

VECTORS FOR PRODUCING VIRUS-LIKE PARTICLES AND USES THEREOF

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This PCT application claims the priority benefit of U.S. Provisional Application Nos. 63/003,281, filed March 31, 2020, and 63/124,397, filed December 11, 2020, which are incorporated herein by reference in their entireties.

FIELD OF DISCLOSURE

[0002] The present disclosure provides vectors for producing virus-like particles (VLPs) and methods of treating subjects with the same.

BACKGROUND

[0003] Despite numerous advances in vaccine technologies, viral infections remain a prevalent health concern that are often under limited control. For example, the COVID-19 coronavirus pandemic became unlike anything the world had seen in over a century, both in terms of global spread and economic impact. It resulted in repeated shutdowns in much of the developed world, with continuously increasing death tolls and new infections.

[0004] COVID-19 causes a respiratory infection, along with acute respiratory distress syndrome in severe cases. Pre/asymptomatic airborne transmission and high viral titre early in the course of the disease significantly increase the infectiousness of COVID-19 compared to other coronaviruses such as SARS-CoV, making the development of vaccines critical for management of the pandemic.

[0005] VLPs represent potent vaccine candidates that mimic viral physicochemical properties and structure without potentiating viral growth (Cimica, V., & Galarza, J. M., *Clin. Immunol.* 183: 99–108 (2017)). As such, they confer strong humoral responses, but often limited cell-mediated responses against the 'whole virus' as they remain exogenously administered antigens. Furthermore, their production, purification, and storage are costly.

- [0006] Existing vaccines have often shown limited cross-protection among different viral strains, complicated by the fact that viruses continue to mutate their genomes in response to evolutionary pressures.
- [0007] There is a need for improved VLPs and methods of treating viral infections.

BRIEF SUMMARY

- [0008] The present disclosure is directed to an expression vector comprising: an expression cassette that comprises a nucleic acid sequence encoding a recombinant protein comprising a conserved amino acid sequence from a virus fused to an immunogenic amino acid sequence, a target sequence for a first recombinase flanking each side of the expression cassette, and one or more additional target sequences for one or more additional recombinases integrated within non-binding regions of the target sequence for the first recombinase, wherein protein expressed intracellularly from the expression cassette is capable of forming a virus-like particle (VLP).
- [0009] In some aspects, the immunogenic amino acid sequence is from the same virus as the conserved amino acid sequence. In some aspects, the conserved amino acid sequence is from a viral glycoprotein. In some aspects, the immunogenic amino acid sequence is from the same viral glycoprotein.
- [0010] In some aspects, the expression cassette further comprises a nucleic acid sequence encoding a viral envelope protein and/or a nucleic acid sequence encoding a viral matrix protein. In some aspects, the viral envelope protein and/or the viral matrix protein are from the same virus as the conserved amino acid sequence.
- [0011] In some aspects, the conserved amino acid sequence, the immunogenic amino acid sequence, the viral envelope protein, and/or the viral matrix protein is a consensus sequence.
- [0012] In some aspects, the recombinant protein is capable of stimulating an immune response against the virus comprising neutralizing antibodies.
- [0013] In some aspects, the recombinant protein is capable of stimulating a Th1 cell-mediated immune response against the virus.
- [0014] In some aspects, the immune response is cross-reactive to a related virus or strain.

- [0015] In some aspects, the recombinant protein excludes amino acid sequences from the virus that stimulate an immune response comprising non-neutralizing antibodies and/or that stimulate a Th2 cell-mediated immune response.
- [0016] In some aspects, the expression cassette comprises a single open reading frame comprising a nucleic acid sequence encoding a self-cleaving peptide between each nucleic acid sequence encoding a protein.
- [0017] In some aspects, the virus is a coronavirus, an influenza virus, a human immunodeficiency virus, a human papillomavirus, a hepatitis virus, or an oncolytic virus.
- [0018] In some aspects, the virus is a coronavirus. In some aspects, the coronavirus is COVID-19.
- [0019] In some aspects, the expression cassette comprises nucleic acid sequences encoding a coronavirus Membrane (M) protein, a coronavirus Envelope (E) protein, and a recombinant protein comprising a conserved amino acid sequence and an immunogenic amino acid sequence from a coronavirus Spike (S) protein. In some aspects, the conserved amino acid sequence is from the S protein S2' cleavage site and internal fusion peptide (IFP).
- [0020] In some aspects, the conserved amino acid sequence comprises SEQ ID NO:12.
- [0021] In some aspects, the immunogenic amino acid sequence is from the S protein receptor-binding domain (RBD).
- [0022] In some aspects, the immunogenic amino acid sequence is at least about 90% identical to SEQ ID NO:11.
- [0023] In some aspects, the recombinant protein further comprises a transmembrane (TM) domain sequence from the S protein.
- [0024] In some aspects, the recombinant protein excludes amino acid sequences from the S protein that stimulate an immune response comprising non-neutralizing antibodies and/or that stimulate a Th2 cell-mediated immune response.
- [0025] In some aspects, the amino acid sequence of the recombinant protein is at least about 90% identical to SEQ ID NO:55.
- [0026] In some aspects, the expression cassette comprises a single open reading frame translated as an amino acid sequence at least about 90% identical to SEQ ID NO:57.
- [0027] In some aspects, the recombinant protein is capable of stimulating an immune response against COVID-19.

- [0028] In some aspects, the recombinant protein is capable of stimulating a Th1 cell-mediated immune response against COVID-19.
- [0029] In some aspects, the immune response is cross-reactive to other coronaviruses. In some aspects, the immune response is cross-reactive to other severe acute respiratory syndrome coronaviruses and/or human betacoronaviruses.
- [0030] In some aspects, the target sequence for the first recombinase and the one or more additional target sequences for the one or more additional recombinases are selected from the group consisting of the PY54 pal site, the N15 telRL site, the loxP site, ϕ K02 telRL site, the FRT site, the phiC31 attP site, and the λ attP site. In some aspects, the expression vector comprises each of the target sequences. In some aspects, the expression vector comprises the Tel recombinase pal site and the telRL, loxP, and FRT recombinase target binding sequences integrated within the pal site.
- [0031] In some aspects, the expression vector is for producing a bacterial sequence-free vector. In some aspects, the bacterial sequence-free vector has circular covalently closed ends. In some aspects, the bacterial sequence-free vector has linear covalently closed ends.
- [0032] In some aspects, the expression vector further comprises at least one enhancer sequence flanking each side of the target sequence for the first recombinase. In some aspects, the at least one enhancer sequence is at least two enhancer sequences. In some aspects, the at least one enhancer sequence is a SV40 enhancer sequence.
- [0033] The present disclosure is directed to a vector production system comprising recombinant cells designed to encode at least a first recombinase under the control of an inducible promoter, wherein the cells comprise any of the above expression vectors. In some aspects, the inducible promoter is thermally-regulated, chemically-regulated, IPTG regulated, glucose-regulated, arabinose inducible, T7 polymerase regulated, cold-shock inducible, pH inducible, or combinations thereof. In some aspects, the first recombinase is selected from telN and tel, and the expression vector incorporates the target sequence for at least the first recombinase. In some aspects, the recombinant cells have been further designed to encode a nuclease genome editing system, and wherein the expression vector further comprises a backbone sequence containing a cleavage site for the nuclease genome editing system. In some aspects, the nuclease genome editing system is a

CRISPR nuclease system comprising a Cas nuclease and gRNA, and the expression vector comprises a target sequence for the gRNA within the backbone sequence.

[0034] The present disclosure is directed to a method of producing a bacterial sequence-free vector comprising incubating any of the above vector production systems under suitable conditions for expression of the first recombinase.

[0035] The present disclosure is directed to a method of producing a bacterial sequence-free vector comprising incubating any of the above vector production systems that comprise recombinant cells designed to encode a nuclease genome editing system under suitable conditions for expression of the first recombinase and the nuclease genome editing system.

[0036] In some aspects, any of the above methods of producing a bacterial sequence-free vector further comprise harvesting the bacterial sequence-free vector.

[0037] The present disclosure is directed to a bacterial sequence-free vector produced by any of the above methods of producing a bacterial sequence-free vector.

[0038] The present disclosure is directed to a bacterial sequence-free vector comprising an expression cassette that comprises a nucleic acid sequence encoding a recombinant protein comprising a conserved amino acid sequence from a virus fused to an immunogenic amino acid sequence, wherein protein expressed intracellularly from the expression cassette is capable of forming a VLP.

[0039] In some aspects, the immunogenic amino acid sequence is from the same virus as the conserved amino acid sequence. In some aspects, the conserved amino acid sequence is from a viral glycoprotein. In some aspects, the immunogenic amino acid sequence is from the same viral glycoprotein.

[0040] In some aspects, the expression cassette further comprises a nucleic acid sequence encoding a viral envelope protein and/or a nucleic acid sequence encoding a viral matrix protein. In some aspects, the viral envelope protein and/or the viral matrix protein are from the same virus as the conserved amino acid sequence.

[0041] In some aspects, the conserved amino acid sequence, the immunogenic amino acid sequence, the viral envelope protein, and/or the viral matrix protein is a consensus sequence.

[0042] In some aspects, the recombinant protein is capable of stimulating an immune response against the virus comprising neutralizing antibodies.

- [0043] In some aspects, the recombinant protein is capable of stimulating a Th1 cell-mediated immune response against the virus.
- [0044] In some aspects, the immune response is cross-reactive to a related virus or strain.
- [0045] In some aspects, the recombinant protein excludes amino acid sequences from the virus that stimulate an immune response comprising non-neutralizing antibodies and/or that stimulate a Th2 cell-mediated immune response.
- [0046] In some aspects, the expression cassette comprises a single open reading frame comprising a nucleic acid sequence encoding a self-cleaving peptide between each nucleic acid sequence encoding a protein.
- [0047] In some aspects, the virus is a coronavirus, an influenza virus, a human immunodeficiency virus, a human papillomavirus, a hepatitis virus, or an oncolytic virus.
- [0048] In some aspects, the virus is a coronavirus. In some aspects, the coronavirus is COVID-19.
- [0049] In some aspects, the expression cassette comprises nucleic acid sequences encoding a coronavirus M protein, a coronavirus E protein, and a recombinant protein comprising a conserved amino acid sequence and an immunogenic amino acid sequence from a coronavirus S protein. In some aspects, the conserved amino acid sequence is from the S protein S2' cleavage site and IFP.
- [0050] In some aspects, the conserved amino acid sequence comprises SEQ ID NO:12.
- [0051] In some aspects, the immunogenic amino acid sequence is from the S protein RBD.
- [0052] In some aspects, the immunogenic amino acid sequence is at least about 90% identical to SEQ ID NO:11.
- [0053] In some aspects, the recombinant protein further comprises a TM domain sequence from the S protein.
- [0054] In some aspects, the recombinant protein excludes amino acid sequences from the S protein that stimulate an immune response comprising non-neutralizing antibodies and/or that stimulate a Th2 cell-mediated immune response.
- [0055] In some aspects, the amino acid sequence of the recombinant protein is SEQ ID NO:55.
- [0056] In some aspects, the expression cassette comprises a single open reading frame translated as an amino acid sequence at least about 90% identical to SEQ ID NO:57.

- [0057] In some aspects, the recombinant protein is capable of stimulating an immune response against COVID-19.
- [0058] In some aspects, the recombinant protein is capable of stimulating a Th1 cell-mediated immune response against COVID-19.
- [0059] In some aspects, the immune response is cross-reactive to other coronaviruses. In some aspects, the immune response is cross-reactive to other severe acute respiratory syndrome coronaviruses and/or human betacoronaviruses.
- [0060] In some aspects, the bacterial sequence-free vector further comprises at least one enhancer sequence flanking each side of the expression cassette. In some aspects, the at least one enhancer sequence is at least two enhancer sequences. In some aspects, the at least one enhancer sequence is a SV40 enhancer sequence.
- [0061] In some aspects, the bacterial sequence-free vector comprises circular covalently closed ends.
- [0062] In some aspects, the bacterial sequence-free vector comprises linear covalently closed ends.
- [0063] The present disclosure is directed to a polynucleotide encoding an amino acid sequence at least about 90% identical to SEQ ID NO:57.
- [0064] The present disclosure is directed to a recombinant cell comprising any of the above expression vectors or any of the above bacterial sequence-free vectors.
- [0065] In some aspects, the present disclosure is directed to a method of producing a VLP, comprising culturing the recombinant cell under suitable conditions for production of the VLP from the expression vector or the bacterial sequence-free vector.
- [0066] In some aspects, the method of producing a VLP further comprises isolating the VLP. In some aspects, the isolating is by affinity purification. In some aspects, the VLP is produced by any of the above expression vectors or any of the above bacterial sequence-free vectors wherein the virus is a coronavirus. In some aspects, the affinity purification comprises an angiotensin-converting enzyme 2 (ACE2) receptor peptide or an anti-S protein monoclonal antibody. In some aspects, the ACE2 receptor peptide comprises an amino acid sequence that is at least about 90% identical to the amino acid sequence of SEQ ID NO:70. In some aspects, the ACE2 receptor peptide comprises a biotin acceptor peptide (BAP) tag at the C-terminus or N-terminus of the peptide. In some aspects, the BAP tag comprises an amino acid sequence at least about 90% identical to the amino acid

sequence of SEQ ID NO:71. In some aspects, the ACE2 receptor peptide or anti-S protein monoclonal antibody is biotinylated and immobilized on a streptavidin-coated bead. In some aspects, the affinity purification comprises microfluidics and/or chromatography. In some aspects, the present disclosure is directed to a VLP produced by any of the methods of producing a VLP.

[0067] The present disclosure is directed to a VLP comprising a recombinant protein comprising a conserved amino acid sequence from a virus fused to an immunogenic amino acid sequence. In some aspects, the immunogenic amino acid sequence is from the same virus as the conserved amino acid sequence. In some aspects, the conserved amino acid sequence is from a viral glycoprotein. In some aspects, the immunogenic amino acid sequence is from the same viral glycoprotein.

[0068] In some aspects, the VLP further comprises a viral envelope protein and/or a viral matrix protein. In some aspects, the viral envelope protein and/or the viral matrix protein are from the same virus as the conserved amino acid sequence.

[0069] In some aspects, the conserved amino acid sequence, the immunogenic amino acid sequence, the viral envelope protein, and/or the viral matrix protein is a consensus sequence.

[0070] In some aspects, the recombinant protein is capable of stimulating an immune response against the virus comprising neutralizing antibodies.

[0071] In some aspects, the recombinant protein is capable of stimulating a Th1 cell-mediated immune response against the virus.

[0072] In some aspects, the immune response is cross-reactive to a related virus or strain.

[0073] In some aspects, the recombinant protein excludes amino acid sequences from the virus that stimulate an immune response comprising non-neutralizing antibodies and/or that stimulate a Th2 cell-mediated immune response.

[0074] In some aspects, the virus is a coronavirus, an influenza virus, a human immunodeficiency virus, a human papillomavirus, a hepatitis virus, or an oncolytic virus. In some aspects, the virus is a coronavirus.

[0075] In some aspects, the coronavirus is COVID-19.

[0076] In some aspects, the VLP comprises a coronavirus Membrane (M) protein, a coronavirus Envelope (E) protein, and a recombinant protein comprising a conserved

amino acid sequence and an immunogenic amino acid sequence from a coronavirus Spike (S) protein.

- [0077] In some aspects, the conserved amino acid sequence is from the S protein S2' cleavage site and internal fusion peptide (IFP).
- [0078] In some aspects, the conserved amino acid sequence comprises SEQ ID NO:12.
- [0079] In some aspects, the immunogenic amino acid sequence is from the S protein receptor-binding domain (RBD).
- [0080] In some aspects, the immunogenic amino acid sequence is at least about 90% identical to SEQ ID NO:11.
- [0081] In some aspects, the recombinant protein further comprises a transmembrane (TM) domain sequence from the S protein.
- [0082] In some aspects, the recombinant protein excludes amino acid sequences from the S protein that stimulate an immune response comprising non-neutralizing antibodies and/or that stimulate a Th2 cell-mediated immune response.
- [0083] In some aspects, the amino acid sequence of the recombinant protein is at least about 90% identical to SEQ ID NO:55.
- [0084] The present disclosure is directed to a VLP comprising a recombinant protein at least about 90% identical to SEQ ID NO:55, an M protein at least about 90% identical to SEQ ID NO:1, and an E protein at least about 90% identical to SEQ ID NO:3.
- [0085] In some aspects, the recombinant protein is capable of stimulating an immune response against COVID-19.
- [0086] In some aspects, the recombinant protein is capable of stimulating a Th1 cell-mediated immune response against COVID-19.
- [0087] In some aspects, the immune response is cross-reactive to other coronaviruses.
- [0088] In some aspects, the immune response is cross-reactive to other severe acute respiratory syndrome coronaviruses and/or human betacoronaviruses.
- [0089] The present disclosure is directed to a composition comprising any of the above expression vectors, any of the above bacterial sequence-free vectors, or any of the above virus-like particles. In some aspects, the composition further comprises a delivery agent. In some aspects, the delivery agent is a nanoparticle. In some aspects, the delivery agent comprises a targeting ligand. In some aspects, the targeting ligand comprises a S protein

peptide. In some aspects, the S protein peptide comprises an amino acid sequence at least about 90% identical to any one of SEQ ID NOs:76-99.

- [0090] The present disclosure is directed to a method of treating a viral infection in a subject, comprising administering to the subject any of the above expression vectors, any of the above bacterial sequence-free vectors, any of the above VLPs, or any of the above compositions, wherein intracellular expression of the expression vector or the bacterial sequence-free vector produces a VLP.
- [0091] In some aspects, the administering is by parenteral or non-parenteral administration. In some aspects, the administering is by oral, pulmonary, intranasal, intravenous, epidermal, transdermal, subcutaneous, intramuscular, or intraperitoneal administration, or by inhalation.
- [0092] In some aspects, the VLP stimulates an immune response in the subject comprising neutralizing antibodies against the viral infection.
- [0093] In some aspects, the VLP stimulates a Th1 cell-mediated immune response in the subject against the viral infection.
- [0094] In some aspects, the immune response is cross-reactive to a related virus or strain.
- [0095] In some aspects, the VLP does not stimulate an immune response comprising non-neutralizing antibodies in the subject and/or does not stimulate a Th2 cell-mediated immune response in the subject.
- [0096] In some aspects, the VLP cross-competes with the infecting virus for binding to a viral receptor.
- [0097] In some aspects, the VLP cross-competes with a related virus or strain for binding to the viral receptor.
- [0098] In some aspects, the viral infection is a coronavirus, an influenza virus, a human immunodeficiency virus, a human papillomavirus, a hepatitis virus, or an oncolytic virus.
- [0099] In some aspects, the viral infection is a coronavirus. In some aspects, the viral infection is COVID-19.
- [0100] In some aspects, the VLP stimulates an immune response in the subject comprising neutralizing antibodies against COVID-19.
- [0101] In some aspects, the VLP stimulates a Th1 cell-mediated immune response in the subject against COVID-19.
- [0102] In some aspects, the immune response is cross-reactive to other coronaviruses.

- [0103] In some aspects, the immune response is cross-reactive to other severe acute respiratory syndrome coronaviruses and/or human betacoronaviruses.
- [0104] In some aspects, the VLP does not stimulate an immune response comprising non-neutralizing antibodies in the subject and/or does not stimulate a Th2 cell-mediated immune response in the subject.
- [0105] In some aspects, the administering is by inhalation.
- [0106] In some aspects, the VLP cross-competes with COVID-19 for binding to ACE2 receptor, neuropilin-1, or other receptors.
- [0107] In some aspects, the VLP cross-competes with other coronaviruses for binding to ACE2 receptor, neuropilin-1, and/or other receptors.
- [0108] In some aspects, the VLP cross-competes with other severe acute respiratory syndrome coronaviruses and/or human betacoronaviruses for binding to ACE2 receptor, neuropilin-1, and/or other receptors.

BRIEF DESCRIPTION OF THE DRAWINGS

- [0109] **Figure 1** shows a schematic representation of an exemplary expression cassette for producing a coronavirus VLP containing simian virus 40 enhancers (SV40E); a cytomegalovirus promoter (P_{CMV}); a sequence encoding a coronavirus Envelope (E) protein; a sequence encoding a coronavirus Membrane (M) protein; a sequence encoding a recombinant protein containing sequences from the receptor-binding domain (RBD), the second subunit cleavage domain and internal fusion peptide (S2'IFP), and transmembrane (TM) domain of a coronavirus S protein (referred to herein as a recombinant Spike (S) protein, RBD::S2'IFP::TM); sequences encoding 2A self-cleaving peptides from porcine teschovirus-1 (P2A) to separate the protein-encoding sequences of the expression cassette; and a polyadenylation (pA) signal.
- [0110] **Figure 2** shows a vector map of an exemplary expression vector (pGL2-SS-CMV-VLP-BGH-SS) containing an expression cassette as described in Figure 1, in which the pA signal is from bovine growth hormone.
- [0111] **Figure 3** shows in vitro expression of genes and protein from the expression vector of Figure 2. **(A)** shows a bar graph depicting relative expression of genes encoding the E protein, M protein, and recombinant S protein (RBD::S2'IFP::TM) as described in Figure 1 from cells containing the expression vector of Figure 2 (VLP) as well as control

cells without the expression vector (CTL). *** = $p < 0.001$ and **** = $p < 0.0001$. **(B)** shows a representative Western blot depicting expression of the recombinant S protein using an antibody that binds to the RBD (α -Spike (RBD)). Detection of beta-actin with the α -beta-actin antibody served as a loading control. Control = protein from cells without the expression vector. VLP = protein from cells containing the expression vector of Figure 2. **(C)** shows the relative mean intensity of recombinant S protein expression from Western blots (n=3) as described in **(B)**.

[0112] **Figure 4** shows an exemplary msDNA-VLP (msDNA VLP Cov 19-BGH poly) as described herein that is encoded by the expression vector of Figure 2.

[0113] **Figure 5** shows the concentration (ng/mL) of antibodies that bind to the S1 subunit of the COVID-19 Spike protein (Spike AB) in serum from C57 mice at days 0, 7, 14, 21, 28, 35, 42, and 49 following intramuscular injection with the expression vector of Figure 2 at day 0 and day 14 (booster). **(A)** and **(B)** show a line graph and a bar graph of the antibody concentration, respectively.

[0114] **Figure 6** shows a sequence conservation analysis of representative COVID-19 genomes. **(A)** shows a bar plot in which the horizontal bars indicate the genomic positions on the x-axis of each of the COVID-19 genes listed on the y-axis as per the Wuhan reference genome (NC_045512.2). **(B)** shows a histogram in which bar heights correspond to the percentage of 3928 representative COVID-19 genomes that differed from the Wuhan reference genome at each genomic position.

[0115] **Figure 7** and **Figure 8** show histograms in which bar heights correspond to the percentage of analyzed genomes that differed from the Wuhan reference genome at each genomic position, with the analyzed genomes being: **(7)** 3928 representative COVID-19 genomes, 120 severe acute respiratory syndrome coronaviruses (SARS-CoV) genomes, and 257 Middle East respiratory syndrome coronaviruses (MERS-CoV) genomes, **(8A)** 233 COVID-19 genomes of variant strain B.1.1.7, **(8B)** 104 COVID-19 genomes of variant strain B.1.351, **(8C)** 39 COVID-19 genomes of variant strain P.1, and **(8D)** 62 COVID-19 genomes of variant strain B.1.427/429.

[0116] **Figure 9** shows an exemplary eukaryotic expression vector (pFastBac™ Dual-VLP) for VLP production in eukaryotic cells as described herein, containing the E, M, and recombinant S proteins as described in Figure 1.

DETAILED DESCRIPTION

[0117] The present disclosure provides expression vectors and bacterial sequence-free vectors (*e.g.*, ministring DNA (msDNA)) for producing virus-like particles (VLPs), vector production systems, and VLPs, as well as compositions and methods thereof. Some aspects of the present disclosure are directed to treating viral infections in a subject (*e.g.*, coronavirus infections in a human subject, such as COVID-19).

[0118] All publications cited herein are hereby incorporated by reference in their entireties, including without limitation all journal articles, books, manuals, patent applications, and patents cited herein, to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference.

I. Terms

[0119] In order that the present disclosure can be more readily understood, certain terms are first defined. As used in this application, except as otherwise expressly provided herein, each of the following terms shall have the meaning set forth below. Additional definitions are set forth throughout the application.

[0120] It is to be noted that the term "a" or "an" entity refers to one or more of that entity; for example, "a nucleotide sequence," is understood to represent one or more nucleotide sequences. As such, the terms "a" (or "an"), "one or more," and "at least one" can be used interchangeably herein.

[0121] The term "and/or" where used herein is to be taken as specific disclosure of each of the two specified features or components with or without the other. Thus, the term "and/or" as used in a phrase such as "A and/or B" herein is intended to include "A and B," "A or B," "A" (alone), and "B" (alone). Likewise, the term "and/or" as used in a phrase such as "A, B, and/or C" is intended to encompass each of the following aspects: A, B, and C; A, B, or C; A or C; A or B; B or C; A and C; A and B; B and C; A (alone); B (alone); and C (alone).

[0122] It is understood that wherever aspects are described herein with the language "comprising," otherwise analogous aspects described in terms of "consisting of" and/or "consisting essentially of" are also provided.

[0123] The terms "about" or "comprising essentially of" refer to a value or composition that is within an acceptable error range for the particular value or composition as

determined by one of ordinary skill in the art, which will depend in part on how the value or composition is measured or determined, *i.e.*, the limitations of the measurement system. For example, "about" or "comprising essentially of" can mean within 1 or more than 1 standard deviation per the practice in the art. Alternatively, "about" or "comprising essentially of" can mean a range of up to 10%. Furthermore, particularly with respect to biological systems or processes, the terms can mean up to an order of magnitude or up to 5-fold of a value. When particular values or compositions are provided in the application and claims, unless otherwise stated, the meaning of "about" or "comprising essentially of" should be assumed to be within an acceptable error range for that particular value or composition.

[0124] As described herein, any concentration range, percentage range, ratio range, or integer range is to be understood to include the value of any integer within the recited range and, when appropriate, fractions thereof (such as one tenth and one hundredth of an integer), unless otherwise indicated. Numeric ranges are inclusive of the numbers defining the range.

[0125] Unless defined otherwise, 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 is related. For example, the Concise Dictionary of Biomedicine and Molecular Biology, Juo, Pei-Show, 2nd ed., 2002, CRC Press; The Dictionary of Cell and Molecular Biology, 5th ed., 2013, Academic Press; and the Oxford Dictionary Of Biochemistry And Molecular Biology, 2006, Oxford University Press, provide one of skill with a general dictionary of many of the terms used in this disclosure.

[0126] Units, prefixes, and symbols are denoted in their Système International de Unites (SI) accepted form.

[0127] Unless otherwise indicated, nucleotide sequences are written left to right in 5' to 3' orientation. Amino acid sequences are written left to right in amino to carboxy orientation.

[0128] The headings provided herein are not limitations of the various aspects of the disclosure, which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification in its entirety.

[0129] "Amino acid" is a molecule having the structure wherein a central carbon atom (the alpha-carbon atom) is linked to a hydrogen atom, a carboxylic acid group (the carbon atom of which is referred to herein as a "carboxyl carbon atom"), an amino group (the nitrogen atom of which is referred to herein as an "amino nitrogen atom"), and a side chain group, R. When incorporated into a peptide, polypeptide, or protein, an amino acid loses one or more atoms of its amino acid carboxylic groups in the dehydration reaction that links one amino acid to another. As a result, when incorporated into a protein, an amino acid is referred to as an "amino acid residue."

[0130] "Protein" or "polypeptide" refers to any polymer of two or more individual amino acids (whether or not naturally occurring) linked via a peptide bond, and occurs when the carboxyl carbon atom of the carboxylic acid group bonded to the alpha-carbon of one amino acid (or amino acid residue) becomes covalently bound to the amino nitrogen atom of amino group bonded to the non alpha-carbon of an adjacent amino acid. The term "protein" is understood to include the terms "polypeptide" and "peptide" (which, at times may be used interchangeably herein) within its meaning. In addition, proteins comprising multiple polypeptide subunits will also be understood to be included within the meaning of "protein" as used herein. Similarly, fragments of proteins and polypeptides are also within the scope of the disclosure and may be referred to herein as "proteins." In one aspect of the disclosure, a polypeptide comprises a chimera of two or more parental peptide segments. The term "polypeptide" is also intended to refer to and encompass the products of post-translation modification ("PTM") of the polypeptide, including without limitation disulfide bond formation, glycosylation, carbamylation, lipidation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, modification by non-naturally occurring amino acids, or any other manipulation or modification, such as conjugation with a labeling component. A polypeptide can be derived from a natural biological source or produced by recombinant technology, but is not necessarily translated from a designated nucleic acid sequence. It can be generated in any manner, including by chemical synthesis. An "isolated" polypeptide or a fragment, variant, or derivative thereof refers to a polypeptide that is not in its natural milieu. No particular level of purification is required. For example, an isolated polypeptide can simply be removed from its native or natural environment. Recombinantly produced polypeptides and proteins expressed in host cells are considered

isolated for the purpose of the disclosure, as are native or recombinant polypeptides which have been separated, fractionated, or partially or substantially purified by any suitable technique.

[0131] "Domain" as used herein can be used interchangeably with the term "peptide segment" and refers to a portion or fragment of a larger polypeptide or protein. A domain need not on its own have functional activity, although in some instances, a domain can have its own biological activity.

[0132] "Fused," "operably linked," and "operably associated" are used interchangeably herein when referring to two or more domains to broadly refer to any chemical or physical coupling of the two or more domains in the formation of a recombinant polypeptide as disclosed herein. In one embodiment, a recombinant polypeptide as disclosed herein is a chimeric polypeptide comprising a plurality of domains from two or more different polypeptides.

[0133] Recombinant polypeptides (*i.e.*, recombinant proteins) comprising two or more domains and/or proteins as disclosed herein can be encoded by a single coding sequence that comprises polynucleotide sequences encoding each domain and/or protein. Unless stated otherwise, the polynucleotide sequences encoding each domain and/or protein are "in frame" such that translation of a single mRNA comprising the polynucleotide sequences results in a single polypeptide comprising each domain and/or protein. Typically, the domains and/or proteins in a recombinant polypeptide as described herein will be fused directly to one another or will be separated by a peptide linker. Various polynucleotide sequences encoding peptide linkers are known in the art and include, for example, self-cleaving peptides.

[0134] "Polynucleotide" or "nucleic acid" as used herein refers to a polymeric form of nucleotides. In some instances, a polynucleotide comprises a sequence that is either not immediately contiguous with the coding sequences or is immediately contiguous (on the 5' end or on the 3' end) with the coding sequences in the naturally occurring genome of the organism from which it is derived. The term therefore includes, for example, a recombinant DNA that is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (*e.g.*, a cDNA) independent of other sequences. The nucleotides of the disclosure can be ribonucleotides, deoxyribonucleotides, or modified forms of either

nucleotide. A polynucleotide as used herein refers to, among others, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. The term polynucleotide encompasses genomic DNA or RNA (depending upon the organism, *i.e.*, RNA genome of viruses), as well as mRNA encoded by the genomic DNA, and cDNA. In certain embodiments, a polynucleotide comprises a conventional phosphodiester bond or a non-conventional bond (*e.g.*, an amide bond, such as found in peptide nucleic acids (PNA)). By "isolated" nucleic acid or polynucleotide is intended a nucleic acid molecule, *e.g.*, DNA or RNA, which has been removed from its native environment. For example, a nucleic acid molecule comprising a polynucleotide encoding a recombinant polypeptide contained in a vector is considered "isolated" for the purposes of the present disclosure. Further examples of an isolated polynucleotide include recombinant polynucleotides maintained in heterologous host cells or purified (partially or substantially) from other polynucleotides in a solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of polynucleotides of the present disclosure. Isolated polynucleotides or nucleic acids according to the present disclosure further include polynucleotides and nucleic acids (*e.g.*, nucleic acid molecules) produced synthetically.

[0135] As used herein, a "coding region" or "coding sequence" is a portion of a polynucleotide, which consists of codons translatable into amino acids. Although a "stop codon" (TAG, TGA, or TAA) is typically not translated into an amino acid, it may be considered to be part of a coding region, but any flanking sequences, for example promoters, ribosome binding sites, transcriptional terminators, introns, and the like, are not part of a coding region. The boundaries of a coding region are typically determined by a start codon at the 5' terminus, encoding the amino-terminus of the resultant polypeptide, and a translation stop codon at the 3' terminus, encoding the carboxyl-terminus of the resulting polypeptide.

[0136] As used herein, the term "expression control region" refers to a transcription control element that is operably associated with a coding region to direct or control expression of the product encoded by the coding region, including, for example, promoters, enhancers, operators, repressors, ribosome binding sites, translation leader

sequences, introns, polyadenylation recognition sequences, RNA processing sites, effector binding sites, stem-loop structures, and transcription termination signals. For example, a coding region and a promoter are "operably associated" (*i.e.*, "operably linked") if induction of promoter function results in the transcription of mRNA comprising a coding region that encodes the product, and if the nature of the linkage between the promoter and the coding region does not interfere with the ability of the promoter to direct the expression of the product encoded by the coding region or interfere with the ability of the DNA template to be transcribed. Expression control regions include nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding region, and which influence the transcription, RNA processing, stability, or translation of the associated coding region. If a coding region is intended for expression in a eukaryotic cell, a polyadenylation signal and transcription termination sequence will usually be located 3' to the coding sequence.

[0137] As used herein, the terms "host cell" and "cell" can be used interchangeably and can refer to any type of cell or a population of cells, *e.g.*, a primary cell, a cell in culture, or a cell from a cell line, that harbors or is capable of harboring a nucleic acid molecule (*e.g.*, a recombinant nucleic acid molecule). Host cells can be a prokaryotic cell, or alternatively, the host cells can be eukaryotic, for example, fungal cells, such as yeast cells, and various animal cells, such as insect cells or mammalian cells.

[0138] "Culture," "to culture" and "culturing," as used herein, means to incubate cells under *in vitro* conditions that allow for cell growth or division or to maintain cells in a living state. "Cultured cells," as used herein, means cells that are propagated *in vitro*.

[0139] A "subject" includes any human or nonhuman animal. The term "nonhuman animal" includes, but is not limited to, vertebrates such as mammals, avians, pets, farm animals, nonhuman primates, sheep, cows, goats, pigs, chickens, dogs, cats, and rodents such as mice, rats, and guinea pigs. In preferred aspects, the subject is a human. The terms, "subject" and "patient" are used interchangeably herein.

[0140] "Administering" refers to the physical introduction of a therapeutic agent to a subject, using any of the various methods and delivery systems known to those skilled in the art.

[0141] The terms "treat," "treating," "treatment," or "therapy" of a subject as used herein, refer to any type of intervention or process performed on, or administering an active agent

to, the subject with the objective of reversing, alleviating, ameliorating, inhibiting, or slowing down or preventing the progression, development, severity or recurrence of a symptom, complication, condition or biochemical indicia associated with a disease or enhancing overall survival. Treatment can be of a subject having a disease or a subject who does not have a disease (*e.g.*, for prophylaxis, such as vaccination).

[0142] The term "effective dose" "effective dosage," or "effective amount" is defined as an amount sufficient to achieve or at least partially achieve a desired effect. A "therapeutically effective amount" or "therapeutically effective dosage" of a drug or therapeutic agent is any amount of the drug that, when used alone or in combination with another therapeutic agent, promotes disease regression evidenced by a decrease in severity of disease symptoms, an increase in frequency and duration of disease symptom-free periods, an increase in overall survival (the length of time from either the date of diagnosis or the start of treatment for a disease that patients diagnosed with the disease are still alive), or a prevention of impairment or disability due to the disease affliction. A therapeutically effective amount or dosage of a drug includes a "prophylactically effective amount" or a "prophylactically effective dosage", which is any amount of the drug that, when administered alone or in combination with another therapeutic agent to a subject at risk of developing a disease or of suffering a recurrence of disease, inhibits the development or recurrence of the disease. The ability of a therapeutic agent to promote disease regression or inhibit the development or recurrence of the disease can be evaluated using a variety of methods known to the skilled practitioner, such as in human subjects during clinical trials, in animal model systems predictive of efficacy in humans, or by assaying the activity of the agent in *in vitro* assays.

[0143] Various aspects of the disclosure are described in further detail in the following subsections.

II. Vectors for producing VLPs

[0144] Bacterial sequence-free vectors and their production are described in U.S. Patent Nos. 9,290,778 and 9,862,954; Nafissi and Slavcev, *Microbial Cell Factories* 11:154 (2012); and Nafissi *et al.*, *Nucleic Acids* 3(6):e165 (2014), incorporated by reference herein in their entireties. These bacterial sequence-free vectors are produced from an expression vector (*e.g.*, a plasmid) that contains specialized "Super Sequence" ("SS") sites comprising target sequences for recombinases. The SS sites flank an expression

cassette containing a nucleic acid(s) of interest. When the expression vector is present in a recombinant cell that expresses an appropriate recombinase, bacterial sequence-free vector containing the expression cassette is separated from the backbone DNA of the expression vector. To produce a circular covalently closed (CCC) bacterial sequence-free vector, a production system is used in which the recombinant cell expresses a Cre or Flp recombinase, for example, and the expression vector contains corresponding target sequences for the recombinases. To produce a linear covalently closed (LCC) bacterial sequence-free vector, also referred to herein as a ministring DNA (msDNA), a production system is used in which the recombinant cell expresses a TelN or Tel recombinase, for example, and the expression vector contains corresponding target sequences for the recombinases. The bacterial sequence-free vector can then be purified from the cells and used directly as a delivery vector. *See* U.S. Patent Nos. 9,290,778 and 9,862,954, Nafissi and Slavcev, and Nafissi *et al.*

[0145] msDNA vectors with LCC ends are torsion-free and not subject to gyrase-directed negative supercoiling during their production in *E. coli*. Exemplary msDNA vectors carry an expression cassette with a eukaryotic promoter, gene of interest (GOI), intron, and polyA sequence, and nuclear translocation enhancing sequences (Nafissi and Slavcev, and Nafissi *et al.*). Furthermore, due to its double stranded LCC topology, integration of msDNA into a cell's chromosome causes a chromosomal break, thereby eliminating the cell from the population. Thus, msDNA eliminates any risk of insertional mutagenesis, protecting patients who are administered the msDNA from potential genotoxicity and cancer (Nafissi *et al.*).

[0146] In some aspects, bacterial sequence-free vectors for producing VLPs as disclosed herein include CCC or LCC vectors produced according to any other method known in the art.

A. Expression vectors, expression cassettes, and vector production systems for producing bacterial sequence-free vectors and VLPs

[0147] Provided herein is an expression vector comprising: an expression cassette that comprises a nucleic acid sequence encoding a recombinant protein comprising a conserved amino acid sequence from a virus fused to an immunogenic amino acid sequence, wherein protein expressed intracellularly from the expression cassette is capable of forming a VLP.

- [0148] Provided herein is an expression vector comprising: an expression cassette that comprises a nucleic acid sequence encoding a recombinant protein comprising a conserved amino acid sequence from a virus fused to an immunogenic amino acid sequence, a target sequence for a first recombinase flanking each side of the expression cassette, and one or more additional target sequences for one or more additional recombinases integrated within non-binding regions of the target sequence for the first recombinase, wherein protein expressed intracellularly from the expression cassette is capable of forming a VLP.
- [0149] Conserved and immunogenic amino acid sequences include those known in the art as well as those determined through known techniques. For example, genome-based reverse vaccinology can be applied towards comparative genomics analysis, a field of biological research that can be used to compare genomic sequences between different pathogenic strains (*see, e.g., Sieb et al., Clin. Microbiol. Infect. 18*(Suppl. 5):109-116 (2012)). Other sequencing, structural, and computational approaches can also be used (*see, e.g., Liljeroos et al., J. Immunol. Res. 2015: 156241; Sette and Rappuoli, Immunity 33*(4):530-541 (2010)).
- [0150] In some aspects, the immunogenic amino acid sequence is from the same virus as the conserved amino acid sequence. In some aspects, the conserved amino acid sequence is from a viral glycoprotein. In some aspects, the immunogenic amino acid sequence is from the same viral glycoprotein.
- [0151] In some aspects, the expression cassette further comprises a nucleic acid sequence encoding a viral envelope protein and/or a nucleic acid sequence encoding a viral matrix protein. In some aspects, the viral envelope protein and/or the viral matrix protein are from the same virus as the conserved amino acid sequence.
- [0152] In some aspects, the conserved amino acid sequence, the immunogenic amino acid sequence, the viral envelope protein, and/or the viral matrix protein is a consensus sequence.
- [0153] In some aspects, the recombinant protein is capable of stimulating an immune response against the virus comprising neutralizing antibodies. Conserved sites, for example, are often recognized by broadly neutralizing antibodies and are susceptible to antibody inactivation (*see, e.g., Nabel, N. Engl. J. Med. 368*(6): 551-560 (2013)).

- [0154] In some aspects, the recombinant protein is capable of stimulating a Th1 cell-mediated immune response against the virus. Cell-mediated immunity is the process by which cytotoxic T cells recognize antigen infected cells, to induce cell lysis.
- [0155] In some aspects, the immune response is cross-reactive to a related virus or strain. For example, conserved sequences among different viral serotypes/strains can be utilized to provide protection against multiple serotypes/strains, including as a universal vaccine.
- [0156] In some aspects, the recombinant protein excludes amino acid sequences from the virus that stimulate an immune response comprising non-neutralizing antibodies and/or that stimulate a Th2 cell-mediated immune response.
- [0157] In some aspects, the expression cassette comprises a single open reading frame comprising a nucleic acid sequence encoding a self-cleaving peptide between each nucleic acid sequence encoding a protein such that the translation product of the expression cassette is cleaved intracellularly into two or more proteins. In some aspects, the self-cleaving peptide is a 2A self-cleaving peptide. In some aspects, the 2A self-cleaving peptide is P2A from porcine teschovirus-1. In some aspects, the 2A self-cleaving peptide is T2A from thosela asigna virus 2A.
- [0158] In some aspects, the expression cassette comprises a nucleic acid sequence encoding a self-cleaving peptide between nucleic acid sequences encoding a viral matrix protein and a viral envelope protein, between nucleic acid sequences encoding a viral matrix protein and the recombinant protein, and/or between nucleic acid sequences encoding a viral envelope protein and the recombinant protein. In some aspects, the expression cassette comprises nucleic acid sequences from 5' to 3' encoding a viral matrix protein, a self-cleaving peptide, a viral envelope protein, a self-cleaving peptide, and the recombinant protein. In some aspects, the expression cassette comprises nucleic acid sequences from 5' to 3' encoding a viral envelope protein, a self-cleaving peptide, a viral matrix protein, a self-cleaving peptide, and the recombinant protein.
- [0159] In some aspects, the expression cassette further comprises a nucleic acid sequence encoding a marker for gene expression. In some aspects, the marker for gene expression is a fluorescent reporter gene, such as green fluorescent protein (GFP), red fluorescent protein (RFP), yellow fluorescent protein (YFP), or near-infrared fluorescent protein (iRFP); a bioluminescent reporter genes such as luciferase; a selectable antibiotic marker; or LacZ. In some aspects, the expression cassette comprises a nucleic acid sequence

encoding a self-cleaving peptide between the nucleic acid sequence encoding a marker for gene expression and any other nucleic acid sequence encoding a protein.

[0160] The expression cassette can contain any expression control region known to those of skill in the art operably linked to the protein-encoding nucleic acid sequence(s). In some aspects, the expression control region is a promoter, enhancer, operator, repressor, ribosome binding site, translation leader sequence, intron, polyadenylation recognition sequence, RNA processing site, effector binding site, stem-loop structure, transcription termination signal, or combination thereof.

[0161] In some aspects, the target sequence for the first recombinase and the one or more additional target sequences for the one or more additional recombinases are selected from the group consisting of the PY54 pal site, the N15 telRL site, the loxP site, ϕ K02 telRL site, the FRT site, the phiC31 attP site, and the λ attP site. In some aspects, the expression vector comprises each of the target sequences. In some aspects, the expression vector comprises the Tel recombinase pal site and the telRL, loxP, and FRT recombinase target binding sequences integrated within the pal site.

[0162] In some aspects, the expression vector is for producing a bacterial sequence-free vector. In some aspects, the bacterial sequence-free vector has circular covalently closed ends. In some aspects, the bacterial sequence-free vector has linear covalently closed ends.

[0163] In some aspects, the expression vector further comprises at least one enhancer sequence flanking each side of the target sequence for the first recombinase. In some aspects, the at least one enhancer sequence is at least two enhancer sequences. In some aspects, the at least one enhancer sequence is a SV40 enhancer sequence.

[0164] The source of the conserved amino acid sequence, the immunogenic amino acid sequence, and/or a viral protein as disclosed herein can be any virus associated with human or animal infection.

[0165] In some aspects, the virus is a coronavirus, an influenza virus, a human immunodeficiency virus, a human papillomavirus, a hepatitis virus, or an oncolytic virus.

[0166] In some aspects, the influenza virus is an influenza A virus. In some aspects, the influenza A virus is H1N1, H5N1, or H3N2.

[0167] In some aspects, the influenza virus is an influenza B virus.

- [0168] In some aspects, the coronavirus is a human coronavirus such as, but not limited to, HCoV-229E, HCoV-NL63, HCoV-OC43, HCoV-HKU1, SARS-CoV-1, SARS-CoV-2 (*i.e.*, COVID-19)), and/or MERS-CoV.
- [0169] In some aspects, the coronavirus is COVID-19 (*i.e.*, Wuhan-Hu-1 or a variant thereof such as, but not limited to, U.K. variant B.1.1.7, South African variant B.1.351, Brazilian variant P.1, or Californian variant B.1.427/429).
- [0170] Provided herein is a vector production system comprising recombinant cells designed to encode at least a first recombinase under the control of an inducible promoter, wherein the cells comprise an expression vector as disclosed herein comprising a target for the at least first recombinase. In some aspects, the inducible promoter is thermally-regulated, chemically-regulated, IPTG regulated, glucose-regulated, arabinose inducible, T7 polymerase regulated, cold-shock inducible, pH inducible, or combinations thereof. In some aspects, the at least first recombinase is selected from telN and tel, and the expression vector incorporates the target sequence for the at least first recombinase. In some aspects, the at least first recombinase is selected from Cre or Flp, and the expression vector incorporates the target sequence for the at least first recombinase. In some aspects, the recombinant cells have been further designed to encode a nuclease genome editing system, and the expression vector further comprises a backbone sequence containing a cleavage site for the nuclease genome editing system. In some aspects, the nuclease genome editing system is a CRISPR nuclease system comprising a Cas nuclease and gRNA, and the expression vector comprises a target sequence for the gRNA within the backbone sequence.
- [0171] Provided herein is a method of producing a bacterial sequence-free vector comprising incubating a vector production system as described herein under suitable conditions for expression of the at least first recombinase or the first recombinase and the nuclease genome editing system. In some aspects, the method further comprises harvesting the bacterial sequence-free vector. The present disclosure is also directed to a bacterial sequence-free vector produced by the method.

A.1 Expression cassettes comprising coronavirus sequences

- [0172] Coronaviruses include any virus of the family *Coronaviridae*, including the subfamily *Coronavirinae*, and including the genera *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus*. *See, e.g.*, Fung and Liu (2019).

Coronaviruses include human coronaviruses (HCoVs), such as HCoV-229E, HCoV-NL63, HCoV-OC43, HCoV-HKU1, severe acute respiratory syndrome coronaviruses (SARS-CoV, *e.g.*, SARS-CoV-1 and SARS-CoV-2 (*i.e.*, COVID-19)), Middle East respiratory syndrome coronaviruses (MERS-CoV), zoonotic coronaviruses (*e.g.*, SARS-CoVs and MERS-CoVs), bat coronaviruses (BtCoVs), *Avian coronavirus*, *Murine coronavirus*, and bulbo coronavirus (BuCoV).

- [0173] Coronavirus genomes are positive-sense, nonsegmented, single-stranded RNA ranging from about 27 to 32 kilobases (*see, e.g.*, Fung and Liu, *Annu. Rev. Microbiol.* 73:529-557 (2019)). For example, the complete genome of COVID-19 (also termed Wuhan-Hu-1 coronavirus (WHCV), SARS-CoV-2, and 2019-nCoV) has a size of 29.9 kb, compared to SARS-CoV and MERS-CoV with genomes of 27.9 kb and 30.1 kb, respectively (Zhou *et al.*, *Nature* 579: 270–273 (2020)). The COVID-19 genome has been found to be 96.2% identical to the Bat CoV RaTG13 genome, which is a type of SARS-CoV-2 found in bats and is likely the source of the virus transmitted to humans via unknown intermediate hosts.
- [0174] Coronaviruses have a membrane (M) protein, which is the most abundant structural protein that supports the viral envelope and embeds in the envelope with three transmembrane domains. The M protein is essential for virus assembly and budding.
- [0175] Envelope (E) protein is a small transmembrane protein in coronaviruses that is also present in the envelope at a lower amount than M protein. E protein is also engaged in virus assembly and egress.
- [0176] The nucleocapsid (N) protein in coronaviruses binds to the RNA genome like beads-on-a-string, forming the helically symmetric nucleocapsid.
- [0177] The virion surface of coronaviruses is decorated with the trimeric Spike (S) protein. Some betacoronaviruses also have dimeric hemagglutinin-esterase (HE) protein that make up shorter projections on the virion surface. S and HE protein each are type I transmembrane proteins with a large ectodomain and a short endodomain.
- [0178] The S protein contains two subunits, S1 and S2, and is anchored in the viral envelope at its C-terminus. The S1 subunit of COVID-19, for example, contains the N-terminal domain (NTD) and receptor-binding domain (RBD), while the S2 subunit contains the fusion peptide (FP), internal fusion peptide (IFP), heptad repeat 1/2 (HR1/2), and the transmembrane domain (TM). The S protein's large ectodomain trimerizes and

forms the characteristic coronavirus spikes at the virion's surface. The S protein is responsible for receptor binding and virion entry to host cells (Fehr and Perlman, Coronaviruses: An Overview of Their Replication and Pathogenesis. In: Maier H., Bickerton E., Britton P. (eds) Coronaviruses. Methods in Molecular Biology, vol 1282. Humana Press, New York, NY; Wall *et al.*, *Cell* 180: 1–12 (2020)).

[0179] Fusion proteins from many viruses require a proteolytic event near a fusion peptide to enable the pathogen's entry into the target cell. For example, the S protein from COVID-19 possesses two cleavage sites, the first of which sits at the S1/S2 boundary but is not closely linked to the fusion peptide. A second cleavage site (S2') exposes the internal fusion peptide (IFP), a motif just downstream of S2' that is highly conserved across all sequenced coronaviruses. The sequence of IFP is SFIEDLLFNKVTADAGF (SEQ ID NO:7), within which the bolded LLF residues are critical for membrane fusion and infectivity (Madu *et al.*, *J. Virol.* 83(15): 7411-7421 (2009)). COVID-19 demonstrates the presence of a canonical furin-like cleavage motif at the S1/S2 site not found in other coronaviruses in the same clade, but similarly found in particularly virulent forms of influenza (H5N1). Cleavage via other proteases such as furin at the S1/S2 interface likely widens the tropism of the virus, making animal to human transmission more likely (Coutard *et al.*, *Antiviral Res.* 176:104742 (2020)).

[0180] In some aspects, the expression cassette comprises nucleic acid sequences encoding a coronavirus Membrane (M) protein, a coronavirus Envelope (E) protein, and a recombinant protein comprising a conserved amino acid sequence and an immunogenic amino acid sequence from a coronavirus Spike (S) protein. The M, E, and S proteins can be interchangeably referred to herein as M, E, and S glycoproteins.

[0181] In some aspects, the M protein comprises an amino acid sequence at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:1. In some aspects, the M protein comprises SEQ ID NO:1. In some aspects, the M protein is SEQ ID NO:1.

[0182] In some aspects, the nucleic acid sequence encoding the M protein is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:2. In some aspects, the nucleic acid sequence

encoding the M protein comprises SEQ ID NO:2. In some aspects, the nucleic acid sequence encoding the M protein is SEQ ID NO:2.

- [0183]** In some aspects, the E protein comprises an amino acid sequence at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:3. In some aspects, the E protein comprises SEQ ID NO:3. In some aspects, the E protein is SEQ ID NO:3. In some aspects, the E protein comprises a replacement of the proline located at amino acid number 71 in SEQ ID NO:3 (*i.e.*, at P71 in SEQ ID NO:3) with another amino acid. In some aspects, the replacement at P71 in SEQ ID NO:3 is a change from proline to leucine (*i.e.*, P71L).
- [0184]** In some aspects, the nucleic acid sequence encoding the E protein is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:4. In some aspects, the nucleic acid sequence encoding the E protein comprises SEQ ID NO:4. In some aspects, the nucleic acid sequence encoding the E protein is SEQ ID NO:4. In some aspects, the nucleic acid sequence encoding the E protein comprises a replacement of the codon for proline at nucleotide numbers 211-213 in SEQ ID NO:4 with a codon for another amino acid. In some aspects, the codon for proline at nucleotide numbers 211-213 in SEQ ID NO:4 is replaced with a codon for leucine.
- [0185]** In some aspects, the conserved amino acid sequence is from the S1 subunit or the S2 subunit of the S protein, the RBD of the S protein, the S protein S2' cleavage site and internal fusion peptide (IFP) of the S protein (referred to herein as S2'IFP), the M protein, or the E protein.
- [0186]** In some aspects, the conserved amino acid sequence is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to any one of SEQ ID NOs:12-54. In some aspects, the conserved amino acid sequence comprises any one of SEQ ID NOs:12-54. In some aspects, the conserved amino acid sequence is any one of SEQ ID NOs:12-54.
- [0187]** In some aspects, the conserved amino acid sequence is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%,

at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:7. In some aspects, the conserved amino acid sequence comprises SEQ ID NO:7. In some aspects, the conserved amino acid sequence is SEQ ID NO:7.

[0188] In some aspects, the nucleic acid sequence encoding the conserved amino acid sequence of the recombinant protein is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:8. In some aspects, the nucleic acid sequence encoding the conserved amino acid sequence of the recombinant protein comprises SEQ ID NO:8. In some aspects, the nucleic acid sequence encoding the conserved amino acid sequence of the recombinant protein is SEQ ID NO:8.

[0189] In some aspects, the immunogenic amino acid sequence is from the S protein receptor-binding domain (RBD).

[0190] In some aspects, the immunogenic amino acid sequence is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:11. In some aspects, the immunogenic amino acid sequence comprises SEQ ID NO:11. In some aspects, the immunogenic amino acid sequence is SEQ ID NO:11. In some aspects, the immunogenic protein comprises a replacement of one or more of: lysine located at amino acid number 88 (*i.e.*, K88), leucine located at amino acid number 123 (*i.e.*, L123), glutamate located at amino acid number 155 (*i.e.*, E155), or asparagine located at amino acid number 172 (*i.e.*, N172) in SEQ ID NO:11 (corresponding to K417, L452, E484, and N501 in SEQ ID NO:5, respectively) with another amino acid. In some aspects, the replacement at K88 is K88N (*i.e.*, a change from lysine to asparagine). In some aspects, the replacement at K88 is K88T (*i.e.*, a change from lysine to threonine). In some aspects, the replacement at L123 is L123R (*i.e.*, a change from leucine to arginine). In some aspects, the replacement at E155 is E155K (*i.e.*, a change from glutamate to lysine). In some aspects, the replacement at N172 is N172Y (*i.e.*, a change from asparagine to tyrosine).

[0191] In some aspects, the nucleic acid sequence encoding the immunogenic amino acid sequence of the recombinant protein is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%,

at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:101. In some aspects, the nucleic acid sequence encoding the immunogenic amino acid sequence of the recombinant protein comprises SEQ ID NO:101. In some aspects, the nucleic acid sequence encoding the immunogenic amino acid sequence of the recombinant protein is SEQ ID NO:101. In some aspects, the nucleic acid sequence encoding the immunogenic protein comprises a replacement of one or more of: the codon for lysine at nucleotide numbers 262-264 of SEQ ID NO:101 with a codon for another amino acid, the codon for leucine at nucleotide numbers 367-369 of SEQ ID NO:101 with a codon for another amino acid, the codon for glutamate at nucleotide numbers 463-465 of SEQ ID NO:101 with a codon for another amino acid, or the codon for asparagine at nucleotide numbers 514-516 of SEQ ID NO:101 with a codon for another amino acid. In some aspects, the codon for lysine at nucleotide numbers 262-264 is replaced with a codon for asparagine or threonine. In some aspects, the codon for leucine at nucleotide numbers 367-369 is replaced with a codon for arginine. In some aspects, the codon for glutamate at nucleotide numbers 463-465 is replaced with a codon for lysine. In some aspects, the codon for asparagine at nucleotide numbers 514-516 is replaced with a codon for tyrosine.

[0192] In some aspects, the recombinant protein further comprises a transmembrane (TM) domain sequence from the S protein.

[0193] In some aspects, the TM domain sequence comprises an amino acid sequence at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:102. In some aspects, the TM domain sequence comprises SEQ ID NO:102. In some aspects, the TM domain sequence is SEQ ID NO:102.

[0194] In some aspects, the nucleic acid sequence encoding the TM domain sequence of the recombinant protein is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:103. In some aspects, the nucleic acid sequence encoding the TM domain sequence of the recombinant protein comprises SEQ ID NO:103. In some aspects, the nucleic acid sequence encoding the TM domain sequence of the recombinant protein is SEQ ID NO:103.

- [0195] In some aspects, the recombinant protein comprises a conserved amino acid sequence from S2'IFP, an immunogenic amino acid sequence from the RBD, and a TM domain sequence of the S protein.
- [0196] In some aspects, the recombinant protein excludes amino acid sequences from the S protein that stimulate an immune response comprising non-neutralizing antibodies and/or that stimulate a Th2 cell-mediated immune response.
- [0197] In some aspects, the amino acid sequence of the recombinant protein is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:55. In some aspects, the amino acid sequence of the recombinant protein comprises SEQ ID NO:55. In some aspects, the amino acid sequence of the recombinant protein is SEQ ID NO:55. In some aspects, the recombinant protein comprises a replacement of one or more of K88, L123, E155, or N172 in SEQ ID NO:55 with another amino acid. In some aspects, the replacement at K88 is K88N. In some aspects, the replacement at K88 is K88T. In some aspects, the replacement at L123 is L123R. In some aspects, the replacement at E155 is E155K. In some aspects, the replacement at N172 is N172Y.
- [0198] In some aspects, the nucleic acid sequence encoding the recombinant protein is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:56. In some aspects, the nucleic acid sequence encoding the recombinant protein comprises SEQ ID NO:56. In some aspects, the nucleic acid sequence encoding the recombinant protein is SEQ ID NO:56. In some aspects, the nucleic acid sequence encoding the recombinant protein comprises a replacement of one or more of: the codon for lysine at nucleotide numbers 262-264 of SEQ ID NO:56 with a codon for another amino acid, the codon for leucine at nucleotide numbers 367-369 of SEQ ID NO:56 with a codon for another amino acid, the codon for glutamate at nucleotide numbers 463-465 of SEQ ID NO:56 with a codon for another amino acid, or the codon for asparagine at nucleotide numbers 514-516 of SEQ ID NO:56 with a codon for another amino acid. In some aspects, the codon for lysine at nucleotide numbers 262-264 is replaced with a codon for asparagine or threonine. In some aspects, the codon for leucine at nucleotide numbers 367-369 is replaced with a codon for arginine. In some

aspects, the codon for glutamate at nucleotide numbers 463-465 is replaced with a codon for lysine. In some aspects, the codon for asparagine at nucleotide numbers 514-516 is replaced with a codon for tyrosine.

[0199] In some aspects, the expression cassette comprises a single open reading frame translated as an amino acid sequence at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:57. In some aspects, the expression cassette comprises a single open reading frame translated as an amino acid sequence comprising SEQ ID NO:57. In some aspects, the expression cassette comprises a single open reading frame translated as an amino acid sequence that is SEQ ID NO:57. In some aspects, the expression cassette comprises a single open reading frame translated as an amino acid sequence that comprises a replacement of one or more of P71, K423, L458, E490, or N507 in SEQ ID NO:57 with another amino acid. In some aspects, the replacement at P71 is P71L. In some aspects, the replacement at K423 is K423N. In some aspects, the replacement at K423 is K423T. In some aspects, the replacement at L458 is L458R. In some aspects, the replacement at E490 is E490K. In some aspects, the replacement at N507 is N507Y.

[0200] In some aspects, the expression cassette comprises a single open reading frame that is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:58. In some aspects, the expression cassette comprises a single open reading frame that comprises SEQ ID NO:58. In some aspects, the expression cassette comprises a single open reading frame that is SEQ ID NO:58. In some aspects, the expression cassette comprises a single open reading frame that comprises a replacement of one or more of: the codon for proline at nucleotide numbers 211-213 in SEQ ID NO:58 with a codon for another amino acid, the codon for lysine at nucleotide numbers 1267-1269 of SEQ ID NO:58 with a codon for another amino acid, the codon for leucine at nucleotide numbers 1372-1374 of SEQ ID NO:58 with a codon for another amino acid, the codon for glutamate at nucleotide numbers 1468-1470 of SEQ ID NO:58 with a codon for another amino acid, or the codon for asparagine at nucleotide numbers 1519-1521 of SEQ ID NO:58 with a codon for another amino acid. In some aspects, the codon for proline at nucleotide numbers 211-213 in SEQ

ID NO:58 is replaced with a codon for leucine. In some aspects, the codon for lysine at nucleotide numbers 1267-1269 is replaced with a codon for asparagine or threonine. In some aspects, the codon for leucine at nucleotide numbers 1372-1374 is replaced with a codon for arginine. In some aspects, the codon for glutamate at nucleotide numbers 1468-1470 is replaced with a codon for lysine. In some aspects, the codon for asparagine at nucleotide numbers 1519-1521 is replaced with a codon for tyrosine.

[0201] In some aspects, the expression cassette is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to the nucleic acid sequence of any one of SEQ ID NOs:59-62. In some aspects, the expression cassette comprises the nucleic acid sequence of any one of SEQ ID NOs:59-62. In some aspects, the expression cassette is the nucleic acid sequence of any one of SEQ ID NOs:59-62.

[0202] In some aspects, the recombinant protein is capable of stimulating an immune response against COVID-19.

[0203] In some aspects, the recombinant protein is capable of stimulating a Th1 cell-mediated immune response against COVID-19.

[0204] In some aspects, the immune response against COVID-19 is against Wuhan-Hu-1 and/or one or more variants such as, but not limited to, the U.K. variant B.1.1.7, the South African variant B.1.351, the Brazilian variant P.1, or the Californian variant B.1.427/429.

[0205] In some aspects, the immune response is cross-reactive to other coronaviruses. In some aspects, the immune response is cross-reactive to other severe acute respiratory syndrome coronaviruses and/or human betacoronaviruses.

[0206] Provided herein is a polynucleotide encoding an amino acid sequence at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:57. In some aspects, the polynucleotide encodes an amino acid sequence comprising SEQ ID NO:57. In some aspects, the polynucleotide encodes an amino acid sequence that is SEQ ID NO:57. In some aspects, the polynucleotide encodes an amino acid sequence that comprises a replacement of one or more of P71, K423, L458, E490, or N507 in SEQ ID NO:57 with another amino acid. In some aspects, the replacement at P71 is P71L. In some aspects, the replacement at K423

is K423N. In some aspects, the replacement at K423 is K423T. In some aspects, the replacement at L458 is L458R. In some aspects, the replacement at E490 is E490K. In some aspects, the replacement at N507 is N507Y.

[0207] Provided herein is a polynucleotide comprising a nucleic acid sequence at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:58. In some aspects, the polynucleotide comprises SEQ ID NO:58. In some aspects, the polynucleotide is SEQ ID NO:58. In some aspects, the polynucleotide comprising a nucleic acid sequence that comprises a replacement of one or more of: the codon for proline at nucleotide numbers 211-213 in SEQ ID NO:58 with a codon for another amino acid, the codon for lysine at nucleotide numbers 1267-1269 of SEQ ID NO:58 with a codon for another amino acid, the codon for leucine at nucleotide numbers 1372-1374 of SEQ ID NO:58 with a codon for another amino acid, the codon for glutamate at nucleotide numbers 1468-1470 of SEQ ID NO:58 with a codon for another amino acid, or the codon for asparagine at nucleotide numbers 1519-1521 of SEQ ID NO:58 with a codon for another amino acid. In some aspects, the codon for proline at nucleotide numbers 211-213 in SEQ ID NO:58 is replaced with a codon for leucine. In some aspects, the codon for lysine at nucleotide numbers 1267-1269 is replaced with a codon for asparagine or threonine. In some aspects, the codon for leucine at nucleotide numbers 1372-1374 is replaced with a codon for arginine. In some aspects, the codon for glutamate at nucleotide numbers 1468-1470 is replaced with a codon for lysine. In some aspects, the codon for asparagine at nucleotide numbers 1519-1521 is replaced with a codon for tyrosine.

B. Bacterial sequence-free vectors

[0208] A bacterial sequence-free vector of the present disclosure can include any expression cassette of the present disclosure.

[0209] Provided herein is a bacterial sequence-free vector comprising an expression cassette that comprises a nucleic acid sequence encoding a recombinant protein comprising a conserved amino acid sequence from a virus fused to an immunogenic amino acid sequence, wherein protein expressed intracellularly from the expression cassette is capable of forming a VLP.

- [0210] In some aspects, the immunogenic amino acid sequence is from the same virus as the conserved amino acid sequence. In some aspects, the conserved amino acid sequence is from a viral glycoprotein. In some aspects, the immunogenic amino acid sequence is from the same viral glycoprotein.
- [0211] In some aspects, the expression cassette further comprises a nucleic acid sequence encoding a viral envelope protein and/or a nucleic acid sequence encoding a viral matrix protein. In some aspects, the viral envelope protein and/or the viral matrix protein are from the same virus as the conserved amino acid sequence.
- [0212] In some aspects, the conserved amino acid sequence, the immunogenic amino acid sequence, the viral envelope protein, and/or the viral matrix protein is a consensus sequence.
- [0213] In some aspects, the recombinant protein is capable of stimulating an immune response against the virus comprising neutralizing antibodies.
- [0214] In some aspects, the recombinant protein is capable of stimulating a Th1 cell-mediated immune response against the virus.
- [0215] In some aspects, the immune response is cross-reactive to a related virus or strain.
- [0216] In some aspects, the recombinant protein excludes amino acid sequences from the virus that stimulate an immune response comprising non-neutralizing antibodies and/or that stimulate a Th2 cell-mediated immune response.
- [0217] In some aspects, the expression cassette comprises a single open reading frame comprising a nucleic acid sequence encoding a self-cleaving peptide between each nucleic acid sequence encoding a protein. Expression cassettes and self-cleaving peptides include those discussed above with respect to expression vectors.
- [0218] In some aspects, the virus is a coronavirus, an influenza virus, a human immunodeficiency virus, a human papillomavirus, a hepatitis virus, or an oncolytic virus.
- [0219] In some aspects, the influenza virus is an influenza A virus. In some aspects, the influenza A virus is H1N1, H5N1, or H3N2.
- [0220] In some aspects, the influenza virus is an influenza B virus.
- [0221] In some aspects, the coronavirus is a human coronavirus such as, but not limited to, HCoV-229E, HCoV-NL63, HCoV-OC43, HCoV-HKU1, SARS-CoV-1, SARS-CoV-2 (*i.e.*, COVID-19), and/or MERS-CoV.

- [0222] In some aspects, the coronavirus is COVID-19 (*i.e.*, Wuhan-Hu-1 or a variant thereof such as, but not limited to, U.K. variant B.1.1.7, South African variant B.1.351, Brazilian variant P.1, or Californian variant B.1.427/429).
- [0223] In some aspects, the expression cassette comprises nucleic acid sequences encoding a coronavirus M protein, a coronavirus E protein, and a recombinant protein comprising a conserved amino acid sequence and an immunogenic amino acid sequence from a coronavirus S protein.
- [0224] In some aspects, the M protein comprises an amino acid sequence at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:1. In some aspects, the M protein comprises SEQ ID NO:1. In some aspects, the M protein is SEQ ID NO:1.
- [0225] In some aspects, the nucleic acid sequence encoding the M protein is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:2. In some aspects, the nucleic acid sequence encoding the M protein comprises SEQ ID NO:2. In some aspects, the nucleic acid sequence encoding the M protein is SEQ ID NO:2.
- [0226] In some aspects, the E protein comprises an amino acid sequence at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:3. In some aspects, the E protein comprises SEQ ID NO:3. In some aspects, the E protein is SEQ ID NO:3. In some aspects, the E protein comprises a replacement of P71 in SEQ ID NO:3 with another amino acid. In some aspects, the replacement at P71 in SEQ ID NO:3 is P71L.
- [0227] In some aspects, the nucleic acid sequence encoding the E protein is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:4. In some aspects, the nucleic acid sequence encoding the E protein comprises SEQ ID NO:4. In some aspects, the nucleic acid sequence encoding the E protein is SEQ ID NO:4. In some aspects, the nucleic acid sequence encoding the E protein comprises a replacement of the codon for proline at

nucleotide numbers 211-213 in SEQ ID NO:4 with a codon for another amino acid. In some aspects, the codon for proline at nucleotide numbers 211-213 in SEQ ID NO:4 is replaced with a codon for leucine.

[0228] In some aspects, the conserved amino acid sequence is from the S1 subunit or the S2 subunit of the S protein, the RBD of the S protein, the S protein S2' cleavage site and internal fusion peptide (IFP) of the S protein (referred to herein as S2'IFP), the M protein, or the E protein.

[0229] In some aspects, the conserved amino acid sequence is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to any one of SEQ ID NOs:12-54. In some aspects, the conserved amino acid sequence comprises any one of SEQ ID NOs:12-54. In some aspects, the conserved amino acid sequence is any one of SEQ ID NOs:12-54.

[0230] In some aspects, the conserved amino acid sequence is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:7. In some aspects, the conserved amino acid sequence comprises SEQ ID NO:7. In some aspects, the conserved amino acid sequence is SEQ ID NO:7.

[0231] In some aspects, the nucleic acid sequence encoding the conserved amino acid sequence of the recombinant protein is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:8. In some aspects, the nucleic acid sequence encoding the conserved amino acid sequence of the recombinant protein comprises SEQ ID NO:8. In some aspects, the nucleic acid sequence encoding the conserved amino acid sequence of the recombinant protein is SEQ ID NO:8.

[0232] In some aspects, the immunogenic amino acid sequence is from the S protein receptor-binding domain (RBD).

[0233] In some aspects, the immunogenic amino acid sequence is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:11. In some aspects, the immunogenic amino acid sequence

comprises SEQ ID NO:11. In some aspects, the immunogenic amino acid sequence is SEQ ID NO:11. In some aspects, the immunogenic amino acid sequence comprises a replacement of one or more of: K88, L123, E155, or N172 in SEQ ID NO:11 with another amino acid. In some aspects, the replacement at K88 is K88N. In some aspects, the replacement at K88 is K88T. In some aspects, the replacement at L123 is L123R. In some aspects, the replacement at E155 is E155K. In some aspects, the replacement at N172 is N172Y.

[0234] In some aspects, the nucleic acid sequence encoding the immunogenic amino acid sequence of the recombinant protein is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:101. In some aspects, the nucleic acid sequence encoding the immunogenic amino acid sequence of the recombinant protein comprises SEQ ID NO:101. In some aspects, the nucleic acid sequence encoding the immunogenic amino acid sequence of the recombinant protein is SEQ ID NO:101. In some aspects, the nucleic acid sequence encoding the immunogenic amino acid sequence comprises a replacement of one or more of: the codon for lysine at nucleotide numbers 262-264 of SEQ ID NO:101 with a codon for another amino acid, the codon for leucine at nucleotide numbers 367-369 of SEQ ID NO:101 with a codon for another amino acid, the codon for glutamate at nucleotide numbers 463-465 of SEQ ID NO:101 with a codon for another amino acid, or the codon for asparagine at nucleotide numbers 514-516 of SEQ ID NO:101 with a codon for another amino acid. In some aspects, the codon for lysine at nucleotide numbers 262-264 is replaced with a codon for asparagine or threonine. In some aspects, the codon for leucine at nucleotide numbers 367-369 is replaced with a codon for arginine. In some aspects, the codon for glutamate at nucleotide numbers 463-465 is replaced with a codon for lysine. In some aspects, the codon for asparagine at nucleotide numbers 514-516 is replaced with a codon for tyrosine.

[0235] In some aspects, the recombinant protein further comprises a transmembrane (TM) domain sequence from the S protein.

[0236] In some aspects, the TM domain sequence comprises an amino acid sequence at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at

least about 99% identical to SEQ ID NO:102. In some aspects, the TM domain sequence comprises SEQ ID NO:102. In some aspects, the TM domain sequence is SEQ ID NO:102.

[0237] In some aspects, the nucleic acid sequence encoding the TM domain sequence of the recombinant protein is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:103. In some aspects, the nucleic acid sequence encoding the TM domain sequence of the recombinant protein comprises SEQ ID NO:103. In some aspects, the nucleic acid sequence encoding the TM domain sequence of the recombinant protein is SEQ ID NO:103.

[0238] In some aspects, the recombinant protein comprises a conserved amino acid sequence from S2'IFP, an immunogenic amino acid sequence from the RBD, and a TM domain sequence of the S protein.

[0239] In some aspects, the recombinant protein excludes amino acid sequences from the S protein that stimulate an immune response comprising non-neutralizing antibodies and/or that stimulate a Th2 cell-mediated immune response.

[0240] In some aspects, the amino acid sequence of the recombinant protein is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:55. In some aspects, the amino acid sequence of the recombinant protein comprises SEQ ID NO:55. In some aspects, the amino acid sequence of the recombinant protein is SEQ ID NO:55. In some aspects, the amino acid sequence of the recombinant protein comprises a replacement of one or more of K88, L123, E155, or N172 in SEQ ID NO:55 with another amino acid. In some aspects, the replacement at K88 is K88N. In some aspects, the replacement at K88 is K88T. In some aspects, the replacement at L123 is L123R. In some aspects, the replacement at E155 is E155K. In some aspects, the replacement at N172 is N172Y.

[0241] In some aspects, the nucleic acid sequence encoding the recombinant protein is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:56. In some aspects, the nucleic acid sequence encoding the recombinant protein comprises SEQ ID NO:56. In some aspects, the nucleic

acid sequence encoding the recombinant protein is SEQ ID NO:56. In some aspects, the nucleic acid sequence encoding the recombinant protein comprises a replacement of one or more of: the codon for lysine at nucleotide numbers 262-264 of SEQ ID NO:56 with a codon for another amino acid, the codon for leucine at nucleotide numbers 367-369 of SEQ ID NO:56 with a codon for another amino acid, the codon for glutamate at nucleotide numbers 463-465 of SEQ ID NO:56 with a codon for another amino acid, or the codon for asparagine at nucleotide numbers 514-516 of SEQ ID NO:56 with a codon for another amino acid. In some aspects, the codon for lysine at nucleotide numbers 262-264 is replaced with a codon for asparagine or threonine. In some aspects, the codon for leucine at nucleotide numbers 367-369 is replaced with a codon for arginine. In some aspects, the codon for glutamate at nucleotide numbers 463-465 is replaced with a codon for lysine. In some aspects, the codon for asparagine at nucleotide numbers 514-516 is replaced with a codon for tyrosine.

[0242] In some aspects, the expression cassette comprises a single open reading frame translated as an amino acid sequence at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:57. In some aspects, the expression cassette comprises a single open reading frame translated as an amino acid sequence comprising SEQ ID NO:57. In some aspects, the expression cassette comprises a single open reading frame translated as an amino acid sequence that is SEQ ID NO:57. In some aspects, the expression cassette comprises a single open reading frame translated as an amino acid sequence that comprises a replacement of one or more of P71, K423, L458, E490, or N507 in SEQ ID NO:57 with another amino acid. In some aspects, the replacement at P71 is P71L. In some aspects, the replacement at K423 is K423N. In some aspects, the replacement at K423 is K423T. In some aspects, the replacement at L458 is L458R. In some aspects, the replacement at E490 is E490K. In some aspects, the replacement at N507 is N507Y.

[0243] In some aspects, the expression cassette comprises a single open reading frame that is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:58. In some aspects, the expression cassette comprises a single open reading frame that comprises SEQ ID NO:58. In some

aspects, the expression cassette comprises a single open reading frame that is SEQ ID NO:58. In some aspects, the expression cassette comprises a single open reading frame that comprises a replacement of one or more of: the codon for proline at nucleotide numbers 211-213 in SEQ ID NO:58 with a codon for another amino acid, the codon for lysine at nucleotide numbers 1267-1269 of SEQ ID NO:58 with a codon for another amino acid, the codon for leucine at nucleotide numbers 1372-1374 of SEQ ID NO:58 with a codon for another amino acid, the codon for glutamate at nucleotide numbers 1468-1470 of SEQ ID NO:58 with a codon for another amino acid, or the codon for asparagine at nucleotide numbers 1519-1521 of SEQ ID NO:58 with a codon for another amino acid. In some aspects, the codon for proline at nucleotide numbers 211-213 in SEQ ID NO:58 is replaced with a codon for leucine. In some aspects, the codon for lysine at nucleotide numbers 1267-1269 is replaced with a codon for asparagine or threonine. In some aspects, the codon for leucine at nucleotide numbers 1372-1374 is replaced with a codon for arginine. In some aspects, the codon for glutamate at nucleotide numbers 1468-1470 is replaced with a codon for lysine. In some aspects, the codon for asparagine at nucleotide numbers 1519-1521 is replaced with a codon for tyrosine.

[0244] In some aspects, the expression cassette is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to any one of SEQ ID NOs:59-62. In some aspects, the expression cassette comprises any one of SEQ ID NOs:59-62. In some aspects, the expression cassette is any one of SEQ ID NOs:59-62.

[0245] In some aspects, the recombinant protein is capable of stimulating an immune response against COVID-19.

[0246] In some aspects, the recombinant protein is capable of stimulating a Th1 cell-mediated immune response against COVID-19.

[0247] In some aspects, the immune response against COVID-19 is against Wuhan-Hu-1 and/or one or more variants such as, but not limited to, the U.K. variant B.1.1.7, the South African variant B.1.351, the Brazilian variant P.1, or the Californian variant B.1.427/429.

[0248] In some aspects, the immune response is cross-reactive to other coronaviruses. In some aspects, the immune response is cross-reactive to other severe acute respiratory syndrome coronaviruses and/or human betacoronaviruses.

- [0249] In some aspects, the bacterial sequence-free vector further comprises at least one enhancer sequence flanking each side of the expression cassette. In some aspects, the at least one enhancer sequence is at least two enhancer sequences. In some aspects, the at least one enhancer sequence is a SV40 enhancer sequence.
- [0250] In some aspects, the bacterial sequence-free vector comprises circular covalently closed ends.
- [0251] In some aspects, the bacterial sequence-free vector comprises linear covalently closed ends. In some aspects, the bacterial sequence-free vector is a msDNA as disclosed herein. A vector map for an exemplary msDNA is shown in Figure 4.
- [0252] In some aspects, the bacterial sequence-free vector is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:104. In some aspects, the bacterial sequence-free vector comprises SEQ ID NO:104. In some aspects, the bacterial sequence-free vector is SEQ ID NO:104.

III. VLPs

- [0253] In some aspects, a VLP as disclosed herein is produced from the expression cassette of an expression vector and/or the expression cassette of a bacterial sequence-free vector as described herein.
- [0254] Provided herein is a recombinant cell comprising an expression vector or a bacterial sequence-free vector as described herein.
- [0255] In some aspects, the recombinant cell is a yeast, bacteria, archaeobacteria, fungi, insect, or animal cell, including a mammalian cell. In some aspects, recombinant cells include *Drosophila melanogaster* cells, *Saccharomyces cerevisiae* or other yeasts, *E. coli*, *Bacillus subtilis*, Sf9 cells, C129 cells, HEK293 cells, Neurospora, BHK, CHO, COS, HeLa cells, Hep G2 cells, and human cells and cell lines.
- [0256] In some aspects, the expression vector is for expression in a human cell or cell line such as the exemplary vector shown in Figure 2.
- [0257] In some aspects, the expression vector is a baculovirus vector such as the exemplary vector shown in Figure 5 and the cell type is an insect cell (e.g., Sf9 cells).
- [0258] In some aspects, the present disclosure is directed to a method of producing a VLP, comprising culturing the recombinant cell comprising the expression vector or the

bacterial sequence-free vector under suitable conditions for production of the VLP from the expression vector or the bacterial sequence-free vector.

- [0259] In some aspects, the method of producing a VLP further comprises isolating the VLP. In some aspects, the VLP produced by any of the above expression vectors or any of the above bacterial sequence-free vectors wherein the virus is a coronavirus.
- [0260] In some aspects, the VLP is isolated from a cell lysate.
- [0261] In some aspects, the isolating is by affinity purification. In some aspects, the affinity purification comprises microfluidics and/or chromatography.
- [0262] In some aspects, the affinity purification comprises an angiotensin-converting enzyme 2 (ACE2) receptor peptide or an anti-S protein monoclonal antibody.
- [0263] In some aspects, the ACE2 receptor peptide comprises an amino acid sequence that is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:70. In some aspects, the ACE2 receptor peptide comprises SEQ ID NO:70. In some aspects, the ACE2 receptor peptide is SEQ ID NO:70.
- [0264] In some aspects, the ACE2 receptor peptide comprises a biotin acceptor peptide (BAP) tag at the C-terminus or N-terminus of the peptide. In some aspects, the BAP tag comprises an amino acid sequence at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:71. In some aspects, the BAP tag comprises SEQ ID NO:71. In some aspects, the BAP tag is SEQ ID NO:71.
- [0265] In some aspects, the ACE2 receptor peptide or anti-S protein monoclonal antibody is biotinylated and immobilized on a streptavidin-coated bead. In some aspects, the affinity purification comprises microfluidics and/or chromatography.
- [0266] In some aspects, the present disclosure is directed to a VLP produced by the method.
- [0267] Provided herein is a VLP comprising a recombinant protein comprising a conserved amino acid sequence from a virus fused to an immunogenic amino acid sequence.

- [0268] In some aspects, the immunogenic amino acid sequence is from the same virus as the conserved amino acid sequence.
- [0269] In some aspects, the conserved amino acid sequence is from a viral glycoprotein. In some aspects, the immunogenic amino acid sequence is from the same viral glycoprotein.
- [0270] In some aspects, the VLP further comprises a viral envelope protein and/or a viral matrix protein. In some aspects, the viral envelope protein and/or the viral matrix protein are from the same virus as the conserved amino acid sequence.
- [0271] In some aspects, the conserved amino acid sequence, the immunogenic amino acid sequence, the viral envelope protein, and/or the viral matrix protein is a consensus sequence.
- [0272] In some aspects, the recombinant protein is capable of stimulating an immune response against the virus comprising neutralizing antibodies.
- [0273] In some aspects, the recombinant protein is capable of stimulating a Th1 cell-mediated immune response against the virus.
- [0274] In some aspects, the immune response is cross-reactive to a related virus or strain.
- [0275] In some aspects, the recombinant protein excludes amino acid sequences from the virus that stimulate an immune response comprising non-neutralizing antibodies and/or that stimulate a Th2 cell-mediated immune response.
- [0276] In some aspects, the virus is a coronavirus, an influenza virus, a human immunodeficiency virus, a human papillomavirus, a hepatitis virus, or an oncolytic virus.
- [0277] In some aspects, the influenza virus is an influenza A virus. In some aspects, the influenza A virus is H1N1, H5N1, or H3N2.
- [0278] In some aspects, the influenza virus is an influenza B virus.
- [0279] In some aspects, the coronavirus is a human coronavirus such as, but not limited to, HCoV-229E, HCoV-NL63, HCoV-OC43, HCoV-HKU1, SARS-CoV-1, SARS-CoV-2 (*i.e.*, COVID-19)), and/or MERS-CoV.
- [0280] In some aspects, the coronavirus is COVID-19 (*i.e.*, Wuhan-Hu-1 or a variant thereof such as, but not limited to, U.K. variant B.1.1.7, South African variant B.1.351, Brazilian variant P.1, or Californian variant B.1.427/429).

- [0281] In some aspects, the VLP comprises a coronavirus M protein, a coronavirus E protein, and a recombinant protein comprising a conserved amino acid sequence and an immunogenic amino acid sequence from a coronavirus S protein.
- [0282] In some aspects, the M protein comprises an amino acid sequence at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:1. In some aspects, the M protein comprises SEQ ID NO:1. In some aspects, the M protein is SEQ ID NO:1.
- [0283] In some aspects, the E protein comprises an amino acid sequence at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:3. In some aspects, the E protein comprises SEQ ID NO:3. In some aspects, the E protein is SEQ ID NO:3. In some aspects, the E protein comprises a replacement of P71 in SEQ ID NO:3 with another amino acid. In some aspects, the replacement at P71 in SEQ ID NO:3 is P71L.
- [0284] In some aspects, the conserved amino acid sequence is from the S1 subunit or the S2 subunit of the S protein, the RBD of the S protein, the S protein S2' cleavage site and internal fusion peptide (IFP) of the S protein, the M protein, or the E protein.
- [0285] In some aspects, the conserved amino acid sequence is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to any one of SEQ ID NOs:12-54. In some aspects, the conserved amino acid sequence comprises any one of SEQ ID NOs:12-54. In some aspects, the conserved amino acid sequence is any one of SEQ ID NOs:12-54.
- [0286] In some aspects, the conserved amino acid sequence is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:7. In some aspects, the conserved amino acid sequence comprises SEQ ID NO:7. In some aspects, the conserved amino acid sequence is SEQ ID NO:7.
- [0287] In some aspects, the immunogenic amino acid sequence is from the S protein receptor-binding domain (RBD).

- [0288]** In some aspects, the immunogenic amino acid sequence is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:11. In some aspects, the immunogenic amino acid sequence comprises SEQ ID NO:11. In some aspects, the immunogenic amino acid sequence is SEQ ID NO:11. In some aspects, the immunogenic amino acid sequence comprises a replacement of one or more of: K88, L123, E155, or N172 in SEQ ID NO:11 with another amino acid. In some aspects, the replacement at K88 is K88N. In some aspects, the replacement at K88 is K88T. In some aspects, the replacement at L123 is L123R. In some aspects, the replacement at E155 is E155K. In some aspects, the replacement at N172 is N172Y.
- [0289]** In some aspects, the recombinant protein further comprises a transmembrane (TM) domain sequence from the S protein.
- [0290]** In some aspects, the TM domain sequence comprises an amino acid sequence at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:102. In some aspects, the TM domain sequence comprises SEQ ID NO:102. In some aspects, the TM domain sequence is SEQ ID NO:102.
- [0291]** In some aspects, the recombinant protein comprises a conserved amino acid sequence from S2'IFP, an immunogenic amino acid sequence from the RBD, and a TM domain sequence of the S protein.
- [0292]** In some aspects, the recombinant protein excludes amino acid sequences from the S protein that stimulate an immune response comprising non-neutralizing antibodies and/or that stimulate a Th2 cell-mediated immune response.
- [0293]** In some aspects, the amino acid sequence of the recombinant protein is at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:55. In some aspects, the amino acid sequence of the recombinant protein comprises SEQ ID NO:55. In some aspects, the amino acid sequence of the recombinant protein is SEQ ID NO:55. In some aspects, the amino acid sequence of the recombinant protein comprises a replacement of one or more of K88, L123, E155,

or N172 in SEQ ID NO:55 with another amino acid. In some aspects, the replacement at K88 is K88N. In some aspects, the replacement at K88 is K88T. In some aspects, the replacement at L123 is L123R. In some aspects, the replacement at E155 is E155K. In some aspects, the replacement at N172 is N172Y.

[0294] Provided herein is a VLP comprising a recombinant protein at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:55, an M protein at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:1, and an E protein at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to SEQ ID NO:3.

[0295] Provided herein is a VLP comprising a recombinant protein that comprises SEQ ID NO:55, an M protein that comprises SEQ ID NO:1, and an E protein that comprises SEQ ID NO:3.

[0296] Provided herein is a VLP comprising the recombinant protein of SEQ ID NO:55, the M protein of SEQ ID NO:1, and the E protein of SEQ ID NO:3.

[0297] In some aspects, the recombinant protein is capable of stimulating an immune response against COVID-19.

[0298] In some aspects, the recombinant protein is capable of stimulating a Th1 cell-mediated immune response against COVID-19.

[0299] In some aspects, the immune response against COVID-19 is against Wuhan-Hu-1 and/or one or more variants such as, but not limited to, the U.K. variant B.1.1.7, the South African variant B.1.351, the Brazilian variant P.1, or the Californian variant B.1.427/429

[0300] In some aspects, the immune response is cross-reactive to other coronaviruses.

[0301] In some aspects, the immune response is cross-reactive to other severe acute respiratory syndrome coronaviruses and/or human betacoronaviruses.

IV. Compositions

[0302] Provided herein is a composition comprising any of the expression vectors, bacterial sequence-free vectors, or VLPs as described herein.

- [0303]** In some aspects, the composition further comprises a physiologically acceptable carrier, excipient, or stabilizer. *See, e.g., Remington: The Science and Practice of Pharmacy*, 22nd ed. (2013). Acceptable carriers, excipients, or stabilizers can include those that are nontoxic to a subject. In some aspects, the composition or one or more components of the composition are sterile. A sterile component can be prepared, for example, by filtration (*e.g.*, by a sterile filtration membrane) or by irradiation (*e.g.*, by gamma irradiation).
- [0304]** An excipient of the present invention can be described as a "pharmaceutically acceptable" excipient when added to a pharmaceutical composition, meaning that the excipient is a compound, material, composition, salt, and/or dosage form which is, within the scope of sound medical judgment, suitable for contact with the tissues of human beings and animals without excessive toxicity, irritation, allergic response, or other problematic complications over the desired duration of contact commensurate with a reasonable benefit/risk ratio. In some aspects, the term "pharmaceutically acceptable" means approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized international pharmacopeia for use in animals, and more particularly in humans. Various excipients can be used. In some aspects, the excipient can be, but is not limited to, an alkaline agent, a stabilizer, an antioxidant, an adhesion agent, a separating agent, a coating agent, an exterior phase component, a controlled-release component, a solvent, a surfactant, a humectant, a buffering agent, a filler, an emollient, or combinations thereof. Excipients in addition to those discussed herein can include excipients listed in, though not limited to, *Remington: The Science and Practice of Pharmacy*, 22nd ed. (2013). Inclusion of an excipient in a particular classification herein (*e.g.*, "solvent") is intended to illustrate rather than limit the role of the excipient. A particular excipient can fall within multiple classifications.
- [0305]** A pharmaceutical composition of the disclosure is formulated to be compatible with its intended route of administration. Exemplary routes of administration include enteral, topical, parenteral, oral, pulmonary, intranasal, intravenous, epidermal, transdermal, subcutaneous, intramuscular, or intraperitoneal administration, or inhalation. "Parenteral administration" as used herein means modes of administration other than enteral and topical administration, usually by injection or infusion, and includes, without limitation, intravenous, intramuscular, intraarterial, intrathecal, intralymphatic,

intralesional, intracapsular, intraorbital, intracardiac, intradermal, intraperitoneal, transtracheal, subcutaneous, subcuticular, intraarticular, subcapsular, subarachnoid, intraspinal, epidural, intrapleural, and intrasternal injection and infusion, as well as in vivo electroporation. In some aspects, the formulation is administered via a non-parenteral route, in some aspects, orally. Other non-parenteral routes include a topical, epidermal, or mucosal route of administration, for example, intranasally, vaginally, rectally, sublingually or topically.

- [0306] In some aspects, the pharmaceutical composition is lyophilized.
- [0307] A variety of methods are known in the art and are suitable for introduction of nucleic acids into a cell. Examples include, but are not limited to, electroporation, calcium phosphate mediated transfer, nucleofection, sonoporation, heat shock, magnetofection, liposome mediated transfer, microinjection, microprojectile mediated transfer (nanoparticles), cationic polymer mediated transfer (DEAE-dextran, polyethylenimine, polyethylene glycol (PEG), and the like), or cell fusion.
- [0308] Nanoparticle carriers such as liposomes, micelles, and polymeric nanoparticles have been investigated for improving bioavailability and pharmacokinetic properties of therapeutics via various mechanisms, for example, the enhanced permeability and retention (EPR) effect.
- [0309] Further improvement can be achieved by conjugation of targeting ligands onto nanoparticles to achieve selective delivery to a target cell. For example, receptor-targeted nanoparticle delivery has been shown to improve therapeutic responses both *in vitro* and *in vivo*. Targeting ligands that have been investigated include folate, transferrin, antibodies, peptides, and aptamers. Additionally, multiple functionalities can be incorporated into the design of nanoparticles, *e.g.*, to enable imaging and to trigger intracellular drug release.
- [0310] In some aspects, the composition further comprises a delivery agent. In some aspects, the delivery agent is a nanoparticle. In some aspects, the delivery agent is selected from the group consisting of liposomes, non-lipid polymeric molecules, endosomes, and any combination thereof.
- [0311] In some aspects, the delivery agent (*e.g.*, a nanoparticle) comprises a targeting ligand.

- [0312] In some aspects, the targeting ligand comprises a S protein peptide with binding affinity to the ACE2 receptor (*e.g.*, for delivery of an expression vector, bacterial sequence-free vector, or VLP comprising coronavirus sequences).
- [0313] In some aspects, the S protein peptide is from a conserved region of the S protein. In some aspects, the length of the S protein peptide is from 3 amino acids to 100 amino acids, including any length or range of lengths therein, such as 3 amino acids to 90, 80, 70, 60, 50, 40, 30, 20, or 10 amino acids.
- [0314] In some aspects, the S protein peptide comprises an amino acid sequence at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identical to any one of SEQ ID NOs:76-99. In some aspects, the S protein peptide comprises any one of SEQ ID NOs:76-99. In some aspects, the S protein peptide is any one of SEQ ID NOs:76-99.

V. Therapeutic Uses and Methods

- [0315] The expression vectors, bacterial sequence-free vectors (*e.g.*, msDNA), VLPs, and compositions as described herein can be utilized for prophylactic or therapeutic treatment of a subject in need thereof, including as a vaccine against a viral infection (*e.g.*, a coronavirus infection such as COVID-19) infection or as a treatment for individuals infected with a virus.
- [0316] Provided herein is a vaccine for a viral infection comprising an expression vector, bacterial sequence-free vector, VLP, or composition as described herein.
- [0317] Provided herein is a method of treating a viral infection in a subject, comprising administering to the subject an expression vector, bacterial sequence-free vector, VLP, or composition as described herein, wherein intracellular expression of the expression vector or the bacterial sequence-free vector in the subject produces a VLP.
- [0318] Provided herein is an expression vector, bacterial sequence-free vector, VLP, or composition as described herein for use in treating a viral infection in a subject, wherein intracellular expression of the expression vector or the bacterial sequence-free vector in the subject produces a VLP.
- [0319] Provided herein is use of an expression vector, bacterial sequence-free vector, VLP, or composition for treating a viral infection in a subject, wherein intracellular

expression of the expression vector or the bacterial sequence-free vector in the subject produces a VLP.

[0320] Provided herein is use of an expression vector, bacterial sequence-free vector, VLP, or composition for the preparation of a medicament for treating a viral infection in a subject, wherein intracellular expression of the expression vector or the bacterial sequence-free vector in the subject produces a VLP.

[0321] The expression vector, bacterial sequence-free vector, or composition can be administered to a subject by any route of administration that is effective in treating the viral infection.

[0322] In some aspects, the administering is by enteral, topical, parenteral, oral, pulmonary, intranasal, intravenous, epidermal, transdermal, subcutaneous, intramuscular, or intraperitoneal administration, or inhalation.

[0323] In some aspects, the administering is by parenteral or non-parenteral administration.

[0324] In some aspects, the parenteral administration is by injection or infusion.

[0325] In some aspects, the parenteral administration is by intravenous, intramuscular, intraarterial, intrathecal, intralymphatic, intralesional, intracapsular, intraorbital, intracardiac, intradermal, intraperitoneal, transtracheal, subcutaneous, subcuticular, intraarticular, subcapsular, subarachnoid, intraspinal, epidural, intrapleural, or intrasternal injection or infusion, or by in vivo electroporation.

[0326] In some aspects, the non-parenteral administration is oral, topical, epidermal, mucosal, intranasal, vaginal, rectal, or sublingual.

[0327] In some aspects, the administering is by oral, pulmonary, intranasal, intravenous, epidermal, transdermal, subcutaneous, intramuscular, or intraperitoneal administration, or by inhalation.

[0328] In some aspects, the administering is by the route of viral infection and transmission.

[0329] In some aspects, the route of viral infection and transmission is mucosal.

[0330] In some aspects, the administering is by oral, nasal, or pulmonary administration for a respiratory tract infection. In some aspects, the administering is by nasal administration.

- [0331] Applying the inhalation and intranasal routes of administration provide a powerful opportunity to generate supporting immune responses via lungs and nasopharyngeal-associated lymphoid tissues (NALT) in addition to efficient, targeted, and non-invasive delivery of a VLP as described herein to lower respiratory tract tissue.
- [0332] In some aspects, the administering is vaginal administration for a sexually transmitted infection.
- [0333] In some aspects, the administering is by intramuscular, subcutaneous, or intradermal administration where both the site and depth of injection effect the immune response. Intramuscular injection offers a powerful alternative and commonly used technique for vaccine administration, particularly as it is validated and readily re-administered.
- [0334] Administering can be performed, for example, once, a plurality of times, and/or over one or more extended periods. In some aspects, the administering is one time, two times (*e.g.*, a first administration followed by a second administration about 1, about 2, about 3, about 4 or more weeks later), once about every week, once about every month, once about every 2 months, once about every 3 months, once about every 4 months, once about every 6 months, once about every year, or once about every decade.
- [0335] The expression cassette as described herein provides a VLP conferring a robust humoral immune response with the benefits of a DNA vaccine for internal processing of intracellular pathogen epitopes for T-cell presentation and cell-mediated immunity. In some aspects, immunodominance is successfully conferred to the conserved amino acid sequence of the recombinant protein, and the vaccine generates universal coronavirus immunity.
- [0336] In some aspects, VLPs that self-assemble intracellularly from translation products of the expression cassette (whether from the expression vector or a bacterial sequence-free vector as described herein) generate a Th1 cell-mediated response as presented in: 1) an MHC-I context to prime specific cytotoxic T-cell activity against virally infected cells; 2) an MHC-II context in phagocytic antigen presenting cells (APCs) for complementary humoral and cell-mediated support.
- [0337] In some aspects, intracellular assembly of VLP from the expression cassettes as described herein eliminates potential vaccine-mediated TH2 immunopathology and any associated requirement for adjuvant therapy.

- [0338] In some aspects, the VLP stimulates an immune response in the subject comprising neutralizing antibodies against the viral infection.
- [0339] In some aspects, the VLP stimulates a Th1 cell-mediated immune response in the subject against the viral infection.
- [0340] In some aspects, the immune response is cross-reactive to a related virus or strain.
- [0341] In some aspects, the VLP does not stimulate an immune response comprising non-neutralizing antibodies in the subject and/or does not stimulate a Th2 cell-mediated immune response in the subject.
- [0342] In some aspects, the VLP induces antibodies that block viral receptor binding, viral genome uncoating, and/or genome injection.
- [0343] In some aspects, the VLP cross-competes with the infecting virus for binding to a viral receptor.
- [0344] In some aspects, the VLP cross-competes with a related virus or strain for binding to the viral receptor.
- [0345] In some aspects, the viral infection is a coronavirus, an influenza virus, a human immunodeficiency virus, a human papillomavirus, a hepatitis virus, or an oncolytic virus.
- [0346] In some aspects, the influenza virus is an influenza A virus. In some aspects, the influenza A virus is H1N1, H5N1, or H3N2.
- [0347] In some aspects, the influenza virus is an influenza B virus.
- [0348] In some aspects, the coronavirus is a human coronavirus such as, but not limited to, HCoV-229E, HCoV-NL63, HCoV-OC43, HCoV-HKU1, SARS-CoV-1, SARS-CoV-2 (*i.e.*, COVID-19), and/or MERS-CoV.
- [0349] In some aspects, the coronavirus is COVID-19 (*i.e.*, Wuhan-Hu-1 or a variant thereof such as, but not limited to, U.K. variant B.1.1.7, South African variant B.1.351, Brazilian variant P.1, or Californian variant B.1.427/429).
- [0350] In some aspects, the VLP stimulates an immune response in the subject comprising neutralizing antibodies against COVID-19.
- [0351] In some aspects, the VLP stimulates a Th1 cell-mediated immune response in the subject against COVID-19.
- [0352] In some aspects, the immune response against COVID-19 is against Wuhan-Hu-1 and/or one or more variants such as, but not limited to, the U.K. variant B.1.1.7, the South African variant B.1.351, the Brazilian variant P.1, or the Californian variant B.1.427/429.

- [0353] In some aspects, the immune response is cross-reactive to other coronaviruses.
- [0354] In some aspects, the immune response is cross-reactive to other severe acute respiratory syndrome coronaviruses and/or human betacoronaviruses.
- [0355] In some aspects, the VLP does not stimulate an immune response comprising non-neutralizing antibodies in the subject and/or does not stimulate a Th2 cell-mediated immune response in the subject.
- [0356] In some aspects, the administering is by inhalation.
- [0357] The cellular ligand for COVID-19 and many other coronaviruses is the ACE2 receptor found in the lower respiratory tract of humans, which regulates both cross-species and human-to-human transmission. The ACE2 receptor is bound by the S glycoprotein on the surface of coronavirus that, upon fusion, forms a replication-transcription complex in a double membrane vesicle (Letko *et al.*, *Nat. Microbiol.* 5(4): 562–569 (2020); Wan *et al.*, *J. Virol.* 4(7) e00127-20 (2020)). The continuous replication and synthesis of nested sets of subgenomic RNAs encode accessory proteins and structural proteins for the viral particles to bud. This causes the virion-containing vesicles to fuse with plasma membrane ultimately releasing the virus into the host (Fehr and Perlman). Hypertensive patients on adrenergic blocking agents (beta-blockers) to control blood pressure are particularly susceptible to infection as beta blockers stimulate ACE2 receptor over-expression in the respiratory tract facilitating viral binding and infection. Susceptibility has also been noted in patients underlying medical conditions such as COPD, diabetes, and cardiovascular disease (Guan *et al.*, *Eur. Resp. Journal*, 2000547; DOI: 10.1183/13993003.00547-2020 (2020)).
- [0358] In some aspects, a VLP against coronavirus (*e.g.*, COVID-19) as described herein not only delivers a therapeutic DNA vaccine, but also competes for available coronavirus receptor sites in respiratory tissue, attenuating further infection.
- [0359] In some aspects, the extrusion of functional VLPs (expressing surface RBD) from cells further promotes competitive interference for available ACE2 receptors on target cells and promotes interaction with B-cells to ensure a robust neutralizing humoral response.
- [0360] In some aspects, the S2'IFP domain for presentation exposes the highly conserved site and confers immuno-dominance to the determinant via hapten-carrier response.

- [0361] In some aspects, the VLP cross-competes with COVID-19 for binding to ACE2 receptor, neuropilin-1, and/or other receptors.
- [0362] In some aspects, the VLP cross-competes with other coronaviruses for binding to ACE2 receptor, neuropilin-1, and/or other receptors.
- [0363] In some aspects, the VLP cross-competes with other severe acute respiratory syndrome coronaviruses and/or human betacoronaviruses for binding to ACE2 receptor, neuropilin-1, and/or other receptors.
- [0364] The following examples are offered by way of illustration and not by way of limitation.

EXAMPLES

Example 1

A. **Generation of expression vectors for producing bacterial sequence-free vectors and VLPs**

- [0365] Four expression vectors were produced by cloning sequences derived from COVID-19 into the multicloning site between two specialized supersequence (SS) sites in a ministring expression vector (Mediphage Bioceuticals, Inc., Toronto, CA) as described in U.S. Patent Nos. 9,290,778 and 9,862,954, incorporated by reference herein in their entireties.
- [0366] The sequences derived from COVID-19 included sequences encoding Envelope (E) protein (GenBank Accession No. QHD43418.1; SEQ ID NO:3) and Membrane (M) protein (GenBank Accession No. QHD43419.1; SEQ ID NO:1). Additionally, a sequence encoding a recombinant Spike (S) protein was produced that contained a fusion of sequences associated with the receptor-binding domain (RBD), the S2' cleavage site and internal fusion peptide (S2'IFP), and the transmembrane (TM) domain (RBD::S2'IFP::TM; SEQ ID NO:55) of the COVID-19 S protein (GenBank Accession No. QHD43416.1; SEQ ID NO:5). The recombinant S protein was engineered to exclude amino acid sequences from the S protein that stimulate an immune response comprising non-neutralizing antibodies and to exclude amino acid sequences that stimulate a Th2 cell-mediated immune response.
- [0367] The expression cassettes of three of the expression vectors contained the E protein, the M protein, and the recombinant S protein fused into a single polynucleotide

(SEQ ID NO:58) via sequences encoding the self-cleaving peptide P2A from porcine teschovirus-1 2A under the control of a cytomegalovirus (CMV) promoter. **Figure 1** illustrates an exemplary expression cassette.

[0368] One of the three expression vectors contained the expression cassette "CMV-E-P2A-M-P2A-RBD::S2'IFP::TM-bGHPolyA" (SEQ ID NO:60), which contained a bovine growth hormone (bGH) polyadenylation (polyA) signal. A map of the expression vector containing the expression cassette is shown in **Figure 2** (pGL2-SS-CMV-VLP-BGH-SS, SEQ ID NO:63).

[0369] Another of the three expression vectors contained the expression cassette "CMV-E-P2A-M-P2A-RBD::S2'IFP::TM-SV40polyA" (SEQ ID NO:59), which contained a simian virus 40 (SV40) polyA.

[0370] Another of the three expression vectors contained the expression cassette "CMV-E-P2A-M-P2A-RBD::S2'IFP::TM-T2A-GFP-SV40polyA" (SEQ ID NO:61), which contained a green fluorescent protein (GFP) fused to the COVID-19 sequences via a sequence encoding the self-cleaving peptide T2A from thosea asigna virus 2A and a SV40 polyA.

[0371] A fourth expression vector contained the expression cassette "CMV-E-P2A-M-T2A-MCS-bGHPolyA" (SEQ ID NO:62), which contained a single polynucleotide having the E protein and the M protein fused to one another via a sequence encoding P2A in turn fused to a multiple cloning site (MCS) via a sequence encoding T2A. The expression cassette also contained a CMV promoter and a bGH polyA. The MCS is for insertion of additional sequences, such as recombinant proteins comprising conserved and immunogenic sequences as disclosed herein.

[0372] The expression vectors containing the expression cassettes of SEQ ID NOs:59-62 are the same as the expression vector of **Figure 2** and SEQ ID NO:63 except for the different expression cassette.

B. Expression of COVID-19 genes

[0373] Human lung A549 cells (1×10^6) were electroporated with 1 μ g of the expression vector shown in **Figure 2**, or no expression vector. Total RNA was extracted after 48 hours after electroporation and converted to cDNA libraries. 1 μ L of cDNA was used as template for Real Time qRT-PCR for E, M, and RBD::S2'IFP::TM transgenes using the

gene-specific primers for E, M, and RBD, respectively, shown below in **Table 1**.

Expression of the transgenes was normalized to β -actin expression.

Table 1. Primer Sequences

Gene	Forward Primer	Reverse Primer
E	ACTGCTGCAACATCGTGAA C (SEQ ID NO:64)	TGCTAGAATTCAGGTTCTTC ACC (SEQ ID NO:65)
M	TTCCTGTGGCTGCTGTGG (SEQ ID NO:66)	ATGACCAGCTCGCTTTCCA G (SEQ ID NO:67)
RBD::S2'IFP:: TM	ATCAGCACAGAGATCTACC AGG (SEQ ID NO:68)	AGCACCACCACTCTGTAAG G (SEQ ID NO:69)

[0374] As shown in **Figure 3(A)**, each of the transgenes was detected in cDNA libraries from cells electroporated with the expression vector ("VLP") but not in cDNA libraries from control cells ("CTL"). The relative gene expression shown in the figure was calculated by $\Delta\Delta CT$ method. Statistical analysis was performed using 1-way ANOVA (***) = $p < 0.001$, **** = $p < 0.0001$).

C. Expression of recombinant Spike protein

[0375] HEK 293 cells (1×10^6) were transfected with 2 μ g of the expression vector of Figure 2 using Lipofectamine® 3000 Reagent (Invitrogen). Protein samples were collected 48 hours after transfection. Western blots were prepared by loading 50 μ g of whole protein lysate from transfected cells as well as from control cells that were not transfected. A rabbit polyclonal anti-RBD antibody was used to in the detection of recombinant S protein, while a rabbit polyclonal anti-beta-actin antibody was used in the detection of beta-actin as a loading control. An anti-rabbit-horse radish peroxidase (HRP) antibody and chemiluminescence imaging was used for signal detection. A representative Western blot is shown in **Figure 3(B)**, showing that recombinant S protein was detected in protein isolated from cells transfected with the expression vector ("VLP") but not in protein isolated from control cells.

[0376] The relative mean protein intensity of recombinant S protein expression in transfected and control cells was determined by densitometry analysis of Western blot images (n=3). *See Figure 3(C).*

Example 2

Stimulation of antibody production by VLP expression vectors

[0377] The expression vector of **Figure 2** was encapsulated in lipid nanoparticles (Entos Pharmaceuticals) and administered to C57 mice at a dose of 100 µg via intramuscular injection at day 0 followed by a booster dose of 100 µg via intramuscular injection at day 14. Serum was collected via tail vein every 7 days through day 49.

[0378] Antibody concentrations in mouse serum were assessed by indirect ELISA by binding to purified S1 protein (Abclonal, Inc.).

[0379] Serum was diluted to 1% in PBS and then added to ELISA plates containing the S1 protein. Mouse serum antibodies that bound to the S1 protein were detected by anti-mouse IgG SULFO-TAGTM conjugated antibody (Meso Scale Diagnostics, LLC).

[0380] Antibody concentrations are shown in **Figures 5A** and **5B**. Concentrations peaked at day 21 at about 5000 ng/mL, with consistent expression maintained at about 3000 ng/mL through day 49.

Example 3

A. Characterization of COVID-19 genomic sequence conservation

[0381] A total of 3928 representative complete COVID-19 genomes were downloaded from the GISAID database (<https://www.gisaid.org>). Collection dates for the genomes ranged from December 2019 to February 2021 and contained all major variant strains as well as the Wuhan reference genome (NC_045512.2). Genomes were aligned to the Wuhan reference genome using the MAFFT multiple sequence alignment program. Sequence conservation and nucleotide frequency analyses were performed.

[0382] **Figure 6** shows a sequence conservation analysis of the 3928 representative COVID-19 genomes. **(A)** Horizontal tracks indicate the genomic positions (indicated on the x-axis) of all COVID-19 genes (depicted on the y-axis) as per the Wuhan reference genome. **(B)** The bar heights in the histogram correspond to the percent of genomes that

differed from the Wuhan reference genome in each given genomic position. The bar plot and histogram were generated in R version 3.6.1 using the ggplot2 package.

[0383] As shown in **Figure 6**, the COVID-19 genome has a relatively high level of sequence conservation with few key genomic variants. Ignoring variable 5' and 3' end regions, only three genomic positions were found to differ from the reference genome in >50% of sequences. Two of these single nucleotide polymorphisms (SNPs) were found within ORF1ab (the first (C241T) in an intergenic region and the second (C14408T -> L4715)) within a coding region, and the third (D614G) within the Spike (S) protein.

B. Characterization of human beta coronavirus genomic sequence conservation

[0384] In addition to the 3928 representative complete COVID-19 genomes discussed in part A of this example, 120 SARS-CoV (the virus responsible for SARS) genomes and 257 MERS-CoV (the virus responsible for MERS) genomes were downloaded from the NCBI GenBank® database. Genomes were aligned to the COVID-19 Wuhan reference genome using the MAFFT multiple sequence alignment program. The comparison was possible due to similar genomic organization across these three viral genomes. Sequence conservation and nucleotide frequency analyses were performed.

[0385] **Figure 7** shows a histogram in which the bar heights correspond to the percent of genomes that differed from the Wuhan reference genome in each given genomic position. The histogram was generated in R version 3.6.1 using the ggplot2 package.

[0386] As shown in **Figure 7**, the genomes of other prominent human beta coronaviruses (SARS-CoV and MERS-CoV) also have relatively high levels of sequence conservation as compared to the COVID-19 genome.

C. Identification of functionally relevant mutations in prominent variant COVID-19 strains

[0387] The 3928 COVID-19 sequences discussed in part A were filtered for those belonging to key variant strains (U.K. variant B.1.1.7 (n=233), South African variant B.1.351 (n=104), Brazilian variant P.1 (n=39), and Californian variant B.1.427/429 (n=62)). Genomes of the four variant strains were independently aligned to the SARS-CoV-2 Wuhan reference genome (NC_045512.2) using the MAFFT multiple sequence alignment program. Sequence conservation and nucleotide frequency analyses were performed. Functional importance was determined via assessment of BLOSUM 62 matrix score, surface exposure analysis (via PyMol), and literature review.

[0388] **Figures 8A-D** shows histograms in which the bar heights correspond to the percent of the variant genomes (B.1.1.7 in **8A**, B.1.351 in **8B**, P.1 in naïve, and B.1.427/429 in **8D**) that differed from the Wuhan reference genome in each given genomic position. The histograms were generated in R version 3.6.1 using the ggplot2 package.

[0389] **Table 2** shows a summary of the identified SNPs from variant COVID-19 strains located in regions of the COVID-19 genome contained within the expression cassette shown in **Figure 1**.

Table 2. Summary of Identified SNPs

Expression Cassette Sequences	COVID-19 Variants			
	U.K. (B.1.1.7)	South Africa (B.1.357)	Brazil (P.1)	California (B.1.427/429)
RBD	AAT > TAT → N501Y	AAT > TAT → N501Y GAA > AAA → E484K AAG > AAT → K417N	AAT > TAT → N501Y GAA > AAA → E484K AAG > ACG → K417T	CTG > CGG → L452R
S2'IFP	-	-	-	-
E	-	CCT > CTT → P71L	-	-
M	-	-	-	TTC > TTT → F53F

[0390] SNPs identified in the receptor-binding domain (RBD) region of the Spike (S) protein of the variant COVID-19 strains were mapped onto a referenced Protein Data Bank (PDB) structure (PDB ID: 6VXX) to assess surface exposure. The N501, K417, and L452 residues were determined to be surface exposed and therefore of potentially greater consequence. The E484 residue was determined not to be surface exposed.

[0391] Surface exposure of SNPs identified in the Envelope (E) protein of the variant COVID-19 strains were assessed via structural information in Bianchi *et al.*, *BioMed*

Research International, <https://doi.org/10.1155/2020/4389089> (2020). The P71 residue was determined to be surface exposed and therefore of potentially greater consequence.

[0392] The SNP identified in the membrane (M) protein results in a synonymous mutation and therefore functional analysis was not performed.

[0393] Overall, the analysis showed that sequences selected for the VLP expression cassette as shown in **Figure 1** are relatively robust against COVID-19 variants, especially the S2'IFP site which is completely conserved across all key variant strains as well as in other coronaviruses (SARS-CoV and MERS-CoV).

Example 4

A. Generation of bacterial sequence-free vectors for producing VLP

[0394] DNA ministrings for producing VLP (msDNA-VLP) are produced in inducible *E. coli* cells from the expression vectors described in Example 1 according to methods described in U.S. Patent Nos. 9,290,778 and 9,862,954.

[0395] msDNA-VLP is purified and concentrated, with quality control testing for purity and sequence.

B. Complexation of bacterial sequence-free vectors with nanoparticles

[0396] The purified msDNA-VLP and a control msDNA (msDNA-control) expressing a marker protein (*e.g.*, GFP) are complexed with nanoparticles (*e.g.*, lipid nanoparticles (LNPs)). In other studies, commercial LNPs have demonstrated strong transfection efficiency in lung in vivo with msDNA (unpublished data). Commercial LNPs are used as in vitro controls. Commercial JetPEI (<https://www.polyplus-transfection.com/products/cgmp-grade-in-vivo-jetpei/>) is used as an in vivo control.

[0397] The msDNA nanoparticles are lyophilized for in vitro and in vivo tests.

C. In vitro VLP formation and immune responses from bacterial sequence-free vectors

[0398] The msDNA nanoparticles (*i.e.*, as described in part B of this example) as well as naked msDNA (*i.e.*, msDNA-VLP as described in part A of this example and msDNA-control that are not complexed with nanoparticles) are delivered into a human cell line expressing ACE2 receptors (*e.g.*, A549 cells (ATCC CCL-185)), vascular endothelial cell, or alveolar epithelial cells (Yen, T.-T., *et al.*, *Journal of Virology* 80(6): 2684-2693

(2006); Qian, Z. *et al.*, *American Journal of Respiratory Cell and Molecular Biology* 48(6): 742-748 (2013)). Efficiency of the delivery and mean fluorescence are assessed.

- [0399] Intracellular VLP formation is assessed by transmission electron microscopy.
- [0400] Cytokine storm and over-activity of inflammation response would be assessed in cell cultures using immune assay techniques.

D. Production of VLP in vitro in a eukaryotic expression system

- [0401] A eukaryotic expression vector is produced comprising M-P2A-E and RBD::S2'::TM under control of a promoter for VLP production in eukaryotic cells. An exemplary baculoviral expression vector for VLP production in Sf9 cells is shown in **Figure 9**. VLP is produced in vitro and purified using standard techniques.

E. In vivo VLP production and immune responses from bacterial sequence-free vectors

- [0402] The msDNA nanoparticles (*i.e.*, as described in part B of this example) are administered by inhalation, intranasal, or intramuscular routes in an animal model. Cytokine profiles, immunoglobulin profiles, and protective effects against COVID-19 are determined.
- [0403] For inhalation and intranasal routes, the following administrations are performed: (1) lyophilized msDNA-VLP or msDNA-control nanoparticles are administered by inhalation in one or multiple doses (e.g., dosing at 1, 2, 3, and/or 4 weeks; dosing at 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and/or 12 months; and/or annual intervals); (2) lyophilized msDNA-VLP or msDNA-control nanoparticles are administered by inhalation in one or multiple doses (e.g., dosing at 1, 2, 3, and/or 4 weeks; dosing at 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and/or 12 months; and/or annual intervals), followed by intranasal administration of a booster of purified VLP (*i.e.*, as described in part D of this example) in one or multiple doses (e.g., dosing at 1, 2, 3, and/or 4 weeks; dosing at 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and/or 12 months; and/or annual intervals); or (3) intranasal administration of purified VLP in one or multiple doses (e.g., dosing at 1, 2, 3, and/or 4 weeks; dosing at 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and/or 12 months; and/or annual intervals).
- [0404] For intramuscular routes, the following administrations are performed: (1) msDNA-VLP or msDNA-control nanoparticles are administered by injection in one or multiple doses (e.g., dosing at 1, 2, 3, and/or 4 weeks; dosing at 2, 3, 4, 5, 6, 7, 8, 9, 10,

11, and/or 12 months; and/or annual intervals); (2) msDNA-VLP or msDNA-control nanoparticles are administered by injection in one or multiple doses (e.g., dosing at 1, 2, 3, and/or 4 weeks; dosing at 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and/or 12 months; and/or annual intervals), followed by injection of a booster of purified VLP in one or multiple doses (e.g., dosing at 1, 2, 3, and/or 4 weeks; dosing at 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and/or 12 months; and/or annual intervals); or (3) injection of a booster of purified VLP in one or multiple doses (e.g., dosing at 1, 2, 3, and/or 4 weeks; dosing at 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and/or 12 months; and/or annual intervals).

Example 5

Affinity purification of VLPs

- [0405] A 64-residue ACE2 receptor peptide ("ACE2-64") was identified as a sufficient interaction interface for binding coronavirus S protein following analysis of four co-crystal structures of S protein and ACE2 receptor as well as one co-crystal structure of lipoprotein E and ACE2 receptor. The amino acid sequence of ACE2-64 is:
STIEEQAKTFLDKFNHEAEDLFYQSSLASWNYNTNITEENVQNMNNAGDKWSAF
LKEQSTLAQMY (SEQ ID NO:70).
- [0406] The peptide is encoded on an expression plasmid encoding a biotin acceptor peptide (BAP) tag (e.g., GLNDIFEAQKIEWHE (SEQ ID NO:71)) at the C-terminus or N-terminus of ACE2-64 (i.e., SEQ ID NO:72, encoded by SEQ ID NO:73, or SEQ ID NO:74, encoded by SEQ ID NO:75, respectively). The expression plasmid is transformed into a *BirA* positive *E. coli* strain, which results in one-step *in vivo* biotinylation of ACE2-64. The cells are lysed, and the biotinylated ACE2-64 peptides are purified by a commercially available kit and mixed with streptavidin-coated magnetic microbeads.
- [0407] A commercial monoclonal antibody against the COVID-19 S protein ("S-Ab") is biotinylated *in vitro* and mixed with streptavidin-coated magnetic microbeads.
- [0408] Beads with immobilized ACE2-64 or immobilized S-Ab are washed and equilibrated in an inert Tris buffer (e.g., 20 mM Tris pH 8.0, 150 mM NaCl).
- [0409] Recombinant cells expressing VLPs from msDNA-VLPs, such as the eukaryotic cells of Example 2(D), are lysed.
- [0410] Beads with immobilized ACE2-64 or immobilized S-Ab and the cell lysate containing VLPs are added to a microfluidic device and mixed. VLPs captured by the

ACE2-64 or S-Ab coated beads are separated from the cell lysate. The beads are then washed three times with a buffer of moderate salinity (*e.g.*, 20 mM Tris pH 8.0, 300 mM NaCl). The VLPs are then purified in a buffer of high salinity (*e.g.*, 20 mM Tris pH 8.0, 1.5 M NaCl), which results in the dissociation of VLPs from the beads. The purified VLPs are collected. Quality control assays, such as agarose gel electrophoresis to detect RNA and episomal DNA, qPCR to assess gDNA levels, and electron microscopy, are performed to confirm the identity and purity of the VLPs.

Example 6

Production of targeting ligands for nanoparticle formulations

- [0411] A peptide library is derived from the conserved regions of coronavirus S protein and produced by peptide synthesis. Exemplary peptides are SEQ ID NOs:76-99.
- [0412] Recombinant ACE2 protein is purchased from a commercial source.
- [0413] The following portion of the COVID-19 S protein is provided as a control for binding to ACE2, with the bolded and underlined residues being directly involved in ACE2 binding:
- RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASF
STFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTG**K**IADYNYKLPPDDF
TGCVIAWNSNNLDSKVGGN**Y**NYL**Y**R**L**FRKSNLKPFERDISTEIYQAGSTPCNGV**E**
G**F**N**C**Y**F**PL**Q**SYGF**Q**P**T**NGVGY**Q** (SEQ ID NO:100).
- [0414] An *in vitro* fluorescence polarization (FP) assay or similar technique is performed according to standard procedures to determine the affinity of each peptide to the recombinant ACE2 protein.
- [0415] Ligands (*i.e.*, peptides) with the strongest affinities to ACE2 receptor are selected and attached to nanoparticles (*e.g.*, LNPs).
- [0416] The ability of single ligand and dual-ligand nanoparticles to target ACE2 receptor is determined. For example, the targeting ability of nanoparticles containing the ligand with the highest affinity to ACE2 receptor is compared to nanoparticles containing two different ligands having the highest affinities to ACE2 receptor.
- [0417] Multiple ligand targeting is also tested using nanoparticles with one ligand that targets ACE2 receptor (*e.g.*, to facilitate ACE2 receptor-mediated endocytosis) and a

second ligand that is a nuclear localization signal (NLS) (e.g., to facilitate proper intracellular delivery via nuclear targeting).

SEQUENCES

SEQ ID NO:1 membrane protein, amino acid sequence

MADSNGTITVEELKKLLEQWNLVIGFLFLTWICLLQFAYANRNRFLYIIKLI FLWLLWPVTLACF
VLAAYRINWITGGIAIAMACLVGLMWLSYFIASFRLFARTRSMWSFNPETNILLNVPLHGTILT
RPLLESELVIGAVILRGHLRIAGHHLGRCDIKDLPKEITVATSRTL SY YKLGASQ RVAGD
SGFAAYSRYRIGNYKLNTDHSSSSDNIALLVQ

SEQ ID NO:2 membrane protein, nucleic acid sequence

atggcagattccaacggtactattaccggtgaagagcttaaaaagctccttgaacaatggaacct
agtaatagggtttcctattccttacatggattgtctctacaatttgcctatgccaacaggaa
taggtttttgtatataattaagttaattttcctctggctgttatggccagtaactttagcttggt
ttgtgcttgctgctggttacagaataaattggatcaccggtggaattgctatcgcaatggcttggt
cttgtaggcttgatgtggctcagctacttcattgcttcttccagactggttgcgcgctacgcgttc
catgtggtcattcaatccagaaactaacattcttctcaacgtgccactccatggcactattctga
ccagaccgcttctagaaagtgaactcgtaatcggagctgtgatccttcgtggacatcttc
gtattgctggacaccatctaggacgctgtgacatcaaggacctgcctaaagaaatcactggttgct
acatcacgaacgctttcttattacaaattgggagcttcgcagcgtgtagcaggtgactcaggttt
tgctgcatacagtcgctacaggattggcaactataaattaaacacagaccattccagtagcagtg
acaatattgctttgcttgctacagtaa

SEQ ID NO:3 envelope protein, amino acid sequence

MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKNL
NSSRVPDLLV

SEQ ID NO:4 envelope protein, nucleic acid sequence

atgtactcattcggttcggaagagacaggtacgttaatagttaatagcgtacttctttttcttgct
tttcgtggatattcttgctagttacactagccatccttactgcgcttcgattgtgtgctgactgct
gcaatattgttaacgtgagtccttgtaaaaccttctttttacggtttactctcggtgttaaaaatctg
aattcttctagagttcctgatcttctggtctaa

SEQ ID NO:5 spike protein, amino acid sequence

MFVFLVLLPLVSSQCVNLTTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWF
HAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLD SKTQSLI VNNATNVVIKV
CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFK
NIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTL LALHRSYLT PGDSSSGWTA
GAAAYYVGYLQPRTFLLKY NENG TITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTES
IVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSP TKLNDL
CFTNVYADSFVIRGDEV RQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRL
FRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHA
PATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEIL

DITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGNSNVFQTRAGCL
IGAHEVNNSEYCDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNNSIAIP
TNFTISVTTEILPVSMTKTSVDCTMYICGDSTECNLLLQYGSFCTQLNRALTGIAVEQDKNTQE
VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIA
ARDLICAQKFNGLTVLPLLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIG
VTQNVLYENQKLIANQFNSAIGKIQDSLSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAISS
VLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRV
DFCGKGYHLSMFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHGKAHFPREGVVFVSNGTHWFVT
QRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVLGDIS
GINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLC
CMTSCCSCCLKGCCSCGSCCKFDEDDSEPVLKGVLHYT

SEQ ID NO:6 spike protein, nucleic acid sequence

atgtttgtttttcttgttttattgccactagtcctctagtcagtggttaattcttacaaccagaac
tcaattacccccctgcatacactaattctttcacacgtggtgtttattaccctgacaaagtttt
cagatcctcagttttacattcaactcaggacttggtcttacctttcttttccaatgttacttggt
tccatgctatacatgtctctggaaccaatggtactaagaggtttgataaccctgtcctaccattt
aatgatggtgtttattttgcttccactgagaagtctaacataataagaggctggatttttggtac
tactttagattcgaagaccagtcctacttattgttaataacgctactaatgttggtattaaag
tctgtgaatttcaattttgtaatgatccatttttggtgtttattaccacaaaaacaacaaagt
tggatggaaagtgaagttcagagtttattctagtgcaataattgcacttttgaatatgtctctca
gccttttcttatggaccttgaaggaaaacagggtaatttcaaaaatcttagggaatttggtgtta
agaatatgtatgggtattttaaaatatattctaaagcacacgcctattaatttagtgctgatctc
cctcaggggtttttcggcttttagaaccattggtagatttgccaataggtattaacatcactaggtt
tcaaactttacttgctttacatagaagttatttgactcctggtgattcttcttcaggttgacag
ctggtgctgcagcttattatgtgggttatcttcaacctaggacttttctattaaaatataatgaa
aatggaaccattacagatgctgtagactgtgcacttgaccctctcagaaacaaagtgtacgtt
gaaatccttactgtagaaaaaggaatctatcaaacttctaacttttagagtccaaccaacagaat
ctattgttagattttcctaataattacaaacttggtgcccttttggtgaagtttttaacgccaccaga
tttgcatctgtttatgcttggaacaggaagagaatcagcaactgtgttgctgattattctgtcct
atataattccgcatc

SEQ ID NO:7 internal fusion peptide, amino acid sequence

SFIEDLLFNKVTLADAGF

SEQ ID NO:8 internal fusion peptide, nucleic acid sequence

tcattttattgaagatctacttttcaacaaagtgcacttgcatgctggcttc

SEQ ID NO:9 receptor-binding domain, amino acid sequence

PNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLVNSASFSTFKCYGVSPTKLNDLCFTN
VYADSFVIRGDEVQRQIAPGQTGKIADYNYKLDDFTGCVIAWNSNNLDSKVGNGYNYLYRLFRKS
NLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHAPATV
CGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLE

SEQ ID NO:10 receptor-binding domain, nucleic acid sequence

cctaataattacaaaacttgtgcccttttgggtgaagtttttaacgccaccagatttgcattctgttta
tgcttggaacaggaagagaatcagcaactgtgttgctgattattctgtcctatataattccgcat
cattttccacttttaagtgttatggagtgtctcctactaaattaaatgatctctgctttactaat
gtctatgcagattcatttgaattagaggtgatgaagtcagacaaatcgctccagggcaaactgg
aaagattgctgattataattataaaattaccagatgattttacaggctgcgttatagcttggaatt
ctaacaatcttgattctaagggttggtggttaattataattacctgtatagattgtttaggaagtct
aatctcaaacccttttgagagagatatttcaactgaaatctatcaggccggttagcacaccttgtaa
tggtgttggaaggttttaattgttacttttctttacaatcatatgggtttccaaccactaatggtg
ttggttaccaaccatacagagtagtagtacttttcttttgaacttctacatgcaccagcaactggt
tgtggacctaaaaagtctactaatttggttaaaaaacaaatgtgtcaatttcaacttcaatggttt
aacaggcacaggtgttcttactgagcttaacaaaaagtttctgcctttccaacaatttggcgagag
acattgctgacactactgatgctgtccgtgatccacagacacttgag

SEQ ID NO:11 immunogenic sequence, amino acid sequence

PNITNLCPFGEVFNATRFASVYAWNKRKISNCVADYSVLVNSASFSTFKCYGVSPTKLNDLCFTN
VYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPPDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKS
NLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYPYRVVVLSEFLLHAP

SEQ ID NO:12 conserved amino acid sequence

SFIEDL

SEQ ID NO:13 conserved amino acid sequence

GVYYP

SEQ ID NO:14 conserved amino acid sequence

FLPF

SEQ ID NO:15 conserved amino acid sequence

VLPF

SEQ ID NO:16 conserved amino acid sequence

SLLI

SEQ ID NO:17 conserved amino acid sequence

LPIGI

SEQ ID NO:18 conserved amino acid sequence

AAYYV

SEQ ID NO:19 conserved amino acid sequence

TFLI

SEQ ID NO:20 conserved amino acid sequence

AVDC

SEQ ID NO:21 conserved amino acid sequence

IVRFP

SEQ ID NO:22 conserved amino acid sequence

ISNC

SEQ ID NO:23 conserved amino acid sequence

LCFT

SEQ ID NO:24 conserved amino acid sequence

YNYKL

SEQ ID NO:25 conserved amino acid sequence

IAWN

SEQ ID NO:26 conserved amino acid sequence

VVVLSP

SEQ ID NO:27 conserved amino acid sequence

CVNF

SEQ ID NO:28 conserved amino acid sequence

GLTG

SEQ ID NO:29 conserved amino acid sequence

VAVLY

SEQ ID NO:30 conserved amino acid sequence

GCLT

SEQ ID NO:31 conserved amino acid sequence

GICA

SEQ ID NO:32 conserved amino acid sequence

FTIS

SEQ ID NO:33 conserved amino acid sequence

SVDC

SEQ ID NO:34 conserved amino acid sequence

YGSFC

SEQ ID NO:35 conserved amino acid sequence

FNFS

SEQ ID NO:36 conserved amino acid sequence

RDLICAQ

SEQ ID NO:37 conserved amino acid sequence

VLPPLL

SEQ ID NO:38 conserved amino acid sequence

IPFA

SEQ ID NO:39 conserved amino acid sequence

YRFN

SEQ ID NO:40 conserved amino acid sequence

KLQDVVN

SEQ ID NO:41 conserved amino acid sequence

GAISS

SEQ ID NO:42 conserved amino acid sequence

EVQIDRLI

SEQ ID NO:43 conserved amino acid sequence

YVTQQL

SEQ ID NO:44 conserved amino acid sequence

HLMSF

SEQ ID NO:45 conserved amino acid sequence

GVVHLF

SEQ ID NO:46 conserved amino acid sequence

WFVT

SEQ ID NO:47 conserved amino acid sequence

INAS

SEQ ID NO:48 conserved amino acid sequence

LLQF

SEQ ID NO:49 conserved amino acid sequence

LWLLWP

SEQ ID NO:50 conserved amino acid sequence

LMWL

SEQ ID NO:51 conserved amino acid sequence

SFRLF

SEQ ID NO:52 conserved amino acid sequence

FNPETN

SEQ ID NO:53 conserved amino acid sequence

ITVA

SEQ ID NO:54 conserved amino acid sequence

LRLC

SEQ ID NO:55 recombinant spike protein, amino acid sequence

PNITNLCPFGEVFNATRFASVYAWNKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTN
VYADSFVIRGDEVQRQIAPGQTGKIADYNYKL PDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKS
NLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYPYRVVVL SFELLHAPGGG
GGGSFIEDLLFNKVT LADAGFGGGGGWPWYIWLGFIAGLIAIVMTIML

SEQ ID NO:56 recombinant spike protein, nucleic acid sequence

ccaaacattaccaacctgtgcccttcggcgaggtgttcaacgccacacgggtcgccagcgtgta
cgcttggaacagaaagcggatcagcaactgcgtggccgactacagtgtcctgtataactccgcca
gcttttctacattcaagtgtctacggcgtctccctaccaagctgaacgacctgtgcttcaccaat

gtgtacgccgattcttttcgtgatcagaggcgacgaggtgcggcagatcgccccctggccagaccgg
aaagatcgctgattacaactacaagctgcctgatgacttcaccggctgcgtgatcgctggaact
ccaacaacctggacagcaaggtggggggcaactacaactacctgtacagactgttcagaaagagc
aatctgaagccttttcgagagagatatcagcacagagatctaccaggccggcagcacccttgttaa
tggcggttgagggttcaattgctactttccactgcagagctatggctttcagcctacaaacggcg
tgggctaccaaccttacagagtgggtgggtgctgtctttcgagctgctgcacgccccctggcggagga
ggaggcggatctttcatcgaggacctgctgttcaacaaggtgacctggccgacgcccgttttgg
cgggtggcggcgggcggctggccttgggtacatctggtgggcttcatcgccggactgatcgccatcg
tgatggtcaccatcatgctgtga

SEQ ID NO:57 single open reading frame for coronavirus VLP, amino acid sequence

MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKNL
NSSRVPDLLVATNFSLLKQAGDVEENPGPMADSNGTITVEELKKLLEQWNLVIGFLFTWICLLQ
FAYANRNRFlyIIKLIFLWLLWPVTLACFVLAAYRINWITGGIAIAMACLVGLMWLSYFIASFR
LFARTRSMWSFNPETNILLNVPLHGTILTRPILLESELVIGAVILRGHLRIAGHHLGRCDIKDLPK
EITVATSRTLsYYKLGSQVRVAGDSGFAAYSRYRIGNYKLNTHSSSSDNIALLVQATNFSLLKQ
AGDVEENPGPPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSP
TKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGNY
NYLYRLFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLS
FELLHAPGGGGGGSFIEDLLFNKVTLADAGFGGGGGGWPWYIWLGFIAGLIAIVMTIML

SEQ ID NO:58 single open reading frame for coronavirus VLP, nucleic acid sequence

atgtactcttttcgtgtctgaggaaaccggcaccctgatcgatgaacagcgtgctgctgtttctggc
cttcgtggttttctctgctgggtcaccctcgccatcctgaccgccccctgcggctgtgcgctactgct
gcaacatcgtgaacgtgtctctgggtcaaacctagcttctacgtgtatagccgggtgaagaacctg
aattctagcagggtgcccagacctgctgggtggccaccaacttcagcctgctgaaacaggctggcga
tgtggaagagaacctggacctatggccgatagcaacggcaccattacagtggaggaactcaaaa
agctgctggaacagtggaatcttgtgatcggttctctgttctgacctggatctgctgctgcag
ttcgctacgccaaccgcaacagattcctgtacatcatcaaactgatcttctctgtggctgctgtg
gccccgtgacctggcttgtttctgtgctggctgctgtttatagaatcaactggatcacaggcggca
tcgcaatcgccatggcctgtctgggtgggctgatgtggctgagctacttcatcgccagctttaga
ctgttcgctagaacaagaagcatgtggtcctttaaccccgagacaaacatcctcctgaatgtgcc
actgcatggcaccatcctgacaagacctctgctggaaagcgagctggtcatcggcgcccgtgatcc
tgccggggccacctgagaatcgctggccaccacctgggcagatgtgacatcaaggacctgcccag
gaaatcactgtggccacaagcagaacctcagctactacaagctgggagcctctcagagagtggc
cggcgacagcggcttcgcccgcctacagccgggtaccggattggcaattacaaactgaacaccgacc
acagctccagcagcgacaacatcgctctgctagtgcaggccaccaatttcagcctgctgaagcaa
gctggagatgtggaagaaaaccccgccctccaaacattaccaacctgtgcccccttcggcgaggt
gttcaacgccacacgggttcgccagcgtgtacgcctggaacagaaagcggatcagcaactgcgtgg
ccgactacagtgtcctgtataactccgccagcttttctacattcaagtgtacggcgtctccct
accaagctgaacgacctgtgcttcaccaatgtgtacgccgattcttttcgtgatcagaggcgacga
ggtgcggcagatcgccccctggccagaccggaaagatcgctgattacaactacaagctgctgatg
acttcaccggctgcgtgatcgctggaactccaacaacctggacagcaaggtggggggcaactac
aactacctgtacagactgttcagaaagagcaatctgaagcctttcgagagagatatcagcacaga
gatctaccaggccggcagcacccttgtaatggcggttgagggttcaattgctactttccactgc
agagctatggctttcagcctacaaacggcgtgggctaccaaccttacagagtgggtgggtgctgtct
ttcgagctgctgcacgccccctggcggaggaggaggcgatctttcatcgaggacctgctgttcaa

caaggtgaccctggccgacgccggttttggcgggtggcgggcgggctggccttggtacatctggc
tgggcttcacgcgcggactgatcgccatcgtgatggtcaccatcatgctgtga

SEQ ID NO:59 expression cassette for VLP, nucleic acid sequence

cgttacataacttacggtaaatggcccgctggctgaccgccccaacgacccccgccattgacgt
caataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggag
tatttacggtaaaactgccacttggcagtagcatcaagtgtatcatatgccaagtacgccccctat
tgacgtcaatgacggtaaatggcccgctggcattatgccagtagcatgaccttatgggactttc
ctacttggcagtagcatctacgtattagtcacgtattaccatgggtgatgcggttttggcagtag
atcaatgggcgtggatagcgggtttgactcacggggatttccaagtctccacccattgacgtcaa
tgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccat
tgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatataagcagagctgggttagtgaa
cgtcagatccgctagcgcaccatgtactcttctgtgtctgaggaaaccggcacccctgatcgtga
acagcgtgctgctgtttctggccttctgtggttttctgtgtggtcaccctcgccatcctgaccgce
ctgcggtgtgtgcgctactgctgcaacatcgtgaacgtgtctctggtcaaacctagcttctacgt
gtatagccgggtgaagaacctgaattctagcaggggtgccgacctgctgggtggccaccaacttca
gctgctgaaacaggctggcgatgtggaagagaacctggacctatggccgatagcaacggcacc
attacagtgagggaactcaaaaagctgctggaacagtggaatcttgatgctggcttctgttct
gacctggatctgctgctgcagttcgcctacgccaaccgcaacagattctgtacatcatcaaac
tgatcttctgtggtgctgtgtggcccgtagacctgggttctgtgtgctggctgctgtttataga
atcaactggatcacaggcggcatcgcaatcgccatggcctgtctgggtgggctgatgtggctgag
ctacttcacgcagctttagactgttctgctagaacaagaagcatgtggtcctttaaccccgaga
caaacctcctcctgaatgtgccactgcatggcaccatcctgacaagacctgctggaaagcgag
ctggatcatcggcgctgatcctgcggggacacctgagaatcgctggccaccacctgggcagatg
tgacatcaaggacactgcccaaggaaatcactgtggccacaagcagaacctcagctacataagc
tgggagcctctcagagagtgcccggcgacagggctcgcgcctacagccggtaccggtattggc
aattacaaaactgaacaccgaccagctccagcagcgacaacatcgctctgctagtgcaggccac
caatttcagcctgctgaagcaagctggagatgtggaagaaaaccccgccctccaaacattacca
acctgtgccccttcggcgaggtgttcaacgccacacgggttcgccagcgtgtacgcctggaacaga
aagcggatcagcaactgcgtggccgactacagtgctgtataactccgccagcttttctacatt
caagtgtacggcgtctccctaccaagctgaacgacctgtgcttcaccaatgtgtacgccgatt
cttctgtgatcagaggcgacgaggtgcggcagatcgccctggccagaccggaaagatcgctgat
tacaactacaagctgcctgatgacttcaccggctgcgtgatcgccctggaactccaacaacctgga
cagcaagggtggggggcaactacaactacctgtacagactgttcagaaagagcaatctgaagcctt
tcgagagagatatcagcacagagatctaccaggccggcagcacccttgtaatggcgttgagggc
ttcaattgctactttccactgcagagctatggctttcagcctacaaacggcgtgggctaccaacc
ttacagagtggtgggtgctgtcttctgagctgctgcacgcccctggcgaggaggaggcggtatctt
tcacgagagacctgctgttcaacaaggtagacctggccgacgcccgttttggcggtggcgggc
ggctggccttggtacatctggctgggcttcacgcggactgatcgccatcgtgatggtcaccat
catgctgtgaacggccggctgatcataatcagccataaccacattttagagaggttttacttgcttt
aaaaaacctccacacctccccctgaacctgaaacataaaatgaatgcaattgttggtgttaact
tgtttattgcagcttataatgggttacaataaagcaatagcatcacaatttcacaaataaagca
tttttttactgcattctagttgtggtttgtccaaactcatcaatgtatctta

SEQ ID NO:60 expression cassette for VLP, nucleic acid sequence

cgttacataacttacggtaaatggcccgctggctgaccgccccaacgacccccgccattgacgt
caataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggag
tatttacggtaaaactgccacttggcagtagcatcaagtgtatcatatgccaagtacgccccctat

tgacgtcaatgacggtaaatggcccgctggcattatgccagtacatgaccttatgggactttc
ctacttggcagtacatctacgtatttagtcatcgctattaccatgggtgatgcggttttggcagtac
atcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaa
tgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccat
tgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatataagcagagctggtttagtgaac
cgtcagatccgctagcgccaccatgtactctttcgtgtctgaggaaaccggcaccctgatcgta
acagegtgctgctgtttctggccttcgtgggttttctgctggtcaccctcgccatcctgaccgcc
ctgcggtgtgctgctactgctgcaacatcgtaacgtgtctctgggtcaaacctagcttctacgt
gtatagccgggtgaagaacctgaattctagcaggggtgcccgacctgctgggtggccaccaacttca
gcctgctgaaacaggctggcgatgtggaagagaacctggacctatggccgatagcaacggcacc
attacagtggaggaactcaaaaagctgctggaacagtggaatcttgtgatcggttctctgttct
gacctggatctgctgctgcagttcgctacgccaaccgcaacagattctgtacatcatcaaac
tgatcttctgtggctgctgtggcccgtagacctggcttgtttcgtgctggctgctgtttataga
atcaactggatcacaggcgcatcgcaatcgccatggcctgtctgggtgggctgatgtggctgag
ctacttcatcgccagcttttagactgttcgctagaacaagaagcatgtgggtcctttaaccccgaga
caaacatcctcctgaatgtgccactgcatggcaccatcctgacaagacctctgctggaaagcgag
ctgggtcatcggcgcgtgatcctgcggggccacctgagaatcgctggccaccacctgggcagatg
tgacatcaaggacctgcccaggaaatcactgtggccacaagcagaacctcagctactacaagc
tgggagcctctcagagagtggccggcgacagcggttcgccgctacagccggtaccggattggc
aattacaaactgaacaccgaccacagctccagcagcgacaacatcgctctgctagtgcaggccac
caatttcagcctgctgaagcaagctggagatgtggaagaaaaccccgccctccaaacattacca
acctgtgccccttcggcgaggtgttcaacgccacacgggttcgccagcgtgtacgcctggaacaga
aagcggtcagcaactgctggccgactacagtgctcctgtataactccgccagcttttctacatt
caagtgtacggcgtctccctaccagctgaacgacctgtgcttcaccaatgtgtacgccgatt
ctttcgtgatcagaggcgacgaggtgcggcagatcgccctggccagaccggaaagatcgctgat
tacaactacaagctgctgatgacttcacgggtgctgctgatcgctggaactccaacaacctgga
cagcaaggtggggggcaactacaactacctgtacagactgttcagaaagagcaatctgaagcctt
tcgagagagatatcagcacagagatctaccaggccggcagcacccttgtaatggcgttgagggc
ttcaattgctactttccactgcagagctatggcttcagcctacaaacggcgtgggtaccaacc
ttacagagtgggtggtgctgtctttcgagctgctgcaacgcccctggcgaggaggaggcggtcctt
tcatcgaggacctgctgttcaacaaggtgacctggccgacgcccgttttggcggtggcgggc
ggctggccttggtacatctggctgggttcctcgccggactgatcgccatcgctgatggtcaccat
catgctgtgactgtgccttctagttgccagccatctgttgtttgcccctcccccgcttctctt
gacctggaaggtgccactcccactgtcctttcctaataaaaatgaggaaattgcatcgcatgtc
tgagtaggtgtcattctattctggggggtgggggtggggcaggacagcaagggggaggattgggaa
gacaatagcaggcatgctggggatgcggtgggctctatgg

SEQ ID NO:61 expression cassette for VLP, nucleic acid sequence

cgttacataacttacggtaaatggcccgctggctgaccgcccacgacccccgcccattgacgt
caataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggag
tatttacggtaaaactgccacttggcagtacatcaagtgtatcatatgccaagtacgccccctat
tgacgtcaatgacggtaaatggcccgctggcattatgccagtacatgaccttatgggactttc
ctacttggcagtacatctacgtatttagtcatcgctattaccatgggtgatgcggttttggcagtac
atcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaa
tgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccat
tgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatataagcagagctggtttagtgaac
cgtcagatccgctagcgccaccatgtactctttcgtgtctgaggaaaccggcaccctgatcgta
acagegtgctgctgtttctggccttcgtgggttttctgctggtcaccctcgccatcctgaccgcc

cgttacataacttacggtaaatggcccgctggctgaccgcccacgacccccgccattgacgt
caataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggag
tatttacggtaaaactgcccacttggcagtacatcaagtgtatcatatgccaaagtacgcccctat
tgacgtcaatgacggtaaatggcccgctggcattatgccagtacatgaccttatgggactttc
ctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtac

atcaatgggcgtggatagcgggtttgactcacggggatttccaagtctccaccccattgacgtcaa
tgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtatacaactccgccccat
tgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatataagcagagctggtttagtgaac
cgtcagatccgctagcgccaccatgtactctttcgtgtctgaggaaaccgggcaccctgatcgtga
acagcgtgctgctgtttctggccttcgtgggttttccctgctggtcacccctcgccatcctgaccgcc
ctgcggtgtgctgctactgctgcaacatcgtgaacgtgtctctgggtcaaacctagcttctacgt
gtatagccgggtgaagaacctgaattctagcaggggtgcccgacctgctgggtggccaccaacttca
gcctgctgaaacaggctggcgatgtggaagagaacctggacctgccgatagcaacggcaccatt
acagtggaggaaactcaaaaagctgctggaacagtggaatcttgtgatcggttccctgttccctgac
ctggatctgctgctgctgagttcgctacgccaaccgcaacagattcctgtacatcatcaaacctga
tcttccctgtggctgctgtggcccgtagacctggcttgtttcgtgctggctgctgtttatagaatc
aactggatcacaggcgccatcgcaatcgccatggcctgtctgggtgggcctgatgtggctgagcta
cttcatcgccagcttttagactgttcgctagaacaagaagcatgtgggtcctttaaccccgagacaa
acatcctcctgaatgtgccactgcatggcaccatcctgacaagacctcctgctggaaagcgagctg
gtcatcgcgccgctgatcctgcggggccacctgagaatcgctggccaccacctgggcagatgtga
catcaaggacctgccaaggaaatcactgtggccacaagcagaacctcagctactacaagctgg
gagcctctcagagagtggccggcgacagcggttcgccgcctacagccgggtaccggatttggaat
tacaactgaacaccgaccacagctccagcagcgacaacatcgctctgctagtgcaggagggcgag
gggaagtcttctaactgcggggacgtggaggaaaatcccggcccaagaccaagctggctagcc
tcgagtctagagggcccggtttaaacccgctgatcagcctcgaggtaccggatccgcggcccgcat
atctctagactgtgccttctagttgccagccatctgttgtttgccccctcccccgctgccttccctg
acctggaaggtgccactcccactgtcctttcctaataaaaatgaggaaattgcatcgcatgtct
gagtaggtgtcattctattctggggggtgggtggggcaggacagcaagggggaggattgggaag
acaatagcaggcatgctggggatgcggtgggtctatgg

SEQ ID NO:63 expression vector with expression cassette for VLP, nucleic acid sequence

cccgagggtaccgagctcttacgcgtgctagaattaaagtaaccaatcagcacacaattgcc
ttatacgcgctataatggactattgtgtgctgataaacctatttcagcatactacgcgctagt
atgctgaaataggtagctagaagttcctatactttctagagaataggaaactcataacttcgtat
aatgtatgctatacgaagtattgggttacttttaatttggttgctgactaattgagatgcatgctt
tgcatacttctgcctgctggggagcctggggactttccacacctgggttgctgactaattgagatg
catgctttgcatacttctgcctgctggggagcctggggactttccacacctgattctgtggat
aacggtattaccgccatgcattagttattaatagtaatcaattacgggggtcattagttcatagcc
catatatggagtccgcgttacataaacttacggtaaatggccgcctgggtgaccgccaacgac
ccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattg
acgtcaatgggtggagtattttacggtaaaactgcccacttggcagtacatcaagtgtatcatatgc
caagtacgccccctattgacgtcaatgacggtaaatggccgcctggcattatgccagtacatg
accttatgggacttttccacttggcagtacatctacgtattagtcacgctattaccatgggtgat
gcggttttggcagtacatcaatgggctggatagcgggtttgactcacggggatttccaagtctcc
accccatgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgt
aacaactccgccccattgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatataagcag
agctgggttagtgaaccgtcagatccgctagcgccaccatgtactctttcgtgtctgaggaaacc
ggcaccctgatcgtgaacagcgtgctgctgtttctggccttcgtgggttttccctgctggtcaccc
cgccatcctgaccgccccgctgctgctgctactgctgcaacatcgtgaacgtgtctctgggtca
aacctagcttctacgtgtatagccgggtgaagaacctgaattctagcaggggtgcccgacctgctg
gtggccaccaacttcagcctgctgaaacaggctggcgatgtggaagagaacctggacctatggc
cgatagcaacggcaccattacagtggaggaaactcaaaaagctgctggaacagtggaaatcttgtga
tcggcttccctgttccctgacctggatctgcctgctgacgttcgcctacgccaaccgcaacagattc

ctgtacatcatcaaactgatcttctgtggctgctgtggcccgtagaccctggcttgtttcgtgct
ggctgctgtttatagaatcaactggatcacaggcgccatcgcaatcgccatggcctgtctgggtg
gcctgatgtggctgagctacttcatcgccagctttagactgttcgctagaacaagaagcatgtgg
tcctttaaccccgagacaaacatcctcctgaatgtgccactgcatggcaccatcctgacaagacc
cctgctggaaagcgagctgggtcatcggcgcctgtagctcctgcgggggccacctgagaatcgctggcc
accacctgggcagatgtgacatcaaggacctgcccaggaaatcactgtggccacaagcagaacc
ctcagctactacaagctgggagcctctcagagagtggccggcgacagcggttcgccgcctacag
ccggtaccggattggcaattacaaactgaacaccgaccacagctccagcagcgacaacatcgctc
tgctagtgcaggccaccaatctcagcctgctgaagcaagctggagatgtggaagaaaaccccggc
cctccaaacattaccaacctgtgcccccttcggcgaggtgttcaacgccacacgggttcgccagcgt
gtacgcctggaacagaaagcggatcagcaactgctggccgactacagtgctcctgtataactccg
ccagcttttctacattcaagtgtacggcgtctccccctaccaagctgaacgacctgtgcttcacc
aatgtgtacgccgattctttcgtgatcagaggcgacgaggtgcggcagatcgccccctggccagac
cggaaagatcgctgattacaactacaagctgcctgatgacttcaccggctgctgtagcctgga
actccaacaacctggacagcaaggtggggggcaactacaactacctgtacagactgttcagaaag
agcaatctgaagcctttcgagagagatatcagcacagagatctaccaggccggcgagcacccttg
taatggcgttgagggttcaattgctactttccactgcagagctatggctttcagcctacaaacg
gcggtgggtaccaaccttacagagtgggtgctgtctttcgagctgctgcacgccccctggcgga
ggaggaggcggtatctttcatcgaggacctgctgttcaacaaggtgaccttggccgacgcgggttt
tggcggtggcgggcggtggccttggtacatctggctgggcttcatcgccggactgatcgcca
tcgtgatggtcaccatcatgctgtgactgtgccttctagtgtccagccatctgttgtttgccct
cccccgctgcttcttgacctggaaggtgccactcccactgtcctttcctaataaaaatgaggaa
attgcatcgattgtctgagtaggtgtcattctattctggggggtgggggtggggcaggacagcaa
gggggaggattgggaagacaatagcaggcatgctggggatgcgggtgggctctatggaagcttacg
cgtggccgctcgagacgcaattcggttgggtgtggaaagtccccaggctccccagcaggcagaag
tatgcaaagcatgcatctcaattagtgcagcaaccaggtgtggaaagtccccaggctccccagcag
gcagaagtatgcaaagcatgcatctcaattagtgcagcaaccaaatataagtaaacccataacttcg
tatagcatacattatacgaagtattgaagttcctattctctagaaagtataggaacttctagtca
cctatttcagcatactacgcgcgtagtatgctgaaatagggttatcagcacacaatagtccatta
tacgcgcgtataatggcaattgtgtgctgattgggttactttaatttggtatcgctcgaccgatgc
ccttgagagccttcaaccagtcagctccttcgggtgggcgcggggcatgactatcgctcgccga
cttatgactgtcttctttatcatgcaactcgtaggacaggtgcccggcagcgctcttcgcttct
cgctcactgactcgctgcgctcggtcggttcggctgcccgcagcggtatcagctcactcaaaggcg
gtaatacgggtatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaaggccagca
aaaggccaggaaccgtaaaaaggccgcttgctggcgtttttccataggctccgccccctgacg
agcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccag
gcgtttccccctggaagctccctcgctgcgtctcctgttccgacctgcccgttaccggatacct
gtccgcctttctcccttcgggaagcgtggcgtttctcatagctcacgctgtaggtatctcagtt
cggtgtaggtcgttcgctccaagctgggtgtgtgcacgaacccccgttcagcccagccgtgc
gccttatccggtaactatcgctcttgagtcacacccggtaagacacgacttatcgccactggcagc
agccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtgg
ggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttacc
ttcggaaaaagagttggtagctcttgatccggcaaaacaaaccacgctggtagcggtgggttttt
tggttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttcta
cggggtctgacgctcagtggaacgaaaactcacgttaagggttttggtcatgagattatcaaaa
aggatcttcacctagatccttttaaattaaaaatgaagtttaaataaatctaaagtatatatga
gtaaacttggtctgacagttaccaatgcttaatacagtgaggcacctatctcagcgatctgtctat
ttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggttacca
tctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaat

aaaccagccagccggaagggccgagcgcagaagtgggtcctgcaactttatccgcctccatccagt
ctattaattgttgcgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttggt
gccattgctacagggcatcggtgtcacgctcgctgtttggtatggcttcattcagctccgggtc
ccaacgatcaaggcgagttacatgatcccccattgtgtgcaaaaaagcgggttagctccttcggtc
ctccgatcgttgtcagaagtaagttggccgcagtggttatcactcatgggttatggcagcactgcat
aattctcttactgtcatgccatccgtaagatgcttttctgtgactgggtgagtactcaaccaagtc
attctgagaatagtgtatgcggcgaccgagttgctcttgcggcggtcaatacgggataataccg
cgccacatagcagaactttaaaagtgtcatcattggaaaaacgttcttcggggcgaaaactctca
aggatcttaccgctgttgagatccagttcgatgtaaccactcgtgcacccaactgatcttcagc
atcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaaatgccgcaaaaaagg
gaataaggcgacacggaaatgttgaatactcatactcttctttttcaatattattgaagcatt
tatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaatatagg
ggttccgcgcacatttccccgaaaagtgccacctgacgcgcctgtagcggcgcatthaagcgcgg
cgggtgtgggtgttacgcgcagcgtgaccgctacacttgcagcgccttagcgcggcgtcctttc
gctttcttcccttcccttctcgcacggttcgcgggtttccccgtcaagctctaaatcggggggt
ccctttagggttccgatttagtgctttacggcacctcgacccccaaaaaacttgattagggtgatg
gttcacgtagtgggcatcgccctgatagacgggttttgcgctttgacggttgagtcacggttc
tttaatagtggactcttgttccaaactggaacaacactcaaccctatctcgggtctattcttttga
tttataagggttttgcgatttcggcctattgggttaaaaaatgagctgatttaacaaaaattta
acgcgaattttaacaaaaatattaacgcttacaatttgccattcgccattcagggtgcgcaactgt
tggaaggcgatcgggtgcgggcctcttcgctattacgccagcccaagctaccatgataagtaag
taatattaaggtagctggaggttttacttgctttaaaaaacctcccacacctccccctgaacctg
aacataaaatgaatgcaattgttgttgaacttggtttattgcagcttataatggttacaata
aagcaatagcatcacaatttcacaaataaagcatttttttactgcattcttagttgtggtttgt
ccaaactcatcaatgtatcttatggtactgttaactgagctaacataa

SEQ ID NO:64 Forward Primer, envelope protein, nucleic acid sequence

actgctgcaacatcgtagaac

SEQ ID NO:65 Reverse Primer, envelope protein, nucleic acid sequence

tgctagaattcaggttcttcacc

SEQ ID NO:66 Forward Primer, membrane protein, nucleic acid sequence

ttcctgtggctgctgtgg

SEQ ID NO:67 Reverse Primer, membrane protein, nucleic acid sequence

atgaccagctcgctttccag

SEQ ID NO:68 Forward Primer, receptor-binding domain, nucleic acid sequence

atcagcacagagatctaccagg

SEQ ID NO:69 Reverse Primer, receptor-binding domain, nucleic acid sequence

agcaccaccactctgtaagg

SEQ ID NO:70 ACE2 receptor peptide, amino acid sequence

STIEEQAKTFLDKFNHEAEDLFYQSSLASWNYNTNITEENVQNMNNAGDKWSAFLKEQSTLAQMY

SEQ ID NO:71 BAP tag, amino acid sequence

GLNDIFEAQKIEWHE

SEQ ID NO:72 ACE2 receptor peptide with C-terminal BAP tag, amino acid sequence

STIEEQAKTFLDKFNHEAEDLFYQSSLASWNYNTNITEENVQNMNNAGDKWSAFLKEQSTLAQMY
GLNDIFEAQKIEWHE

SEQ ID NO:73 ACE2 receptor peptide with C-terminal BAP tag, nucleic acid sequence

tccactattgaagaacaggcacaagactttcttgacaaaattcaaccacgaggccgaagacttggt
ctatcaaagttcccttgcgagttggaattacaatacgaatatcaccgaagaaaacgttcagaata
tgaacaatgcaggcgacaaaatggtccgcctttttgaaagaacaaagtaccctggcccagatgtac
ggtcttaatgacatctttgaagcgcaaaagatcgagtggcacgaa

SEQ ID NO:74 ACE2 receptor peptide with N-terminal BAP tag, amino acid sequence

GLNDIFEAQKIEWHESTIEEQAKTFLDKFNHEAEDLFYQSSLASWNYNTNITEENVQNMNNAGDK
WSAFLKEQSTLAQMY

SEQ ID NO:75 ACE2 receptor peptide with N-terminal BAP tag, nucleic acid sequence

ggtcttaatgacatctttgaagcgcaaaagatcgagtggcacgaatccactattgaagaacaggc
aaagactttcttgacaaaattcaaccacgaggccgaagacttggtctatcaaagttcccttgcca
gttggaattacaatacgaatatcaccgaagaaaacgttcagaatatgaacaatgcaggcgacaaa
tggtccgcctttttgaaagaacaaagtaccctggcccagatgtac

SEQ ID NO:76 ACE2 binding peptide, amino acid sequence

QSYGFQPTN

SEQ ID NO:77 ACE2 binding peptide, amino acid sequence

LQSYGFQPTN

SEQ ID NO:78 ACE2 binding peptide, amino acid sequence

QSYGFQPTNGVGY

SEQ ID NO:79 ACE2 binding peptide, amino acid sequence

QPTNGVGY

SEQ ID NO:80 ACE2 binding peptide, amino acid sequence

FQPTNGVGY

SEQ ID NO:81 ACE2 binding peptide, amino acid sequence

QPTN

SEQ ID NO:82 ACE2 binding peptide, amino acid sequence

FQPTN

SEQ ID NO:83 ACE2 binding peptide, amino acid sequence

FQPTNGV

SEQ ID NO:84 ACE2 binding peptide, amino acid sequence

TNGVGY

SEQ ID NO:85 ACE2 binding peptide, amino acid sequence

FNCYFPLQ

SEQ ID NO:86 ACE2 binding peptide, amino acid sequence

GFNCYFPLQ

SEQ ID NO:87 ACE2 binding peptide, amino acid sequence

EGFN

SEQ ID NO:88 ACE2 binding peptide, amino acid sequence

VEGFNCY

SEQ ID NO:89 ACE2 binding peptide, amino acid sequence

EGFNCYFPLQ

SEQ ID NO:90 ACE2 binding peptide, amino acid sequence

YNYLY

SEQ ID NO:91 ACE2 binding peptide, amino acid sequence

NYNYLYR

SEQ ID NO:92 ACE2 binding peptide, amino acid sequence

SFIEDLLFNKVTLADAGF

SEQ ID NO:93 ACE2 binding peptide, amino acid sequence

SFIEDLLFNKVTLADAGFMKQYGCCKKKK

SEQ ID NO:94 ACE2 binding peptide, amino acid sequence

SFIEDLLF

SEQ ID NO:95 ACE2 binding peptide, amino acid sequence

SFIEDLLFGCGKKKK

SEQ ID NO:96 ACE2 binding peptide, amino acid sequence

SFIEDLLFNKVTLADAGFMKQY

SEQ ID NO:97 ACE2 binding peptide, amino acid sequence

SFIEDAAAGCGKKKK

SEQ ID NO:98 ACE2 binding peptide, amino acid sequence

SFIEDAAA

SEQ ID NO:99 ACE2 binding peptide, amino acid sequence

TRYYYLNINYTTGY

SEQ ID NO:100 ACE2 binding control peptide, amino acid sequence

RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRRKRISNCVADYSVLVNSASFSTFKCYGVS
PTKLNLDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGN
YNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQ

SEQ ID NO:101 immunogenic sequence, nucleic acid sequence

cctaataattacaaacttggtgcccttttggtgaagtttttaacgccaccagatttgcatctgttta
tgcttggaacaggaagagaatcagcaactgtgttgctgattattctgtcctatataattccgcat
cattttccacttttaagtgttatggagtgtctcctactaaattaaatgatctctgctttactaat
gtctatgcagattcatttgtaattagaggtgatgaagtcagacaaatcgctccagggcaaactgg
aaagattgctgattataattataaattaccagatgattttacaggctgcgttatagcttggaatt
ctaacaatcttgattctaagggttggtggaattataattacctgtatagattgttttaggaagtct
aatctcaaacccttttgagagagatatctcaactgaaatctatcaggccggttagcacaccttgtaa
tggtgttggaaggttttaattgttactttcctttacaatcatatgggtttccaaccactaatgggtg
ttgggtaccaaccatacagagtagtagtactttcttttgaaacttctacatgcacca

SEQ ID NO:102 transmembrane domain, amino acid sequence

WPWYIWLGFIAGL

SEQ ID NO:103 transmembrane domain, nucleic acid sequence

tggccatggtacatttggttaggttttatagctggcttga

SEQ ID NO:104 bacterial sequence-free vector, nucleic acid sequence

cgcgcgtagtatgctgaaataggtgactagaagttcctatacttttctagagaataggaacttcat
aacttcgtataatgtatgctatacgaagttatgggttactttaatttggttgctgactaattgag
atgcatgctttgcatacttctgctgctggggagcctggggactttccacacctggttgctgact
aattgagatgcatgctttgcatacttctgctgctggggagcctggggactttccacacccctga
ttctgtggataaccgtattaccgccatgcattagttattaatagtaatcaattacggggtcatta
gttcatagcccatatatggagttccgcgttacataaacttacggtaaatggcccgctggctgacc
gcccacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaataggga
ctttccattgacgtcaatgggtggagtatttacggtaaaactgcccacttggcagtacatcaagt
tatcatatgccaaagtacgccccctattgacgtcaatgacggtaaatggcccgctggcattatgc
ccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcacgctatta
ccatggtgatgcggttttggcagtacatcaatgggcgtggatagcgggttgactcacggggattt
ccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttcc
aaaatgtcgtacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtct
atataagcagagctgggttagtgaaccgtcagatccgctagcgccaccatgtactcttctgtgct
tgaggaaaccggcaccctgatcgtgaacagcgtgctgctgttctggccttcgtgggttttctgc
tggtcaccctcgccatcctgaccgcccctgcggtgtgctgcctactgctgcaacatcgtgaacgtg
tctctgggtcaaacctagcttctacgtgtatagccgggtgaagaacctgaattctagcaggggtgcc
cgacctgctgggtggccaccaacttcagcctgctgaaacaggctggcgatgtggaagagaacctg
gacctatggcgatagcaacggcaccattacagtgagggaactcaaaaagctgctggaacagtgg
aatcttggtgatcggttctctgttctgacctggatctgctgctgcagttcgccacgccaaccg
caacagattcctgtacatcatcaaaactgatcttctctgtggctgctgtggcccgctgacctggctt
gttctgtgctggctgctgtttatagaatcaactggatcacaggcggcatcgcaatcgccatggcc
tgtctgggtggcctgatgtggctgagctacttcatcgccagcttttagactgttcgctagaacaag
aagcatgtggtcctttaaccccgagacaaacatcctcctgaatgtgccactgcatggcaccatcc
tgacaagaccctgctggaaagcgagctggtcatcggcgcctgatcctgcggggcccacctgaga
atcgctggccaccacctgggcagatgtgacatcaaggacctgcccaggaatcactgtggccac
aagcagaacctcagctactacaagctgggagcctctcagagagtggccggcgacagcggttcg
ccgcctacagccggtaccggattggcaattacaaactgaacaccgaccacagctccagcagcgac
aacatcgctctgctagtgcaggccaccaatttcagcctgctgaagcaagctggagatgtggaaga
aaaccccgccctccaaacattaccaacctgtgccccttcggcgaggtgttcaacgccacacgggt
tcgccagcgtgtacgcctggaacagaaagcggatcagcaactgctggccgactacagtgtcctg
tataactccgccagcttttctacattcaagtgtacggcgtctcccctaccaagctgaacgacct
gtgcttcaccaatgtgtacgcccattcttctgtgatcagaggcgacgaggtgcggcagatcgccc
ctggccagaccggaaaagatcgctgattacaactacaagctgcctgatgacttcaccggctgcgtg
atcgctggaactccaacaacctggacagcaaggtggggggcaactacaactacctgtacagact
gttcagaaagagcaatctgaagcctttcgagagagatatcagcacagagatctaccaggccggca
gcacccttgtaatggcggttgagggttcaattgctactttccactgcagagctatggctttcag
cctacaaacggcggtgggtaccacacattacagagtggtggtgctgtcttccgagctgctgcacgc
ccctggcgaggaggaggcggtatcttctacatcgaggacctgctgttcaacaaggtgacctggccg
acgcgggttttggcggtggcgggcggtggtggtacatctggctgggcttcacgcggga
ctgatcgccatcgatgggtcaccatcatgctgtgactgtgccttctagtggcagccatctgtt
gtttgcccctcccccgctgccttcttgacctggaaggtgccactcccactgcttctcctaata
aaatgaggaaattgcattgcattgtctgagtaggtgtcattctattctggggggtgggtggggc
aggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcgggtgggtctatg
gaagcttacgctggccgctcgagacgcaattcggttgggtgtggaaagtccccaggctccccag
caggcagaagtatgcaaagcatctcaattagtcagcaaccaggtgtggaaagtccccaggc
tcccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaaaattaaagtaacc

cataacttcgtatagcatacattatacgaagttatgaagttcctattctctagaaagtataggaa
cttctagtcacctatttcagcatactacgcgcg

* * *

[0418] The disclosure is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the disclosure in addition to those described will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

Other embodiments are within the following claims.