked with 2% milk (w/v) in 1x PBS, 0.05% Tween 20 for 1 hour at room temperature with rocking. Anti-His HRP conjugated antibody (R&D systems MAB050H) was diluted 1:7000 in 2% milk (w/v) in 1x PBS, 0.05% Tween 20 and incubated with the membrane for 1 hour at room temperature with rocking. The membrane was then washed three times for five minutes using 1x PBS, 0.05% Tween 20, and His-tagged protein was detected using (BioRx TMB Enhanced One Component HRP Membrane Substrate, Surmodics) and captured on BioRad ChemiDoc MP imaging system.

[0290] IMAC purification. NA protein samples were purified directly from HEK Expi293F cell supernatant by immobilized metal affinity chromatography (IMAC). Ni²⁺-Agarose resin (Cube Biotech, Indigo) was washed with 5 column volumes (CV) of water followed by 5 CV of equilibration buffer (20 mM Tris pH 8.0, 250 mM NaCl, 5% glycerol, 30 mM imidazole), then resuspended in 1 CV of equilibration buffer. Supernatants were clarified by centrifugation at 4,000 x g followed by filtration using 0.2 mm vacuum filtration units. Resin suspension was added to each supernatant at a ratio of 4 μ L of resin per mL of supernatant. The supernatant-resin slurries were incubated with gentle rocking at 4 °C for ~16-18 hours, then the resin was collected by application to empty gravity columns (Biorad, 7321010). The resin beds were washed with 20 CV equilibration buffer, and proteins were eluted using 6 CV elution buffer (20 mM Tris pH 8.0, 250 mM NaCl, 5%

glycerol, 300 mM imidazole) in three 2 CV increments with 30-minute incubations between each application. After IMAC purification, protein samples were either further purified by size exclusion chromatography (SEC) or dialyzed into “RBD buffer 1” (20 mM Tris pH 8.0, 150 mM NaCl), “RBD buffer 2” (5% glycerol, 20 mM Tris pH 8.0, 250 mM NaCl, 5% glycerol), or “RBD buffer 3” (50 mM MOPS, 150 mM NaCl, pH 7.4).

[0291] SEC purification. Each IMAC elution for antigen-CompA samples was injected onto either a Superdex 200 Increase 10/300 GL column (Cytiva) equilibrated in RBD buffer 1, RBD buffer 2, or RBD buffer 3 and eluted with 1.2 column volumes of the respective running buffer to further purify the antigens by size exclusion chromatography (SEC). For assembled protein nanostructure preparations, samples were applied to a Superose 6 Increase 10/300 GL column (Cytiva) equilibrated in 20 mM Tris Ph 8.0, 150 mM NaCl, 100 mM L-Arginine and 5% glycerol and eluted with 1.2 column volumes of the same buffer. Following analysis, purified samples were flash frozen in liquid nitrogen and storage at -80 °C.

[0292] Biolayer interferometry. ACE2-Fc, CR3022 (Ref 12) or S309 (Ref 16) normalized in concentration to 10 ug/mL in BLI assay buffer (PBS, 0.5% BSA, 0.05% Tween 20, pH 7.4) in a sufficient volume to load 200 µL per well of a black 96-well microplate (Greiner, 655209). Briefly, on an Octet Red96e instrument, pre-hydrated Protein G biosensors (Sartorius, 18-5022) were dipped into assay buffer for 60 seconds to achieve a baseline. Next, the biosensors were dipped into ACE2-Fc, CR3022 or S309 for 60 seconds or 120 seconds respectively for immobilization, followed by an additional baseline step. The immobilized antibodies allowed to associate with 200 µL of purified antigen-CompA fusions (at 10 ug/mL in BLI assay buffer) or cell supernatants for 120 to 700 seconds, and then the biosensors were dipped back into assay buffer for 120 to 700 seconds to observe any possible dissociation. Maximum shift values during the association phase are reported unless otherwise noted.

[0293] NanoDSF using ITF. Protein thermal stability and fluorescent baseline measurements were determined using intrinsic protein fluorescence (ITF) on a nanoDSF instrument (UNchained Laboratories, UNcle). Antigen concentrations were normalized to 0.5-1mg/mL. Each sample was applied to a quartz capillary cassette (UNchained Laboratories, UNi) for ITF measurements over a temperature ramp (15 – 95 °C, 0.5-1 °C/min) after a 5-minute equilibration at 15 °C. Melting temperatures were determined for representative single samples from barycentric mean calculations using Uncle Analysis V6.0 software with default settings, or with a custom python script. The python script

calculated melting temperature from the second derivative of the temperature-dependent barycentric mean. The onset temperature was calculated from the intersection of a linear fit of the pre-transition baseline and the melting transition.

[0294] NanoDSF using SYPRO. Thermal melt profiles NA constructs were determined by normalizing sample concentrations to 0.5-1 mg/mL. SYPRO™ Orange Protein Gel Stain (Invitrogen S6651) was then diluted 1:20 in each sample buffer and 3.5 µL was added to 31.5 µL of normalized NA constructs. Each sample was applied to a quartz capillary cassette (UNchained Laboratories, UNi) for ITF measurements over a temperature ramp (15 – 95 °C, 0.5-1 °C/min) after a 5-minute equilibration at 15 °C. Melting temperatures were determined using fluorescent area measurements from representative single samples using Uncle Analysis V6.0 software with default settings. Baseline fluorescence near 30 °C was measured from fluorescent intensity integrations based on three replicates by averaging points that were collected at temperatures closest to 30 °C for each replicate.

[0295] Assembly of antigen-CompA samples with CompB. For any protein nanostructure assembly process, purified CompB (I53-50B.4PosT1) was mixed with an antigen-CompA sample by pipetting, with a molar ratio of 1.1 parts antigen-CompA to 1 part CompB. For small-scale assembly reactions, the assembly process was sometimes performed with addition of other concentrated buffer solutions to add excipients to the final solution, including final concentrations of 50-100 mM L-Arginine, 0.75% CHAPS, 4% sucrose or 5% glycerol. Excipients used in each case are detailed in subsequent tables. For scale-up assemblies prior to SEC, the assembly process was performed in a background of 20 mM Tris Ph 8.0, 150 mM NaCl, 100 mM L-Arginine and 5% glycerol. All assembly processes allowed to sit at room temperature for at least 30 minutes prior to SEC or analysis by dynamic light scattering (DLS) using default protocols on a DLS instrument (UNchained Laboratories, UNcle).

[0296] Animal immunizations. Female BALB/c mice were immunized intramuscularly on days 0 or 21 with 0.027-1.0 ug of purified antigen-CompA or antigen-VLP protein mixed 1:1 with Addavax™ adjuvant. Serum samples were collected on day 0 (prior to first immunization), day 21 (prior to second immunization), and day 35. Antigen binding and neutralization assays were performed at Nexelis against respective antigens and pseudoviruses for ancestral and BA.5 strains.

Example 1: Design of SD1resurf mutations and their increased function towards nanostructure assembly

[0297] In a first round of design process, initial features were tested using RBD antigens from the ancestral SARS-CoV-2 strain. All tested antigens were genetically fused to the I53-50A “CompA” trimer, which is a protein-based component that can be assembled *in vitro* into a protein nanostructure when a complementary “CompB” pentamer (in this case, I53-50B.4PosT1) is added (**Ref 2**). Multiple previously described stabilizing mutations (Rpk2, Rpk4, Rpk8, Rpk9, Rpk11 and Rpk15) (**Ref 5**) were tested separately in the context of this RBD-SD1 antigen (**Table 14**). A novel set of mutations to the SD1 domain were designed (named “SD1resurf”) that targeted exposed hydrophobic groups on the SD1 domain, which made these groups more polar to reduce aggregation (L560Q, F562Y, I569S). L560Q and I569S were designed based on other sarbecovirus SD1 sequences containing these polar residues at these positions, while F562Y was designed rationally to make this residue more polar. All RBD-SD1 designs were tested individually with one of five different flexible linkers between the RBD-SD1 antigen and CompA (**Table 15**), which ranged from 16 to 24 amino acids in length. Codon-optimized DNA sequences were generated for all designs with an added secretion signal sequence, which were then cloned into a vector for expression from HEK293 cells using transient transfections.

Table 14. Mutations tested in the first round of designs.

Mutation set name	Mutations
Rpk2	F338L, Y365W
Rpk4	F392W
Rpk8	Y365F, V395I
Rpk9	Y365F, F392W, V395I
Rpk11	F338L, A363L, Y365M
Rpk15	I358F, Y365F, V395I
SD1resurf	L560Q, F562Y, I569S

Table 15. Linkers tested between RBD-SD1 antigens and CompA

Linker length (amino acid residues)	Amino acid sequence
16	GGSGGSGSGGSGGSGS

18	SGGGSGGSGSGGSGGSGS
19	EPEGSGGSGSGGSGGSGS
20	YGGSGGSGGSGSGGSGGSGS
24	GSGGSGGSGGSGGSGGSGGSGS

[0298] Designs were first evaluated from supernatants at small-scale, both for expression using Western blots (**FIG. 2A**), and antigenicity via Bio-Layer Interferometry (BLI) using immobilized ACE2-Fc receptors (**FIG. 2B** and **FIG. 2C**, **Tables 16 and 17**) after downselection. Surprisingly, the majority of samples tested showed higher ACE2-binding signal compared to the “RBD01” control, which featured the RBD fragment fused to CompA (**Ref 14**), with these data signifying improved expression and/or antigenicity. Higher levels of binding were seen to be influenced independently by addition of stabilizing mutations, including multiple Rpk mutations and SD1resurf mutations, and by use of 18- or 20-residue linkers. Based on these data, multiple designs containing the 18-residue linker were scaled-up via HEK293 transfection and purified by immobilized metal ion affinity chromatography (IMAC) for more detailed characterization, and compared against a wild-type RBD antigen without SD1 domain (residues 328-531) genetically fused to CompA. BLI against immobilized ACE2-Fc or the S309 neutralizing monoclonal antibody (**Ref 12**) or the CR3022 non-neutralizing monoclonal antibody (**Ref 16**) was performed, with ACE2-Fc and S309 results used for down selection given their binding to epitopes targeted by neutralizing antibodies. BLI against ACE2-Fc and S309 showed generally similar results (**FIG. 3**, **Table 18**), with some constructs retaining the native isoleucine (I) at position 358 appearing superior by S309 measurements. Nano differential scanning fluorimetry (NanoDSF) based on intrinsic tryptophan fluorescence showed improved melting temperatures from designs containing Rpk4, Rpk8, Rpk9, and Rpk15 stabilizing mutations relative to wild-type antigens with and without SD1 domains (**FIG. 4A to FIG. 4G**, **Table 19**). Melting temperatures were similar for wildtype RBD antigens with and without SD1 domains (**FIG. 4A to FIG. 4B**, **Table 19**), and for the design containing SD1resurf mutations (**FIG. 4G**, **Table 19**). Similar results were seen for NanoDSF as monitored using SYPRO dye, although Rpk4 and Rpk9 designs showed notably lower baselines in comparison to other designs, suggesting an improvement in foldedness that reduces hydrophobicity of these designs (**FIG. 4H to FIG. 4K**, **Table 20**). Finally, assembly was

tested for different designs in diverse excipient conditions through addition of the complementary CompB pentameric component, and monitoring of particulate size using dynamic light scattering (DLS) in ancestral strain (**FIG. 5A and FIG. 5B, Table 21**). When assembled in buffers with more modest excipients (arginine (**FIG. 5B, Table 21**), or arginine and sucrose (**FIG. 5A, Table 21**), significant aggregation was seen for all designs except those containing Rpk9 stabilizing mutations and SD1resurf mutations, which provided DLS signals consistent with successful protein nanostructure assembly (35-45 nm).

[0299] In all, this first round of design demonstrated that the previously-described Rpk9 mutations can successfully improve the thermal stability of RBD-SD1 antigens from the ancestral strain, and further improve their ability to assemble into I53-50 protein nanostructures with minimal aggregation in simpler buffer formulations. Surprisingly, novel SD1resurf mutations were identified which assisted assembly of I53-50 protein nanostructures with minimal aggregation. Both sets of mutations showed at least antigenicity of key epitopes targeted by neutralizing antibodies.

[0300] **Table 16** shows a first round of BLI using immobilized ACE2-Fc against supernatants of ancestral designs expressed at small-scale.

Table: 16

Constructs	Maximum shift (nm)
RBDb001	0.7794
RBDb002	0.8548
RBDb003	0.8513
RBDb004	0.85
RBDb005	0.8797
RBDb006	0.8115
RBDb011	0.9095
RBDb016	1.2278
RBDb021	0.8613
RBDb026	0.783
RBDb031	0.8718
RBDb036	0.8456
RBDb041	0.8785
RBD01 replicate 1	0.7766
RBD01 replicate 2	0.7968

RBD01 replicate 3	0.7721
-------------------	--------

[0301] Table 17 shows a second round of BLI using immobilized ACE2-Fc against supernatants of ancestral designs expressed at small-scale.

Table 17:

Constructs	Maximum shift (nm)
RBDb016	1.1147
RBDb017	0.8861
RBDb018	0.8502
RBDb019	0.8952
RBDb020	0.859
RBDb021	0.833
RBDb022	0.8914
RBDb023	0.8497
RBDb024	0.8868
RBDb025	0.8685
RBDb036	0.8031
RBDb037	0.8782
RBDb038	0.8251
RBDb039	0.8861
RBDb040	0.867
RBD01 replicate 1	0.7022
RBD01 replicate 2	0.7005
RBD01 replicate 3	0.6929

[0302] Table 18 shows BLI against purified antigen-CompA proteins using immobilized ACE2-Fc, CR3022 or S309.

Table 18:

Constructs	ACE2, maximum shift (nm)	CR3022, maximum shift (nm)	S309, maximum shift (nm)
RBDb001 replicate 1	0.3426	1.0786	0.8241
RBDb001 replicate 2	0.3416	1.0706	0.8127
RBDb002	0.3267	0.5924	0.7878
RBDb012	0.3355	1.4242	0.8286

RBDb016	0.3434	1.3782	0.7426
RBDb017	0.333	0.5919	0.696
RBDb022	0.329	1.3828	0.7234
RBDb032	0.3376	1.343	0.5848
RBDb037	0.3267	1.3211	0.7265
RBD01 replicate 1	0.3154	0.8185	0.7775
RBD01 replicate 2	0.3153	0.8147	0.7691
Buffer (negative control)	0.0098	-0.0542	0.0281

[0303] Table 19. NanoDSF analysis of purified ancestral-based antigen-CompA samples using intrinsic tryptophan fluorescence. Each measurement reported as an average of three replicates.

Table 19:

Constructs	Average melting temperature (°C)
RBD01	49.2
RBDb001	49.2
RBDb002	48.8
RBDb012	51.3
RBDb017	50.9
RBDb022	52.2
RBDb032	51.3
RBDb037	48.7

[0304] Table 20. NanoDSF analysis of purified ancestral-based antigen-CompA samples using SYPRO dye. Each measurement reported as an average of three replicates.

Table 20:

Constructs	Average melting temperature (°C)	Average area of SYPRO fluorescence near 30°C (Counts × nm)
RBDb002	44.9	1508685.71
RBDb012	46.6	824854.98

RBDb017	47.4	1353775.56
RBDb022	48.8	902591.02
RBDb032	49.1	1536223.86

[0305] **Table 21** shows DLS results for assembly of purified antigen-CompA samples with CompB.

Table 21:

Constructs	Buffer formulation	Z-Ave. Dia. (nm)	PDI	DLS interpretation
RBDb001	50 mM MOPS, 0.75 % CHAPS, 150 mM NaCl, pH 7.4	35.70	0.112	Assembly with low polydispersity
RBDb001	50 mM MOPS, 150 mM NaCl, 4 % Sucrose, pH 7.4	>1000	0.283	
RBDb001	50 mM MOPS, 150 mM NaCl, pH 7.4	>1000	0.427	
RBDb001	50 mM Arginine, 150 mM NaCl, 4 % Sucrose, pH 7.4	>1000	0.324	
RBDb001	50 mM Arginine, 150 mM NaCl, pH 7.4	>1000	0.246	
RBDb002	50 mM MOPS, 0.75 % CHAPS, 150 mM NaCl, pH 7.4	36.48	0.120	Assembly with low polydispersity
RBDb002	50 mM MOPS, 150 mM NaCl, 4 % Sucrose, pH 7.4	479.98	0.158	
RBDb002	50 mM MOPS, 150 mM NaCl, pH 7.4	>1000	0.541	
RBDb002	50 mM MOPS, 50 mM Arginine, 150 mM NaCl, 4 % Sucrose, pH 7.4	318.83	0.176	

RBDb002	50 mM MOPS, 50 mM Arginine, 150 mM NaCl, pH 7.4	>1000	0.361	
RBDb017	50 mM MOPS, 0.75 % CHAPS, 150 mM NaCl, pH 7.4	141.95	0.111	
RBDb017	50 mM MOPS, 150 mM NaCl, 4 % Sucrose, pH 7.4	>1000	0.364	
RBDb017	50 mM MOPS, 150 mM NaCl, pH 7.4	>1000	0.340	
RBDb017	50 mM MOPS, 50 mM Arginine, 150 mM NaCl, 4 % Sucrose, pH 7.4	83.35	0.103	
RBDb017	50 mM MOPS, 50 mM Arginine, 150 mM NaCl, pH 7.4	>1000	0.320	
RBDb022	50 mM MOPS, 0.75 % CHAPS, 150 mM NaCl, pH 7.4	36.59	0.100	Assembly with low polydispersity
RBDb022	50 mM MOPS, 150 mM NaCl, 4 % Sucrose, pH 7.4	>1000	0.235	
RBDb022	50 mM MOPS, 150 mM NaCl, pH 7.4	42.93	0.151	Assembly with low polydispersity
RBDb022	50 mM MOPS, 50 mM Arginine, 150 mM NaCl, 4 % Sucrose, pH 7.4	38.84	0.107	Assembly with low polydispersity
RBDb022	50 mM MOPS, 50 mM Arginine, 150 mM NaCl, pH 7.4	39.23	0.142	Assembly with low polydispersity

RBDb037	50 mM MOPS, 0.75 % CHAPS, 150 mM NaCl, pH 7.4	37.31	0.099	Assembly with low polydispersity
RBDb037	50 mM MOPS, 150 mM NaCl, 4 % Sucrose, pH 7.4	317.80	0.106	
RBDb037	50 mM MOPS, 150 mM NaCl, pH 7.4	33.91	0.199	Assembly with low polydispersity
RBDb037	50 mM MOPS, 50 mM Arginine, 150 mM NaCl, 4 % Sucrose, pH 7.4	38.68	0.067	Assembly with low polydispersity
RBDb037	50 mM MOPS, 50 mM Arginine, 150 mM NaCl, pH 7.4	39.56	0.156	Assembly with low polydispersity
RBD01	50 mM MOPS, 0.75 % CHAPS, 150 mM NaCl, pH 7.4	33.57	0.151	Assembly with low polydispersity
RBD01	50 mM MOPS, 150 mM NaCl, 4 % Sucrose, pH 7.4	>1000	0.198	
RBD01	50 mM MOPS, 150 mM NaCl, pH 7.4	>1000	0.429	
RBD01	50 mM MOPS, 50 mM Arginine, 150 mM NaCl, 4 % Sucrose, pH 7.4	>1000	0.398	
RBD01	50 mM MOPS, 50 mM Arginine, 150 mM NaCl, pH 7.4	>1000	0.309	
RBD01	Tris, 20 mM Tris, 0.75 % CHAPS, 5 % Glycerol, 250 mM NaCl, pH 8	32.50	0.164	Assembly with low polydispersity

RBD01	Tris, 20 mM Tris, 5 % Glycerol, 250 mM NaCl, 4 % Sucrose, pH 8	>1000	0.437	
RBD01	Tris, 20 mM Tris, 5 % Glycerol, 250 mM NaCl, pH 8	607.76	0.486	
RBD01	Tris, 20 mM Tris, 50 mM Arginine, 5 % Glycerol, 250 mM NaCl, 4 % Sucrose, pH 8	40.55	0.290	Assembly with low polydispersity
RBD01	Tris, 20 mM Tris, 50 mM Arginine, 5 % Glycerol, 250 mM NaCl, pH 8	180.49	0.185	

Example 2: Design of additional stabilizing mutations based on structural analysis of highly conserved regions of the RBD-SD1 antigen

[0306] Constructs containing RBD-SD1 antigen fragments, Rpk9 mutations, SD1resurf mutations and the 18-residue linker, were generated and these combinations were tested with both the ancestral strain and BA.5, an Omicron-derived variant. Purified protein was tested for assembly with CompB, which worked as expected for the ancestral antigen. (**Fig. 6, Table 22**) (**Refs 3, 13**).

[0307] Based on the structural location of added stabilizing mutations and the close proximity of F392W to the RBD-SD1 interface, designs that built on the combination of Rpk8 mutations (same as Rpk9 but lacking F392W) and SD1resurf mutations, while adding further stabilizing mutations to improve stability and assembly, were created. RBD sequences that are less thermally stable, such as Omicron variants, were the basis for other designs. Such lowered stability provides a liability for reliable manufacturing of RBD antigens and derived protein nanostructure-based vaccines, and allows for a more stringent test of novel stabilizing mutations. Due to the global relevance of the Omicron-derived BA.5 variant and its representation in bivalent mRNA vaccines in 2022, BA.5 was selected for testing of novel stabilizing mutations in RBD-SD1 antigens.

[0308] To best select for stabilizing mutations which could maximally assist RBD-SD1 antigens from diverse strains of SARS-CoV-2 or other sarbecoviruses, structural data was analyzed for highly conserved regions of the RBD-SD1 antigen that appeared to lack local stability, such as structural cavities or other non-ideal features. The interface between the RBD and SD1 was highlighted as a region that appeared locally unstable (**FIG. 1C**). Multiple cavities and poorly-formed polar groups were observed between the two domains, and a loop region on the RBD from residues 517-522 appeared to be held in tension against the SD1 domain, and further contained exposed hydrophobic groups. Structural analysis of Spike glycoprotein structures from multiple SARS-CoV-2 variants and from SARS-CoV-1 showed both structural and residue-level conservation of this interface region (PDBs 6VXX, 7UB5, 5XLR). To further confirm this conservation, sequences from diverse representative SARS-CoV-2 variants and sarbecoviruses were aligned and analyzed. The alignment showed strong conservation of residues at this interface (**FIG. 7**). Further, no characterized potent neutralizing antibodies could be identified that targeted any residues within this region, which made it more amenable to mutation without impacting key antigenicities. Based on modeling in PyMol, Rosetta (**Ref 10**) and Colabfold (**Ref 11**), multiple sets of mutations were designed that aimed to reinforce the stability of the RBD-SD1 interface. Some mutation sets aimed to strengthen interactions between the RBD and the SD1 domain by improved packing and/or hydrogen bonding, or addition of a disulfide bond. Others mutated the 517-522 loop to make hydrophobic groups more polar and/or remove tension in the loop, such as by addition of glycine mutations or removal of native prolines. Select combinations of different design strategies were also evaluated. As well, many designs were also tested in other parts of the RBD that were proximal to the RBM (**FIG. 1C**), which aimed to improve stability by improving other noncovalent interactions. DNA sequences were generated for all designs as genetic fusions to CompA using an 18-residue linker. As mentioned previously, Rpk8 mutations were added to all designs to further support stability, and SD1resurf mutations were also added to all designs to assist nanostructure assembly.

[0309] Table 22 shows DLS results for assembly of purified antigen-CompA samples with CompB.

Table 22:

Constructs	Buffer formulation	Z-Ave. Dia. (nm)	PDI	DLS interpretation
RBDb092	20 mM Tris, 100 mM Arginine, 5 % Glycerol, 150 mM NaCl, pH 7.4	38.17	0.083	Assembly with low polydispersity
RBDb093	20 mM Tris, 100 mM Arginine, 5 % Glycerol, 150 mM NaCl, pH 7.4	548.60	0.179	

Example 3: Evaluation of stabilizing mutations and their increased function on antigenicity and thermal stability

[0310] Designs were downselected from small-scale transfections in HEK293 cells, first by Western blot (**FIG. 8**), which removed some designs with RBM-proximal mutations based on poor secretion. Well-expressing designs were further downselected based on BLI signals from supernatants against ACE2-Fc (**Table 23**), to select for designs that were high expressing and/or have improved antigenicity. Multiple designs with higher BLI signals were selected for scale-up and purification, and were compared to a wild-type antigen sequence lacking Rpk8 and SD1resurf mutations. NanoDSF of designs using intrinsic Tryptophan fluorescence showed varied levels of improvements to thermal stability (**FIG. 9A, Table 24**), with improved melting and onset temperatures of up to 6-7°C relative to RBDb046 (wild-type design). The wild-type design showed respective melting and onset temperatures of approximately 37°C and 40.5°C. Such low thermal stability provides risks for manufacturing due to elevated temperatures and associated stress, with stabilizing mutations directly providing a way to mitigate such risks. Surprisingly, BLI against immobilized ACE2-Fc receptors showed improved binding signals from stabilized designs relative to the wild-type design (**FIG. 9B and FIG. 9C, Table 24**), particularly in terms of binding rates. All added stabilizing mutations were structurally distant from the ACE2-binding site on the RBD, which suggests that antigenic improvements are provided through improved foldedness of the RBD-SD1 antigens. Finally, a subset of these designs were found to reduce aggregation during assembly of I53-50 protein nanostructures after addition of CompB in the presence of minor excipients (**FIG. 10, Table 24**), with the wild-type design showing DLS signals that were not consistent with proper protein nanostructure assembly.

[0311] **Table 23** shows BLI binding shifts for supernatant screen of ACE2 binding for BA.5 designs.

Table 23:

Construct	Shift	Construct	Shift
RBDb001	0.229		
RBDb046	0.2865	RBDb074	0.3063
RBDb047	0.3069	RBDb075	0.3179
RBDb048	0.4857	RBDb076	0.3146
RBDb049	0.3398	RBDb077	0.3302
RBDb050	0.3056	RBDb078	0.3689
RBDb051	0.3252	RBDb079	0.2788
RBDb052	0.331	RBDb080	0.3165
RBDb062	0.3073	RBDb081	0.3815
RBDb063	0.3326	RBDb082	0.2954
RBDb065	0.326	RBDb083	0.2663
RBDb066	0.3235	RBDb084	0.3153
RBDb067	0.2971	RBDb085	0.2805
RBDb068	0.2964	RBDb086	0.2606
RBDb069	0.2858	RBDb087	0.3044
RBDb070	0.2968	RBDb088	0.3044
RBDb071	0.2932	RBDb089	0.2712
RBDb072	0.3163	RBDb090	0.349
RBDb073	0.3306	RBDb091	0.347

[0312] **Table 24** provides a summary of *in vitro* analyses for purified BA.5-based antigen-CompA samples. Onset and melting temperatures were averages of three replicates, with values calculated based on a customized analysis method described in Materials and Methods. “N.C.” means “not collected” due to subsequently labelled constructs having other inferior properties. Desired Z-Ave.d values range from 30-45 nm, with larger values suggesting aggregation during nanostructure assembly.

Table 24:

Construct	Onset temperature (°C)	Melting temperature (°C)	Yield (g/L)	k_{eq}	Maximum Shift (nm)	Z-Ave.d (nm)	PDI
RBDb046	36.9	40.5	0.066	0.032	0.271	58.3	0.164
RBDb047	38.9	43.2	0.042	0.106	0.277	32.67	0.25
RBDb052	35.9	40.1	0.099	0.043	0.282	N.C.	N.C.
RBDb063	39.4	42.5	0.092	0.047	0.279	N.C.	N.C.
RBDb066	39.7	43.7	0.049	0.097	0.282	36.33	0.183
RBDb072	39.4	42.9	0.076	0.044	0.280	N.C.	N.C.
RBDb073	40.9	44.9	0.077	0.089	0.288	38.16	0.161
RBDb075	43.1	47.3	0.089	0.067	0.270	36.67	0.157
RBDb076	40.5	44.2	0.072	0.054	0.288	38.53	0.15
RBDb078	41.9	46.1	0.114	0.109	0.287	>1000	0.113
RBDb081	43.8	47.7	0.093	0.116	0.293	>1000	0.252
RBDb084	40.3	43.8	0.095	0.023	0.293	N.C.	N.C.
RBDb087	42.2	46.2	0.091	0.065	0.282	40.65	0.101
RBDb090	42.5	46.2	0.239	0.038	0.271	43.6	0.135

Example 4: Immunogenicity of selected stabilizing mutations

[0313] Immunogenicity was next tested for a small set of assembled protein nanostructures and a larger set of trimeric CompAs. For constructs based on the ancestral RBD, protein nanostructures and CompAs were both tested for RBD01 (WT, no SD1) and RBDb022 (containing stabilizing mutations, with SD1). For constructs based on the BA.5 RBD, only RBDb047 was tested as a protein nanostructure while RBDb046 (WT, with SD1), RBDb047, RBDb075 and RBDb078 (all three containing stabilizing mutations, with SD1) were tested as CompAs. Mice were immunized in a two-dose regimen on days 0 and 21, with protein nanostructure dosages at 0.027-0.030 μ g (matched to equivalent moles of antigen between each protein nanostructure) and CompA dosages at 0.85-1.0 μ g (matched to equivalent moles of antigen between each CompA), with all groups adjuvanted using Addavax™. Binding titers were measured against strain-matched Spike glycoprotein antigens for ancestral and BA.5 strains (FIG. 11A), and pseudovirus neutralization titers were measured against strain-matched ancestral and BA.5 pseudoviruses (FIG. 11B). Bleeds for serological measurements were performed on days

21 (post-prime) and 35 (post-boost). Within ancestral CompAs, RBDb022 maintained slightly higher binding titers than RBD01 on both days 21 and 35. While no neutralizing titers were detectable on day 21, the RBDb022 CompA also showed higher neutralizing titers compared to RBD01 CompA on day 35. This was surprising, as previous comparisons between similar immunogens to RBD01 CompAs and different stabilized RBDs lacking the SD1 domain had not shown differences in published results (**Ref 5**). Binding and neutralizing titers were generally similar between RBD01 and RBDb022 protein nanostructure groups. Surprisingly, striking differences were seen between CompAs for BA.5 antigens. Among the CompAs, only RBDb075 showed clear signal above baseline at day 21. At day 35, binding titers were higher for CompAs with stabilized antigens compared to RBDb046, with RBDb075 shown to have the highest binding titers. While no neutralizing activity was detected for BA.5 CompAs on day 21, dramatic differences were seen on day 35 with RBDb046 showing near-baseline responses while all CompA groups with stabilizing mutations showed higher responses. Most surprisingly, RBDb075 showed higher neutralizing titers than both RBDb047 and RBDb078. These data show that the specific set of added stabilizing mutations provided a unique assist to immunogenicity in addition to improvements in thermal stability, antigenicity and nanostructure assembly, the combination of which could not have been predicted based on the sequence of this antigen using current methods. The protein nanostructure for RBDb047 was more immunogenic than the CompAs in all measurements.

Example 5: Antigenicity, expression, and melting temperature of selected stabilizing mutations applied to diverse variants RBD-SD1 antigens

[0314] To understand how the stabilizing mutations used in RBDb075 function in RBDs from other variants, the equivalent designs for other variant RBDs using the same stabilizing mutations, antigen fragment (start and end points), and linker to CompA were generated. Variants chosen for analysis included the ancestral variant, BA.1, BA.5, XBB and SARS-CoV-1. As controls, these stabilized designs were compared to equivalent designs lacking stabilizing mutations, with RBDb002 and RBDb046 respectively used for the ancestral and BA.5 variants.

[0315] The five stabilized designs and equivalent five wildtype designs were first secreted from HEK293 cells at small-scale, and antigenicity was assessed by BLI against ACE2, S309, and 16A8 (anti-CompA). Ratios of binding signals against ACE2 and 16A8 were compared for each sample. For BA.1, BA.5 and XBB, the ACE2/16A8 binding ratio was notably increased for the modified constructs relative to the wildtype constructs, suggesting

that the stabilizing mutations were improving the foldedness of the antigen to better enable ACE2 binding. The ratios were generally unchanged for constructs derived from the ancestral variant or SARS-CoV-1, suggesting that these RBD variants were already well-folded in their wildtype form (FIG. 12A and Table 25). Similar results were seen for ratios of S309 to 16A8.

[0316] Next, all ten designs were scaled-up and purified by IMAC, and total protein yields were compared between stabilized and wildtype designs from each variant. For BA.1, BA.5 and XBB, the yields were clearly improved from stabilized designs, with particularly large improvements seen with BA.1 and BA.5. Yields were generally similar from wildtype and stabilized designs from ancestral and SARS-CoV-1 variants (FIG. 12B and Table 25).

[0317] Finally, melting temperatures of purified protein were measured by intrinsic tryptophan fluorescence and compared between RBDs from each variant. The wildtype BA.1 construct was not analyzed due to low yields that prohibited further analysis. For each of the four remaining pairs of variant RBDs, the melting temperature was improved for all stabilized antigens. Improvements ranged from 4.7 °C for SARS-CoV-1 to 6.8 °C for BA.5 (FIG. 12C and Table 25).

[0318] In all, the stabilizing design elements established with RBDb075 were successfully generalized to RBDs from diverse variants, showed no deleterious impact on each variant RBD, and showed improvements in antigenicity, yield and/or thermal stability for each variant RBD.

[0319] Table 25 shows Generalization of designs used in RBDb002 (“Wildtype”) and RBDb075 to diverse strain sequences from SARS-CoV-2 and SARS-CoV-1. Average normalized ACE2 binding measurements were BLI-based, and were calculated from two replicate measurements of ACE2 binding signal divided by binding signal to an anti-CompA mAb. Yields were measured after IMAC purification from equivalent volumes of transfected cell culture. Melting temperatures were measured using intrinsic tryptophan fluorescence, and measured based on values calculated by UNcle software. “N.C.” means “not collected”.

Table 25:

Constructs	Strain	Wildtype or stabilized	Average normalized ACE2 binding (nm/nm)	Yield (mg/L)	Melting temperature (°C)
RBDb002	Ancestral	Wildtype	0.128	55	48.0

RBDb213	Ancestral	Stabilized	0.137	53	52.9
RBDb094	BA.1	Wildtype	0.008	5	N.C.
RBDb155	BA.1	Stabilized	0.096	36	43.4
RBDb046	BA.5	Wildtype	0.042	18	40.5
RBDb075	BA.5	Stabilized	0.085	57	47.3
RBDb270	XBB	Wildtype	0.136	33	46.9
RBDb272	XBB	Stabilized	0.212	45	53.0
RBDb329	SARS-CoV-1	Wildtype	0.137	36	56.9
RBDb331	SARS-CoV-1	Stabilized	0.150	33	61.6

[0320] Table 26 shows constructs used in the experiments.

Table 26:

Construct	Strain	Mutations	Linker
RBDb001	Ancestral	None	GGSGGSGSGGSGSGS
RBDb002	Ancestral	None	SGGGSGGSGSGGSGSGS
RBDb003	Ancestral	None	EPEGSGGSGSGGSGGSGS
RBDb004	Ancestral	None	YGGSGGSGGSGSGGSGGS
RBDb005	Ancestral	None	GSGGSGGSGGSGGSGSGGSGGSGS
RBDb006	Ancestral	F338L/Y365W	GGSGGSGSGGSGGSGS
RBDb007	Ancestral	F338L/Y365W	SGGGSGGSGSGGSGGSGS
RBDb008	Ancestral	F338L/Y365W	EPEGSGGSGSGGSGGSGS
RBDb009	Ancestral	F338L/Y365W	YGGSGGSGGSGSGGSGGS
RBDb010	Ancestral	F338L/Y365W	GSGGSGGSGGSGGSGSGGSGGSGS
RBDb011	Ancestral	F392W	GGSGGSGSGGSGGSGS
RBDb012	Ancestral	F392W	SGGGSGGSGSGGSGGSGS

RBDb013	Ancestral	F392W	EPEGGSGGSGSGSGSGGS S
RBDb014	Ancestral	F392W	YGGSGGSGGSGSGSGSGS GS
RBDb015	Ancestral	F392W	GSGGSGGSGGSGGSGSGG SGGSGS
RBDb016	Ancestral	Y365F/V395I	GGSGGSGSGGSGGSGS
RBDb017	Ancestral	Y365F/V395I	SGGGSGGSGSGGSGGSGS
RBDb018	Ancestral	Y365F/V395I	EPEGGSGGSGSGSGSGGS S
RBDb019	Ancestral	Y365F/V395I	YGGSGGSGGSGSGGSGGS GS
RBDb020	Ancestral	Y365F/V395I	GSGGSGGSGGSGGSGSGG SGGSGS
RBDb021	Ancestral	Y365F/F392W/V395I	GGSGGSGSGGSGGSGS
RBDb022	Ancestral	Y365F/F392W/V395I	SGGGSGGSGSGGSGGSGS
RBDb023	Ancestral	Y365F/F392W/V395I	EPEGGSGGSGSGSGSGGS S
RBDb024	Ancestral	Y365F/F392W/V395I	YGGSGGSGGSGSGGSGGS GS
RBDb025	Ancestral	Y365F/F392W/V395I	GSGGSGGSGGSGGSGSGG SGGSGS
RBDb026	Ancestral	F338L/A363L/Y365M	GGSGGSGSGGSGGSGS
RBDb027	Ancestral	F338L/A363L/Y365M	SGGGSGGSGSGGSGGSGS
RBDb028	Ancestral	F338L/A363L/Y365M	EPEGGSGGSGSGSGSGGS S
RBDb029	Ancestral	F338L/A363L/Y365M	YGGSGGSGGSGSGGSGGS GS
RBDb030	Ancestral	F338L/A363L/Y365M	GSGGSGGSGGSGGSGSGG SGGSGS
RBDb031	Ancestral	I358F/Y365F/V395I	GGSGGSGSGGSGGSGS
RBDb032	Ancestral	I358F/Y365F/V395I	SGGGSGGSGSGGSGGSGS

RBDb033	Ancestral	I358F/Y365F/V395I	EPEGGSGGSGSGSGSGGS S
RBDb034	Ancestral	I358F/Y365F/V395I	YGGSGGSGGSGSGSGSGS GS
RBDb035	Ancestral	I358F/Y365F/V395I	GSGGSGGSGGSGGSGSGG SGGSGS
RBDb036	Ancestral	L560Q/F562Y/I569S	GGSGGSGSGGSGGSGS
RBDb037	Ancestral	L560Q/F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb038	Ancestral	L560Q/F562Y/I569S	EPEGGSGGSGSGSGSGGS S
RBDb039	Ancestral	L560Q/F562Y/I569S	YGGSGGSGGSGSGGSGGS GS
RBDb040	Ancestral	L560Q/F562Y/I569S	GSGGSGGSGGSGGSGSGG SGGSGS
RBDb041	Ancestral	L452R	GGSGGSGSGGSGGSGS
RBDb042	Ancestral	L452R	SGGGSGGSGSGGSGGSGS
RBDb043	Ancestral	L452R	EPEGGSGGSGSGSGSGGS S
RBDb044	Ancestral	L452R	YGGSGGSGGSGSGGSGGS GS
RBDb045	Ancestral	L452R	GSGGSGGSGGSGGSGSGG SGGSGS
RBDb046	BA.5	None	SGGGSGGSGSGGSGGSGS
RBDb047	BA.5	Y365F/V395I/L560Q/ F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb048	BA.2	Y365F/V395I/L560Q/ F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb049	BA.2.12.1	Y365F/V395I/L560Q/ F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb050	BA.2.75	Y365F/V395I/L560Q/ F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb051	BA.5	Y365F/V395I/V407L/ L560Q/F562Y/I569S	SGGGSGGSGSGGSGGSGS

RBDb052	BA.5	Y365F/V395I/I402V/V 407L/L560Q/F562Y/I5 69S	SGGGSGGSGSGGSGGSGS
RBDb053	BA.5	Y365F/V395I/I402V/ I 410F/L560Q/F562Y/I5 69S	SGGGSGGSGSGGSGGSGS
RBDb054	BA.5	Y365F/V395I/I402V/ I 410F/A435V/V512I/L5 60Q/F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb055	BA.5	Y365F/V395I/I402V/V 407L/ I410F/A435V/V 512I/L560Q/F562Y/I5 69S	SGGGSGGSGSGGSGGSGS
RBDb056	BA.5	Y365F/V395I/V350L/I 402V/V407L/ I410F/I4 18V/A435V/V512I/L5 60Q/F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb057	BA.5	Y365F/V395I/V350L/I 402V/V407L/ I410F/I4 18V/A435I/V510I/V51 2I/L560Q/F562Y/I569 S	SGGGSGGSGSGGSGGSGS
RBDb058	BA.5	Y365F/V395I/V350L/I 402V/V407L/ I410F/I4 18V/F429W/V433I/A4 35V/V510I/V512F/L56 0Q/F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb059	BA.5	Y365F/V395I/I402V/V 407L/I410F/L560Q/F5 62Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb060	BA.5	Y365F/V395I/I402V/V 407L/I410F/V512I/L56 0Q/F562Y/I569S	SGGGSGGSGSGGSGGSGS

RBDb061	BA.5	Y365F/V395I/I402V/I 410F/V512I/L560Q/F5 62Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb062	BA.5	Y365F/V395I/A348P/ L560Q/F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb063	BA.5	Y365F/V395I/I402V/L 560Q/F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb064	BA.5	Y365F/V395I/S514T/L 560Q/F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb065	BA.5	Y365F/V395I/I402V/F 464Y/L560Q/F562Y/I5 69S	SGGGSGGSGSGGSGGSGS
RBDb066	BA.5	Y365F/V395I/F329Y/ L560Q/F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb067	BA.5	Y365F/V395I/G526S/P 527N/L560Q/F562Y/I5 69S	SGGGSGGSGSGGSGGSGS
RBDb068	BA.5	Y365F/V395I/G526S/P 527N/K528T/L560Q/F 562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb069	BA.5	Y365F/V395I/F329Y/ G526S/P527N/K528T/ L560Q/F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb070	BA.5	Y365F/V395I/F329Y/ G526S/P527N/K528I/ L560Q/F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb071	BA.5	Y365F/V395I/F329K/ K528Q/L560Q/F562Y/ I569S	SGGGSGGSGSGGSGGSGS
RBDb072	BA.5	Y365F/V395I/F329R/ K528Q/L560Q/F562Y/ I569S	SGGGSGGSGSGGSGGSGS

RBDb073	BA.5	Y365F/V395I/L517T/ H519S/L560Q/F562Y/ I569S	SGGGSGGSGSGGSGGSGS
RBDb074	BA.5	Y365F/V395I/L517A/ H519S/L560Q/F562Y/ I569S	SGGGSGGSGSGGSGGSGS
RBDb075	BA.5	Y365F/V395I/L517T/ H519S/A520G/L560Q/ F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb076	BA.5	Y365F/V395I/L517T/ H519S/N544L/L560Q/ F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb077	BA.5	Y365F/V395I/L517T/ H519S/A520C/Q564C/ L560Q/F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb078	BA.5	Y365F/V395I/L517T/ H519G/A520C/Q564C /L560Q/F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb079	BA.5	Y365F/V395I/L517D/ H519T/A520D/P512Q/ N544L/L546V/Q564L/ F562T/L560Q/F562Y/I 569S	SGGGSGGSGSGGSGGSGS
RBDb080	BA.5	Y365F/V395I/L517S/ H519T/A520D/P521Q/ N544L/L546V/Q564L/ F562T/L560Q/F562Y/I 569S	SGGGSGGSGSGGSGGSGS
RBDb081	BA.5	Y365F/V395I/L517S/ H519R/A520H/P521Q/ N544L/L546V/Q564W /F562T/L560Q/F562Y/ I569S	SGGGSGGSGSGGSGGSGS

RBDb082	BA.5	Y365F/V395I/L517S/ H519T/A520D/N544L/ Q564L/F562T/L560Q/ F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb083	BA.5	Y365F/V395I/F329Y/ G526S/P527N/K528T/ L517S/H519T/A520D/ P521Q/N544L/L546V/ Q564L/F562T/L560Q/ F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb084	BA.5	Y365F/V395I/F329Y/ G526S/P527N/K528I/ L517S/H519T/A520D/ P521Q/N544L/L546V/ Q564L/F562T/L560Q/ F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb085	BA.5	Y365F/V395I/F329Y/ G526S/P527N/K528T/ L517D/H519T/A520D/ P512Q/N544L/L546V/ Q564L/F562T/L560Q/ F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb086	BA.5	Y365F/V395I/F329Y/ G526S/P527N/K528T/ L517S/L517S/H519R/ A520H/P521Q/N544L/ L546V/Q564W/F562T/ L560Q/F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb087	BA.5	Y365F/V395I/F329Y/ L517T/ H519S/L560Q/ F562Y/I569S	SGGGSGGSGSGGSGGSGS

RBDb088	BA.5	Y365F/V395I/F329K/ K528Q/ L517T/ H519S /L560Q/F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb089	BA.5	Y365F/V395I/F329Y/ G526S/P527N/K528T/ L517T/H519S/L560Q/ F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb090	BA.5	Y365F/V395I/V367F/ L560Q/F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb091	BA.5	Y365F/V395I/F375Y/ L560Q/F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb092	Ancestral	Y365F/F392W/V395I/ L560Q/F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb093	BA.5	Y365F/F392W/V395I/ L560Q/F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb094	BA.1	None	SGGGSGGSGSGGSGGSGS
RBDb155	BA.1	Y365F/V395I/L517T/ H519S/A520G/L560Q/ F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb213	Ancestral	Y365F/V395I/L517T/ H519S/A520G/L560Q/ F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb270	XBB	None	SGGGSGGSGSGGSGGSGS
RBDb272	XBB	Y365F/V395I/L517T/ H519S/A520G/L560Q/ F562Y/I569S	SGGGSGGSGSGGSGGSGS
RBDb329	SARS-CoV-1	None	SGGGSGGSGSGGSGGSGS
RBDb331	SARS-CoV-1	Y365F/V395I/L517T/ H519S/A520G/L560Q/ F562Y/I569S	SGGGSGGSGSGGSGGSGS

[0321] In some embodiments, signal sequence is post-translationally cleaved when the protein is expressed. In some embodiments, the underlined/italicized signal and poly-His tag sequences are optionally included in the sequence.

[0322] Table 27 provides sequence of constructs used in the experiments (with underlined/italicized signal sequence and tags).

Table 27:

Constructs	AA seq (with sec signal and tags)
RBDb001	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKC YGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFRD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSSGSGSGSGS GGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEKAV AVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSVEQ ARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTEL VKA MKLGHTILKLFPGEVVGPQFVKAMKGPFNVK FVPTGGVNLDN VAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGATE <u>L</u> <u>EHHHHHH</u> (SEQ ID NO: 111)
RBDb002	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKC YGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFRD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSSGSGSGSGS GSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTEL V KAMKLGHTILKLFPGEVVGPQFVKAMKGPFNVK FVPTGGVNL DNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGAT <u>ELEHHHHHH</u> (SEQ ID NO: 112)
RBDb003	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKC YGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFRD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSEPEGSGSGSGS GGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNVK FVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 113)

RBDb004	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKC YGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFRD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSYGGSGGSGGSG SGGSGGSGGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAI EKAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVT SVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTE LVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGV NLDNVAEWFKAGVLA VGVGSALVKGTPDEVREKAKAFVEKIRG ATELEHHHHHH (SEQ ID NO: 114)
RBDb005	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKC YGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFRD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSGSGGSGGSGGS GGSGSGSGSGGSEKA AKAEEAARKMEELFKKHKIVAVLRANS VEEAIEKAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIG AGTVTSVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGV MTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVP TGGVNLDNVAEWFKAGVLA VGVGSALVKGTPDEVREKAKAFV EKIRGATELEHHHHHH (SEQ ID NO: 115)
RBDb006	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPLGEVFNATRFASVYAWNRRKRISNCVADWSVLYNSASFSTFKC YGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFRD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSGSGGSGGSGGS GGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAIEKAV AVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSVEQ ARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKA MKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVNLDN VAEWFKAGVLA VGVGSALVKGTPDEVREKAKAFVEKIRGATEL EHHHHHH (SEQ ID NO: 116)
RBDb007	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPLGEVFNATRFASVYAWNRRKRISNCVADWSVLYNSASFSTFKC YGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFRD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSSGGSGGSGGSGG SGGSGGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL V KAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVNLDN

	DNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGAT <u>ELEHHHHHH</u> (SEQ ID NO: 117)
RBDb008	<u>MGILSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPLGEVFNATRFASVYAWNRKRISNCVADWSVLYNSASFSTFKC YGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVG GNYNYLYRLFRKSNLKP FERD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVG YQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSEPEGSGSGSGS GGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPPNVK FVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 118)
RBDb009	<u>MGILSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPLGEVFNATRFASVYAWNRKRISNCVADWSVLYNSASFSTFKC YGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVG GNYNYLYRLFRKSNLKP FERD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVG YQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSYGGSGSGSGSGS SGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAI EKAVAVFAGGVHLEITFTVPDADTVIKALSVLKEKGAIIGAGTVT SVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTE LVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPPNVK FVPTGGV NLDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRG <u>ATELEHHHHHH</u> (SEQ ID NO: 119)
RBDb010	<u>MGILSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPLGEVFNATRFASVYAWNRKRISNCVADWSVLYNSASFSTFKC YGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVG GNYNYLYRLFRKSNLKP FERD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVG YQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSGSGSGSGSGSGS GGSGSGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANS VEEAIEKAVAVFAGGVHLEITFTVPDADTVIKALSVLKEKGAIIG AGTVTSVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGV MTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPPNVK FVP TGGVNLDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFV EKIRGATELEHHHHHH (SEQ ID NO: 120)
RBDb011	<u>MGILSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFGEVFNATRFASVYAWNRKRISNCVADY SVLYNSASFSTFKC YGVSP TKLNDLCWTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVG GNYNYLYRLFRKSNLKP FERD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVG YQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSGSGSGSGSGSGS GGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEKAV AVFAGGVHLEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSVEQ

	ARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTELVKA MKLGHTILKLFPGEVVGPQFVKAMKGPPPNVKFVPTGGVNLDN VAEWFKAAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGATEL <u>EHHHHHH</u> (SEQ ID NO: 121)
RBDb012	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKC YGVSP TKLNDLCWTVNYADSFVIRGDEVQRQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSSGGSGSGSGS GSGSGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTEL KAMKLGHTILKLFPGEVVGPQFVKAMKGPPPNVKFVPTGGVN DNVAEWFKAAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGAT <u>ELEHHHHHH</u> (SEQ ID NO: 122)
RBDb013	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKC YGVSP TKLNDLCWTVNYADSFVIRGDEVQRQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSEPEGSGSGSGS GGSGSGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPPNVKFVPTGGVN LDNVAEWFKAAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 123)
RBDb014	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKC YGVSP TKLNDLCWTVNYADSFVIRGDEVQRQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSYGGSGSGSGSG SGSGSGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAI EKAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVT SVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTE LVKAMKLGHTILKLFPGEVVGPQFVKAMKGPPPNVKFVPTGGV NLDNVAEWFKAAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRG <u>ATELEHHHHHH</u> (SEQ ID NO: 124)
RBDb015	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKC YGVSP TKLNDLCWTVNYADSFVIRGDEVQRQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSGSGSGSGSGSGS

	GGSGSGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANS VEEAIEKAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIG AGTVTSVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGV MTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVP TGGVNLDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFV EKIRGATE <u>LEHHHHHH</u> (SEQ ID NO: 125)
RBDb016	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRV</u> <u>OPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKRISNCVADFSVLYNSASFSTFKCY GVSPTKLNLDLCFTNIYADSFVIRGDEVRIAPGQTGKIADYNYKL PDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERDIS TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVL SFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNK KFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSSGGSGSGSGSGG SGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEKAVA VFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSVEQA RKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTELVKA MKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVNLDN VAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGATE <u>L</u> <u>EHHHHHH</u> (SEQ ID NO: 126)
RBDb017	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRV</u> <u>OPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKRISNCVADFSVLYNSASFSTFKCY GVSPTKLNLDLCFTNIYADSFVIRGDEVRIAPGQTGKIADYNYKL PDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERDIS TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVL SFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNK KFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSSGGSGSGSGSGG SGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEKA VAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSVE QARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTELVK AMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVNLDN NVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGATE <u>LEHHHHHH</u> (SEQ ID NO: 127)
RBDb018	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRV</u> <u>OPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKRISNCVADFSVLYNSASFSTFKCY GVSPTKLNLDLCFTNIYADSFVIRGDEVRIAPGQTGKIADYNYKL PDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERDIS TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVL SFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNK KFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSEPEGSGSGSGSG SGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTELV KAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVNL DNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGAT <u>ELEHHHHHH</u> (SEQ ID NO: 128)
RBDb019	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRV</u> <u>OPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKRISNCVADFSVLYNSASFSTFKCY GVSPTKLNLDLCFTNIYADSFVIRGDEVRIAPGQTGKIADYNYKL PDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERDIS TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVL

	SFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNK KFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSYGGSGGSGSGS GGSGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLPGEVVGPPQFVKAMKGPPNVKFPVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 129)
RBDb020	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKRISNCVADFSVLYNSASFSTFKCY GVSPTKLNLDLCFTNIYADSFVIRGDEVQRQIAPGQTGKIADYNYKL PDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERDIS TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVL SFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNK KFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSGSGSGSGSGSG SGSGSGSGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVE EAIEKAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAG TVTSVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMT PTTELVKAMKLGHTILKLPGEVVGPPQFVKAMKGPPNVKFPVPTG GVNLDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKI RGATE <u>LEHHHHHH</u> (SEQ ID NO: 130)
RBDb021	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKRISNCVADFSVLYNSASFSTFKCY GVSPTKLNLDLCWTNIYADSFVIRGDEVQRQIAPGQTGKIADYNYKL PDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERDIS TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVL SFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNK KFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSGSGSGSGSGSG SGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAIEKAVA VFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSVEQA RKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTELVKA MKLGHTILKLPGEVVGPPQFVKAMKGPPNVKFPVPTGGVNLDN VAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGATE <u>L</u> <u>EHHHHHH</u> (SEQ ID NO: 131)
RBDb022	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKRISNCVADFSVLYNSASFSTFKCY GVSPTKLNLDLCWTNIYADSFVIRGDEVQRQIAPGQTGKIADYNYKL PDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERDIS TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVL SFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNK KFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSSGGSGSGSGSGG SGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAIEKA VAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSVE QARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTELVK AMKLGHTILKLPGEVVGPPQFVKAMKGPPNVKFPVPTGGVNLD NVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGATE <u>LEHHHHHH</u> (SEQ ID NO: 132)
RBDb023	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKRISNCVADFSVLYNSASFSTFKCY GVSPTKLNLDLCWTNIYADSFVIRGDEVQRQIAPGQTGKIADYNYKL

	PDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERDIS TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVL SFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNK KFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSEPEGSGSGSGSG GSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL KAMKLGHTILKLFPGEVVGPQFVKAMKGPFNVKVFVPTGGVNL DNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGAT <u>ELEHHHHHH</u> (SEQ ID NO: 133)
RBDb024	<u>MGLPSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRKRISNCVADFSVLVNSASFSTFKCY GVSP TKLNDLCWTNIYADSFVIRGDEV RQIAPGQTGKIADYNYKL PDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERDIS TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVL SFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNK KFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSYGGSGSGSGSGS GGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNVKVFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 134)
RBDb025	<u>MGLPSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRKRISNCVADFSVLVNSASFSTFKCY GVSP TKLNDLCWTNIYADSFVIRGDEV RQIAPGQTGKIADYNYKL PDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERDIS TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVL SFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNK KFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSGSGSGSGSGSGS GSGSGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVE EAIEKAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAG TVTSVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMT PTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNVKVFVPTG GVNLDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKI RGATE <u>LEHHHHHH</u> (SEQ ID NO: 135)
RBDb026	<u>MGLPSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPLGEVFNATRFASVYAWNRKRISNCVLDMSVLVNSASFSTFKC YGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERD ISTEYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSGSGSGSGSGSGS GGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEKAV AVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSVEQ ARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKA MKLGHTILKLFPGEVVGPQFVKAMKGPFNVKVFVPTGGVNLDN VAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGATE <u>L</u> <u>EHHHHHH</u> (SEQ ID NO: 136)

RBDb027	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u>RFPNITNL CPLGEVFNATRFASVYAWNRRKRISNCVLDMSVLYNSASFSTFKC YGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFRD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSSGGSGSGSGS GSGSGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMPTTEL KAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVNL DNVAEWFKAGVLA VGVGSALVKGTPDEVREKAKAFVEKIRGAT <u>ELEHHHHHH</u> (SEQ ID NO: 137)
RBDb028	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u>RFPNITNL CPLGEVFNATRFASVYAWNRRKRISNCVLDMSVLYNSASFSTFKC YGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFRD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSEPEGSGSGSGS GGSGSGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVN LDNVAEWFKAGVLA VGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 138)
RBDb029	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u>RFPNITNL CPLGEVFNATRFASVYAWNRRKRISNCVLDMSVLYNSASFSTFKC YGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFRD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSYGGSGSGSGSG SGSGSGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAI EKAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVT SVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMPTTE LVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGV NLDNVAEWFKAGVLA VGVGSALVKGTPDEVREKAKAFVEKIRG <u>ATELEHHHHHH</u> (SEQ ID NO: 139)
RBDb030	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u>RFPNITNL CPLGEVFNATRFASVYAWNRRKRISNCVLDMSVLYNSASFSTFKC YGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFRD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSGSGSGSGSGGS GGSGSGSGSGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANS VEEAIEKAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIG AGTVTSVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGV MPTTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVP

	TGGVNLNDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFV EKIRGATE <u>LEHHHHHH</u> (SEQ ID NO: 140)
RBDb031	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKFSNCVADFSVLYNSASFSTFKC YGVSP TKLNDLCFTNIYADSFVIRGDEV RQIAPGQTGKIADYNYK LPDDFTGCVIAWNSNNLDSKVGGNYYLYRLFRKSNLKPFERDI STEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSGGSGGSGSGS GGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEKAV AVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSVEQ ARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTEL VKA MKLGHTILKLFPGEVVGPQFVKAMKGPFNVK FVPTGGVNLNDN VAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGATE <u>L</u> <u>EHHHHHH</u> (SEQ ID NO: 141)
RBDb032	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKFSNCVADFSVLYNSASFSTFKC YGVSP TKLNDLCFTNIYADSFVIRGDEV RQIAPGQTGKIADYNYK LPDDFTGCVIAWNSNNLDSKVGGNYYLYRLFRKSNLKPFERDI STEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSSGGSGGSGSGS GGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTEL V KAMKLGHTILKLFPGEVVGPQFVKAMKGPFNVK FVPTGGVNL DNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGAT <u>ELEHHHHHH</u> (SEQ ID NO: 142)
RBDb033	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKFSNCVADFSVLYNSASFSTFKC YGVSP TKLNDLCFTNIYADSFVIRGDEV RQIAPGQTGKIADYNYK LPDDFTGCVIAWNSNNLDSKVGGNYYLYRLFRKSNLKPFERDI STEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSEPEGSGGSGS GGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNVK FVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 143)
RBDb034	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKFSNCVADFSVLYNSASFSTFKC YGVSP TKLNDLCFTNIYADSFVIRGDEV RQIAPGQTGKIADYNYK LPDDFTGCVIAWNSNNLDSKVGGNYYLYRLFRKSNLKPFERDI STEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSYGGSGGSGGSG SGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAI EKAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVT

	SVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTLE LVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGV NLDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRG <u>ATELEHHHHHH</u> (SEQ ID NO: 144)
RBDb035	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFGEVFNATRFASVYAWNRKRFSNCVADFSVLYNSASFSTFKC YGVSP TKLNDLCFTNIYADSFVIRGDEV RQIAPGQTGKIADYNYK LPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDI STEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSGSGSGSGSGS GGSGSGSGSGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANS VEEAIEKAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIG AGTVTSVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGV MPTTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVP TGGVNLDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFV EKIRGATE <u>LEHHHHHHH</u> (SEQ ID NO: 145)
RBDb036	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKC YGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERD ISTEYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSGSGSGSGSGG SGSGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAIEKA VAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSVE QARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTELVK AMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVNLD NVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGATE <u>LEHHHHHHH</u> (SEQ ID NO: 146)
RBDb037	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKC YGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERD ISTEYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGSGSGSGS GGSGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHHH</u> (SEQ ID NO: 147)
RBDb038	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKC YGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERD ISTEYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSEPEGSGSGSG

	SGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAI EKAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVT SVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTE LVKAMKLGHTILKLFPGEVVGPQFVKAMKGPPPNVKFVPTGGV NLDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRG ATELEHHHHHH (SEQ ID NO: 148)
RBDb039	<u>MGILSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKC YGVSP TKLNDLCFTN VYADSFVIRGDEV RQIAPGQTGKIADYNY KL PDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERD ISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSYGGSGSGSGS SGSGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEE AIEKAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGT VTSVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTP TELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPPPNVKFVPTGG VNLDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIR GATELEHHHHHH (SEQ ID NO: 149)
RBDb040	<u>MGILSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKC YGVSP TKLNDLCFTN VYADSFVIRGDEV RQIAPGQTGKIADYNY KL PDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERD ISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSGSGSGSGSGG SGSGSGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANS VEEAIEKAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIG AGTVTSVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGV MTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPPPNVKFVP TGGVNLDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFV EKIRGATELEHHHHHH (SEQ ID NO: 150)
RBDb041	<u>MGILSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKC YGVSP TKLNDLCFTN VYADSFVIRGDEV RQIAPGQTGKIADYNY KL PDDFTGCVIAWNSNNLDSKVGGNYNYRYRLFRKSNLKPFERD ISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSGSGSGSGSGS GGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEKAV AVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSVEQ ARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTELVA MKLGHTILKLFPGEVVGPQFVKAMKGPPPNVKFVPTGGVNLDN VAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGATEL EHHHHHH (SEQ ID NO: 151)
RBDb042	<u>MGILSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKC YGVSP TKLNDLCFTN VYADSFVIRGDEV RQIAPGQTGKIADYNY KL PDDFTGCVIAWNSNNLDSKVGGNYNYRYRLFRKSNLKPFERD ISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV

	LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSSGGGSGGSGSG GSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL KAMKLGHTILKLFPGEVVGPQFVKAMKGPPNVKVFPTGGVNL DNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGAT <u>LEHHHHHHH</u> (SEQ ID NO: 152)
RBDb043	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKC YGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNRYRLFRKSNLKPFERD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSEPEGSGSGSGS GSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPNVKVFPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHHH</u> (SEQ ID NO: 153)
RBDb044	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKC YGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNRYRLFRKSNLKPFERD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSYGGSGSGSGSGS GSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE EKAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVT SVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTE LVKAMKLGHTILKLFPGEVVGPQFVKAMKGPPNVKVFPTGGV NLDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRG <u>ATELEHHHHHHH</u> (SEQ ID NO: 154)
RBDb045	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKC YGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNY KLPDDFTGCVIAWNSNNLDSKVGGNYNRYRLFRKSNLKPFERD ISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSGSGSGSGSGSGS GSGSGSGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANS VEEAIEKAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIG AGTVTSVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGV MTPTTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPPNVKVFVP TGGVNLNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFV EKIRGATE <u>LEHHHHHHH</u> (SEQ ID NO: 155)
RBDb046	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFDEVFNATRFASVYAWNRRKRISNCVADYSVLYNFAPFFAFKC YGVSPTKLNDLCFTNVYADSFVIRGNEVSQIAPGQTGNIADYNYK

	LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIYQAGNKPCNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GSGGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL KAMKLGHTILKLFPGEVVGPQFVKAMKGFPNPKFVPTGGVNL DNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGAT <u>ELEHHHHHH</u> (SEQ ID NO: 156)
RBDb047	<u>MGILPSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIYQAGNKPCNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GSGGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGFPNPKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 157)
RBDb048	<u>MGILPSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYLYRLFRKSNLKPFERDI STEIYQAGNKPCNGVAGFNCYFPLRSYGFRPTYGVGHQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GSGGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGFPNPKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 158)
RBDb049	<u>MGILPSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYQYRLFRKSNLKPFERDI STEIYQAGNKPCNGVAGFNCYFPLRSYGFRPTYGVGHQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GSGGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGFPNPKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 159)

RBDb050	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u>RFPNITNL CPFHEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVS GNYNYLYRLFRKSKLKPFERDIS TEIYQAGNKP CNGVAGFNCYFPLQSYGFRPTYGVGHQPYRVVV SFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNK KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGSGSGSGS GSGSGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTELV KAMKLGHTILKLFPGEVVGPQFVKAMKGPPFPNVKFVPTGGVNL DNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGAT <u>ELEHHHHHH</u> (SEQ ID NO: 160)
RBDb051	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u>RFPNITNL CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNELSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVG GNYNYRYRLFRKSNLKPFERDI STEIYQAGNKP CNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGSGSGSGS GSGSGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPFPNVKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 161)
RBDb052	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u>RFPNITNL CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVVRGNELSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVG GNYNYRYRLFRKSNLKPFERDI STEIYQAGNKP CNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGSGSGSGS GSGSGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPFPNVKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 162)
RBDb053	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u>RFPNITNL CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVVRGNEVSQFAPGQTGNIADYNY KLPDDFTGCVIAWNSNKLDSKVG GNYNYRYRLFRKSNLKPFERD ISTEIYQAGNKP CNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVV VLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTES NKKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGSGSGSGS SGGSGSGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAI EKAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVT SVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPT ELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPPFPNVKFVPTGGV

	NLDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRG ATELEHHHHHHH (SEQ ID NO: 163)
RBDb054	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSPTKLNDLCFTNIYADSFVVRGNEVSQFAPGQTGNIADYNY KLPDDFTGCVIVWNSNKLDISKVGGNYNYRYRLFRKSNLKPFERD ISTEIQAGNKPENGAVGVNCFPLQSYGFRPTYGVGHQPYRVVI LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTEILDITPCSSGGGSGGSGS GGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPPNVKFFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA TELEHHHHHHH (SEQ ID NO: 164)
RBDb055	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSPTKLNDLCFTNIYADSFVVRGNELSQFAPGQTGNIADYNY KLPDDFTGCVIVWNSNKLDISKVGGNYNYRYRLFRKSNLKPFERD ISTEIQAGNKPENGAVGVNCFPLQSYGFRPTYGVGHQPYRVVI LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTEILDITPCSSGGGSGGSGS GGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPPNVKFFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA TELEHHHHHHH (SEQ ID NO: 165)
RBDb056	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFDEVFNATRFASLYAWNRRKRISNCVADFSVLYNFAPFFAFKCY GVSPTKLNDLCFTNIYADSFVVRGNELSQFAPGQTGNVADYNYK LPDDFTGCVIVWNSNKLDISKVGGNYNYRYRLFRKSNLKPFERDI STEIQAGNKPENGAVGVNCFPLQSYGFRPTYGVGHQPYRVVI LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTEILDITPCSSGGGSGGSGS GGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPPNVKFFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA TELEHHHHHHH (SEQ ID NO: 166)
RBDb057	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFDEVFNATRFASLYAWNRRKRISNCVADFSVLYNFAPFFAFKCY GVSPTKLNDLCFTNIYADSFVVRGNELSQFAPGQTGNVADYNYK LPDDFTGCVIIWNSNKLDISKVGGNYNYRYRLFRKSNLKPFERDIS TEIQAGNKPENGAVGVNCFPLQSYGFRPTYGVGHQPYRIVILS FELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK FQPYQQFGRDSADTTDAVRDPQTEILDITPCSSGGGSGGSGSGG SGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEKA VAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSVE

	QARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTELVK AMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVNLD NVAEWFKAAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGATE <u>LEHHHHHHH</u> (SEQ ID NO: 167)
RBDb058	<u>MGILPSPGMPALLSLVSLLSVLLMGCVAETGSRV/OPTESIV</u> RFPNITNL CPFDEVFNATRFASLYAWNRRKRISNCVADFSVLYNFAPFFAFKCY GVSP TKLNDLCFTNIYADSFVVRGNELSQFAPGQTGNVADYNYK LPDDWTGCIIVWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIYQAGNKPNGVAGVNCYFPLQSYGFRPTYGVGHQPYRIVF LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVN LDNVAEWFKAAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHHH</u> (SEQ ID NO: 168)
RBDb059	<u>MGILPSPGMPALLSLVSLLSVLLMGCVAETGSRV/OPTESIV</u> RFPNITNL CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVVRGNELSQFAPGQTGNIADYNY KLPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERD ISTEYQAGNKPNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVV VLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTES NKKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS SGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAI EKAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVT SVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTE LVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGV LDNVAEWFKAAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRG <u>ATELEHHHHHHH</u> (SEQ ID NO: 169)
RBDb060	<u>MGILPSPGMPALLSLVSLLSVLLMGCVAETGSRV/OPTESIV</u> RFPNITNL CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVVRGNELSQFAPGQTGNIADYNY KLPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERD ISTEYQAGNKPNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVI LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVN LDNVAEWFKAAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHHH</u> (SEQ ID NO: 170)
RBDb061	<u>MGILPSPGMPALLSLVSLLSVLLMGCVAETGSRV/OPTESIV</u> RFPNITNL CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVVRGNEVSQFAPGQTGNIADYNY KLPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERD ISTEYQAGNKPNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVI LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS

	GGSGGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPPNVKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 171)
RBDb062	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFDEVFNATRFPSVYAWNRRKRISNCVADFSVLYNFAPFFAFKCY GVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYKL PDDFTGCVIAWNSNKLDSKVGGNYNRYRLFRKSNLKPFERDIS TEIYQAGNKPCNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGSGSGSGS GGSGGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPPNVKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 172)
RBDb063	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVVRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNRYRLFRKSNLKPFERDI STEIYQAGNKPCNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGSGSGSGS GGSGGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPPNVKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 173)
RBDb064	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNRYRLFRKSNLKPFERDI STEIYQAGNKPCNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LTFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGSGSGSGS GGSGGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPPNVKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 174)
RBDb065	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVVRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNRYRLFRKSNLKPYPERDI STEIYQAGNKPCNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV

	LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTEILDITPCSSGGSGSGSGS GGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPNVK FVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 175)
RBDb066	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RYPNITN LCPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNRYRLFRKSNLKPFERDI STEIYQAGNKP CNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTEILDITPCSSGGSGSGSGS GGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPNVK FVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 176)
RBDb067	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNRYRLFRKSNLKPFERDI STEIYQAGNKP CNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFELLHAPATVCSNKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTEILDITPCSSGGSGSGSGS GGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPNVK FVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 177)
RBDb068	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNRYRLFRKSNLKPFERDI STEIYQAGNKP CNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFELLHAPATVCSNKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTEILDITPCSSGGSGSGSGS GGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPNVK FVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 178)
RBDb069	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RYPNITN LCPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK

	LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIYQAGNKPNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFELLHAPATVCSNTKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 179)
RBDb070	<u>MGILPSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRPNITN</u> LCPFDEVFNATRFASVYAWNRRKRISNCVADFSVLNFAPFFAFKC YGVSPTKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIYQAGNKPNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFELLHAPATVCSNIKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL KAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVNL DNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGAT <u>ELEHHHHHH</u> (SEQ ID NO: 180)
RBDb071	<u>MGILPSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRPNITN</u> LCPFDEVFNATRFASVYAWNRRKRISNCVADFSVLNFAPFFAFKC YGVSPTKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIYQAGNKPNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFELLHAPATVCGPQKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 181)
RBDb072	<u>MGILPSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRRPNITN</u> LCPFDEVFNATRFASVYAWNRRKRISNCVADFSVLNFAPFFAFKC YGVSPTKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIYQAGNKPNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFELLHAPATVCGPQKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 182)

RBDb073	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u>RFPNITNL CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIQAGNKP CNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFETLSAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GSGGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPPNVKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 183)
RBDb074	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u>RFPNITNL CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIQAGNKP CNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFEALSAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GSGGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPPNVKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 184)
RBDb075	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u>RFPNITNL CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIQAGNKP CNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFETLSGPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GSGGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPPNVKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 185)
RBDb076	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u>RFPNITNL CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIQAGNKP CNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFETLSAPATVCGPKKSTNLVKNKCVNFNGLTGTGVLTESNK KFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGSG GSGGSGSEKA AKAEEAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPPNVKFVPTGGVNL

	DNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGAT <u>EEHHHHHHH</u> (SEQ ID NO: 186)
RBDb077	<u>MGILSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLNFPFFAFKC YGVSPTKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIQAGNKPNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFETLSCPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNK KFQPYQCFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGSG GGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTLV KAMKLGHTILKLFPGEVVGPQFVKAMKGPPFNVKFVPTGGVNL DNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGAT <u>EEHHHHHHH</u> (SEQ ID NO: 187)
RBDb078	<u>MGILSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLNFPFFAFKC YGVSPTKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIQAGNKPNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFETLSCPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQCFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTLV VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPFNVKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TEHHHHHHH</u> (SEQ ID NO: 188)
RBDb079	<u>MGILSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLNFPFFAFKC YGVSPTKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIQAGNKPNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFEDLTDQATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPTQLFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTLV VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPFNVKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TEHHHHHHH</u> (SEQ ID NO: 189)
RBDb080	<u>MGILSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLNFPFFAFKC YGVSPTKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIQAGNKPNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFESLTDQATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPTQLFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS

	VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPNVKFFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 190)
RBDb081	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRV/OPTESIV</u> RFPNITNL CPFDEVFNATRFASVYAWNRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIQAGNKPNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFESLRHQATVCGPKKSTNLVKNKCVNFNFLGVTGTGVLTESN KKFQPTQWFRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GGSGGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPNVKFFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 191)
RBDb082	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRV/OPTESIV</u> RFPNITNL CPFDEVFNATRFASVYAWNRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIQAGNKPNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFESLTDPATVCGPKKSTNLVKNKCVNFNFLGLTGTGVLTESNK KFQPTQLFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGSG GSGGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL KAMKLGHTILKLFPGEVVGPQFVKAMKGPPNVKFFVPTGGVNL DNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGAT <u>ELEHHHHHH</u> (SEQ ID NO: 192)
RBDb083	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRV/OPTESIV</u> RYPNITN LCPFDEVFNATRFASVYAWNRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIQAGNKPNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFESLTDQATVCSNTKSTNLVKNKCVNFNFLGVTGTGVLTESN KKFQPTQLFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GGSGGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPNVKFFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 193)
RBDb084	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRV/OPTESIV</u> RYPNITN LCPFDEVFNATRFASVYAWNRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIQAGNKPNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFESLTDQATVCSNIKSTNLVKNKCVNFNFLGVTGTGVLTESNK KFQPTQLFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGSG

	GSGGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTLV KAMKLGHTILKLFPGEVVGPQFVKAMKGPPPNVKFVPTGGVNL DNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGAT <u>ELEHHHHHH</u> (SEQ ID NO: 194)
RBDb085	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVPTESIVRYPNITN</u> LCPFDEVFNATRFASVYAWNRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIYQAGNKP CNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFEDLTDQATVCSNTKSTNLVKNKCVNFNFGVLTGTGVLTESN KKFQPTQLFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GGSGGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTLV VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPPNVKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 195)
RBDb086	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVPTESIVRYPNITN</u> LCPFDEVFNATRFASVYAWNRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIYQAGNKP CNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFESLRHQATVCSNTKSTNLVKNKCVNFNFGVLTGTGVLTESN KKFQPTQWFGGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GGSGGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTLV VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPPNVKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 196)
RBDb087	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVPTESIVRYPNITN</u> LCPFDEVFNATRFASVYAWNRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIYQAGNKP CNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFETLSAPATVCGPKKSTNLVKNKCVNFNFGVLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGS GGSGGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTLV VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPPNVKFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 197)
RBDb088	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVPTESIVRKPNITN</u> LCPFDEVFNATRFASVYAWNRKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYRYRLFRKSNLKPFERDI STEIYQAGNKP CNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV

	LSFETLSAPATVCGPQKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGSGSGSGS GGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPNVK FVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 198)
RBDb089	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RYPNITN LCPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKC YGVSPTKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNRYRLFRKSNLKPFERDI STEIYQAGNKPCNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFETLSAPATVCSNTKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGSGSGSGS GGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL KAMKLGHTILKLFPGEVVGPQFVKAMKGPPNVK FVPTGGVN DNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGAT <u>ELEHHHHHH</u> (SEQ ID NO: 199)
RBDb090	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFFAFKCY GVSPTKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYKL PDDFTGCVIAWNSNKLDSKVGGNYNRYRLFRKSNLKPFERDIS TEIYQAGNKPCNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGSGSGSGS GGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPNVK FVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 200)
RBDb091	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFDEVFNATRFASVYAWNRRKRISNCVADFSVLYNFAPFYAFKC YGVSPTKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNRYRLFRKSNLKPFERDI STEIYQAGNKPCNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGSGSGSGS GGSGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPPNVK FVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 201)
RBDb092	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKRISNCVADFSVLYNSASFSTFKCY GVSPTKLNDLCWTNIYADSFVIRGDEVRIAPGQTGKIADYNYKL

	PDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDIS TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVL SFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNK KFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGSG GSGGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL KAMKLGHTILKLFPGEVVGPQFVKAMKGPFNVKVFVPTGGVNL DNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGAT <u>ELEHHHHHH</u> (SEQ ID NO: 202)
RBDb093	<u>MGILPSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFDEVFNATRFASVYAWNKRISNCVADFSVLYNFAPFFAFKC YGVSP TKLNDLCWTNIYADSFVIRGNEVSQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVGGNYNYLYRLFRKSNLKPFERDI STEIYQAGNKPCNGVAGVNCYFPLQSYGFRPTYGVGHQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGSG GSGGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIE KAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTS VEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKAMKLGHTILKLFPGEVVGPQFVKAMKGPFNVKVFVPTGGVN LDNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGA <u>TELEHHHHHH</u> (SEQ ID NO: 203)
RBDb094	<u>MGILPSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFDEVFNATRFASVYAWNKRISNCVADYSVLYNLAPFFTFKC YGVSP TKLNDLCFTNVIYADSFVIRGDEV RQIAPGQTGNIADYNY KL PDDFTGCVIAWNSNKLDSKVSGNYNYLYRLFRKSNLKPFERD ISTEYQAGNKPCNGVAGFNCYFPLRSYSFRPTYGVGHQPYRVVV LSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESN KKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSSGGGSGGSGSG GSGGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL KAMKLGHTILKLFPGEVVGPQFVKAMKGPFNVKVFVPTGGVNL DNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGAT <u>ELEHHHHHH</u> (SEQ ID NO: 265)
RBDb155	<u>MGILPSPGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIVRFPNITNL</u> CPFDEVFNATRFASVYAWNKRISNCVADFSVLYNLAPFFTFKC YGVSP TKLNDLCFTNIYADSFVIRGDEV RQIAPGQTGNIADYNYK LPDDFTGCVIAWNSNKLDSKVSGNYNYLYRLFRKSNLKPFERDIS TEIYQAGNKPCNGVAGFNCYFPLRSYSFRPTYGVGHQPYRVVVL SFETLSGPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNK KFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGGSGGSGSG GSGGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL KAMKLGHTILKLFPGEVVGPQFVKAMKGPFNVKVFVPTGGVNL DNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGAT <u>ELEHHHHHH</u> (SEQ ID NO: 266)

RBDb213	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFGEVFNATRFASVYAWNRRKISNCVADFSVLYNSASFSTFKCY GVSPTKLNDLCFTNIYADSFVIRGDEVQRQIAPGQTGKIADYNYKL PDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERDIS TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVL SFETLSGPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNK KFQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGSGSGSGSG SGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL KAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVNL DNVAEWFKAGVLAAGVGVSALVKGTPDEVREKAKAFVEKIRGAT <u>ELEHHHHHH</u> (SEQ ID NO: 267)
RBDb270	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFHEVFNATTFASVYAWNRRKISNCVADYSVIYNFAPFFAFKCY GVSPTKLNDLCFTNVYADSFVIRGNEVSQIAPGQTGNIADYNYKL PDDFTGCVIAWNSNKLDSKPSGNYNLYRLFRKSKLKPFERDIST EIQAGNKPENGAVAGSNCSPLQSYGFRPTYGVGHQPYRVVVL FELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK FLPFQQFGRDIADTTDAVRDPQTLEILDITPCSSGGSGSGSGSGGS GGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEKAV AVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSVEQ ARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VKA MKLGHITILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVNL DN VAEWFKAGVLAAGVGVSALVKGTPDEVREKAKAFVEKIRGATE <u>LEHHHHHH</u> (SEQ ID NO: 268)
RBDb272	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVOPTESIV</u> RFPNITNL CPFHEVFNATTFASVYAWNRRKISNCVADFSVIYNFAPFFAFKCY GVSPTKLNDLCFTNIYADSFVIRGNEVSQIAPGQTGNIADYNYKL PDDFTGCVIAWNSNKLDSKPSGNYNLYRLFRKSKLKPFERDIST EIQAGNKPENGAVAGSNCSPLQSYGFRPTYGVGHQPYRVVVL FETLSGPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKK FQPYQQFGRDSADTTDAVRDPQTLEILDITPCSSGGSGSGSGSGSG SGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEKA VAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSVE QARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL VK AMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVNL DN NVAEWFKAGVLAAGVGVSALVKGTPDEVREKAKAFVEKIRGATE <u>LEHHHHHH</u> (SEQ ID NO: 269)
RBDb329	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVPSGDV</u> RFPNITN LCPFGEVFNATKFPSVYAWERKKISNCVADYSVLYNSTFFSTFKC YGVSATKLNLCFSNVYADSFVVKGDDVRQIAPGQTGVIADYN YKLPDDFMGCVLAWNTRNIDATSTGNYNKYRYLRHGLRPF RDISNVPFSPDGKPTPALNCYWPLNDYGFYTTTGIGYQPYRVV VLSFELLNAPATVCGPKLSTDLIKNCVNFNFNGLTGTGVLTPSS KRFQPFQQFGRDVSDFDTSVRDPKTSEILDISPCSSGGSGSGSGSG SGSGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEK AVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTTEL V KAMKLGHTILKLFPGEVVGPQFVKAMKGPFNPVKFVPTGGVNL

	DNVAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGAT <u>ELEHHHHHH</u> (SEQ ID NO: 270)
RBDb331	<u>MGILPSGMPALLSLVSLLSVLLMGCVAETGSRVYPSGDVYRFPNITN</u> LCPFGEVFNATKFPSVYAWERKKISNCVADFSVLYNSTFFSTFKC YGVSATKLNDLCSNIYADSFVVKGDDVRQIAPGQTGVIADYNY KLPDDFMGCVLAWNTRNIDATSTGNYNKYRYLRHGKLRPFER DISNVPFSPDGKPTPPALNCYWPLNDYGFYTTTGIGYQPYRVVV LSFETLSGPATVCGPKLSTDLIKNCVNFNFNGLTGTGVLTTPSSKR FQPYQQFGRDSSDFTDSVRDPKTSEILDISPCSSGGSGSGSGSGS GGSGSEKAAKAEAAARKMEELFKKHKIVAVLRANSVEEAIEKAV AVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSVEQ ARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMPTTELVKA MKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDN VAEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGATEL <u>EHHHHHH</u> (SEQ ID NO: 271)

References

1. Ahmed, S., et al., A Stabilized, Monomeric, Receptor Binding Domain Elicits High-Titer Neutralizing Antibodies Against All SARS-CoV-2 Variants of Concern, *Front. Immunol.*, 12:765211 (2021), available at <https://pubmed.ncbi.nlm.nih.gov/34956193/>.
2. Bale, J.B., et al., Accurate design of megadalton-scale two-component icosahedral protein complexes, *Science*, 353(6297):389-394 (2016), available at <https://pubmed.ncbi.nlm.nih.gov/27463675/>.
3. Cao, Y., et al., Characterization of the enhanced infectivity and antibody evasion of Omicron BA.2.75. *Cell Host Microbe.*, 30(11):1527-1539 (2022), available at <https://pubmed.ncbi.nlm.nih.gov/36270286/>.
4. Dalvie, N., et al., Engineered SARS-CoV-2 receptor binding domain improves manufacturability in yeast and immunogenicity in mice, *Proc. Natl. Acad. Sci.*, 118(38):e2106845118 (2021), available at <https://pubmed.ncbi.nlm.nih.gov/34493582/>.
5. Ellis, D., et al., Stabilization of the SARS-CoV-2 Spike glycoprotein Receptor-Binding Domain Using Deep Mutational Scanning and Structure-Based Design, *Front Immunol.*, 12:710263 (2021), available at <https://pubmed.ncbi.nlm.nih.gov/34267764/>.
6. Fougereous, C., et al., Capsid-like particles decorated with the SARS-CoV-2 receptor-binding domain elicit strong virus neutralization activity, *Nat. Commu.*, 12(1):324 (2021), available at <https://pubmed.ncbi.nlm.nih.gov/33436573/>.

7. Heinze, K., et al., Protein Nanocontainers from Nonviral Origin: Testing the Mechanics of Artificial and Natural Protein Cages by AFM, *J. Phys. Chem.*, 120(26):5945-5952 (2016), available at <https://pubmed.ncbi.nlm.nih.gov/27187612/>.
8. King, N., et al., Computational design of self-assembling protein nanomaterials with atomic level accuracy, *Science*, 336(6085):1171-1174 (2012), available at <https://pubmed.ncbi.nlm.nih.gov/22654060/>.
9. King, N., et al., Accurate design of co-assembling multi-component protein nanomaterials, *Nature*, 510(7503):103-108 (2014), available at <https://pubmed.ncbi.nlm.nih.gov/24870237/>.
10. Leman, J. K., et al., Macromolecular modeling and design in Rosetta: recent methods and frameworks, *Nat. Methods*, 17(7):665-680 (2020), available at <https://pubmed.ncbi.nlm.nih.gov/32483333/>.
11. Mirdita, M., et al., ColabFold: making protein folding accessible to all, *Nat. Methods*, 19(6):679-682 (2022), available at <https://pubmed.ncbi.nlm.nih.gov/35637307/>.
12. Pinto, D., et al., Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody, *Nature*, 583(7815):290-295 (2020), available at <https://pubmed.ncbi.nlm.nih.gov/32422645/>.
13. Stalls, V., et al., Cryo-EM structures of SARS-CoV-2 Omicron BA.2 Spike glycoprotein, *Cell Rep.*, 39(13):111009 (2022), available at <https://pubmed.ncbi.nlm.nih.gov/35732171/>.
14. Walls, A.C., et al., Elicitation of Potent Neutralizing Antibody Responses by Designed Protein Nanoparticle Vaccines for SARS-CoV-2, *Cell*, 183(5):1367-1382 (2020), available at <https://pubmed.ncbi.nlm.nih.gov/33160446/>.
15. Wrapp, D., et al., Cryo-EM structure of the 2019-nCoV Spike glycoprotein in the prefusion conformation, *Science*, 367(6483):1260-1263 (2020), available at <https://pubmed.ncbi.nlm.nih.gov/32075877/>.
16. Yuan, M., et al., A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV, *Science*, 368(6491):630-633 (2020), available at <https://pubmed.ncbi.nlm.nih.gov/32245784/>.

EMBODIMENTS

The disclosure further provides the following enumerated embodiments:

Embodiment 1. A recombinant polypeptide, comprising an antigenic fragment of a coronavirus Spike glycoprotein, said antigenic fragment comprising a receptor-binding domain (RBD) and optionally, a coronavirus subdomain 1 (SD1).

Embodiment 2. The polypeptide of Embodiment 1, wherein the antigenic fragment comprises RBD-SD1.

Embodiment 3. The polypeptide of Embodiment 1 or Embodiment 2, wherein the SD1 comprises at least one substitution of a surface-exposed, non-polar amino acid residue with a polar amino acid residue.

Embodiment 4. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises a substitution of one or more, two or more, or three amino acids at positions 560, 562, or 569 with a polar residue, wherein the substitutions are relative to a reference sequence according to SEQ ID NO: 1.

Embodiment 5. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises a substitution of one or more, two or more, or three of the amino acid substitutions L560Q, F562T, F562Y, or I569S relative to a reference sequence according to SEQ ID NO: 1.

Embodiment 6. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises a substitution of one or more, two or more, or three or more amino acids at positions 544, 546, 560, 562, 564, 569, or 582 relative to a reference sequence according to SEQ ID NO: 1.

Embodiment 7. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acids at positions N544L, N544M, N544Q, L546V, L560Q, F562T, F562Y, Q564C, Q564L, Q564N, Q564W, I569S, or L582S relative to a reference sequence according to SEQ ID NO: 1.

Embodiment 8. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acids at positions 338, 358, 363, 365, 392, or 395 relative to a reference sequence according to SEQ ID NO: 1.

Embodiment 9. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acids at positions F338L, I358F, A363L, Y365F, Y365M, Y365W, F392W, or V395I relative to a reference sequence according to SEQ ID NO: 1.

Embodiment 10. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises amino acid substitutions comprising F338L/Y365W; F392W; Y365F/V395I; Y365F/ F392W /V395I; F338L/A363L/Y365M; and/or I358F/Y365F/V395I.

Embodiment 11. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acids at positions 329, 348, 350, 367, 375, 402, 407, 410, 418, 429, 433, 435, 452, 464, 510, 512, 514, 517, 518, 519, 520, 522, 527, or 528 relative to a reference sequence according to SEQ ID NO: 1.

Embodiment 12. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acids at positions F329K, F329R, F329Y, A348P, V350L, V367F, F375Y, I402V, V407L, I410F, I418V, F429W, V433I, A435I, A435V, L452R, F464Y, P512Q, V510I, V512F, V512I, S514T, L517A, L517D, L517S, L517T, L518N, L518Q, L518V, H519D, H519G, H519R, H519S, H519T, A520C, A520D, A520G, A520H, P521A, P521D, P521N, P521Q, P521S, A522G, A522I, P527N, K528I, K528Q, K528T, L518G, or L518S relative to a reference sequence according to SEQ ID NO: 1.

Embodiment 13. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises amino acid substitutions at positions 365, 395, 560, 562, and 569; and one or more, two or more, or three or more amino acid substitutions at positions 348, 402, 464, 514, 520, 526, and 527, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1.

Embodiment 14. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises the amino acid substitutions comprising A348P/Y365F/V395I/L560Q/F562Y/I569S; Y365F/V395I/I402V/L560Q/F562Y/I569S; Y365F/V395I/S514T/L560Q/F562Y/I569S; Y365F/V395I/I402V/F464Y/L560Q/F562Y/I569S; and/or Y365F/V395I/G526S/P527N/L560Q/F562Y/I569S.

Embodiment 15. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises amino acid substitutions at positions 365, 395, 560, 562, and 569; and one or more, two or more, or three or more amino acid substitutions at positions 329, 517, 519, 520, and 544, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1.

Embodiment 16. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises the amino acid substitutions comprising Y365F/V395I/L560Q/F562Y/I569S; Y365F/V395I/F329Y/L560Q/F562Y/I569S; Y365F/V395I/L517T/H519S/L560Q/F562Y/I569S; Y365F/V395I/L517T/H519S/A520G/L560Q/F562Y/I569S; Y365F/V395I/L517T/H519S/N544L/L560Q/F562Y/I569S; and/or Y365F/V395I/F329Y/L517T/H519S/L560Q/F562Y/I569S.

Embodiment 17. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises amino acid substitutions at positions 365, 392, 395, 560, 562, and 569; and one or more, two or more, or three or more amino acid substitutions at positions 329, 517, 519, 520, and 544, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1.

Embodiment 18. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises the amino acid substitutions comprising Y365F/F392W/V395I/L560Q/F562Y/I569S; Y365F/F392W/V395I/F329Y/L560Q/F562Y/I569S; Y365F/F392W/V395I/L517T/H519S/L560Q/F562Y/I569S; Y365F/F392W/V395I/L517T/H519S/A520G/L560Q/F562Y/I569S; Y365F/F392W/V395I/L517T/H519S/N544L/L560Q/F562Y/I569S; and/or Y365F/F392W/V395I/F329Y/L517T/H519S/L560Q/F562Y/I569S.

Embodiment 19. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises amino acid substitutions at positions 365, 395, 517, 519, 520, 560, 562, 564, and 569; and one or more, two or more, or three or more amino acid substitutions at positions 521, 544, and 546, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1.

Embodiment 20. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises the amino acid substitutions comprising Y365F/V395I/L517T/H519G/A520C/Q564C/L560Q/F562Y/I569S; and/or Y365F/V395I/L517S/H519R/A520H/P521Q/N544L/L546V/Q564W/F562T/L560Q/F562Y/I569S.

Embodiment 21. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises amino acid substitutions at positions 365, 395, 517, 519, and 520; and one or more, or two or more amino acid substitutions at positions 518 and 392, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1.

Embodiment 22. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises the amino acid substitutions comprising Y365F/V395I/L517T/H519S/A520G; Y365F/V395I/L517T/L518G/H519D/A520G; Y365F/V395I/L517T/L518S/H519S/A520G; Y365F/F392W/V395I/L517T/H519S/A520G; Y365F/F392W/V395I/L517T/L518G/H519D/A520G; and/or Y365F/F392W/V395I/L517T/L518S/H519S/A520G.

Embodiment 23. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises amino acid substitutions at positions 329, 365, 395, 517, 519, 520, 560, 562, and 569; and one or more, two or more, or three or more amino acid substitutions at positions 392, 544, and 564, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1.

Embodiment 24. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises the amino acid substitutions comprising Y365F/V395I/L517T/H519S/A520G/N544M/L560Q/F562Y/I569S/F329Y; Y365F/V395I/L517T/H519S/A520G/N544Q/L560Q/F562Y/I569S/F329Y; Y365F/V395I/L517T/H519S/A520G/Q564N/L560Q/F562Y/I569S/F329Y; Y365F/F392W/V395I/L517T/H519S/A520G/N544M/L560Q/F562Y/I569S/F329Y; Y365F/F392W/V395I/L517T/H519S/A520G/N544Q/L560Q/F562Y/I569S/F329Y; and/or Y365F/F392W/V395I/L517T/H519S/A520G/Q564N/L560Q/F562Y/I569S/F329Y.

Embodiment 25. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises the amino acid substitutions comprising Y365F/V395I/L517T/H519S/A520G/N544M/L560Q/F562Y/I569S; Y365F/V395I/L517T/H519S/A520G/N544Q/L560Q/F562Y/I569S; Y365F/V395I/L517T/H519S/A520G/Q564N/L560Q/F562Y/I569S; Y365F/F392W/V395I/L517T/H519S/A520G/N544M/L560Q/F562Y/I569S; Y365F/F392W/V395I/L517T/H519S/A520G/N544Q/L560Q/F562Y/I569S; and/or Y365F/F392W/V395I/L517T/H519S/A520G/Q564N/L560Q/F562Y/I569S.

Embodiment 26. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises amino acid substitutions at positions 365, 395, 517, 519, 520, 521, 504, 560, 562, 569, and 329; and one or more amino acid substitutions at positions 392, the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1

Embodiment 27. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises the amino acid substitutions comprising

Y365F/V395I/L517T/H519G/P521A/A520C/Q564C/L560Q/F562Y/I569S/F329Y;
 Y365F/V395I/L517T/H519G/P521N/A520C/Q564C/L560Q/F562Y/I569S/F329Y;
 Y365F/F392W/V395I/L517T/H519G/P521A/A520C/Q564C/L560Q/F562Y/I569S/F329Y
 ; and/or
 Y365F/F392W/V395I/L517T/H519G/P521N/A520C/Q564C/L560Q/F562Y/I569S/F329Y
 .

Embodiment 28. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises a plurality of amino acid substitutions at positions 329, 365, 395, 526, 527, 528, 560, 562, and 569.

Embodiment 29. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises the amino acid substitutions comprising F329Y/Y365F/V395I/G526S/P527N/K528T/L560Q/F562Y/I569S; and/or F329Y/Y365F/V395I/G526S/P527N/K528I/L560Q/F562Y/I569S.

Embodiment 30. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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%, or at least 99% identical to any one of SEQ ID NOs: 11-19.

Embodiment 31. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises a polypeptide sequence identical to any one of SEQ ID NOs: 11-19.

Embodiment 32. The polypeptide of any one of the previous Embodiments, wherein the polypeptide expresses at a higher level than a reference polypeptide without the amino acid substitutions, optionally a reference polypeptide according to SEQ ID NOs: 2 or 10.

Embodiment 33. The polypeptide of any one of the previous Embodiments, wherein said expressed polypeptide is more thermally stable or more antigenic than SEQ ID NOs: 2 or 10.

Embodiment 34. The polypeptide of any one of the previous Embodiments, wherein the polypeptide comprises a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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%, or at least 99% identical to SEQ ID NOs: 2-10.

Embodiment 35. A self-assembling protein nanostructure, comprising a first component comprising a recombinant coronavirus polypeptide according to any one of the Embodiments 1 to 34; and optionally a second component comprising a second protein.

Embodiment 36. The protein nanostructure of Embodiment 35, wherein the recombinant coronavirus polypeptide comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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%, or at least 99% identical to SEQ ID NOs: 11-19.

Embodiment 37. The protein nanostructure of Embodiment 35 or Embodiment 36, wherein the recombinant coronavirus polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acids relative to a reference sequence according to SEQ ID NO: 1.

Embodiment 38. The protein nanostructure of any one of Embodiments 35 to 37, wherein a protein nanostructure component polypeptide and the recombinant coronavirus polypeptide are non-covalently coupled.

Embodiment 39. The protein nanostructure of any one of Embodiments 35 to 38, wherein a protein nanostructure component polypeptide and the recombinant coronavirus polypeptide are covalently coupled.

Embodiment 40. The protein nanostructure of any one of Embodiments 35 to 39, wherein the protein nanostructure comprises a fusion protein, the fusion protein comprising the recombinant coronavirus polypeptide, a linker, and a protein nanostructure component polypeptide.

Embodiment 41. The protein nanostructure of any one of Embodiments 35 to 40, wherein the protein nanostructure comprises a fusion protein, the fusion protein comprising, in N- to C-terminal order, the recombinant coronavirus polypeptide, a linker, and the protein nanostructure component polypeptide.

Embodiment 42. The protein nanostructure of any one of Embodiments 35 to 41, wherein the recombinant coronavirus polypeptide comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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%, or at least 99% identical to SEQ ID NOs: 11-19.

Embodiment 43. The protein nanostructure of any one of Embodiments 35 to 42, wherein the fusion protein comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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%, or at least 99% identical to SEQ ID NOs: 11-19.

Embodiment 44. The protein nanostructure of any one of Embodiments 35 to 43, wherein the first component comprises I53-50A.

Embodiment 45. The protein nanostructure of any one of Embodiments 35 to 44, wherein the protein nanostructure is an I53-50AB protein nanostructure; or an I3-01/MI3 protein nanostructure.

Embodiment 46. The protein nanostructure of any one of Embodiments 35 to 45, wherein the protein nanostructure is a ferritin protein nanostructure; an encapsulin protein nanostructure; a CP3 bacteriophage capsid protein nanostructure; a Q β bacteriophage capsid protein nanostructure; or an AP205 bacteriophage capsid protein nanostructure.

Embodiment 47. The protein nanostructure of any one of Embodiments 35 to 46, wherein the recombinant polypeptide comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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%, or at least 99% identical to SEQ ID NOs: 11-19.

Embodiment 48. The protein nanostructure of any one of Embodiments 35 to 47, wherein the protein nanostructure component polypeptide comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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%, or at least 99% identical to SEQ ID NO: 20-21.

Embodiment 49. The protein nanostructure of any one of Embodiments 35 to 48, wherein the protein nanostructure comprises a second component polypeptide.

Embodiment 50. The protein nanostructure of any one of Embodiments 35 to 49, wherein the second component polypeptide is I53-50B.

Embodiment 51. The protein nanostructure of any one of Embodiments 35 to 50, wherein the second component polypeptide comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at

least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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%, or at least 99% identical to SEQ ID NO: 22.

Embodiment 52. The protein nanostructures of any one of Embodiments 35 to 51, wherein the linker comprises between 8 and 24 amino acid residues.

53. The protein nanostructures of any one of Embodiments 35 to 52, wherein the linker is selected from the group consisting of GGSGSGSGSGSGSGS; SGGSGSGSGSGSGSGS; EPEGSGSGSGSGSGSGS; YGGSGSGSGSGSGSGSGS; and GSGSGSGSGSGSGSGSGSGSGS.

Embodiment 54. The protein nanostructures of any one of Embodiments 35 to 53, wherein the recombinant polypeptide comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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%, or at least 99% identical to polypeptide sequence shown in Table 27.

Embodiment 55. A polynucleotide encoding the polypeptide of any one of Embodiment 1 to 34 or the protein nanostructure of any one of Embodiment 35 to 54.

Embodiment 56. The polynucleotide of Embodiment 55, wherein the polynucleotide is mRNA.

Embodiment 57. A pharmaceutical composition comprising the polypeptide of any one of Embodiment 1 to 34, the protein nanostructure of any one of Embodiment 35 to 54, or the polynucleotide of any one of Embodiment 55 to 56.

Embodiment 58. A vaccine comprising the polypeptide of any one of Embodiment 1 to 34, the protein nanostructure of any one of Embodiment 35 to 54, or the polynucleotide of any one of Embodiment 55 to 56.

Embodiment 59. The vaccine of Embodiment 58, further comprising an adjuvant.

Embodiment 60. The vaccine of Embodiment 58 or Embodiment 59, wherein the adjuvant comprises squalene.

Embodiment 61. The vaccine of any one of Embodiment 58 to 60, wherein the vaccine is a bivalent vaccine.

Embodiment 62. The vaccine of any one of Embodiments 58 to 61, wherein the bivalent vaccine comprises the polypeptide of any one of Embodiment 1 to 34 or the protein nanostructure of any one of Embodiment 35 to 54.

Embodiment 63. A method of preventing coronavirus disease in a subject in need thereof, comprises administering to the subject the polypeptide of any one of Embodiment 1 to 34, the protein nanostructure of any one of Embodiment 35 to 54, the polynucleotide of any one of Embodiment 55 to 56, the pharmaceutical component of c Embodiment 57 , the vaccine of any one of Embodiment 58 to 62.

Embodiment 64. A method of immunizing a subject in need thereof against infection by coronavirus, the method comprising administering to the subject the vaccine of any one of Embodiment 58 to 62.

Embodiment 65. A kit comprising the polypeptide of any one of Embodiment 1 to 34, the protein nanostructure of any one of Embodiment 35 to 54, the polynucleotide of any one of Embodiment 55 to 56, the pharmaceutical component of Embodiment 57, the vaccine of any one of Embodiment 58 to 62, and instructions for use thereof

While the invention has been described in connection with proposed specific embodiments thereof, it will be understood that it is capable of further modifications, and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles described herein, with such departures that may come within known or customary practice within the art to which the invention pertains.

CLAIMS

What is claimed is:

1. A recombinant polypeptide, comprising an antigenic fragment of a coronavirus Spike glycoprotein, said antigenic fragment comprising a receptor-binding domain (RBD) and optionally, a coronavirus subdomain 1 (SD1).
2. The polypeptide of claim 1, wherein the antigenic fragment comprises RBD-SD1.
3. The polypeptide of claim 1, wherein the SD1 comprises at least one substitution of a surface-exposed, non-polar amino acid residue with a polar amino acid residue.
4. The polypeptide of claim 3, wherein the polypeptide comprises a substitution of one or more, two or more, or three amino acids at positions 560, 562, or 569 with a polar residue, wherein the substitutions are relative to a reference sequence according to SEQ ID NO: 1.
5. The polypeptide of claim 3, wherein the polypeptide comprises a substitution of one or more, two or more, or three of the amino acid substitutions L560Q, F562T, F562Y, or I569S relative to a reference sequence according to SEQ ID NO: 1.
6. The polypeptide of claim 3, wherein the polypeptide comprises a substitution of one or more, two or more, or three or more amino acids at positions 544, 546, 560, 562, 564, 569, or 582 relative to a reference sequence according to SEQ ID NO: 1.
7. The polypeptide of claim 3, wherein the polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acids at positions N544L, N544M, N544Q, L546V, L560Q, F562T, F562Y, Q564C, Q564L, Q564N, Q564W, I569S, or L582S relative to a reference sequence according to SEQ ID NO: 1.
8. The polypeptide of claim 1, wherein the polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acids at positions 338, 358, 363, 365, 392, or 395 relative to a reference sequence according to SEQ ID NO: 1.
9. The polypeptide of claim 1, wherein the polypeptide comprises a substitution of one or more, two or more, three or more, four or more

amino acids at positions F338L, I358F, A363L, Y365F, Y365M, Y365W, F392W, or V395I relative to a reference sequence according to SEQ ID NO: 1.

10. The polypeptide of claim 1, wherein the polypeptide comprises amino acid substitutions comprising:
 - F338L/Y365W;
 - F392W;
 - Y365F/V395I;
 - Y365F/ F392W /V395I;
 - F338L/A363L/Y365M; and/or
 - I358F/Y365F/V395I.
11. The polypeptide of claim 1, wherein the polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acids at positions 329, 348, 350, 367, 375, 402, 407, 410, 418, 429, 433, 435, 452, 464, 510, 512, 514, 517, 518, 519, 520, 522, 527, or 528 relative to a reference sequence according to SEQ ID NO: 1.
12. The polypeptide of claim 1, wherein the polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acids at positions F329K, F329R, F329Y, A348P, V350L, V367F, F375Y, I402V, V407L, I410F, I418V, F429W, V433I, A435I, A435V, L452R, F464Y, P512Q, V510I, V512F, V512I, S514T, L517A, L517D, L517S, L517T, L518N, L518Q, L518V, H519D, H519G, H519R, H519S, H519T, A520C, A520D, A520G, A520H, P521A, P521D, P521N, P521Q, P521S, A522G, A522I, P527N, K528I, K528Q, K528T, L518G, or L518S relative to a reference sequence according to SEQ ID NO: 1.
13. The polypeptide of claim 1, wherein the polypeptide comprises amino acid substitutions at positions 365, 395, 560, 562, and 569; and
 - one or more, two or more, or three or more amino acid substitutions at positions 348, 402, 464, 514, 520, 526, and 527,
 - the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1.
14. The polypeptide of claim 1, wherein the polypeptide comprises the amino acid substitutions comprising:

- A348P/Y365F/V395I/L560Q/F562Y/I569S;
Y365F/V395I/I402V/L560Q/F562Y/I569S;
Y365F/V395I/S514T/L560Q/F562Y/I569S;
Y365F/V395I/I402V/F464Y/L560Q/F562Y/I569S; and/or
Y365F/V395I/G526S/P527N/L560Q/F562Y/I569S.
15. The polypeptide of claim 1, wherein the polypeptide comprises
amino acid substitutions at positions 365, 395, 560, 562, and 569;
and
one or more, two or more, or three or more amino acid substitutions
at positions 329, 517, 519, 520, and 544,
the amino acid substitutions relative to a reference sequence
according to SEQ ID NO: 1.
16. The polypeptide of claim 1, wherein the polypeptide comprises the amino
acid substitutions comprising:
Y365F/V395I/L560Q/F562Y/I569S;
Y365F/V395I/F329Y/L560Q/F562Y/I569S;
Y365F/V395I/L517T/H519S/L560Q/F562Y/I569S;
Y365F/V395I/L517T/H519S/A520G/L560Q/F562Y/I569S;
Y365F/V395I/L517T/H519S/N544L/L560Q/F562Y/I569S; and/or
Y365F/V395I/F329Y/L517T/H519S/L560Q/F562Y/I569S.
17. The polypeptide of claim 1, wherein the polypeptide comprises
amino acid substitutions at positions 365, 392, 395, 560, 562, and
569; and
one or more, two or more, or three or more amino acid substitutions
at positions 329, 517, 519, 520, and 544,
the amino acid substitutions relative to a reference sequence
according to SEQ ID NO: 1.
18. The polypeptide of claim 1, wherein the polypeptide comprises the amino
acid substitutions comprising:
Y365F/F392W/V395I/L560Q/F562Y/I569S;
Y365F/F392W/V395I/F329Y/L560Q/F562Y/I569S;
Y365F/F392W/V395I/L517T/H519S/L560Q/F562Y/I569S;
Y365F/F392W/V395I/L517T/H519S/A520G/L560Q/F562Y/I569S;

- Y365F/F392W/V395I/L517T/H519S/N544L/L560Q/F562Y/I569S;
and/or
Y365F/F392W/V395I/F329Y/L517T/H519S/L560Q/F562Y/I569S.
19. The polypeptide of claim 1, wherein the polypeptide comprises
amino acid substitutions at positions 365, 395, 517, 519, 520, 560,
562, 564, and 569; and
one or more, two or more, or three or more amino acid substitutions
at positions 521, 544, and 546,
the amino acid substitutions relative to a reference sequence
according to SEQ ID NO: 1.
20. The polypeptide of claim 1, wherein the polypeptide comprises the amino
acid substitutions comprising:
Y365F/V395I/L517T/H519G/A520C/Q564C/L560Q/F562Y/I569S; and/or
Y365F/V395I/L517S/H519R/A520H/P521Q/N544L/L546V/Q564W/F562
T/L560Q/F562Y/I569S.
21. The polypeptide of claim 1, wherein the polypeptide comprises
amino acid substitutions at positions 365, 395, 517, 519, and 520;
and
one or more, or two or more amino acid substitutions at positions
518 and 392,
the amino acid substitutions relative to a reference sequence
according to SEQ ID NO: 1.
22. The polypeptide of claim 1, wherein the polypeptide comprises the amino
acid substitutions comprising:
Y365F/V395I/L517T/H519S/A520G;
Y365F/V395I/L517T/L518G/H519D/A520G;
Y365F/V395I/L517T/L518S/H519S/A520G;
Y365F/F392W/V395I/L517T/H519S/A520G;
Y365F/F392W/V395I/L517T/L518G/H519D/A520G; and/or
Y365F/F392W/V395I/L517T/L518S/H519S/A520G.
23. The polypeptide of claim 1, wherein the polypeptide comprises
amino acid substitutions at positions 329, 365, 395, 517, 519, 520,
560, 562, and 569; and

one or more, two or more, or three or more amino acid substitutions at positions 392, 544, and 564,

the amino acid substitutions relative to a reference sequence according to SEQ ID NO: 1.

24. The polypeptide of claim 1, wherein the polypeptide comprises the amino acid substitutions comprising:

Y365F/V395I/L517T/H519S/A520G/N544M/L560Q/F562Y/I569S/F329Y;

Y365F/V395I/L517T/H519S/A520G/N544Q/L560Q/F562Y/I569S/F329Y;

Y365F/V395I/L517T/H519S/A520G/Q564N/L560Q/F562Y/I569S/F329Y;

Y365F/F392W/V395I/L517T/H519S/A520G/N544M/L560Q/F562Y/I569S/F329Y;

Y365F/F392W/V395I/L517T/H519S/A520G/N544Q/L560Q/F562Y/I569S/F329Y; and/or

Y365F/F392W/V395I/L517T/H519S/A520G/Q564N/L560Q/F562Y/I569S/F329Y.

25. The polypeptide of claim 1, wherein the polypeptide comprises the amino acid substitutions comprising:

Y365F/V395I/L517T/H519S/A520G/N544M/L560Q/F562Y/I569S;

Y365F/V395I/L517T/H519S/A520G/N544Q/L560Q/F562Y/I569S;

Y365F/V395I/L517T/H519S/A520G/Q564N/L560Q/F562Y/I569S;

Y365F/F392W/V395I/L517T/H519S/A520G/N544M/L560Q/F562Y/I569S;

Y365F/F392W/V395I/L517T/H519S/A520G/N544Q/L560Q/F562Y/I569S; and/or

Y365F/F392W/V395I/L517T/H519S/A520G/Q564N/L560Q/F562Y/I569S.

26. The polypeptide of claim 1, wherein the polypeptide comprises amino acid substitutions at positions 365, 395, 517, 519, 520, 521, 504, 560, 562, 569, and 329; and one or more amino acid substitutions at positions 392,

the amino acid substitutions relative to a reference sequence
according to SEQ ID NO: 1

27. The polypeptide of claim 1, wherein the polypeptide comprises the amino acid substitutions comprising:
 Y365F/V395I/L517T/H519G/P521A/A520C/Q564C/L560Q/F562Y
 /I569S/F329Y;
 Y365F/V395I/L517T/H519G/P521N/A520C/Q564C/L560Q/F562Y
 /I569S/F329Y;
 Y365F/F392W/V395I/L517T/H519G/P521A/A520C/Q564C/L560
 Q/F562Y/I569S/F329Y; and/or
 Y365F/F392W/V395I/L517T/H519G/P521N/A520C/Q564C/L560
 Q/F562Y/I569S/F329Y.
28. The polypeptide of claim 1, wherein the polypeptide comprises a plurality of amino acid substitutions at positions 329, 365, 395, 526, 527, 528, 560, 562, and 569.
29. The polypeptide of claim 1, wherein the polypeptide comprises the amino acid substitutions comprising:
 F329Y/Y365F/V395I/G526S/P527N/K528T/L560Q/F562Y/I569S;
 and/or
 F329Y/Y365F/V395I/G526S/P527N/K528I/L560Q/F562Y/I569S.
30. The polypeptide of claim 1, wherein the polypeptide comprises a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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%, or at least 99% identical to any one of SEQ ID NOs: 11-19.
31. The polypeptide of claim 1, wherein the polypeptide comprises a polypeptide sequence identical to any one of SEQ ID NOs: 11-19.
32. The polypeptide of claim 1, wherein the polypeptide expresses at a higher level than a reference polypeptide without the amino acid substitutions, optionally a reference polypeptide according to SEQ ID NOs: 2 or 10.
33. The polypeptide of claim 1, wherein said expressed polypeptide is more thermally stable or more antigenic than SEQ ID NOs: 2 or 10.

34. The polypeptide of claim 1, wherein the polypeptide comprises a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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%, or at least 99% identical to SEQ ID NOs: 2-10.
35. A self-assembling protein nanostructure, comprising:
a first component comprising a recombinant coronavirus polypeptide according to any one of claims 1 to 34; and
optionally a second component comprising a second protein.
36. The protein nanostructure of claim 35, wherein the recombinant coronavirus polypeptide comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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%, or at least 99% identical to SEQ ID NOs: 11-19.
37. The protein nanostructure of claim 35, wherein the recombinant coronavirus polypeptide comprises a substitution of one or more, two or more, three or more, four or more amino acids relative to a reference sequence according to SEQ ID NO: 1.
38. The protein nanostructure of claim 35, wherein a protein nanostructure component polypeptide and the recombinant coronavirus polypeptide are non-covalently coupled.
39. The protein nanostructure of claim 35, wherein a protein nanostructure component polypeptide and the recombinant coronavirus polypeptide are covalently coupled.
40. The protein nanostructure of claim 39, wherein the protein nanostructure comprises a fusion protein, the fusion protein comprising the recombinant coronavirus polypeptide, a linker, a protein nanostructure component polypeptide.
41. The protein nanostructure of claim 39, wherein the protein nanostructure comprises a fusion protein, the fusion protein comprising, in N- to C-

- terminal order, the recombinant coronavirus polypeptide, a linker, and the protein nanostructure component polypeptide.
42. The protein nanostructure of claim 39, wherein the recombinant coronavirus polypeptide comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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%, or at least 99% identical to SEQ ID NOs: 11-19.
43. The protein nanostructure of claim of 40, wherein the fusion protein comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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%, or at least 99% identical to SEQ ID NOs: 11-19.
44. The protein nanostructure of claim 39, wherein the first component comprises I53-50A.
45. The protein nanostructure of claim 35, wherein the protein nanostructure is
an I3-01/MI3 protein nanostructure; or
an I53-50AB protein nanostructure;
46. The protein nanostructure of claim 35, wherein the protein nanostructure is
a ferritin protein nanostructure;
an encapsulin protein nanostructure;
a CP3 bacteriophage capsid protein nanostructure;
a Q β bacteriophage capsid protein nanostructure; or
an AP205 bacteriophage capsid protein nanostructure.
47. The protein nanostructure of claim 45, wherein the recombinant polypeptide comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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%, or at least 99% identical to SEQ ID NOs: 11-19.

48. The protein nanostructure of claim 35, wherein the protein nanostructure component polypeptide comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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%, or at least 99% identical to SEQ ID NO: 20-21.
49. The protein nanostructure of claim 35, wherein the protein nanostructure comprises a second component polypeptide.
50. The nanoparticle of claim 48, wherein the second component polypeptide is I53-50B.
51. The protein nanostructure of claim 49, wherein the second component polypeptide comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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%, or at least 99% identical to SEQ ID NO: 22.
52. The protein nanostructures of claim 35, wherein the linker comprises between 8 and 24 amino acid residues.
53. The protein nanostructures of claim 35, wherein the linker is selected from the group consisting of:
GGSGGSGSGSGSGSGS;
SGGGSGGSGSGSGSGS;
EPEGGSGGSGSGSGSGS;
YGGSGGSGGSGSGSGSGS; and
GSGGSGGSGGSGGSGSGSGSGS.
54. The protein nanostructures of claim 35, wherein the recombinant polypeptide comprises a polypeptide segment having a polypeptide sequence at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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%, or at least 99% identical to polypeptide sequence shown in Table 18.

55. A polynucleotide encoding the polypeptide of any one of claims 1 to 34 or the protein nanostructure of any one of claims 35 to 54.
56. The polynucleotide of claim 55, wherein the polynucleotide is mRNA.
57. A pharmaceutical composition comprising the polypeptide of any one of claims 1 to 34, the protein nanostructure of any one of claims 35 to 54, or the polynucleotide of any one of claims 55 to 56.
58. A vaccine comprising the polypeptide of any one of claims 1 to 34, the protein nanostructure of any one of claims 35 to 54, or the polynucleotide of any one of claims 55 to 56.
59. The vaccine of claim 58, further comprising an adjuvant.
60. The vaccine of claim 59, wherein the adjuvant comprises squalene.
61. The vaccine of any one of claims 58 to 60, wherein the vaccine is a bivalent vaccine.
62. The vaccine of claim 61, wherein the bivalent vaccine comprises the polypeptide of any one of claims 1 to 34 or the protein nanostructure of any one of claims 35 to 54.
63. A method of preventing coronavirus disease in a subject in need thereof, comprises administering to the subject the polypeptide of any one of claims 1 to 34, the protein nanostructure of any one of claims 35 to 54, the polynucleotide of any one of claims 55 to 56, the pharmaceutical component of claim 57, the vaccine of any one of claims 58 to 62.
64. A method of immunizing a subject in need thereof against infection by coronavirus, the method comprising administering to the subject the vaccine of any one of claims 58 to 62.
65. A kit comprising the polypeptide of any one of claims 1 to 34, the protein nanostructure of any one of claims 35 to 54, the polynucleotide of any one of claims 55 to 56, the pharmaceutical component of claim 57, the vaccine of any one of claims 58 to 62, and instructions for use thereof.