



(51) International Patent Classification:

A61K 39/215 (2006.01)

(21) International Application Number:

PCT/US2021/048750

(22) International Filing Date:

01 September 2021 (01.09.2021)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/075,022 04 September 2020 (04.09.2020) US
63/225,783 26 July 2021 (26.07.2021) US

(71) Applicant: HAWAII BIOTECH, INC. [US/US]; 650 Iwilei Road, Suite 204, Honolulu, Hawaii 96817 (US).

(72) Inventors: CLEMENTS, David E.; c/o HAWAII BIOTECH, INC., 650 Iwilei Road, Suite 204, Honolulu, Hawaii 96817 (US). SENDA, James T.; c/o HAWAII BIOTECH, INC., 650 Iwilei Road, Suite 204, Honolulu, Hawaii 96817 (US). HORTON, Jaime S.; c/o HAWAII BIOTECH, INC., 650 Iwilei Road, Suite 204, Honolulu, Hawaii 96817 (US).

(74) Agent: HAILE, Lisa A. et al.; DLA PIPER LLP US, 4365 Executive Drive, Suite 1100, San Diego, California 92121-2133 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ,

CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

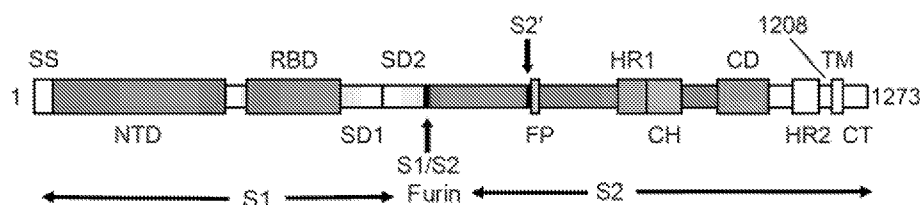
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,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

Published:

— without international search report and to be republished upon receipt of that report (Rule 48.2(g))

(54) Title: RECOMBINANT SARS-COV-2 SPIKE PROTEIN SUBUNITS, EXPRESSION AND USES THEREOF

FIGURE 1



(57) Abstract: The present invention is directed to the expression and secretion recombinant SARS-CoV-2 spike protein subunits. Various subunits have been designed and expressed as secreted products into the culture medium of transformed insect cell lines. The design of subunits is focused on the production of products that provide the ability to induce focused immune responses without inducing immune enhancing responses. The expressed and purified products are suitable as vaccine candidates to protect against disease caused by SARS-CoV-2.

RECOMBINANT SARS-CoV-2 SPIKE PROTEIN SUBUNITS, EXPRESSION AND USES THEREOF

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Patent Application Nos. 63/225,783, filed July 26, 2021 and 63/075,022, filed September 4, 2020, which are herein incorporated by reference in their entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under contract numbered 3R44AI118017-03S1 (NIH). The government has certain rights in the invention.

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

[0003] The invention relates generally to the design of optimized SARS-CoV-2 virus spike glycoprotein genes and expression thereof and more specifically to the development of a COVID-19 vaccine.

BACKGROUND INFORMATION

[0004] In the past 18 years there have been 3 significant Coronavirus outbreaks that have greatly impacted global health. First, there was the Severe Acute Respiratory Syndrome (SARS) in 2002, then the Middle East Respiratory Syndrome (MERS) in 2012, and most recently the 2019 novel Coronavirus (2019 nCoV). The 2019 nCoV is now officially named SARS-CoV-2 and the disease it causes is referred to as COVID-19. On March 11, 2020, as a result of the rapid global spread of the virus, the WHO declared that COVID-19 had reached pandemic levels. As of June 10, 2021, there are now greater than 175,000,000 confirmed cases worldwide, and the United States has over 33,400,000 confirmed cases, the most reported by any single country.

[0005] The emergence of novel coronaviruses necessitates the need to develop effective medical countermeasures to effectively respond to the public health threat posed by these viruses. Vaccines are one of the most effective methods to combat infectious disease threats such as that posed by the SARS-CoV-2.

[0006] Currently, there is no licensed vaccine for SARS-CoV-2; however, three SARS-CoV-2 vaccines have received Emergency Use Authorization (EUA) from the U.S. FDA as of April 2021. These include the mRNA-based Pfizer-BioNTech BNT162b2 and Moderna

mRNA-1273 vaccines, as well as Janssen's Ad26.COVS adenovirus vector-based vaccine. According to the World Health Organization there are currently 276 vaccine candidates in development (WHO 27Apr2021), underscoring the ongoing need for continued vaccine development. Ninety-two of these vaccine candidates are currently in various stages of clinical trials and the other 184 vaccine candidates are in various stages of preclinical development. These approaches employ methods that include inactivated virus, attenuated virus nucleic acid-based, both mRNA and DNA, viral vectored, VLP based, and recombinant subunit proteins. While any one of these may progress to a viable candidate in the coming months to years, there are knowledge gaps that could impact the ultimate realization of developing a safe and effective SARS-CoV-2 vaccine. First, the selection of ideal antigen(s) is not yet defined. While the spike protein (S) is the leading candidate, it is yet to be determined as to which portion of S is best: full length, ectodomain, S1, or receptor binding domain (RBD) only. Second, preclinical studies with SARS and MERS vaccine candidates have raised safety concerns due to enhancement of disease as a direct result of skewed antibody responses. Third, there are no clear correlates of protection for SARS-CoV-2. While they can be inferred from SARS and MERS studies, correlates still need to be established for SARS-CoV-2. Lastly, there are concerns about the duration of immunity elicited by vaccines. It is critical that these various aspects which will have an impact on the outcome of any vaccine candidate be addressed in the vaccine development process.

[0007] In addition to the concerns related to the immune responses elicited by current vaccine candidates, there are also concerns related to each of the platform technologies that are being utilized. These include product stability, product delivery methods, potential for cGMP scale up, and regulatory pathways. For example, mRNA vaccines have unique challenges related to the lipid nanoparticle vehicle employed (stability and reactogenicity/toxicity). While DNA vaccines have ease of manufacturing and scalability in their favor, delivery methods and weak immunogenicity in humans still pose significant challenges. Replication-competent vaccines have advantages of rapid immune responses with a single dose; however, manufacturing processes that rely on viral replication become more pronounced with scale-up and some require more extensive cold chain requirements. For example, the current Merck VSV-based Ebola vaccine requires -60°C storage conditions. It has also been reported that the mRNA vaccines will require -20°C storage conditions. If an inactivated vaccine approach is pursued, there will be safety concerns with the scaling of live SARS-CoV-2 and the need for

BSL3 containment. Alternative vaccine manufacturing platforms that can avoid some, if not all, of these gaps/challenges are of great value. Recombinant subunit SARS-CoV-2 vaccines provide an opportunity to address many of these concerns. Furthermore, it has become clear that having more than one technology platform/manufacturer for emerging infectious disease is important to ensure adequate vaccine supplies.

[0008] The choice of a recombinant protein expression system to employ is dependent on the desired application. The system of choice must meet key criteria such as proper folding and processing, consistency, and productivity (cost effectiveness) of the desired protein product. Insect cell-based expression systems have the potential to meet capacity requirements based on ease of culture, higher tolerance to osmolality and by-product concentrations during large scale culture, and generally higher expression levels.

[0009] Recently, the use of expression systems based on insect cells has become more common. These systems provide most of the characteristics desired of eukaryotic systems but have added benefits such as lower cost of goods. Insect cell systems are either based on infection of host cells with insect virus vectors (e.g., baculovirus) or on the generation of stable cell lines by integration of expression plasmids into the genome of the host cells.

[0010] The baculovirus expression system (BES) has emerged as the primary insect cell culture system utilized for recombinant protein expression. This system is based on the use of vectors derived from the insect viruses known as baculovirus. These vectors are used to generate recombinant viruses that encode the desired protein product. The recombinant viruses are used to infect host insect cells that then express the desired recombinant proteins. While there are advantages to this system in regard to ease of cloning and “time to product”, there are also several disadvantages. The primary challenge in the use of BES is that it is based on the viral infection of the host cells. This results in cellular lysis and cell death 72-96 hrs post infection. As a result, during the late stages of infection the processing machinery of the insect cells is compromised to the extent that the processing of the desired product is also compromised. This limits the time that the cells can produce product and possibly more importantly leads to altered forms of the product being produced. Furthermore, the lysis of cells releases large amounts of cellular debris and enzymes that can impact the quality of the desired product.

[0011] The use of stably transformed insect cells for the expression of recombinant proteins is an alternative to the use of BES. Expression systems based on stably transformed insect cell

lines are non-lytic and provide for steady long-term production of secreted products that require proper folding and post translational modifications. The secretion of the product into the culture medium provides a cleaner starting material for the purification process and allows for the final protein product to be purified with basic methods. This leads to products that are of higher quality.

[0012] The development of a recombinant subunit vaccine for SARS-CoV-2 requires the selection of appropriate gene sequences from the SARS-CoV-2 genome that encode proteins that are the target of neutralizing antibodies (nAbs). Like other members of the coronavirus family, the spike glycoprotein of SARS-CoV-2 is the primary target of nAbs. In addition to selection of an appropriate SARS-CoV-2 gene sequence, efforts to optimize the expression of the selected gene sequences is also desirable to enhance the ability to effectively express the selected sequences such that the resultant products are soluble, stable and conformationally relevant.

[0013] The combination of multiple optimizations directed at different aspects of a protein's gene sequence and or structure in an appropriate manner such that an additive benefit is achieved can further enhance the utility of the optimized protein product. For SARS-CoV-2 S glycoprotein subunit proteins, examples exist of expression of various truncated products as soluble recombinant proteins which indicate various levels of optimization; however, these examples are only reporting moderate expression levels and employ limited approaches to optimization. Given the need for a vast number of vaccine doses to combat the SARS-CoV-2 pandemic, efforts to further the optimal expression of high levels of SARS-CoV-2 S glycoprotein subunits are needed. There are no examples of where multiple components are optimized to further enhance the expression and secretion of soluble SARS-CoV-2 S glycoproteins in insect cells. Therefore, the technical problems to be solved are: (1) identification of translational, posttranslational, or structural components of the spike protein or associated components that when optimized result in improved expression levels and potentially enhance structural quality of the protein such that it is a more potent immunogen; (2) the design of synthetic components where possible to aid in the optimization; and (3) determining the optimal truncations points to increase in the productivity and quality of protein expression. Further improvements in the expression of the SARS-CoV-2 S glycoprotein S protein subunits in insect cells provide for effective immunogens at an improved cost of goods which would bolster the ability to manufacture recombinant proteins suitable for use in

vaccines to combat the threat posed by SARS-CoV-2. The use of such improvements could also be applied to other members of the coronavirus family.

SUMMARY OF THE INVENTION

[0014] The invention provides optimized expression of soluble recombinant SARS-CoV-2 S subunit proteins that result in high levels of expression of a native-like or biologically relevant proteins; and is therefore, an effective immunogen for the production of nAbs. Specifically, the invention is directed to expression of the optimized SARS-CoV-2 S gene sequences when *Drosophila melanogaster* S2 cells are used as the host cell.

[0015] The invention also provides methods for utilizing the products encoded by the optimized SARS-CoV-2 S gene sequences in vaccine formulations for protecting against disease caused by infection with SARS-CoV-2.

[0016] In one embodiment, the invention provides an isolated nucleic acid sequence selected from SEQ ID NO:1, 2, 3, 4, 5 and 6.

[0017] In another embodiment, the invention provides an isolated amino acid sequence encoded by a nucleic acid sequence selected from SEQ ID NO:1, 2, 3, 4, 5 and 6.

[0018] In one aspect, the amino acid sequence includes SEQ ID NO: 9, 10, 11, 12, 13 or 14.

[0019] In an additional embodiment, the invention provides an expression vector including a nucleic acid sequence encoding a SARS-CoV-2 spike (S) protein, wherein the nucleic acid sequence includes SEQ ID NO: 1, 2, 3, 4, 5 or 6.

[0020] In one aspect, the vector is a *Drosophila melanogaster* expression vector. In some aspects, the vector has a nucleic acid sequence including SEQ ID NO:7. In another aspect, the SARS-CoV-2 S protein has an amino acid sequence including SEQ ID NO: 9, 10, 11, 12, 13 or 14.

[0021] In one embodiment, the invention provides a method of producing a protein *in vitro* including an expression vector with an operably-linked nucleic acid sequence encoding a SARS-CoV-2 spike (S) protein, wherein the nucleic acid sequence includes SEQ ID NO: 1, 2, 3, 4, 5 or 6 using *Drosophila melanogaster* cells and culturing the cells under conditions to produce the protein.

[0022] In one aspect, the *Drosophila melanogaster* cells are Schneider 2 (S2) cells.

[0023] In another embodiment, the invention provides a vaccine composition including (a) an effective amount of a SARS-CoV-2 spike (S) protein, wherein S protein is encoded by a

nucleic acid sequence including SEQ ID NO: 1, 2, 3, 4, 5 or 6 and (b) an effective amount of an adjuvant selected from the group consisting of GPI-0100, synthetic lipid A (SLA) in a stable oil-in water emulsion (SE) (SLA-SE), QS21, QS21 combined with SLA to form a liposome formulation (SLA-LSQ), and QS21+CpG.

[0024] In one aspect, the S protein is a S1 subunit protein. In some aspects, the S1 subunit protein is encoded by a nucleic acid sequence of SEQ ID NO: 3. In another aspect, the adjuvant is SLA-SE. In one aspect, the vaccine composition includes the S1 subunit protein with the amino acid sequence of SEQ ID NO: 10 and the adjuvant SLA-SE. In another aspect, the S1 subunit protein is recombinantly produced and expressed in insect host cells. In some aspects, the vaccine composition further includes a pharmaceutically acceptable excipient or carrier.

[0025] In one embodiment, the invention provides a method of preventing SARS-CoV-2 entry into a target cell including contacting a subject with the target cell of SARS-CoV-2 with a therapeutically or prophylactically effective amount of composition including (a) an effective amount of a SARS-CoV-2 spike (S) protein wherein the S protein is encoded by a nucleic acid sequence including SEQ ID NO: 1, 2, 3, 4, 5 or 6, and (b) an effective amount of an adjuvant, thereby preventing SARS-CoV-2 entry into the cell.

[0026] In another embodiment, the invention provides a method of stimulating a protective immune response in a subject including administering to the subject a therapeutically or prophylactically effective amount of composition including (a) an effective amount of a SARS-CoV-2 spike (S) protein, wherein the S protein is encoded by a nucleic acid sequence including SEQ ID NO: 1, 2, 3, 4, 5 or 6, and (b) an effective amount of an adjuvant, thereby stimulating a protective immune response in the subject.

[0027] In one aspect, the immune response is a balanced immune response. In another aspect, a balanced immune response is characterized by an IgG2a:IgG1 ratio that is equal or greater than 1.

[0028] In an additional embodiment, the invention provides a method of inhibiting a SARS-CoV-2 infection in a subject including administering to the subject a therapeutically or prophylactically effective amount of composition including (a) an effective amount of a SARS-CoV-2 spike (S) protein, wherein the S protein is encoded by a nucleic acid sequence including SEQ ID NO: 1, 2, 3, 4, 5 or 6, and (b) an effective amount of an adjuvant, thereby inhibiting SARS-CoV-2 infection.

[0029] In another embodiment the invention provides a method of inhibiting transmission of a SARS-CoV-2 infection by a subject including administering to the subject a therapeutically or prophylactically effective amount of composition including (a) an effective amount of a SARS-CoV-2 spike (S) protein, wherein S protein is encoded by a nucleic acid sequence including SEQ ID NO: 1, 2, 3, 4, 5 or 6, and (b) an effective amount of an adjuvant selected from the group consisting of GPI-0100, synthetic lipid A (SLA) in a stable oil-in water emulsion (SE) (SLA-SE), QS21, QS21 combined with SLA to form a liposome formulation (SLA-LSQ), and QS21+CpG, thereby inhibiting transmission of a SARS-CoV-2 infection.

[0030] In one aspect, the S protein is a S1 subunit protein. In some aspects, the S1 subunit protein is encoded by a nucleic acid sequence of SEQ ID NO: 3. In another aspect, the adjuvant is SLA-SE. In one aspect, the vaccine composition including the S1 subunit protein with the amino acid sequence of SEQ ID NO: 10 and the adjuvant SLA-SE. In another aspect, the vaccine induces the production of nAbs in the subject. In some aspects, the nAbs prevent the binding of a SARS-CoV-2 to a target cell and/or target receptor. In various aspects, the target receptor is an ACE2 receptor. In one aspect, administering includes injecting two doses to the subject at a 3-weeks interval. In another aspect, administering includes injecting intramuscularly. In some aspects, a dose comprises about 0.5-50 µg of purified S protein. In other aspects, administering the vaccine to the subject increases the subject survival. In some aspects, administering the vaccine prevents the development of COVID-19 disease in the subject.

[0031] Another aspect of the present invention is to provide a method to elicit an immune response that provides protection against disease caused by SARS-CoV-2 infection. The method includes administering to a subject in need thereof a composition that includes a soluble spike protein subunit expressed and secreted by an expression vector that includes a codon optimized DNA sequence that includes SEQ ID NO: 1, 2, 3, 4, 5 or 6.

[0032] In one aspect, the composition includes an adjuvant to enhance the immune response.

[0033] Other aspects and advantages of the invention will be apparent from the following description and the appended claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0034] **FIGURE 1.** SARS-CoV-2 Spike Protein Linear Map. Schematic of SARS-CoV-2 Spike protein primary structure with the various domains labeled, SS signal sequence; S2', S2'

protease cleavage site; FP, fusion peptide; HR1, heptad repeat 1; CH, central helix; CD, connector domain; HR2, heptad repeat 2; TM, transmembrane domain; CT, cytoplasmic tail. Domains that were excluded from the ectodomain expression construct or could not be visualized in the final map are colored white. Adapted from Wrapp et al (2020).

[0035] FIGURE 2. SARS-CoV-2 Spike Protein Structure. Crystal structure of single pre-fusion protomer from SARS-CoV-2 Spike protein trimer with the RBD in the up conformation is shown. The main domains are labeled, S1, S2, and S1 subdomains, NTD N-terminal domain; RBD receptor binding domain. Adapted from Wrapp et al (2020).

[0036] FIGURE 3. Alignment of the amino acid sequences from all 2019 nCoV Spike protein subunits expressed in *Drosophila* S2 cells relative to the 2019 nCoV Spike full length protein sequence (SEQ ID NO 8). Modifications relative to the native sequence are indicated in bold and underlined sequences.

[0037] FIGURE 4. Diagram of the SARS-COV-2 spike subunit proteins expressed in *Drosophila* S2 cells relative to the full-length spike protein. Spike 1-1273 is SEQ ID NO:8 (aa 1-1273); Ecto aa 14-1147 of SEQ ID NO:8 (SEQ ID NO:1); Ecto-F 14-1147 of SEQ ID NO:8 with L-Fold (SEQ ID NO:2); S1 aa 14-594 of SEQ ID NO:8 (SEQ ID NO:3); RBD aa 318-594 of SEQ ID NO:8 (SEQ ID NO:4); RBD-F aa 318-594 of SEQ ID NO:8 with L-Fold (SEQ ID NO:5); and NTD aa 14-305 of SEQ ID NO:8 (SEQ ID NO:6).

[0038] FIGURE 5. The wild type Wuhan-Hu-1 spike nucleotide sequence (SEQ ID NO:15) along with translation (SEQ ID NO:8).

[0039] FIGURE 6. Expression of nCoV Spike Ectodomain. Coomassie stained gel and Western Blot with unconcentrated culture medium from parental S2 cell lines expressing the nCoV-S-SSDQ-Ecto-CO, nCoV-S-SQ-Ecto-CO and nCoV-S-WT-Ecto. Arrows indicate the position of the S-Ecto protein.

[0040] FIGURE 7. Expression of nCoV-S-S1-CO. Coomassie stained gel and Western Blot with unconcentrated culture medium of S2 cell lines compared to purified protein. Arrow indicates the position of the nCoV-S-S1 protein.

[0041] FIGURE 8. Binding of recombinant hACE2-His-tagged protein to expressed recombinant SARS-CoV-2 spike proteins. Two different lots of each the expressed SARS-CoV-2 spike protein subunits, RBD, S1, and Ecto, were coated on ELISA plates and detected with recombinant hACE2 protein. RBD-Mo lot 1009 and S1 lot 1014 show strong binding of hACE2 protein indicating structural and functional integrity.

[0042] **FIGURE 9.** Expression of nCoV-S-RBD-CO. Coomassie stained gel and Western Blot with unconcentrated culture medium of S2 cell lines compared to purified protein. Arrow indicates the position of the nCoV-S-RBD protein.

[0043] **FIGURE 10.** Expression of nCoV-S-RBD-CO-foldon and nCoV S SSDQ Ecto CO-foldon. Coomassie stained gel and Western Blot with unconcentrated culture medium of S2 cell lines compared to subunits lacking the foldon domain and purified protein. Arrows indicates the position of the nCoV-S-RBD-foldon and nCoV-S-Ecto-foldon protein.

[0044] **FIGURE 11.** Expression of nCoV-S-NTD-CO. Coomassie stained gel and Western Blot with unconcentrated culture medium of S2 cell lines compared to purified protein. Arrow indicates the position of the nCoV-S-NTD protein.

[0045] **FIGURE 12.** Immunogenicity of nCoV-S-RBD-foldon in mice. ELISA analysis of serum after 2 and 3 doses. The dilution of serum that results in half maximal binding (EC₅₀) is reported as geometric mean (GMT) for each group.

[0046] **FIGURE 13.** Immunogenicity of nCoV-S-RBD-foldon in mice. Micro-neutralization analysis of serum post-dose 2 (PD2) and post-dose 3 (PD3). The dilution of serum that results in 50% neutralization (MN₅₀) is determined for each serum sample. The MN₅₀ geometric mean (GMT) for each group is reported.

[0047] **FIGURE 14.** Immunogenicity of nCoV-S-RBD-foldon in mice. RBD blocking analysis of serum post-dose 2 (PD2) and post-dose 3 (PD3). The percent blocking (reduction) for each of serum samples in each group are plotted as reciprocal dilutions.

[0048] **FIGURES 15A-15B.** Results of blocking assay and the MN assay for the serum samples for study M-001. **FIGURE 15A.** Blocking assay results are reported as geometric mean (GMT) of the EC₅₀ of blocking for individual serum samples. **FIGURE 15B.** Micro-neutralization (MN₅₀) titers are reported for serum pools for each group tested in duplicate.

[0049] **FIGURE 16A-16B.** Results of blocking assay and the MN assay for the serum samples for study M-002. **FIGURE 16A.** Blocking assay results are reported as geometric mean (GMT) of the EC₅₀ of blocking for individual serum samples. **FIGURE 16B.** Micro-neutralization (MN₅₀) titers are reported for serum pools for each group tested in duplicate.

[0050] **FIGURE 17.** ELISA results reported as IgG2a/IgG1 ratio for Immunogenicity Study M-002.

[0051] **FIGURE 18.** IgG2a/IgG1 ratios for select groups from Immunogenicity Studies M-000, M-001, and M-002. P values determined using one-way ANOVA test as compared to the

Alum Group. If P value is not shown then it was not significant, otherwise ** = $P < 0.01$, *** = $P < 0.001$, **** = $P < 0.0001$.

[0052] FIGURES 19A-19B. Results for the M-003 challenge study: Percent of body weight change post challenge and a summary table of results that includes, pre-challenge MN_{50} results and percent survival for each group.

[0053] FIGURE 20A-20B. Results for the M-005 challenge study: Percent of body weight change post challenge and a summary table of results that includes, pre-challenge and post-challenge MN_{50} results, and percent survival for each group.

DETAILED DESCRIPTION OF THE INVENTION

[0054] The present invention is based on the seminal discovery that truncated and codon-optimized nucleic acid sequences encoding novel SARS-Cov-2 spike protein subunits can be expressed and formulated in vaccine compositions comprising such novel SARS-Cov-2 spike protein subunits, and for methods of use thereof.

[0055] Before the present compositions and methods are described, it is to be understood that this invention is not limited to particular compositions, methods, and experimental conditions described, as such compositions, methods, and conditions may vary. It is also to be understood that the terminology used herein is for purposes of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only in the appended claims.

[0056] As used in this specification and the appended claims, the singular forms “a”, “an”, and “the” include plural references unless the context clearly dictates otherwise. Thus, for example, references to “the method” includes one or more methods, and/or steps of the type described herein which will become apparent to those persons skilled in the art upon reading this disclosure and so forth.

[0057] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

[0058] 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 invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, it will be understood that modifications and variations are encompassed within the spirit and scope of the instant

disclosure. The preferred methods and materials are now described.

[0059] The invention provides optimized SARS-CoV-2 spike gene sequences for expression of soluble and stable spike protein subunits that are composed of contiguous sequences that have been codon optimized, contain an optimized signal peptide cleavage site, and have optimized N-terminal and C-terminal ends that enhance expression and stability of the expressed spike subunit proteins. The optimized gene sequences are inserted into a *Drosophila* S2 cell expression vector which drives the expression of high levels of high-quality SARS-CoV-2 spike subunit proteins in S2 cells that have been stably transformed with the expression vectors carrying the optimized gene sequences. The use of the optimized gene sequences results in an increase in the productivity and quality of the expressed SARS-CoV-2 spike subunit proteins. The enhanced expression of the SARS-CoV-2 spike subunit proteins provides for production of effective immunogens at an improved cost of goods which can bolster the ability to manufacture recombinant proteins suitable for use in vaccines to combat the spread of the SARS-CoV-2.

[0060] SARS-CoV-2 Spike optimized proteins, polynucleotides encoding them and expression vectors thereof

[0061] In one embodiment, the invention provides an isolated nucleic acid sequence selected from SEQ ID NO:1, 2, 3, 4, 5 and 6. As used herein, the term “nucleic acid” or “oligonucleotide” refers to polynucleotides such as deoxyribonucleic acid (DNA) or ribonucleic acid (RNA). Nucleic acids include but are not limited to genomic DNA, cDNA, mRNA, iRNA, miRNA, tRNA, ncRNA, rRNA, and recombinantly produced and chemically synthesized molecules such as aptamers, plasmids, anti-sense DNA strands, shRNA, ribozymes, nucleic acids conjugated and oligonucleotides. According to the invention, a nucleic acid may be present as a single-stranded or double-stranded and linear or covalently circularly closed molecule. A nucleic acid can be isolated. The term “isolated nucleic acid” means, that the nucleic acid (i) was amplified *in vitro*, for example via polymerase chain reaction (PCR), (ii) was produced recombinantly by cloning, (iii) was purified, for example, by cleavage and separation by gel electrophoresis, (iv) was synthesized, for example, by chemical synthesis, or (vi) extracted from a sample. A nucleic acid might be employed for introduction into, i.e. transfection of, cells, in particular, in the form of RNA which can be prepared by *in vitro* transcription from a DNA template. The RNA can moreover be modified before application by stabilizing sequences, capping, and polyadenylation.

[0062] The terms "sequence identity" or "percent identity" of polynucleotides or polypeptides are used interchangeably herein. To determine the percent identity of two polypeptide molecules or two polynucleotide sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first polypeptide or polynucleotide for optimal alignment with a second polypeptide or polynucleotide sequence). The amino acids or nucleotides at corresponding amino acid or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = number of identical positions / total number of positions (i.e., overlapping positions) x 100). In some embodiments the length of a reference sequences (e.g. SEQ ID NOs:1-6) aligned for comparison purposes is at least 80% of the length of the comparison sequence, and in some embodiments is at least 90% or 100%. In an embodiment, the two sequences are the same length.

[0063] Ranges of desired degrees of sequence identity are approximately 80% to 100% and integer values in between. Percent identities between a disclosed sequence and a claimed sequence can be at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, or at least 99.9%. In general, an exact match indicates 100% identity over the length of the reference sequences (e.g., SEQ ID NOs:1-6).

[0064] Polypeptides and polynucleotides that are about 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 99.5% or more identical to polypeptides and polynucleotides described herein are embodied within the disclosure. For example, a polynucleotide can have 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to SEQ ID NOs:1-6.

[0065] The term "gene sequence" refers to a sequence of DNA that is transcribed into an RNA molecule that may function directly or be translated into an amino acid chain.

[0066] The term "optimized" refers to sequences that were derived from naturally occurring sequences and have been altered to enhance their functions.

[0067] The term "codon optimized" refers to a nucleic acid coding region that has been adapted for expression in the cells of a given host by replacing at least one, or more than one, or a significant number, of codons with one or more codons that are more frequently used in the genes of that host.

[0068] One aspect of the present invention is to provide a codon optimized DNA sequence encoding a SARS-CoV-2 S protein subunit. The expression of the DNA sequence results in secretion of a soluble S protein subunit in the culture medium.

[0069] The coronavirus (CoV) genome encodes four major structural proteins: nuclear, membrane, small envelope, and spike (S). The S protein is the major viral surface protein and mediates viral entry. Recent studies indicate that human angiotensin-converting enzyme 2 (hACE2) protein is the major cellular receptor for the S protein for SARS-CoV-2 (Wall et al, 2020). The S protein of SARS-CoV and SARS-CoV-2 has also been shown to be the target of virus nAbs (He et al, 2006; Wall et al, 2020). The S protein is a class 1 fusion protein related to HIV, influenza and Ebola. The S protein is divided into the S1 and S2 domains. The S1/S2 junction is delineated by a furin protease cleavage site. S1 contains the receptor binding domain (RBD) and the S2 contains the fusion peptide, 2 helical domains, transmembrane domain, and cytoplasmic tail. **FIGURE 1** shows the primary structure of the S protein and the various features. The S protein forms a homotrimer that make up the spikes on the surface of the virus. The crystal structure of the SARS-CoV-2 S glycoprotein has recently been solved (Wall et al, 2020; Wrapp et al, 2020). **FIGURE 2** shows a single protomer of the S trimer.

[0070] As indicated above, the S glycoprotein has been determined to be the target of virus nAb; therefore, it is an obvious choice for vaccine development. There are several factors that need to be considered in an effort to define which portions of the S glycoprotein comprises a suitable vaccine candidate when expressed as a recombinant subunit. First, the expression of a soluble product requires the exclusion of the transmembrane domain, thus the need to define an appropriate carboxy terminal truncation point. Second, the S glycoprotein contains a furin protease cleavage site which can impact the integrity of an expressed protein product. Third, as the goal of a vaccine candidate based on the S glycoprotein is to promote the production of nAb in immunized subjects, the relevant regions and structural constraints need to be considered. Fourth, while the focus is on regions of spike that elicit nAbs, there is also benefits in excluding certain segments of the protein that may distract from the ability to elicit nAbs. Fifth, there are concerns that various portions of the S glycoprotein can contribute to immune enhancement or immunopathology and, therefore, selection of the portions of the protein that avoid this potential is preferred. In order to develop SARS-COV-2 S glycoprotein subunit vaccine candidate, it is important to consider the factors above so that a suitable vaccine candidate can be produced. Various methods can be applied to enhance the quality and

productivity of expressed recombinant S glycoprotein subunit proteins such that they will result in effective vaccine candidates.

[0071] While there are examples of coronavirus spike gene sequences being expressed, there are no clear examples of expression-enhancing sequence optimization, as most examples utilize the naturally occurring (wild type) sequences of the parent viruses. In an effort to produce selected recombinant proteins as vaccine candidates, it is important to optimize conditions to express these proteins at high levels and in proper conformations such that they serve as effective vaccines. Various means exist to optimize expression. However, each selected sequence requires several rounds of experimentation to determine which methods or combination of methods will result in the most effective expression in a given expression system while maintaining the appropriate native, or biologically relevant characteristics. These efforts to optimize selected sequences are a means to enhance the immunogenic potential to induce a nAb response, in this case, efforts are focused on the SARS-CoV-2 spike gene sequence.

[0072] In the course of expressing optimized gene sequences for the production of desired recombinant protein subunits, it is important to utilize biologically functional methods to assess the potential of the expressed products to assess to potential of the products. For vaccine candidates targeting viral diseases, the primary criteria are to assess the ability of the proteins to elicit potent and specific antibody responses when used to immunize animals. The most basic assay is the Enzyme Linked Immuno-Sorbent Assay (ELISA) which measures total antibody binding to a specific target. This provides a general measure of the strength of the immune response following a course of immunization. A more specific assay is one in which the ability of serum to neutralize virus from infecting cells in culture. There are various forms of these assays, based on how viral infection of the cells are measured. One form of the assay is the Micro-Neutralization (MN) test which monitors cytopathic effect (CPE) on the cells following the addition of virus and serum mixture. Lastly, when the cell receptor and the receptor binding site on the viral proteins are known and recombinant versions are available, assays that measure the ability to block protein-protein interaction (binding) can be set up. In the case of SARS-CoV-2, the receptor binding domain has been defined and the host cell receptor is known to be ACE2. The use of these assays provide can provide strong evidence that properly designed and expressed recombinant protein subunits are potential vaccine candidates.

[0073] In efforts to express proteins, problems are often encountered with virus envelope

sequences that impact the appropriate native or biologically relevant characteristics which can hamper optimal expression. These problems include non-optimal truncations that define the amino-terminus and the carboxy-terminus of the expressed product, poor or ineffective post-translational processing, and poor matching of the native codon usage with the codon usage of the selected expression system's host cells.

[0074] The efficiency of heterologous protein expression in eukaryotic systems is dependent on many factors, such as promoter and associated regulatory elements, transcription initiation sequences, and poly-adenylation signals. As the expression vectors used in a typical system are optimized for the given host cell utilized, the optimization of the gene sequence of interest is often of great importance to ensure optimal expression of the desired protein product. This is typically done by adaptation of the codon usage of the gene sequence to the typical codon usage of the host cells. While the gene sequence is altered through codon optimization, the amino acid sequence of the encoded protein is not modified through the optimization process (Gustafsson et al, 2004). Basic codon usage optimization involves substituting rare codons in the target gene sequence to ones used more frequently by the host cells. Alternatively, the entire gene sequence can be altered to be in line with the codon usage of the host cells used to express the desired product. With the current efficiency of *de novo* gene synthesis, the latter approach has become the preferred method of codon optimization. As the expression of heterologous proteins is an important part of the biotechnology industry, methods such as codon optimization are often useful in improving expression levels.

[0075] Most proteins that are secreted from cells contain an N-terminal signal sequence that directs the protein into the cell's secretion pathway. Optimization of internal secretion signal or signal peptide sequences that interact with the endoplasmic membrane to initiate the secretion process has the potential to increase the efficiency of processing and hence an increase in protein expression. The eukaryotic signal sequence has been divided into three structural regions, basic, hydrophobic, and polar, starting from the N-terminus and proceeding to the C-terminus respectively (von Heijne, 1986 and Bendtsen et al 2004). Over the years, numerous secretion signals have been identified and used to direct the secretion of recombinant proteins. Although many different signal sequences have been used and shown to be functional, few studies have been reported that define optimal sequences for a given cell type. The general characteristics and rules related to the three structural regions are well established, as detailed by von Heijne (1986) and by Bendtsen et al (2004), however, little comparative experimental

data exist as to what constitutes an optimal secretion signal in a given expression system or a given heterologous protein being expressed. Most published reports deal with the characterization and optimization of gram positive bacterial or yeast secretion signals (Le Loir et al, 2005 and Hofmann and Schultz, 1991). One report that describes the optimization of the IL-2 secretion signal clearly demonstrates the benefits of optimization (Zhang et al, 2005). While the secretion signal peptide is important for optimal expression of secreted proteins, the signal peptidase cleavage site is also an important part in the secretion process. In designing heterologous expression and secretion of selected protein sequences it is important to ensure the signal peptidase cleavage site remains optimal.

[0076] Many eukaryotic proteins are modified by N-linked glycosylation (asparagine-linked). The of number of glycosylation sites and the efficiency of glycosylation by the enzyme oligosaccharyltransferase can vary for each protein expressed and can be based on a number of factors. This can influence its expression and function. N-Linked glycosylation usually occurs at the Asn residues in the Asn-X-Ser/Thr motif, where X is any amino acid except for Pro. However, many Asn-X-Ser/Thr sequences are not glycosylated or are glycosylated inefficiently (Mellquist et al, 1998). Inefficient glycosylation at one or more Asn-X-Ser/Thr sequences in a protein results in the production of heterogeneous glycoprotein products. The work of Mellquist et al has revealed that the amino acid at the Y position (amino acid residue immediately following the Ser or Thr residue) is an important determinant of core glycosylation efficiency. This provides an example of a means to optimize the glycosylation efficiency of proteins expressed in heterologous systems.

[0077] The surface proteins of enveloped viruses form multimeric configurations. The spike protein on the surface of coronaviruses is defined as a class 1 fusion protein as it forms a homotrimer. These trimeric structures are anchored in the virus membrane shell by the transmembrane (TM) domains. Expression of the spike protein ectodomain (lacking the TM) results in monomers. Various methods are available to drive the formation of multimeric forms of expressed proteins. In general, these methods involve the fusion of protein domains to the protein of interest. These protein domains can promote the formation of multimeric states of the protein. A commonly used protein domain used to promote trimer formation consist of a 29 amino acid sequence that is derived from the bacteriophage T4 fibritin protein sequence. This is referred to as the “foldon” sequence. The foldon sequence is located at the C-terminus of the fibritin protein, and naturally brings together three monomers of fibritin via non-covalent

bonding to form a trimeric molecule. The use of the foldon domain provides a means to drive the formation of trimeric molecules when fused to the protein of interest.

[0078] As described above, various structural aspects are required for optimal protein expression. When selected protein sequences are chosen for expression in a selected cell system, these structural aspects can be altered to enhance the expression level of the desired protein product. Alternatively, modification of internal protein sequences to enhance selected epitopes or to remove selected epitopes can be employed to create a more desired product, as in, a protein product that has an enhanced immunogenic potential. As an example, epitopes in the dengue E protein that are believed to generate flavivirus cross-reactivity antibody responses were altered to reduce the potential of immune enhancement (Hughes et al, 2012). These included the immunodominant B cell epitope of the fusion peptide and domain III epitopes. The SARS-CoV S glycoprotein provides another example where the peptide 597-603 has been identified as a B-cell epitope that induces antibodies that enhance infection. (Wang et al, 2016). This same location in the spike protein of SARS-CoV-2 has been identified to contain a mutation, Asp 614 to Gly, that results in higher virus titers, but also may have an immunological role (Korber et al, 2020).

[0079] The crystal structure of the SARS-CoV-2 S glycoprotein has recently been solved (Wall et al, 2020; Wrapp et al, 2020); however, the optimal gene sequence for the expression of S proteins subunits in eukaryotic host cell expression systems are not currently well defined. Current technology and methods provide the potential to assemble gene sequences and make modifications to internal sequences and define new end points that can lead to improvements in structure and function. While the potential exists to make such modifications, it is common knowledge that not all attempts to do so result in success. The modifications or attempts to optimize that work with one protein and in a given expression system do not always work on other proteins or in other expression systems. For example, the removal of the first 58 amino acids from the N-terminus of the West Nile envelope protein ectodomain abolishes expression. In another example, the expression of the C-terminal domain (domain III) as of the West Nile or Tick-borne encephalitis envelope protein is easily expressed, however, these subunit proteins are suboptimal in their ability to prime functional immune responses, though it is not obvious why this occurs. In an example specific to SARS-CoV-2, an attempt to express the spike protein subunit NTD (SEQ. ID. NO:14), resulted in multiple forms of the NTD that were misfolded and were heterogenous in the extent of glycosylation (Example 5). Therefore as

illustrated by these examples, a systematic evaluation is required to determine the potential of various efforts to modify and optimize a given gene sequence such that high levels of high-quality heterologous protein are expressed and that such alterations do not negatively impact the desired functional attributes.

[0080] In the biotechnology field, the ability to efficiently produce recombinant proteins at a favorable cost of goods is a key to success. In order to achieve this goal for a particular protein, in this case the SARS-CoV-2 S glycoprotein using the *Drosophila* S2 cell expression system, further experimentation is required to define the parameters that result in optimal expression of a high-quality protein product.

[0081] The optimized SARS-CoV-2 gene sequences for expression of soluble and stable subunit proteins are composed of a spike gene sequence that has been codon optimized, that contains an optimized signal peptidase cleavage site, and that has selected amino and carboxy truncation points that enhance expression and stability of the S subunit proteins. The optimized gene sequence is contained in an expression vector for use in *Drosophila* S2 cells. The codon optimization of the gene sequence is designed for optimal expression in *Drosophila* S2 cells. The optimized signal peptidase cleavage site is utilized to result in effective post-translational processing at the N-terminus of the subunits. The end points of the gene sequences have been selected to produce optimal N-terminal and C-terminal ends for the spike protein subunits which help to stabilize the expressed subunits. The foldon trimerization domain is added to the C-terminus to further stabilize expressed spike protein subunits. The combination of the optimization methods has resulted in unique gene sequences that provide for the expression of soluble SARS-CoV-2 S proteins at high levels and with enhanced stability. These improved SARS-CoV-2 S protein subunits are suitable for use as vaccine candidates to protect against disease caused by SARS-CoV-2 infection.

[0082] The optimized SARS-CoV-2 S gene sequences of the present invention are capable of high-level expression and secretion of the encoded S subunit proteins into the culture medium of transformed S2 cells. Specifically, the described SARS-CoV-2 S gene sequences have been optimized for 1) codon usage in *Drosophila* S2 cells, 2) a synthetic, optimized secretion signal processing site, and 3) N-terminal and C-terminal truncation points that add stability and function to the expressed products.

[0083] In one aspect, the codon optimized DNA sequence encodes the ectodomain with the optimized signal peptidase cleavage site and includes SEQ ID NO:1.

[0084] In a preferred embodiment of the invention, the SARS-CoV-2 spike subunit proteins have an N-terminus sequence with a synthetic sequence designed to enhance the processing of the secretion signal cleavage site by signal peptidase as the translated protein transits the endoplasmic reticulum. This synthetic cleavage site and N-terminal modification has been optimized to increase the recognition of the signal protease cleavage site. Specifically, the amino acid residues +1, +2, and +3 relative to the signal peptidase cleavage site have been modified. Upon cleavage, these amino acid residues define the N-terminus of the expressed SARS-CoV-2 protein subunit sequences. Following the cleavage by signal peptidase in S2 cells, the expressed SARS-CoV-2 spike protein subunits begin with the amino acid residues Ser, Ser, Asp. The N-terminal Ser, Ser, Asp (SSD) sequence can be seen in the SARS-CoV-2 spike subunit proteins shown in the amino acid alignment shown in **FIGURE 3**.

[0085] In a preferred embodiment of the invention, the C-terminus of the SARS-CoV-2 spike ectodomain has been truncated to stabilize the soluble spike subunit protein that is expressed, SEQ ID NO:9. Specifically, the C-terminus of the soluble SARS-CoV-2 spike ectodomain has been defined as Ser-1147 as shown in **FIGURE 3**. The codon optimized nucleotide sequence that encodes the truncated C-terminal SARS-CoV-2 spike ectodomain is detailed in SEQ ID NO:1.

[0086] In another aspect the codon optimized DNA sequence encodes the ectodomain, the optimized signal peptidase cleavage site, with a foldon domain at the C-terminal end to promote trimerization and includes SEQ ID NO:2.

[0087] In another aspect, the codon optimized DNA sequence encodes the S1 subunit with the optimized signal peptidase cleavage site and includes SEQ ID NO:3.

[0088] In a more preferred embodiment of the invention, the C-terminus of the SARS-CoV-2 spike S1 subunit has been truncated to stabilize the soluble spike subunit protein that is expressed, SEQ ID NO:10. Specifically, the C-terminus of the soluble SARS-CoV-2 spike S1 subunit has been defined as Gly-594 as shown in **FIGURE 3**. The codon optimized nucleotide sequence that encodes the truncated C-terminal SARS-CoV-2 spike S1 subunit is detailed in SEQ ID NO:3.

[0089] In another aspect, the codon optimized DNA sequence encodes the RBD subunit with the optimized signal peptidase cleavage site and includes SEQ ID NO:4.

[0090] In a more preferred embodiment of the invention, the SARS-CoV-2 spike RBD subunit has been truncated to stabilize the soluble spike subunit protein that is expressed, SEQ

ID NO:11. Specifically, the N-terminus and C-terminus of the soluble SARS-CoV-2 spike RBD subunit has been defined as Phe-318 and Gly-594, respectively, as shown in **FIGURE 3**. The codon optimized nucleotide sequences that encode the truncated N- and C-terminal SARS-CoV-2 spike RBD subunits is detailed in SEQ ID NO:4.

[0091] In another aspect the codon optimized DNA sequence encodes the RBD subunit, the optimized signal peptidase cleavage site, with a foldon domain at the C-terminal end to promote trimerization and includes SEQ ID NO:5.

[0092] In a more preferred embodiment of the invention, the truncated the SARS-CoV-2 spike ectodomain and RBD subunits are further modified at the C-terminus by operably linking a foldon domain to the subunit to promote the trimerization of the expressed subunits, SEQ ID NO:12 and SEQ ID NO:13. The ectodomain-foldon and RBD-foldon relative to the ectodomain and RBD lacking the foldon domain are shown in **FIGURE 3**. The codon optimized nucleotide sequences that encodes the SARS-CoV-2 spike ectodomain and RBD subunits with the linked foldon domains are detailed in SEQ ID NO:2 and SEQ ID NO:5.

[0093] Thus, the present invention provides the combination of multiple optimizations directed at different aspects of the SARS-CoV-2 spike gene sequence in such a manner that an additive benefit is achieved and results in high levels of the spike protein subunits being expressed. A summary of the SARS-CoV-2 spike subunit proteins expressed in *Drosophila* S2 cells relative to the full-length spike protein is presented in **FIGURE 4**. The optimized SARS-CoV-2 spike sequence when used to express the defined protein subunits in *Drosophila* S2 cells results in the economic production of large quantities of high-quality proteins. The Examples below show that using the individual optimized elements in the SARS-CoV-2 gene sequence results in improved or enhanced expression of the spike protein subunits.

[0094] In another aspect, the codon optimized DNA sequence encodes the NTD subunit with the optimized signal peptidase cleavage site and includes SEQ ID NO:6.

[0095] In another embodiment, the invention provides an isolated amino acid sequence encoded by a nucleic acid sequence selected from SEQ ID NO:1, 2, 3, 4, 5 and 6.

[0096] The terms “peptide”, “polypeptide” and “protein” are used interchangeably herein and refer to any chain of at least two amino acids, linked by a covalent chemical bound. As used herein polypeptide can refer to the complete amino acid sequence coding for an entire protein or to a portion thereof. A "protein coding sequence" or a sequence that "encodes" a particular polypeptide or peptide, is a nucleic acid sequence that is transcribed (in the case of

DNA) and is translated (in the case of mRNA) into a polypeptide *in vitro* or *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxyl) terminus. A coding sequence can include, but is not limited to, cDNA from prokaryotic or eukaryotic mRNA, genomic DNA sequences from prokaryotic or eukaryotic DNA, and even synthetic DNA sequences. A transcription termination sequence will usually be located 3' to the coding sequence.

[0097] In one aspect, the amino acid sequence includes SEQ ID NO: 9, 10, 11, 12, 13 or 14.

[0098] Polypeptides and polynucleotides that are about 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 99.5% or more identical to polypeptides and polynucleotides described herein are embodied within the disclosure. For example, a polypeptide can have 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to SEQ ID NOs:9-14.

[0099] Variants of the disclosed sequences also include peptides, or full-length protein, that contain substitutions, deletions, or insertions into the protein backbone, that would still leave at least about 70% homology to the original protein over the corresponding portion. A yet greater degree of departure from homology is allowed if like-amino acids, i.e. conservative amino acid substitutions, do not count as a change in the sequence. Examples of conservative substitutions involve amino acids that have the same or similar properties. Illustrative amino acid conservative substitutions include the changes of: alanine to serine; arginine to lysine; asparagine to glutamine or histidine; aspartate to glutamate; cysteine to serine; glutamine to asparagine; glutamate to aspartate; glycine to proline; histidine to asparagine or glutamine; isoleucine to leucine or valine; leucine to valine or isoleucine; lysine to arginine, glutamine, or glutamate; methionine to leucine or isoleucine; phenylalanine to tyrosine, leucine or methionine; serine to threonine; threonine to serine; tryptophan to tyrosine; tyrosine to tryptophan or phenylalanine; valine to isoleucine to leucine.

[0100] In another embodiment, the C-terminal end of the expressed RBD and S1 subunits is defined as Gly 594 to avoid the region of the spike protein with potential for immune enhancement. This modification is included in SEQ ID NO:10, SEQ ID NO:11, and SEQ ID NO:13.

[0101] In an additional embodiment, the invention provides an expression vector including a nucleic acid sequence encoding a SARS-Cov-2 spike (S) protein, wherein the nucleic acid sequence includes SEQ ID NO: 1, 2, 3, 4, 5 or 6.

[0102] The term “synthetic” refers to sequences that are not found to occur naturally.

[0103] The term “operably linked” refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner.

[0104] “Expression cassette” means the combination of promoter elements with other transcriptional and translational regulatory control elements which are operably linked to a gene sequence to be expressed. A gene sequence can be inserted into the expression cassette for the purpose of expression of said gene sequence. The expression cassette is capable of directing transcription which results in the production of an mRNA for the desired gene product which is then translated to protein by the host cell translational systems. The expression cassette is integral to the expression vector (plasmid). Such an expression vector directs expression of the protein encoded by the gene sequence once introduced into host cells.

[0105] In one aspect, the vector is a *Drosophila melanogaster* expression vector.

[0106] The *Drosophila melanogaster* cell expression system (“*Drosophila* expression system”) is an established heterologous protein expression system based on the use of expression vectors containing *Drosophila* promoters and *Drosophila* S2 cells (“S2 cells”) (Schneider, *Embryol. Exp. Morph.* (1972) 27:353-365). S2 cells are transformed with these vectors in order to establish stable cell lines expressing proteins corresponding to the heterologous sequences introduced into the vector (Johansen, H. *et al.*, *Genes Dev.* (1989) 3:882-889; Ivey-Hoyle, M., *Curr. Opin. Biotechnol.* (1991) 2:704-707; Culp, J.S., *et al.*, *Biotechnology (NY)* (1991) 9:173-177; U.S. Patents 5,550,043; 5,681,713; 5,705,359; 6,046,025). This insect cell expression system has been shown to successfully produce a number of proteins from different sources. Examples of proteins that have been successfully expressed in the *Drosophila* S2 cell system include HIV gp120 (Culp, J.S., *et al.*, *Biotechnology (NY)* (1991) 9:173-177; Ivey-Hoyle, M., *Curr. Opin. Biotechnol.* (1991) 2:704-707), human dopamine β -hydrolase (Bin *et al.*, *Biochem. J.* (1996) 313:57-64), human vascular cell adhesion protein (Bernard *et al.*, *Cytotechnol.* (1994) 15:139-144). In each of these examples, expression levels were greater than other expression systems that had been previously utilized.

[0107] In addition to high levels of expression, the *Drosophila* expression system has been shown to be able to express heterologous proteins that maintain native-like biological function

(Bin et al., *Biochem. J.* (1996) 313:57-64), (Incardona and Rosenberry, *Mol. Biol. Cell.* (1996) 7:595-611). More recent examples have shown by means of X-ray crystallography studies that this expression system is capable of producing molecules with native-like structure (Modis et al., *Proc. Natl. Acad. Sci. USA* (2003) 100:6986-6991), (Modis et al., *Nature* (2004) 427:313-319), (Xu et al., *Acta. Crystallogr. D Biol. Crystallogr* (2005) 61:942-950). Two other recent publications have also demonstrated the ability of the *Drosophila* expression system to produce high quality products. In the first report, Schmetzer et al. (*J. Immun.* (2005) 174: 942-952) compares baculovirus-expressed EpCAM protein to *Drosophila*-expressed EpCAM protein for protein folding and native conformation. Specifically, BES-expressed EpCAM and *Drosophila*-expressed EpCAM were compared to denatured *Drosophila*-expressed EpCAM. It was determined that the BES-expressed EpCAM was in a partial folded state relative to the non-denatured and denatured *Drosophila*-expressed EpCAM protein. This indicates that the BES-expressed protein is in an incompletely folded state. The *Drosophila*-expressed EpCAM protein, on the other hand, adopted a more completely folded state. The authors of this paper considered the *Drosophila*-expressed protein to be in the “natural” state while the baculovirus-expressed protein was not. In the second report, Gardsvoll et al. (*Prot. Exp. Purif.* (2004) 34:284-295) demonstrate that the expression of the urokinase-type plasminogen activator receptor (uPAR) in S2 cells results in a more homogeneous product in regard to glycosylation (5 N-linked sites) than uPAR expressed in CHO cells.

[0108] In some aspects, the expression cassette of the *Drosophila* expression vector pHH202 has a nucleic acid sequence including SEQ ID NO:7.

[0109] In another aspect, the SARS-CoV-2 S protein has an amino acid sequence including SEQ ID NO: 9, 10, 11, 12, 13 or 14.

[0110] In another embodiment, the *Drosophila* expression vector pHH202 that includes SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:6 is used to express and secrete the encoded heterologous SARS-CoV-2 S protein subunits, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or SEQ ID NO:14, from cultured insect cells. The expression cassette of the pHH202 vector includes SEQ ID NO:7.

[0111] In one embodiment, the expression vectors are used in *Drosophila* cells.

[0112] In another embodiment, the expression vectors are used in *Drosophila* S2 cells.

[0113] The term “transformed” refers to the DNA-mediated transformation of cells. This

refers to the introduction of plasmid DNA into insect cells in the process of generating stable cell lines following the integration of the introduced DNA into the genome of the cells. This term is used in place of the term “transfection” which is often used in the same context. The term transformation is used for the introduction of plasmid DNA to cultured cells to distinguish from the introduction of viral DNA into cultured cells which was originally referred to as transfection. As there are no viral DNA sequences in the present invention which are introduced into the host cells that results in the production of virus-like particles or cell lysis the term transformed is preferred.

[0114] “Expression” or “expressed” means the production of proteins using expression vectors and host cells, for instance, *Drosophila* S2 cells to produce a recombinant protein product that is readily detectable as a cell-associated product or as a secreted product in the culture medium.

[0115] “Secretion” means secretion of an expressed recombinant protein from cultured host cells into culture medium. The expressed and secreted protein is the result of a given gene sequence being operably linked to an expression cassette such that the sequence codes for the given protein.

[0116] The term “product” refers to any recombinant protein, full length or subunit thereof, which is expressed by a host cell into which an expression vector carrying the gene sequence encoding the product has been introduced.

[0117] Insect cells are an alternative eukaryotic expression system that provide the ability to express properly folded and post-translationally modified proteins while providing simple and relatively inexpensive growth conditions. The use of stably transformed insect cell expression systems provide benefits over those based on baculovirus infection of the host insect cells. On this basis, S2 cells were selected as the insect host cells of choice. As a result, the efforts to optimize the expression vectors for stably transformed insect cells were based on data derived from the analysis of specific *Drosophila* genes as well as the complete *Drosophila* genome.

[0118] In one embodiment, the invention provides a method of producing a protein *in vitro* including an expression vector with an operably-linked nucleic acid sequence encoding a SARS-CoV-2 spike (S) protein, wherein the nucleic acid sequence includes SEQ ID NO: 1, 2, 3, 4, 5 or 6 with a *Drosophila melanogaster* cell and culturing the cell under conditions to produce the protein.

[0119] In one aspect, the *Drosophila melanogaster* cell is an Schneider 2 (S2) cell.

[0120] **Vaccine compositions**

[0121] In another embodiment, the invention provides a vaccine composition including (a) an effective amount of a SARS-CoV-2 spike (S) subunit protein, wherein the S protein is encoded by a nucleic acid sequence including SEQ ID NO: 1, 2, 3, 4, 5 or 6, and (b) an effective amount of an adjuvant.

[0122] According to the invention, the term “vaccine” relates to a pharmaceutical preparation (pharmaceutical composition) or product that upon administration induces an immune response, in particular a cellular immune response, which recognizes and attacks a pathogen such as a virus, or a diseased cell such as an infected cell. A vaccine may be used for the prevention or treatment of a disease.

[0123] The terms “therapeutically effective amount”, “effective dose,” “therapeutically effective dose”, “effective amount,” “prophylactically effective amount”, “prophylactic dose,” “prophylactically effective dose”, “prophylactic amount,” or the like refer to that amount of the subject compound that will elicit the biological or medical response of a tissue, system, animal or human that is being sought by the researcher, veterinarian, medical doctor or other clinician. Generally, the response is either amelioration of symptoms in a patient or a desired biological outcome (e.g., induction of an immune response, prevention of viral infection, and the like). For example, a therapeutic or prophylactic dose of the vaccine composition described herein is a dose or amount of the vaccine composition that is sufficient to induce an immune response in the subject, the immune response being then sufficient to provide immune protection to the subject. The effective or prophylactic amount can be determined as described herein or using methods known to those of skill in the art.

[0124] The vaccine compositions described herein include an adjuvant. Vaccine adjuvants are critical for the effective development of protective responses with many antigens.

[0125] Toll-like receptor (TLR) agonist adjuvants are particularly promising, as they engage the innate immune system to stimulate a more robust and durable adaptive immune response. Ligands for TLR 7/8 (Imiquimod, Resiquimod), TLR-9 (CpG), TLR-5 (Flagellin), and TLR-4 have been evaluated pre-clinically as components of vaccine adjuvants. TLR-4 agonist adjuvants have been shown to be safe and effective in several clinical trials, and the TLR-4 agonist adjuvant MPL is a component of the licensed HPV vaccine Cervarix® (GlaxoSmithKline, Rixensart, Belgium).

[0126] TLR-4 agonist adjuvants can be combined with saponin adjuvants. The use of this combination of adjuvants results in potent and durable immune response.

[0127] An ideal TLR-4 agonist is a fully synthetic lipid A molecule (SLA). The most commonly used TLR-4 agonist used for vaccines is monophosphoryl lipid A (MPL). MPL is prepared from bacterial cell walls. The processes used result in heterogeneous preparations of MPL. The synthetic nature of SLA provides for more defined composition relative to MPL. Furthermore, the structure of SLA has been optimized to bind more effectively to the human TLR-4 receptor. SLA enhances the ability of the immune system to respond to vaccine antigens.

[0128] An ideal saponin adjuvant is a highly purified preparation derived from the Soap bark tree (*Quillaja saponaria*) and contains a water-soluble triterpene glucoside molecule. QS21 is a saponin-based adjuvant of this nature. QS21 is purified from extracts of the tree bark. QS21 enhances the ability of the immune system to respond to vaccine antigens. Another example of saponin derived adjuvant includes GPI-0100.

[0129] In one aspect, liposomes are combined with the SLA and QS21 adjuvants to form a liposome-based formulation. The liposome formulation containing SLA and QS21 is referred to as LSQ. The liposome composition can be either anionic or cationic nature, or more preferably it has a neutral charge. The liposome size range can vary from 20-300 nm, more preferably from 40-200 nm, and most preferably 50-150 nm in size.

[0130] In another aspect, a stable oil-in-water emulsion (SE) which preferably includes squalene is combined with SLA to form a stable oil-based emulsion.

[0131] In some aspects, the adjuvant is selected from the group consisting of GPI-0100, synthetic lipid A (SLA) in a stable oil-in water emulsion (SE) (SLA-SE), QS21, QS21 combined with SLA to form a liposome formulation (SLA-LSQ), and QS21+CpG.

[0132] The vaccine compositions described herein include a S subunit protein that is recombinantly produced and expressed in insect host cells. The host cells are modified (e.g., transformed) to express any one of the expression vectors including the nucleic acid sequences described herein, encoding a S protein (e.g., SEQ ID NO: 1, 2, 3, 4, 5 or 6). The resulting proteins, recombinantly produced by the host cells are used in the formulation of the vaccine compositions described herein.

[0133] In one aspect, the S protein of the vaccine composition is the S1 subunit protein. In some aspects, the S1 subunit protein is encoded by a nucleic acid sequence comprising SEQ

ID NO: 3. In another aspect, the S protein is recombinantly produced and expressed in insect host cells.

[0134] In one aspect, the S protein of the vaccine composition is the S1 subunit protein encoded by a nucleic acid sequence comprising SEQ ID NO: 3 and the adjuvant of the vaccine composition is SLA-SE.

[0135] In some aspects, the vaccine composition further includes a pharmaceutically acceptable excipient or carrier.

[0136] The vaccine formulation of the present invention may further include one or more additional pharmaceutically acceptable diluents, carriers, solubilizers, emulsifiers, preservatives and/or adjuvants. By “pharmaceutically acceptable” it is meant the carrier, diluent, solubilizer, emulsifier, preservative or excipient must be compatible with the other ingredients of the formulation and not deleterious to the recipient thereof. For example, a pharmaceutical composition may contain formulation materials for modifying, maintaining or preserving, for example, the pH, osmolarity, viscosity, clarity, color, isotonicity, odor, sterility, stability, rate of dissolution or release, adsorption or penetration of the composition. Suitable formulation materials include, but are not limited to, amino acids (such as glycine, glutamine, asparagine, arginine or lysine); antimicrobials; antioxidants (such as ascorbic acid, sodium sulfite or sodium hydrogen-sulfite); buffers (such as borate, bicarbonate, Tris-HCl, citrates, phosphates or other organic acids); bulking agents (such as mannitol or glycine); chelating agents (such as ethylenediamine tetraacetic acid (EDTA)); complexing agents (such as caffeine, polyvinylpyrrolidone, beta-cyclodextrin or hydroxypropyl-beta-cyclodextrin); fillers; monosaccharides; disaccharides; and other carbohydrates (such as glucose, mannose or dextrans); proteins (such as serum albumin, gelatin or immunoglobulins); coloring, flavoring and diluting agents; emulsifying agents; hydrophilic polymers (such as polyvinylpyrrolidone); low molecular weight polypeptides; salt-forming counterions (such as sodium); preservatives (such as benzalkonium chloride, benzoic acid, salicylic acid, thimerosal, phenethyl alcohol, methylparaben, propylparaben, chlorhexidine, sorbic acid or hydrogen peroxide); solvents (such as glycerin, propylene glycol or polyethylene glycol); sugar alcohols (such as mannitol or sorbitol); suspending agents; surfactants or wetting agents (such as pluronics, PEG, sorbitan esters, polysorbates such as polysorbate 20, polysorbate 80, triton, tromethamine, lecithin, cholesterol, tyloxapal); stability enhancing agents (such as sucrose or sorbitol); tonicity enhancing agents (such as alkali metal halides, preferably sodium or potassium chloride,

mannitol sorbitol); delivery vehicles; diluents; excipients and/or pharmaceutical adjuvants. (Remington's Pharmaceutical Sciences, 18th Edition, A. R. Gennaro, ed., Mack Publishing Company (1990).

[0137] Methods of use

[0138] In one embodiment, the invention provides a method of preventing a SARS-CoV-2 entry into a cell including contacting the cell with a therapeutically or prophylactically effective amount of composition including (a) an effective amount of a SARS-CoV-2 spike (S) protein, wherein S protein is encoded by a nucleic acid sequence including SEQ ID NO: 1, 2, 3, 4, 5 or 6, and (b) an effective amount of an adjuvant, thereby preventing SARS-CoV-2 entry into the cell.

[0139] SARS-CoV-2 is capable of infecting cells by interacting with human angiotensin converting enzyme 2 (hACE2), which acts as a receptor for the virus S protein. By “preventing entry” of the virus into a cell, it is meant that the immune response elicited by the vaccine composition described herein prevents the interaction of the virus S protein with its receptor at the surface of the cell. Thus, the interaction can be reduced, inhibited or prevented by administration of the vaccine composition described herein.

[0140] In one aspect, the vaccine induces the production of nAbs in the subject.

[0141] By “neutralizing antibody” or “nAb” it is meant that the vaccine induces the production of an antibody that defends a cell from a pathogen or infectious particle by neutralizing any effect it has biologically. Neutralization renders the particle, such as the virus, no longer infectious or pathogenic. Neutralizing antibodies are part of the humoral response of the adaptive immune system against viruses, intracellular bacteria and microbial toxin. By binding specifically to surface structures (such as the spike protein of SARS-CoV-2) on an infectious particle, nAbs prevent the particle from interacting with its host cells and thus, prevent entry of the virus into the host cell. If the immune response, nAbs, are able to eliminate the infectious particles before any infection takes place, this is known as sterilizing immunity.

[0142] In some aspects, the nAbs prevent the binding of a SARS-CoV-2 to a target cell and/or target receptor. In various aspects, the target receptor is an ACE2 receptor.

[0143] In another embodiment, the invention provides a method of stimulating a protective immune response in a subject including administering to the subject a therapeutically or prophylactically effective amount of composition including (a) an effective amount of a SARS-

CoV-2 spike (S) protein wherein S protein is encoded by a nucleic acid sequence including SEQ ID NO: 1, 2, 3, 4, 5 or 6, and (b) an effective amount of an adjuvant, , thereby stimulating a protective immune response.

[0144] The term “subject” as used herein refers to any individual or patient to which the subject methods are performed. Generally, the subject is human, although as will be appreciated by those in the art, the subject may be an animal. Thus other animals, including vertebrate such as rodents (including mice, rats, hamsters and guinea pigs), cats, dogs, rabbits, farm animals including cows, horses, goats, sheep, pigs, chickens, etc., and primates (including monkeys, chimpanzees, orangutans and gorillas) are included within the definition of subject.

[0145] The terms “administration of” and or “administering” should be understood to mean providing a pharmaceutical composition in a therapeutically or prophylactically effective amount to the subject in need of treatment. Administration routes can be enteral, topical or parenteral. As such, administration routes, in general, include but are not limited to intracutaneous, subcutaneous, intravenous, intraperitoneal, intraarterial, intrathecal, intracapsular, intraorbital, intracardiac, intradermal, transdermal, transtracheal, subcuticular, intraarticular, subcapsular, subarachnoid, intraspinal and intrasternal, oral, sublingual, buccal, rectal, vaginal, nasal ocular administrations, as well infusion, inhalation, and nebulization. The phrases “parenteral administration” and “administered parenterally” as used herein means modes of administration other than enteral and topical administration.

[0146] To immunize subjects against SARS-CoV-2-induced disease for example, the vaccine formulation containing the recombinant subunit protein and adjuvant are administered to the subject in conventional immunization protocols involving, usually, multiple administrations of the vaccine. Administration is typically by injection, typically intramuscular or subcutaneous injection; however, other systemic modes of administration may also be employed.

[0147] In one aspect, administering includes injecting two doses to the subject at a 3-weeks interval.

[0148] Many different techniques exist for the timing of the immunizations when a multiple administration regimen is utilized. Generally, it is preferable to use a vaccine more than once to increase the levels and diversities of expression of the immunoglobulin repertoire expressed by the immunized subject. Typically, if multiple immunizations are given, they will be given one to two months apart. To further boost the immune response, a second dose of

vaccine can be administered. The preferred immunization schedule for two doses is 0 and 3 weeks. Other immunizations schedules can also be utilized. For example, alternative immunization schedules such as 0, 1; 0, 2 or 0, 3 months could be used. Additional booster vaccinations may be administered at prescribed intervals such as every 5 to 10 years.

[0149] In another aspect, administering includes injecting intramuscularly. In some aspects, a dose includes about 0.5-50 µg of purified S protein. For example, a dose includes about 0.5, 1, 2, 5, 10, 15, 20, 25, 30, 35, 40, 45, or 50 µg of purified S protein.

[0150] By “stimulating, inducing, or eliciting a protective immune response”, it is meant that the vaccine composition described herein induces an immune response in the subject, which provides the subject with an immune protection against a disease. “Inducing an immune response” may mean that there was no immune response against a particular antigen before induction, but it may also mean that there was a certain level of immune response against a particular antigen before induction and after induction said immune response is enhanced. Thus, “inducing an immune response” also includes “enhancing an immune response”. Preferably, after inducing an immune response in a subject, said subject is protected from developing a disease such as a COVID-19 disease or the disease condition is ameliorated by inducing an immune response.

[0151] The term “immune response” refers to an integrated bodily response to an antigen and preferably refers to a cellular immune response and or a humoral immune response. The immune response may be protective/preventive/prophylactic.

[0152] In one aspect, the immune response is a balanced immune response.

[0153] Immune response includes cellular and humoral immune responses. The humoral immune response is mediated by antibodies secreted by B cells. The antibodies neutralize and opsonize free extracellular pathogens, and prolonged antibody production lasting for years after infection or vaccination provide the first line of defense by the adaptive immune system. The cellular immune response is mediated by antigen specific CD4⁺ and CD8⁺ T cells and cells of the innate immune system (e.g. dendritic cells, NK cells and macrophages). T cells cannot recognize free pathogens, but instead identify infected cells and exert effector functions including direct cytotoxic effect and cytokine release. CD4⁺ T cells can be divided into two major subsets, type 1 helper cells (Th1) that secrete interleukin-2 and interferon-γ, and type 2 helper T cells (Th2) secreting interleukins-4, 5, 6 and 10. The cytokines produced by Th1 cells promote a cell-mediated immune response, whereas the humoral immune response is triggered

and maintained by cytokines secreted by Th2 cells following antigenic exposure. The IgG antibody subclass distribution elicited after vaccination is also indicative of the type of immune response, as the IgG1 subclass in mice is believed to signal a Th2 response whilst the IgG2a subclass indicates more of a Th1 profile.

[0154] A “balanced immune response” is characterized by an immune response that include both a strong humoral and a strong cellular immune response, as can be measured by evaluating a ratio between Th1 and Th2 profiles, such as by establishing a IgG2a:IgG1 ratio.

[0155] In one aspect, a balanced immune response is characterized by a IgG2a:IgG1 ratio that is equal or greater than 1.

[0156] In an additional embodiment, the invention provides a method of inhibiting a SARS-CoV-2 infection in a subject including administering to the subject a therapeutically or prophylactically effective amount of composition including (a) an effective amount of a SARS-CoV-2 spike (S) protein, wherein S protein is encoded by a nucleic acid sequence including SEQ ID NO: 1, 2, 3, 4, 5 or 6, and (b) an effective amount of an adjuvant, thereby inhibiting SARS-CoV-2 infection.

[0157] In another embodiment the invention provides a method of inhibiting transmission of a SARS-CoV-2 infection by a subject including administering to the subject a therapeutically or prophylactically effective amount of composition including (a) an effective amount of a SARS-CoV-2 spike (S) protein, wherein S protein is encoded by a nucleic acid sequence including SEQ ID NO: 1, 2, 3, 4, 5 or 6, and (b) an effective amount of an adjuvant, thereby inhibiting transmission of a SARS-CoV-2 infection.

[0158] By “inhibiting infection”, it is meant that the immune response elicited by the vaccine composition prevents the entry of the virus into cells in the subject, thereby blocking the replication and infection cycle of the virus. As previously discussed, the induction of the production of nAbs by the vaccine prevents the binding of SARS-CoV-2 to its target cell and/or target receptor, and thus prevents the infection of the target cells by the virus.

[0159] By “inhibiting transmission” it is meant that the vaccine composition, by preventing the entry of the virus into cells of a subject inhibits the replication and infection cycle of the virus, and therefore reduces or inhibits the risk of transmitting the virus to another subject, by reducing the viral load in the vaccinated subject.

[0160] In many aspects, the vaccine induces the production of nAbs in the subject. In some aspects, the nAbs prevent the binding of a SARS-CoV-2 to a target cell and/or target receptor. In various aspects, the target receptor is an ACE2 receptor.

[0161] In one aspect, the adjuvant is selected from the group consisting of GPI-0100, QS21 + CpG, SLA-SE, and SLA-LSQ.

[0162] In other aspects, administering the vaccine to the subject increases the subject survival.

[0163] The vaccine composition described herein induces the production of nAbs in the subject and prevents SARS-CoV-2 entry into cells of the subject, thereby inhibiting SARS-CoV-2 infection and transmission in/by the subject. COVID 19 disease, the respiratory disease induced by SARS-CoV-2 infection is a deadly disease. By inhibiting SARS-CoV-2 infection and transmission in/by the subject, the vaccine composition described herein prevents the development of SARS-CoV-2-related disease COVID-19, and therefore increase the survival of the subject.

[0164] In some aspects, administering the vaccine prevents the development of a COVID-19 disease in the subject.

[0165] Although the descriptions presented above and the examples that follow are primarily directed at the use of the optimized expression vectors with *Drosophila* S2 cells, the vectors and methods can be applied to other insect cell lines that result in stable cell lines following transformation of host cells with plasmid DNA.

[0166] Presented below are examples discussing expression of SARS-CoV-2 Spike protein subunits, the optimized polynucleotides encoding them, and expression vector thereof contemplated for use in the discussed applications. The following examples are provided to further illustrate the embodiments of the present invention but are not intended to limit the scope of the invention. While they are typical of those that might be used, other procedures, methodologies, or techniques known to those skilled in the art may alternatively be used.

[0167] The following examples are offered by way of illustration and not by way of limitation.

EXAMPLES

[0168] The following examples describe the development of the optimized SARS-CoV-2 gene sequences for the expression of the envelope protein in insect cells. The examples demonstrate the ability to effectively express the proteins in *Drosophila* S2 cells at levels that

are commercially suitable for product development.

[0169] The examples demonstrate the ability to enhance expression of SARS-CoV-2 proteins in S2 cells and the efforts made to determine the contribution of various changes to enhance the expression and function of the spike protein subunits. The results presented below demonstrate that different modifications and truncations of the spike protein can result in high levels of expression or in products that are not properly folded or glycosylated. Thus, the selection of effective modifications must be determined thorough experimentation. Hence, the invention described herein is unique in that the modifications and truncations described result in distinct protein subunits suitable for use as vaccines to protect against disease caused by infection with SARS-CoV-2.

EXAMPLE 1

EXPRESSION OF WILD TYPE AND CODON OPTIMIZED SARS-CoV-2 SPIKE ECTODOMAIN IN *DROSOPHILA* S2 CELLS

[0170] In an effort to identify optimized gene sequences for driving high levels of high-quality SARS-CoV-2 spike protein in S2 cells, the wild type spike gene ectodomain (ecto) sequence (S-WT-Ecto) was compared to a codon optimized spike gene ectodomain sequence (S-Ecto-CO). Both the wild type and codon optimized sequences were produced synthetically (ATUM, Santa Clara, CA). For codon optimization, the standard *Drosophila melanogaster* codon table was used (Kazusa DNA Research Institute, www.kazusa.or.jp/codon/). As the objective is to improve the efficiency of expression, which is in part controlled by the translation process, a threshold of 10% usage was used in assigning codons (any codon that is used < 10% is excluded). Additionally, based on our analysis of highly expressed proteins in *Drosophila*, we have added the exclusion of the following codons, CGA for Arg, ATA for Ile, and GTA for Val. The synthesized gene sequences included appropriate restriction enzyme sites at the ends and a stop codon was included at the end of the envelope protein coding region. For the expression of the SARS-CoV-2 spike protein subunits, the genomic sequence representing the SARS-CoV-2 spike protein was used. The sequences utilized for expression are based on the SARS-CoV-2 Wuhan-Hu-1 strain (Genbank Accession number NC_045512). The wild type Wuhan-Hu-1 spike nucleotide sequence (SEQ ID NO:15) along with translation (SEQ ID NO:8) is provided in **FIGURE 5**. The sequence for spike protein was truncated at the carboxy end at Ser-1147 to produce the ectodomain (S-Ecto). Additionally, the native furin

cleavage site RRAR (682-685) that defines the S1/S2 junction was mutated to remove the cleavage site and further stabilize the expressed ectodomain. The sequence was mutated to encode GSAR.

[0171] To improve the processing at the signal peptide cleavage site, the three amino acids, Ser-Ser-Asp, were added immediately before the N-terminal Gln (Q) residue that results from the following cleavage of the signal peptide. The sequence that encodes Ser-Ser-Asp (SSD) was fused in frame with the secretion signal of the expression vector. When the S-SSD-Ecto sequence is expressed in the S2 cells, the S-Ecto junction is processed by an S2 encoded signal peptidase. This results in the secretion of an S-Ecto product into the culture medium with an N-terminus of Ser-Ser-Asp-Gln (SSDQ). The codon optimized nCoV-S-SSDQ-Ecto-CO sequence with the enhanced N-terminus and the mutated furin site is detailed in SEQ ID NO:1. The resultant protein product is detailed in SEQ ID NO:9. The alignment of the nCoV S SSDQ Ecto CO protein product relative to the wild type SARS-CoV-2 S protein is shown in **FIGURE 3**.

[0172] As a control, the codon optimized S-Ecto was also fused to the secretion signal of the expression vector using the native N-terminus (Gln) lacking the SSD sequence. This is referred to as nCoV-S-SSD-SQ-Ecto-CO.

[0173] The synthetic DNA fragments were digested with appropriate restriction enzymes and inserted within the expression cassette (SEQ ID NO:7) of the pHH202 expression vector that has been digested with Nhe I and Xho I. The pHH202 expression cassette contains the following elements: metallothionein promoter, optimized Kozak sequence, influenza HA secretion signal, and the SV40 early 3'UTR. The hygromycin encoding gene is also incorporated into the pHH202 expression plasmid downstream of the expression cassette. The pHH202 expression plasmid is designed to allow directional cloning of the gene of interest into unique Nhe I and Xho I sites. The junctions and full inserts of all constructs were sequenced to verify that the various components that have been introduced are correct and that the proper reading frame has been maintained.

[0174] Standard methods of culturing and transformation of S2 cells were utilized (Van der Straten, *Methods in Mol. and Cell Biol.* (1989) 1:1-8; Culp et al., *Biotechnology* (1991) 9:173-177; Kirkpatrick and Shatzman, In *Gene Expression Systems: Using Nature for the Art of Expression*, Eds. Fernandez and Hoeffler, Academic Press, (1999) 289-330). *Drosophila* S2 cells (Schneider, *J. Embryol. Exp. Morph.* (1972) 27:353-365) obtained from ATCC were

utilized. The S2 cells have been adapted to growth in Excell 420 medium (Sigma, St Louis, MO) and all procedures and culturing described herein were in Excell 420 medium. Cultures are typically seeded at a density of 2×10^6 cells/mL and are passed between days 5 and 7. All cultures were incubated at 26° to 27°C. Expression plasmids into which genes of interest were inserted were transformed into S2 cells using the ExpiFectamine Sf reagent (ThermoFisher, Waltham, MA). Following transformation, cells resistant to hygromycin B, 0.3 mg/mL, were selected. Once stable cell lines were selected, they were evaluated for expression of the appropriate products. For the evaluation of expression, 2 mL cultures of selected cell lines were seeded at 2×10^6 cells/mL and cultured in the presence of 0.2 mM copper sulfate at 26°C for 6 days. Cultures were evaluated for expression of recombinant proteins in both the cell associated fractions and the culture medium. Proteins were separated by SDS-PAGE and either stained with Coomassie blue or transferred onto nitrocellulose for Western blot analysis. Expression levels ≥ 5 μ g/mL (5 mg/L) are readily detected in S2 cultures by Coomassie staining of SDS-PAGE gels.

[0175] Parental S2 cell lines expressing the nCoV-S-SSDQ-Ecto-CO, nCoV-S-SQ-Ecto-CO and nCoV-S-WT-Ecto, were established as described above. The expression data for the parental S2 cell lines expressing two codon-optimized S-Ecto constructs and the WT S-Ecto is shown in **FIGURE 6**. Purified nCoV-S-Ecto-Foldon is included for comparison. The expression of the nCoV-S-Ecto products has been confirmed using the conformationally sensitive monoclonal antibody (mAb) CR3022 that recognizes most SARS coronavirus spike proteins. The mAb CR3022 is also used for purification utilizing immunoaffinity chromatography (IAC) methods. The mAb CR3022 based IAC purification methods are based on processes that have been successfully transferred to cGMP manufacturing for other protein subunits produced in S2 cells.

[0176] The use of the codon optimized nCoV-S-SSDQ-Ecto-CO or nCoV-S-SQ-Ecto-CO gene sequence in transformed S2 cells resulted in the expression nCoV-S-Ecto at approximately 5 mg/L, whereas the nCoV-S-WT-Ecto failed to result in detectable expression of the S-Ecto protein in the S2 cells. The lack of expression from the nCoV-S-WT-Ecto and the expression from the two codon optimized sequences demonstrates the benefits of the optimizations to drive improved expression.

EXAMPLE 2**EXPRESSION OF CODON OPTIMIZED SARS-CoV-2 SPIKE S1
SUBUNIT IN *DROSOPHILA* S2 CELLS**

[0177] The SARS-CoV-2 S protein is divided into the S1 and S2 domains as shown in **FIGURE 1**. The S1/S2 junction is delineated by a furin protease cleavage site. In an effort to further truncate the S protein a S1 subunit was designed to further focus the immune response. However, to further enhance the subunit as a vaccine candidate by avoiding sequences that could distract from the most desired epitopes, those that induce nAbs, the S1 subunit was truncated at Gly-594 at the C-terminal end. This is in contrast to truncating the S1 at Pro-681 that immediately precedes the furin cleavage site that defines the S1/S2 junction. The truncation at Gly-594 avoids a region around Asp-614 that has been implicated as a B-cell epitope that induces antibodies that enhance infection in SARS-CoV (Wang et al, 2016). This same location in SARS-CoV-2 has been identified to contain a mutation, Asp 614 to Gly, that results in higher virus titers, but also may have an immunological role (Korber et al, 2020). As with the codon optimized nCoV-S-Ecto subunit, the N-terminus of the S1 subunit is preceded with the SSD amino acids to ensure optimal cleavage during secretion. The defined S1 subunit relative to the full-length S protein can be seen in **FIGURE 3**.

[0178] The codon optimized nucleotide sequence that encodes the C-terminal truncated SARS-CoV-2-S-S1 subunit is detailed in SEQ ID NO:3 and the expressed and stabilized soluble S1 subunit protein is detailed in SEQ ID NO:10.

[0179] Parental S2 cell lines expressing the nCoV-S-S1-CO have been established as described. The expression data for six parental S2 cell lines expressing codon optimized nCoV-S-S1 is shown in **FIGURE 7**. Purified nCoV-S-Ecto-Foldon is included for comparison. The expression of the nCoV-S-S1 product has been confirmed using the conformationally sensitive mAb CR3022. The nCoV-S-S1 protein subunit can also be purified using the mAb CR3022 based IAC methods.

[0180] The use of the codon optimized nCoV-S-S1-CO gene sequence in transformed S2 cells resulted in the expression nCoV-S-S1 at > 50 mg/L. This expression level is approximately 10-fold greater than the expression of the codon optimized nCoV-S-Ecto-CO gene sequence.

[0181] The functionality of nCoV-S-S1 subunit was confirmed in a binding assay with the recombinant hACE2 receptor. The nCoV-S-S1 subunit is bound to the ELISA plate and His-tagged recombinant hACE2 protein is then titrated to determine the binding potential of the

hACE2 to the expressed S1 subunit. The bound hACE2 is detected by alkaline phosphatase conjugated mouse anti-His. The results of the assay are shown in **FIGURE 8**. From this data the functionality of the nCoV-S-S1 in terms of its ability to bind to the hACE2 protein has been confirmed.

EXAMPLE 3

EXPRESSION OF CODON OPTIMIZED SARS-CoV-2 SPIKE RBD SUBUNIT IN *DROSOPHILA* S2 CELLS

[0182] While a main goal is to optimize the expression of the SARS-CoV-2 spike protein as described in Examples 1 and 2, an equally important goal is to define spike protein subunits that can focus the immune response to critical epitopes in terms of defining potential vaccine candidates. For the SARS-CoV-2 spike protein, epitopes eliciting nAb responses are desired. The receptor binding domain (RBD) is a primary target as this domain plays a critical role in viral attachment to cells through the ACE2 receptor. Antibodies that can block the RBD/ACE2 interaction have the potential to block virus infection and thus prevent disease. Therefore, targeting a spike protein RBD subunit is an obvious choice for a vaccine candidate. While the RBD domain has been defined in the context of the full SARS-CoV-2 spike protein, the boundaries of the RBD in terms of the ability to express the domain in various cell-based expression systems is not well-defined.

[0183] In order to stabilize the SARS-CoV-2 spike RBD subunit, the N- and C-terminal ends have been defined to ensure proper folding. Specifically, the N-terminus and C-terminus of the soluble SARS-CoV-2 spike RBD subunit has been defined as Phe-318 and Gly-594, respectively, as shown in **FIGURE 3**. The selection of this segment of the spike protein ensure that 8 cysteine residues are included that form 4 disulfide bonds and that N- and C-terminal lengths are present for stabilization of the RBD domain. As with the optimized nCoV-S-Ecto subunit, the N-terminus of the RBD subunit is preceded with the SSD amino acids to ensure optimal cleavage during secretion. The codon optimized nucleotide sequence that encodes the truncated N- and C-terminal SARS-CoV-2 spike RBD subunit is detailed in SEQ ID NO:4 and the expressed and stabilized soluble spike subunit protein is detailed in SEQ ID NO:11.

[0184] Parental S2 cell lines expressing the nCoV-S-RBD-CO have been established as described. The expression of the nCoV-S-RBD product has been confirmed using the conformationally sensitive mAb CR3022. The expression data for two parental S2 cell lines

expressing codon optimized nCoV-S-RBD is shown in **FIGURE 9**. Purified nCoV-S-RBD-Foldon is included for comparison. The expression of the nCoV-S-S1 product has been confirmed using the conformationally sensitive mAb CR3022. The nCoV-S-RBD protein subunit can also be purified using the mAb CR3022 based IAC methods.

[0185] The use of the codon optimized nCoV-S-RBD-CO gene sequence in transformed S2 cells resulted in the expression of nCoV-S-RBD at approximately 100 mg/L. This expression level is approximately 2-fold greater than the expression of the codon optimized nCoV-S-S1-CO gene sequence and approximately 20-fold greater than the expression of the codon optimized nCoV-S-Ecto-CO gene sequence.

[0186] The functionality of nCoV-S-RBD subunit was confirmed in a binding assay with the recombinant hACE2 receptor. The nCoV-S-RBD subunit is bound to the ELISA plate and His-tagged recombinant hACE2 protein is then titrated to determine the binding potential of the hACE2 to the expressed RBD subunit. The bound hACE2 is detected by alkaline phosphatase conjugated mouse anti-His. The results of the assay are shown in **FIGURE 8**. From this data the functionality of the nCoV-S-RBD in terms of its ability to bind to the hACE2 protein has been confirmed.

EXAMPLE 4

EXPRESSION OF CODON OPTIMIZED SARS-CoV-2 SPIKE ECTODOMAIN AND RBD SUBUNITS WITH FOLDON DOMAINS IN *DROSOPHILA* S2 CELLS

[0187] The surface proteins of enveloped viruses form multimeric configurations. The spike protein on the surface of coronaviruses is defined as a class 1 fusion protein as it forms a homotrimer. These trimeric structures are anchored in the virus membrane shell by the transmembrane (TM) domains. Expression of the spike protein ectodomain (lacking the TM) results in monomers. Various methods are available to drive the formation of multimeric forms of expressed proteins. In general, these methods involve the fusion of protein domains to the protein of interest. These protein domains can promote the formation of multimeric states of the protein. A commonly used protein domain used to promote trimer formation consists of a 29 amino acid sequence that is derived from the bacteriophage T4 fibritin protein sequence. This is referred to as the “foldon” sequence. The foldon sequence is located at the C-terminus of the fibritin protein, and naturally brings together three monomers of fibritin via non-covalent bonding to form a trimeric molecule. The use of the foldon domain provides a means to drive the formation of trimeric molecules when fused to the protein of interest.

[0188] The truncated SARS-CoV-2 spike ectodomain and RBD subunits are further modified at the C-terminus by operably linking a foldon domain to the subunit to promote the trimerization of the expressed subunits, SEQ ID NO:12 and SEQ ID NO:13. The expression of the nCoV-S-Ecto-foldon and nCoV-S RBD-foldon are shown in **FIGURE 10**. The expression of the nCoV-S-Ecto-foldon and nCoV-S RBD-foldon products has been confirmed using the conformationally sensitive mAb CR3022. The nCoV-S-Ecto-foldon and nCoV-S RBD-foldon protein subunits can also be purified using the mAb CR3022 based IAC methods.

[0189] The codon optimized nucleotide sequences that encodes the SARS-CoV-2 spike ectodomain and RBD subunits with the linked foldon domains are detailed in SEQ ID NO:2 and SEQ ID NO:5 and the expressed spike subunit protein with the C-terminal foldon domains are detailed in SEQ ID NO:12 (Ecto-foldon) SEQ ID NO:11 (RBD-Foldon).

[0190] The use of the codon optimized nCoV-S-Ecto-foldon-CO and nCoV-S-RBD-foldon-CO gene sequences in transformed S2 cells resulted in the expression of nCoV-S-Ecto-foldon-CO and nCoV-S-RBD-foldon-CO at approximately 50 mg/mL and 100 mg/L, respectively.

EXAMPLE 5

EXPRESSION OF CODON OPTIMIZED SARS-CoV-2 SPIKE NTD IN *DROSOPHILA* S2 CELLS

[0191] As indicated in Example 3, an important goal is to define spike protein subunits that can focus the immune response to critical epitopes in terms of defining potential vaccine candidates. For the SARS-CoV-2 spike protein, epitopes eliciting nAb responses are desired. While the RBD is a primary target for nAb's, the N-terminal domain (NTD) is also a potential site for nAb epitopes. Therefore, targeting a spike protein NTD subunit is another choice for a potential vaccine candidate.

[0192] The NTD subunit was truncated at Ser-305 for the C-terminal end. As with the optimized nCoV-S-Ecto subunit, the N-terminus of the S1 subunit is preceded with the SSD amino acids to ensure optimal cleavage during secretion. The defined NTD subunit relative to the RBB-Foldon protein is shown in **FIGURE 11**. The NTD subunit was not recognized by mAb CR3022. The codon optimized nucleotide sequence that encodes the SARS-CoV-2-S-NTD subunit is detailed in SEQ ID NO:6 and the expressed and stabilize the soluble NTD subunit protein is detailed in SEQ ID NO:14.

[0193] Initial attempts to express the spike protein subunit NTD resulted in multiple forms of the NTD that were misfolded and were heterogenous in the extent of glycosylation. Expression level in the S2 cells was approximately 20 mg/L.

EXAMPLE 6

IMMUNOGENIC EVALUATION OF CODON OPTIMIZED SARS-CoV-2 SPIKE RBD-FOLDON PROTEINS IN MICE

[0194] The immunogenicity of the *Drosophila* S2 expressed codon optimized nCoV-S-RBD-Foldon subunit was evaluated in both Balb/c mice with 3 different adjuvants to assess immunogenic potential. Mice were immunized intra-muscularly with either two or 3 doses of nCoV-S-RBD-Foldon separated by 3 weeks intervals. A quantity of 5.0 µg of nCoV-S-RBD-Foldon was used for all doses. The 3 adjuvants tested were Alhydrogel, GPI-0100, and SLA-LSQ. Mice were bled 2 weeks after the 2nd or 3rd dose to prepare serum samples for antibody analysis. The design of the immunogenicity study M-000 is presented in **Table 1**.

Table 1. Immunogenicity Study M-000: Codon Optimized nCoV-S-RBD-Foldon

Group	Antigens	SLA-LSQ	GPI-0100	Alhydrogel	# mice
1	5 µg nCoV-RBD-F-2	5 µg / 2 µg	---	---	10
2	5 µg nCoV-RBD-F-2	---	100 µg	---	10
3	5 µg nCoV-RBD-F-2	---	---	120 µg	10

[0195] Five mice were sacrificed and bled two weeks after dose two and 5 mice were sacrificed and bled two weeks after dose three. Serum was then assessed for nCoV-RBD antibody titers by ELISA and for virus nAbs using a MN assay with live SARS-CoV-2. Additionally, serum was tested in an RBD/ACE2 blocking assay.

[0196] The ELISA results for the serum collected following two or three doses of nCoV-RBD-Foldon formulated with the 3 adjuvants are presented in **FIGURE 12**. Serum from individual mice was diluted and were added to a plate coated with nCoV-RBD-Foldon. Bound antibody was detected with goat anti-mouse-AP conjugate. After the addition of substrate, color development was read after 1 hour. The results of the ELISA indicate that the nCoV-RBD-Foldon is immunogenic and the responses are most robust in the GPI-0100 group. The results of the MN assay of the post-dose 2 serum sample is presented in **FIGURE 13**. Based on the 50% neutralization titers (NT – reported as GMT) for each group, the GPI-0100

formulated nCoV-RBD-Foldon results in the strongest response after 2 doses. To further assess the functionality of the serum samples from the nCoV-RBD-Foldon immunized mice, post-dose 2 and post-3 samples were tested for the ability to block the binding of RBD to the hACE2 receptor protein. In this assay, the hACE2 receptor protein, which is His-tagged, is captured on an ELISA plate with anti-His mAb. Dilutions of serum are mixed with a fixed amount of biotinylated nCoV-RBD-Foldon protein, after an incubation period the mixture is added to the plate with the hACE2 protein. Any nCoV-RBD-Foldon that binds to the hACE2 protein is detected with a streptavidin-alkaline phosphatase conjugate. The results of the blocking assay are presented in **FIGURE 14**. The serum from the immunized mice are able to block the binding of biotinylated nCoV-RBD-Foldon protein to varying degrees, with the GPI-0100 group have the greatest blocking capacity. These results correlate with the PRNT results and confirm that the serum produced following immunization with nCoV-RBD-Foldon is able to block the RBD/ACE2 interaction.

EXAMPLE 7

IMMUNOGENIC EVALUATION OF CODON OPTIMIZED SARS-CoV-2 SPIKE

RBD, RBD-FOLDON, S1, AND ECTO-FOLDON PROTEINS IN MICE

[0197] The immunogenicity of the *Drosophila* S2 expressed codon optimized nCoV spike protein subunits, RBD, RBD-Foldon, S1, and Ecto-Foldon, were evaluated in Balb/c mice in 2 different studies. In addition to evaluation of the various nCoV spike protein subunits, varying amounts of proteins, and several different adjuvants were evaluated. Mice were immunized intra-muscularly with 2 doses separated by 3 weeks. A quantity of 10.0 µg or 20.0 µg nCoV spike protein subunits was used. The adjuvants tested were GPI-0100, SLA-LSQ, SLA-SE, QS21, CpG, and QS21+CpG. The design of the immunogenicity study M-001 and M-002 are presented in Table 2 and Table 3.

Table 2. Immunogenicity Study M-001: Codon Optimized nCoV RBD-Foldon, S1, Ecto-Foldon

Group	Antigen	GPI-0100	SLA-LSQ	# mice
1	10 µg nCoV-RBD-F-2	100 µg	---	6
2	20 µg nCoV-RBD-F-2	100 µg	---	6
3	10 µg nCoV-RBD-F-2	200 µg	---	6
4	10 µg nCoV-S1-6	100 µg	---	6
5	12 µg nCoV-S1-6	100 µg	---	5

6	10 µg nCoV-S1-6	200 µg	---	5
7	10 µg nCoV-S1-6	---	5 µg / 2 µg	5
8	10 µg nCoV-Ecto-F-6	100 µg	---	5
9	none	200 µg	---	5

Table 3. Immunogenicity Study M-002: Codon Optimized nCoV RBD, RBD-Foldon, S1, and Ecto-Foldon

Group	Antigens	GPI-0100	QS-21	CpG	SLA-SE	# mice
1	10 µg nCoV-RBD	100 µg	---	---	---	6
2	10 µg nCoV-RBD-F-2	100 µg	---	---	---	6
3	10 µg nCoV-S1-6	100 µg	---	---	---	6
4	20 µg nCoV-RBD	200 µg	---	---	---	6
5	20 µg nCoV-S1-6	200 µg	---	---	---	6
6	10 µg nCoV-S1-6	---	10 µg	---	---	6
7	10 µg nCoV-S1-6	---	10 µg	20 µg	---	6
8	10 µg nCoV-RBD 10 µg nCoV-S1-6	100 µg	---	---	---	6
9	10 µg nCoV-S1-6 10 µg nCoV-RBD	200 µg	---	---	---	6
10	20 µg nCoV-S1-6	---	---	---	20 µg	6
11	none	---	10 µg	20 µg	---	6

[0198] Mice were sacrificed and bled two weeks after the second dose. Serum was then assessed for nCoV-RBD antibody titers by ELISA and for virus nAbs using a MN assay with live SARS-CoV-2. Additionally, serum was tested in an RBD/ACE2 blocking assay.

[0199] For the determination of total IgG, IgG1, and IgG2a by ELISA, serum from individual mice was diluted and added to a plate coated with nCoV-RBD-Foldon. Bound antibody was detected with goat anti-mouse-IgG-AP conjugates. After the addition of substrate, color development was read after 1 hour. For the determination of 50% micro-neutralization titers (MN₅₀), pooled serum for each group was tested in duplicate and the average titer is reported. For the determination of serum ability to block the binding of RBD to the hACE2 receptor protein, serum from individual mice was diluted and mixed with a fixed amount of biotinylated nCoV-RBD-Foldon protein and incubation for a set time, the mixture was then added to a plate with bound hACE2 protein. The amount of bound biotinylated nCoV-RBD-Foldon was detected with a streptavidin-alkaline phosphatase conjugate. The blocking ability of serum was compared to signal generated by the fixed amount of biotinylated nCoV-RBD-Foldon protein with no serum added. The GMT of the serum dilution for each group that

blocks 50% of binding is reported. The results of the blocking assay and MN assay for the serum samples for study M-001 and M-002 are presented in **FIGURE 15A**, **FIGURE 15B**, **FIGURE 16A** and **FIGURE 16B**, respectively. This data indicates that in terms of immunogenic potential, specifically, the ability to elicit high levels of ACE2 blocking Abs and nAbs, the recombinant S protein subunits can be ranked from highest to lowest as follows: S1, Ecto, RBD. In terms of specific subunit/adjuvant combinations and immunogenic potential, the combination of S1/SLA-SE resulted in the strongest immune response.

[0200] The ELISA IgG2a/IgG1 ratio results are reported in **FIGURE 17** for study M-002. These results indicate that only the S1 antigen with the SLA-SE adjuvant, or QS-21/CpG adjuvant combination results in a balanced response as indicated by a IgG2a/IgG1 ratio >1. The IgG2a/IgG1 ratios for select groups from studies M-000, M-001, M-002 are shown in **FIGURE 18**. These results demonstrate the range of responses from Alum at the lowest and SLA-SE at the highest.

[0201] The evaluation of the various nCoV spike protein subunits at different amounts and with a variety of adjuvants all resulted in moderate to robust immune responses after 2 doses. In general, 20 µg of protein resulted in higher responses compared to a similar group with 10 µg of protein. For the various nCoV spike subunit proteins, the responses were highest for S1, followed by Ecto, followed by RBD. The combination of S1 and SLA-SE resulted in the strongest response (Study M-002, group 10). The MN results and the ACE2 blocking results are in good agreement. The combination of S1 and SLA-SE also resulted in the most balanced response as determined by the assessment of a IgG2a/IgG1 ratio.

EXAMPLE 8

PROTECTIVE EFFICACY OF CODON OPTIMIZED SARS-CoV-2 SPIKE RBD AND S1 PROTEINS IN hACE2 TRANSGENIC MICE

[0202] The protective efficacy of *Drosophila* S2 expressed codon optimized nCoV spike protein subunits, RBD and S1 were evaluated in hACE2 transgenic mice (AC70) in 2 different virus challenge studies. In the first study (M-003), RBD and S1 proteins were evaluated. This study utilized 3 adjuvants, GPI-0100, QS21, and QS21 + CpG. The M-003 study design is presented in **Table 4**. In the second study (M-005), the S1 protein was evaluated at 2 amounts, 10 and 20 µg. This study only utilized the adjuvant SLA-SE. The M-005 study design is presented in **Table 5**. Mice were immunized intra-muscularly with 2 doses separated by 3 weeks.

Table 4. Challenge Study M-003: Evaluation of protective efficacy for codon optimized nCoV RBD and S1 with various adjuvants in hACE2 transgenic mice, challenge dose 100xLD₅₀.

Group	Antigens	GPI-0100	QS-21	CpG	# mice
1	10 µg nCoV-RBD	200 µg	---	---	6
2	10 µg nCoV-RBD	---	10 µg	20 µg	6
3	10 µg nCoV-S1-6	---	10 µg	---	6
4	10 µg nCoV-S1-6	---	10 µg	20 µg	6
5	none	---	10 µg	20 µg	6

Table 5. Challenge Study M-005: Evaluation of protective efficacy for codon optimized nCoV S1 with SLA-SE adjuvant in hACE2 transgenic mice, challenge dose 1000xLD₅₀.

Group	Antigens	SLA-SE	# mice
1	10 µg nCoV-S1-6	20 µg	6
2	20 µg nCoV-S1-6	20 µg	6
3	none	---	6

[0203] In both studies the mice were challenged 2 weeks after the second dose with SARS-CoV-2 wild type virus, USA WA1/2020 isolate. In challenge study M-003, mice were challenged with 31xLD₅₀ (100 TCID₅₀). In challenge study M-005, mice were challenged with 1000xLD₅₀ (3.2x10³ TCID₅₀). Following challenge, mice were monitored for signs of disease onset and body weights were recorded. If disease progressed to a defined state, or at a weight loss of greater than 20%, mice were euthanized. The results of percent body weight change and a summary table of results for MN₅₀ titers and percent survival for studies M-003 and M-005 are presented in **FIGURE 19A and FIGURE 19B** and **FIGURE 20A and FIGURE 20B**, respectively.

[0204] In study M-003, both the RBD and the S1 proteins resulted in appreciable MN₅₀ titers regardless of adjuvant used (441 to 1236). The RBD MN₅₀ titer with GPI-0100 is 2-fold greater than the titer with QS21 + CpG, 1050 and 441, respectively. For S1 groups, there is not much difference in the MN₅₀ titer for QS21 alone or QS21+CpG, 981 and 1236, respectively. In terms of percent body weight change, the RBD groups had slight decreases, but recovered, while there was no weight loss in the S1 groups. All 4 groups had 100% survival, while the control group had 33% survival. While the protection in the immunized groups is significant

relative to the control group, the lack of 100% mortality in the control group limits the interpretation of this study.

[0205] In study M-005, the two S1 protein groups resulted in appreciable MN₅₀ titers, with the 10 µg group having a titer greater than 2-fold that of the 20 µg group, 2572 and 981, respectively. These titers suggest that the antigen to adjuvant ratio impacts the titers. To ensure 100% mortality in the control group, the virus challenge was increased to 1000xLD₅₀, which was achieved. Both S1 groups were fully protected from this higher virus challenge. Importantly, the mice in the two S1 groups showed no signs of disease. In the 10 µg group there was no decrease in body weight, whereas in the 20 µg group, there was a slight decrease in body weight though day 2 post-challenge, but recovered after the initial decrease. Furthermore, the MN₅₀ titers post challenge did not increase substantially from the pre-challenge titers, 1.5 and 2.9 fold for the 10 µg and 20 µg groups, respectively. These limited increases in MN₅₀ titers post-challenge indicate that there was limited virus replication post challenge. Overall, study M-005 demonstrates that the combination of the nCoV spike S1 subunit protein in combination with the SLA-SE adjuvant provides solid protection against challenge with a lethal dose of SARS-CoV-2.

[0206] Although the invention has been described with reference to the above examples, it will be understood that modifications and variations are encompassed within the spirit and scope of the invention. Accordingly, the invention is limited only by the following claims.

What is claimed is:

1. An isolated nucleic acid sequence selected from the group consisting of SEQ ID NO: 1, 2, 3, 4, 5 and 6.
2. An isolated amino acid sequence encoded by a nucleic acid sequence of claim 1.
3. The isolated amino acid sequence of claim 2, wherein the sequence comprises SEQ ID NO: 9, 10, 11, 12, 13 or 14.
4. An expression vector comprising a nucleic acid sequence encoding a SARS-CoV-2 spike (S) protein, wherein the nucleic acid sequence comprises SEQ ID NO: 1, 2, 3, 4, 5 or 6.
5. The vector of claim 4, wherein the vector is a *Drosophila melanogaster* expression vector.
6. The vector of claim 4, wherein the vector has a nucleic acid sequence comprising SEQ ID NO: 7.
7. The vector of claim 4, wherein the SARS-CoV-2 S protein has an amino acid sequence as set forth in SEQ ID NO: 9, 10, 11, 12, 13 or 14.
8. A method of producing a protein *in vitro* comprising *Drosophila melanogaster* cells transformed with the vector of claim 4 and culturing the cells under conditions to produce the protein.
9. The method of claim 8, wherein the *Drosophila melanogaster* cells are Schneider 2 (S2) cells.
10. A vaccine composition comprising:
 - (a) an effective amount of a SARS-CoV-2 spike (S) protein, wherein the S protein is encoded by a nucleic acid sequence as set forth in SEQ ID NO: 1, 2, 3, 4, 5 or 6, and
 - (b) an effective amount of an adjuvant.
11. The vaccine of claim 10, wherein the S protein is the S1 subunit protein.
12. The vaccine of claim 11, wherein the S1 subunit protein is encoded by a nucleic acid sequence as set forth in SEQ ID NO: 3.
13. The vaccine of claim 10, wherein the adjuvant is selected from the group consisting of GPI-0100, synthetic lipid A (SLA) in a stable oil-in water emulsion (SE) (SLA-SE), QS21, QS21 combined with SLA to form a liposome formulation (SLA-LSQ), and QS21+CpG.
14. The vaccine of claim 11, wherein the adjuvant is SLA-SE.

15. The vaccine of claim 10, wherein the S protein is a S1 subunit protein encoded by a nucleic acid sequence as set forth in SEQ ID NO: 3, and the adjuvant is SLA-SE.
16. The vaccine of claim 10, wherein the S protein is recombinantly produced and expressed in insect host cells.
17. The vaccine of claim 10, wherein the S protein is recombinantly produced and expressed in *Drosophila melanogaster* Schneider 2 (S2) cells.
18. The vaccine of claim 10, further comprising a pharmaceutically acceptable excipient or carrier.
19. A method of preventing a SARS-CoV-2 entry into a cell comprising contacting the cell with an effective amount of the vaccine of claim 10, thereby preventing SARS-CoV-2 entry into the cell.
20. A method of stimulating a protective immune response in a subject comprising administering to the subject an effective amount of the vaccine of claim 10, thereby stimulating a protective immune response.
21. The method of claim 17, wherein the immune response is a balanced immune response.
22. The method of claim 18, wherein the balanced immune response is characterized by a IgG2a:IgG1 ratio that is equal or greater than 1.
23. A method of inhibiting a SARS-CoV-2 infection in a subject comprising administering to the subject a therapeutically or prophylactically effective amount of the vaccine of claim 10, thereby inhibiting SARS-CoV-2 infection.
24. A method of inhibiting transmission of a SARS-CoV-2 infection by a subject comprising administering to the subject an effective amount of the vaccine of claim 10, thereby inhibiting transmission of a SARS-CoV-2 infection.
25. The method of any one of claims 20 and 23-24, wherein the S protein is a S1 subunit protein.
26. The method of claim 25, wherein the S1 subunit protein is encoded by a nucleic acid sequence comprising SEQ ID NO: 3.
27. The method of any one of claims 20 and 23-24, wherein the adjuvant is selected from the group consisting of GPI-0100, QS21 + CpG, SLA-SE, and SLA-LSQ.
28. The method of claim 27, wherein the adjuvant is SLA-SE.

- 29.** The method of any one of claims 20 and 23-24, wherein the S protein is a S1 subunit protein encoded by a nucleic acid sequence set forth in SEQ ID NO: 3 and the adjuvant is SLA-SE
- 30.** The method of any one of claims 20 and 23-24, wherein the vaccine induces the production of neutralizing antibodies in the subject.
- 31.** The method of claim 30, wherein the neutralizing antibodies prevent the binding of a SARS-CoV-2 to a target cell and/or target receptor.
- 32.** The method of claim 31, wherein the target receptor is an ACE2 receptor.
- 33.** The method of any one of claims 20 and 23-24, wherein administering comprises injecting two doses to the subject at a 3-weeks interval.
- 34.** The method of claim 33, wherein administering comprises injecting intramuscularly.
- 35.** The method of claim 33, wherein a dose comprises about 0.5-50 μ g of purified S protein.
- 36.** The method of any one of claims 20 and 23-24, wherein administering the vaccine to the subject increases the subject survival.
- 37.** The method of any one of claims 20 and 23-24, wherein administering the vaccine prevents the development of COVID-19 disease in the subject.

FIGURE 1

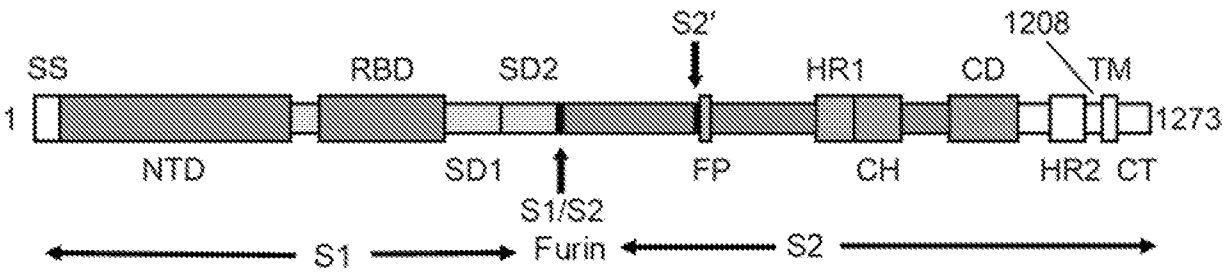
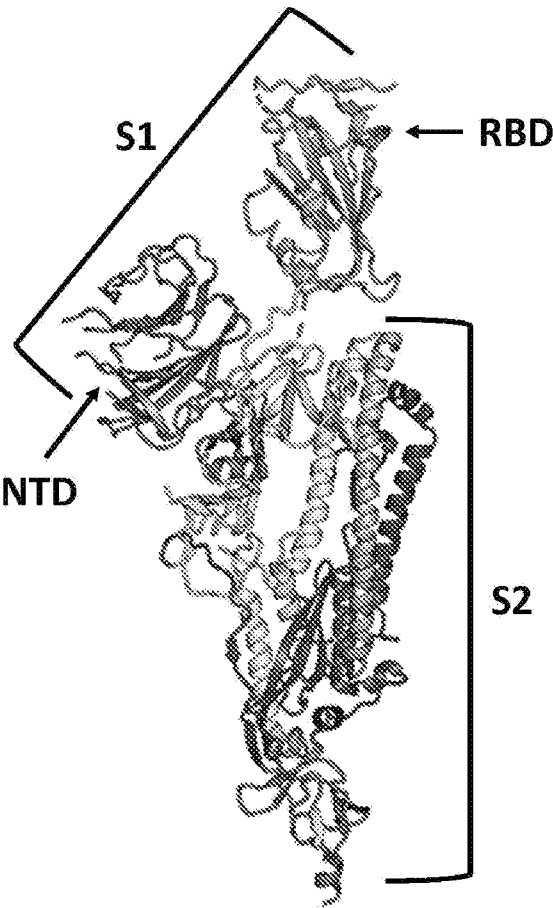


FIGURE 2



2/17

FIGURE 3

	10	20	30	40	50	60	
1	-----+-----+-----+-----+-----+-----+						
	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS						2019 nCoV S
	<u>SSD</u> QCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS						SEQ ID NO 10
	<u>SSD</u> QCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS						SEQ ID NO 14
	<u>SSD</u> QCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS						SEQ ID NO 9
	<u>SSD</u> QCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS						SEQ ID NO 12
	70	80	90	100	110	120	
61	-----+-----+-----+-----+-----+-----+						
	NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLIV						2019 nCoV S
	NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLIV						SEQ ID NO 10
	NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLIV						SEQ ID NO 14
	NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLIV						SEQ ID NO 9
	NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLIV						SEQ ID NO 12
	130	140	150	160	170	180	
121	-----+-----+-----+-----+-----+-----+						
	NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMDLE						2019 nCoV S
	NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMDLE						SEQ ID NO 10
	NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMDLE						SEQ ID NO 14
	NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMDLE						SEQ ID NO 9
	NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMDLE						SEQ ID NO 12
	190	200	210	220	230	240	
181	-----+-----+-----+-----+-----+-----+						
	GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT						2019 nCoV S
	GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT						SEQ ID NO 10
	GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT						SEQ ID NO 14
	GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT						SEQ ID NO 9
	GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT						SEQ ID NO 12
	250	260	270	280	290	300	
241	-----+-----+-----+-----+-----+-----+						
	LLALHRSYLTPGDSSSGWTAGAAAYVGYLQPRTFLLKYNENGITITDAVDCALDPLSETK						2019 nCoV S
	LLALHRSYLTPGDSSSGWTAGAAAYVGYLQPRTFLLKYNENGITITDAVDCALDPLSETK						SEQ ID NO 10
	LLALHRSYLTPGDSSSGWTAGAAAYVGYLQPRTFLLKYNENGITITDAVDCALDPLSETK						SEQ ID NO 14
	LLALHRSYLTPGDSSSGWTAGAAAYVGYLQPRTFLLKYNENGITITDAVDCALDPLSETK						SEQ ID NO 9
	LLALHRSYLTPGDSSSGWTAGAAAYVGYLQPRTFLLKYNENGITITDAVDCALDPLSETK						SEQ ID NO 12
	310	320	330	340	350	360	
301	-----+-----+-----+-----+-----+-----+						
	CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRRKRISN						2019 nCoV S
	CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRRKRISN						SEQ ID NO 10
	CTLKS						SEQ ID NO 14
	CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRRKRISN						SEQ ID NO 9
	CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRRKRISN						SEQ ID NO 12
	<u>SSD</u> FRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRRKRISN						SEQ ID NO 11
	<u>SSD</u> FRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRRKRISN						SEQ ID NO 13

FIGURE 3 (cont.)

	370	380	390	400	410	420	
	-----+	-----+	-----+	-----+	-----+	-----+	
361	CVADYSVLYNSASFSTFKCYGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIAD						2019 nCoV S
	CVADYSVLYNSASFSTFKCYGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIAD						SEQ ID NO 10
	CVADYSVLYNSASFSTFKCYGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIAD						SEQ ID NO 9
	CVADYSVLYNSASFSTFKCYGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIAD						SEQ ID NO 12
	CVADYSVLYNSASFSTFKCYGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIAD						SEQ ID NO 11
	CVADYSVLYNSASFSTFKCYGVSP TKLNDLCFTNVYADSFVIRGDEV RQIAPGQTGKIAD						SEQ ID NO 13
	430	440	450	460	470	480	
	-----+	-----+	-----+	-----+	-----+	-----+	
421	YNYKLPDDFTGCVIAWNSNNLDSKVGGN YNYLYRLFRKSNLKPFFERDISTEIQAGSTPC						2019 nCoV S
	YNYKLPDDFTGCVIAWNSNNLDSKVGGN YNYLYRLFRKSNLKPFFERDISTEIQAGSTPC						SEQ ID NO 10
	YNYKLPDDFTGCVIAWNSNNLDSKVGGN YNYLYRLFRKSNLKPFFERDISTEIQAGSTPC						SEQ ID NO 9
	YNYKLPDDFTGCVIAWNSNNLDSKVGGN YNYLYRLFRKSNLKPFFERDISTEIQAGSTPC						SEQ ID NO 12
	YNYKLPDDFTGCVIAWNSNNLDSKVGGN YNYLYRLFRKSNLKPFFERDISTEIQAGSTPC						SEQ ID NO 11
	YNYKLPDDFTGCVIAWNSNNLDSKVGGN YNYLYRLFRKSNLKPFFERDISTEIQAGSTPC						SEQ ID NO 13
	490	500	510	520	530	540	
	-----+	-----+	-----+	-----+	-----+	-----+	
481	NGVEGFNCYFPLQSYGFQPTNGVG YQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVN						2019 nCoV S
	NGVEGFNCYFPLQSYGFQPTNGVG YQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVN						SEQ ID NO 10
	NGVEGFNCYFPLQSYGFQPTNGVG YQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVN						SEQ ID NO 9
	NGVEGFNCYFPLQSYGFQPTNGVG YQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVN						SEQ ID NO 12
	NGVEGFNCYFPLQSYGFQPTNGVG YQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVN						SEQ ID NO 11
	NGVEGFNCYFPLQSYGFQPTNGVG YQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVN						SEQ ID NO 13
	550	560	570	580	590	600	
	-----+	-----+	-----+	-----+	-----+	-----+	
541	FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVS VITP						2019 nCoV S
	FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGG						SEQ ID NO 10
	FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVS VITP						SEQ ID NO 9
	FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVS VITP						SEQ ID NO 12
	FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGG						SEQ ID NO 11
	FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGG <u>GGSGGP</u>						SEQ ID NO 13
	610	620	630	640	650	660	
	-----+	-----+	-----+	-----+	-----+	-----+	
601	GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGA EHVNN SY						2019 nCoV S
	GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGA EHVNN SY						SEQ ID NO 9
	GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGA EHVNN SY						SEQ ID NO 12
	<u>GGGYIPEAPRDGQAYVRKDGEWVLLSTFL</u>						SEQ ID NO 13

FIGURE 3 (cont.)

	670	680	690	700	710	720	
761	-----+-----+-----+-----+-----+-----+	ECDIPIGAGICASYQTQTNSP RRAR SVASQSIIAYTMSLGAENSVAYSNNNSIAIPTNFTI 2019 nCoV S					
	-----+-----+-----+-----+-----+-----+	ECDIPIGAGICASYQTQTNSP GS ARSVASQSIIAYTMSLGAENSVAYSNNNSIAIPTNFTI SEQ ID NO 9					
	-----+-----+-----+-----+-----+-----+	ECDIPIGAGICASYQTQTNSP GS ARSVASQSIIAYTMSLGAENSVAYSNNNSIAIPTNFTI SEQ ID NO 12					
	730	740	750	760	770	780	
721	-----+-----+-----+-----+-----+-----+	SVTTEILPVSMTKTSVDCTMYICGDSTEC SNLLLQYGSFCTQLNRALTGIAVEQDKNTQE 2019 nCoV S					
	-----+-----+-----+-----+-----+-----+	SVTTEILPVSMTKTSVDCTMYICGDSTEC SNLLLQYGSFCTQLNRALTGIAVEQDKNTQE SEQ ID NO 9					
	-----+-----+-----+-----+-----+-----+	SVTTEILPVSMTKTSVDCTMYICGDSTEC SNLLLQYGSFCTQLNRALTGIAVEQDKNTQE SEQ ID NO 12					
	790	800	810	820	830	840	
781	-----+-----+-----+-----+-----+-----+	VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLDAGFIKQYGDC 2019 nCoV S					
	-----+-----+-----+-----+-----+-----+	VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLDAGFIKQYGDC SEQ ID NO 9					
	-----+-----+-----+-----+-----+-----+	VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLDAGFIKQYGDC SEQ ID NO 12					
	850	860	870	880	890	900	
841	-----+-----+-----+-----+-----+-----+	LGDIAARDLICAQKFENGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM 2019 nCoV S					
	-----+-----+-----+-----+-----+-----+	LGDIAARDLICAQKFENGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM SEQ ID NO 9					
	-----+-----+-----+-----+-----+-----+	LGDIAARDLICAQKFENGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM SEQ ID NO 12					
	910	920	930	940	950	960	
901	-----+-----+-----+-----+-----+-----+	QMAYRFENGIGVTONVLYENQKLIANQFN SAIGKIQDSLSTASALGKLQDVVNQNAQALN 2019 nCoV S					
	-----+-----+-----+-----+-----+-----+	QMAYRFENGIGVTONVLYENQKLIANQFN SAIGKIQDSLSTASALGKLQDVVNQNAQALN SEQ ID NO 9					
	-----+-----+-----+-----+-----+-----+	QMAYRFENGIGVTONVLYENQKLIANQFN SAIGKIQDSLSTASALGKLQDVVNQNAQALN SEQ ID NO 12					
	970	980	990	1000	1010	1020	
961	-----+-----+-----+-----+-----+-----+	TLVKQLSSNFGAISSVLNDILSRDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA 2019 nCoV S					
	-----+-----+-----+-----+-----+-----+	TLVKQLSSNFGAISSVLNDILSRDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA SEQ ID NO 9					
	-----+-----+-----+-----+-----+-----+	TLVKQLSSNFGAISSVLNDILSRDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA SEQ ID NO 12					
	1030	1040	1050	1060	1070	1080	
1021	-----+-----+-----+-----+-----+-----+	SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA 2019 nCoV S					
	-----+-----+-----+-----+-----+-----+	SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA SEQ ID NO 9					
	-----+-----+-----+-----+-----+-----+	SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA SEQ ID NO 12					
	1090	1100	1110	1120	1130	1140	
1081	-----+-----+-----+-----+-----+-----+	ICHHDGKAHFPREGVFVSNNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP 2019 nCoV S					
	-----+-----+-----+-----+-----+-----+	ICHHDGKAHFPREGVFVSNNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP SEQ ID NO 9					
	-----+-----+-----+-----+-----+-----+	ICHHDGKAHFPREGVFVSNNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP SEQ ID NO 12					

5/17

FIGURE 3 (cont.)

```

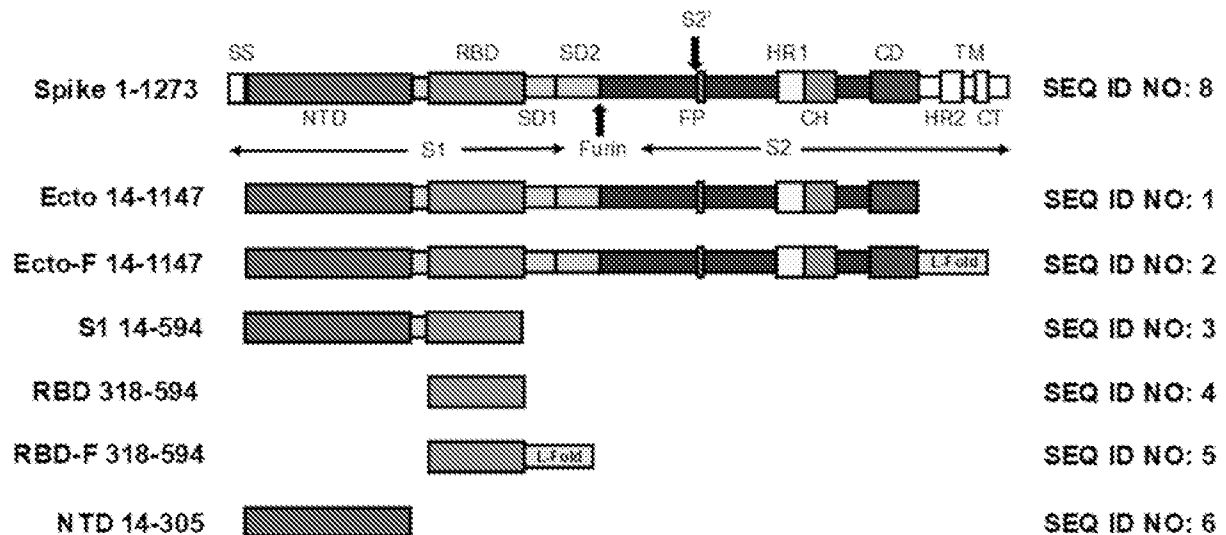
      1150      1160      1170      1180      1190      1200
      -----+-----+-----+-----+-----+-----+
1141 LQPELDSFKEELDKYFKNHTSPDVLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL 2019 nCoV S
      LQPELDS                                           SEQ ID NO 9
      LQPELDSGGSGGPGSGYIPEAPRDGQAYVRKDGEWVLLSTFL      SEQ ID NO 12

      1210      1220      1230      1240      1250      1260
      -----+-----+-----+-----+-----+-----+
1201 QELGKYEQYIKWPWYIWLGFIAGLIAIVMTIMLCCMTSCCSCCLKGCCSCGSCCKFDEDD 2019 nCoV S

      1270
1261 -----+---
      SEPVLKGVKLHYT 2019 nCoV S

```

FIGURE 4



6/17

FIGURE 5

atgtttgtttttcttggttttattgccactagtctctagtcagtggtgtaatcttacaacc	60
M F V F L V L L P L V S S Q C V N L T T	20
agaactcaattacccctgcatacactaattctttcacacgtggtggttattaccctgac	120
R T Q L P P A Y T N S F T R G V Y Y P D	40
aaagttttcagatcctcagttttacattcaactcaggacttggtcttacctttcttttcc	180
K V F R S S V L H S T Q D L F L P F F S	60
aatgttacttggttccatgctatacatgtctctgggaccaatggtactaagaggtttgat	240
N V T W F H A I H V S G T N G T K R F D	80
aaccctgtcctaccattttaatgatggtggttattttgcttccactgagaagtctaacata	300
N P V L P F N D G V Y F A S T E K S N I	100
ataagaggctggatttttgggtactacttttagattcgaagaccagtcctacttattggt	360
I R G W I F G T T L D S K T Q S L L I V	120
aataacgctactaatgttggttatttaaagtctgtgaatttcaattttgtaatgatccattt	420
N N A T N V V I K V C E F Q F C N D P F	140
ttgggtggttattaccacaaaaacaacaaagttggatggaaagtgagttcagagtttat	480
L G V Y Y H K N N K S W M E S E F R V Y	160
tctagtgcgaataattgcacttttgaatatgtctctcagccttttcttatggaccttgaa	540
S S A N N C T F E Y V S Q P F L M D L E	180
ggaaaacagggttaatttcaaaaatcttaggggaatttggtggttaagaatattgatgggttat	600
G K Q G N F K N L R E F V F K N I D G Y	200
tttaaaatatattctaagcacacgcctattaatttagtgcgatctccctcagggtttt	660
F K I Y S K H T P I N L V R D L P Q G F	220
tcggcttttagaaccattggtagatttgccaataggtattaacatcactagggtttcaaact	720
S A L E P L V D L P I G I N I T R F Q T	240
ttacttgctttacatagaagttatttgactcctggtgattcttcttcagggttgacagct	780
L L A L H R S Y L T P G D S S S G W T A	260
ggtgctgcagcttattatgtgggttatcttcaacctaggacttttctattaaaatataat	840
G A A A Y Y V G Y L Q P R T F L L K Y N	280
gaaaatggaaccattacagatgctgtagactgtgcacttgaccctctctcagaaacaaag	900
E N G T I T D A V D C A L D P L S E T K	300
tgtacgttgaaatccttcactgtagaaaaaggaatctatcaaacttctaacttttagagtc	960
C T L K S F T V E K G I Y Q T S N F R V	320
caaccaacagaatctattgttagatttccataatattacaaacttggtgcccttttgggtgaa	1020
Q P T E S I V R F P N I T N L C P F G E	340

7/17

FIGURE 5 (cont.)

gtttttaacgccaccagatttgcacatctgtttatgcttggaacaggaagagaatcagcaac	1080
V F N A T R F A S V Y A W N R K R I S N	360
tgtgttgctgattattctgtcctatataaattccgcatcattttccacttttaagtgttat	1140
C V A D Y S V L Y N S A S F S T F K C Y	380
ggagtgtctcctactaaattaaatgatctctgctttactaatgtctatgcagattcattt	1200
G V S P T K L N D L C F T N V Y A D S F	400
gtaattagagggtgatgaagtcagacaaaatcgctccaggggcaaactggaaagattgctgat	1260
V I R G D E V R Q I A P G Q T G K I A D	420
tataattataaattaccagatgattttacaggctgcgttatagcttggaattctaacaat	1320
Y N Y K L P D D F T G C V I A W N S N N	440
cttgattctaagggttggtggttaattataattacctgtatagattgttttaggaagtcta	1380
L D S K V G G N Y N Y L Y R L F R K S N	460
ctcaaacccttttgagagagatatattcaactgaaatctatcaggccggtagcacaccttgt	1440
L K P F E R D I S T E I Y Q A G S T P C	480
aatggtggtgaagggttttaattgttactttcctttacaatcatatggtttccaaccact	1500
N G V E G F N C Y F P L Q S Y G F Q P T	500
aatggtggttggttaccaaccatacagagtagtagtactttcctttgaacttctacatgca	1560
N G V G Y Q P Y R V V V L S F E L L H A	520
ccagcaactgtttgtggacctaataaagtctactaatttggttaaaaacaaatgtgtcaat	1620
P A T V C G P K K S T N L V K N K C V N	540
ttcaacttcaatgggtttaacaggcacagggtgttcttactgagtcatacaaaaagtgttctg	1680
F N F N G L T G T G V L T E S N K K F L	560
cctttccaacaatttggcagagacattgctgacactactgatgctgtccgtgatccacag	1740
P F Q Q F G R D I A D T T D A V R D P Q	580
acacttgagattccttgacattacaccatgttcttttggtggtgtcagtgttataacacca	1800
T L E I L D I T P C S F G G V S V I T P	600
ggaacaaatacttctaaccagggttgctgttctttatcaggatgttaactgcacagaagtc	1860
G T N T S N Q V A V L Y Q D V N C T E V	620
cctgttgctattcatgcagatcaacttactcctacttggcgtgtttattctacaggttct	1920
P V A I H A D Q L T P T W R V Y S T G S	640
aatgtttttcaaacacgtgcaggctgtttaataggggctgaacatgtcaacaactcatat	1980
N V F Q T R A G C L I G A E H V N N S Y	660
gagtgtgacatacccatgggtgcagggtatatgcgctagttatcagactcagactaattct	2040
E C D I P I G A G I C A S Y Q T Q T N S	680
cctcggcgggcacgtagtgtagctagtcattccatcattgcctacactatgtcacttggt	2100
P R R A R S V A S Q S I I A Y T M S L G	700

FIGURE 5 (cont.)

```

gcagaaaattcagttgcttactctaataactctattgccataccacaaaattttactatt 2160
A E N S V A Y S N N S I A I P T N F T I 720
agtgttaccacagaaattctaccagtggtctatgaccaagacatcagtagattgtacaatg 2220
S V T T E I L P V S M T K T S V D C T M 740
tacatttgtggtgattcaactgaatgcagcaatcttttgttgcaatatggcagtttttgt 2280
Y I C G D S T E C S N L L L Q Y G S F C 760
acacaattaaaccgtgctttaactggaatagctggtgaacaagacaaaaacaccaagaa 2340
T Q L N R A L T G I A V E Q D K N T Q E 780
gtttttgcacaagtcaaacaattttacaaaacaccaccaattaaagattttggtggtttt 2400
V F A Q V K Q I Y K T P P I K D F G G F 800
aatttttcacaaatattaccagatccatcaaaaccaagcaagaggtcattttattgaagat 2460
N F S Q I L P D P S K P S K R S F I E D 820
ctacttttcaacaaagtgcacttgcagatgctggcttcatcaaacaatatggtgattgc 2520
L L F N K V T L A D A G F I K Q Y G D C 840
cttggtgatattgctgctagagacctcatttgtgcacaaaagttaaacggccttactggt 2580
L G D I A A R D L I C A Q K F N G L T V 860
ttgccacctttgctcacagatgaaatgattgctcaatacacttctgcactgttagcgggt 2640
L P P L L T D E M I A Q Y T S A L L A G 880
acaatcacttctggttggacctttggtgcaggtgctgcattacaaataccatttgcctatg 2700
T I T S G W T F G A G A A L Q I P F A M 900
caaatggcttatagggttaatgggtattggagttacacagaatgttctctatgagaaccaa 2760
Q M A Y R F N G I G V T Q N V L Y E N Q 920
aaattgattgccaaccaatttaatagtgcatttggcaaaattcaagactcactttcttcc 2820
K L I A N Q F N S A I G K I Q D S L S S 940
acagcaagtgcacttggaaaacttcaagatgtggtcaacccaaatgcacaagctttaaac 2880
T A S A L G K L Q D V V N Q N A Q A L N 960
acgcttggttaacaacttagctccaattttggtgcaatttcaagtgttttaaatgatatc 2940
T L V K Q L S S N F G A I S S V L N D I 980
ctttcacgtcttgacaaagttgaggctgaagtgc aaattgatagggttgatcacaggcaga 3000
L S R L D K V E A E V Q I D R L I T G R 1000
cttcaaagtttgcagacatatgtgactcaacaattaattagagctgcagaaatcagagct 3060
L Q S L Q T Y V T Q Q L I R A A E I R A 1020
tctgctaattcttgctgctactaaaatgtcagagtggtgtacttggacaatcaaaaagagtt 3120
S A N L A A T K M S E C V L G Q S K R V 1040
gatttttgtggaaagggctatcatcttatgtccttccctcagtcagcacctcatggtgta 3180
D F C G K G Y H L M S F P Q S A P H G V 1060

```

9/17

FIGURE 5 (cont.)

gtcttcttgcatgtgacttatgtccctgcacaagaaaagaacttcacaactgctcctgcc	3240
V F L H V T Y V P A Q E K N F T T A P A	1080
atttgtcatgatggaaaagcacacttttcctcgtgaagggtgtctttgtttcaaattggcaca	3300
I C H D G K A H F P R E G V F V S N G T	1100
cactgggtttgtaacacaaaaggaatttttatgaaccacaaatcattactacagacaacaca	3360
H W F V T Q R N F Y E P Q I I T T D N T	1120
tttgtgtctggtaactgtgatgttgtaataggaattgtcaacaacacagtttatgatcct	3420
F V S G N C D V V I G I V N N T V Y D P	1140
ttgcaacctgaattagactcattcaaggaggagttagataaatattttaagaatcataca	3480
L Q P E L D S F K E E L D K Y F K N H T	1160
tcaccagatgttgatttaggtgacatctctggcattaatgcttcagttgtaaaccattcaa	3540
S P D V D L G D I S G I N A S V V N I Q	1180
aaagaaattgaccgcctcaatgaggttgccaagaatttaaatgaatctctcatcgatctc	3600
K E I D R L N E V A K N L N E S L I D L	1200
caagaacttggaagtatgagcagtatataaaatggccatggtacatttggttaggtttt	3660
Q E L G K Y E Q Y I K W P W Y I W L G F	1220
atagctggcttgattgccatagtaatggtgacaattatgctttgctgtatgaccagttgc	3720
I A G L I A I V M V T I M L C C M T S C	1240
tgtagttgtctcaagggctgttggttcttggtggtatcctgctgcaaatttgatgaagacgac	3780
C S C L K G C C S C G S C C K F D E D D	1260
tctgagccagtgctcaaaggagtc aaattacattacacataa	3822
S E P V L K G V K L H Y T -	1273

10/17

FIGURE 6

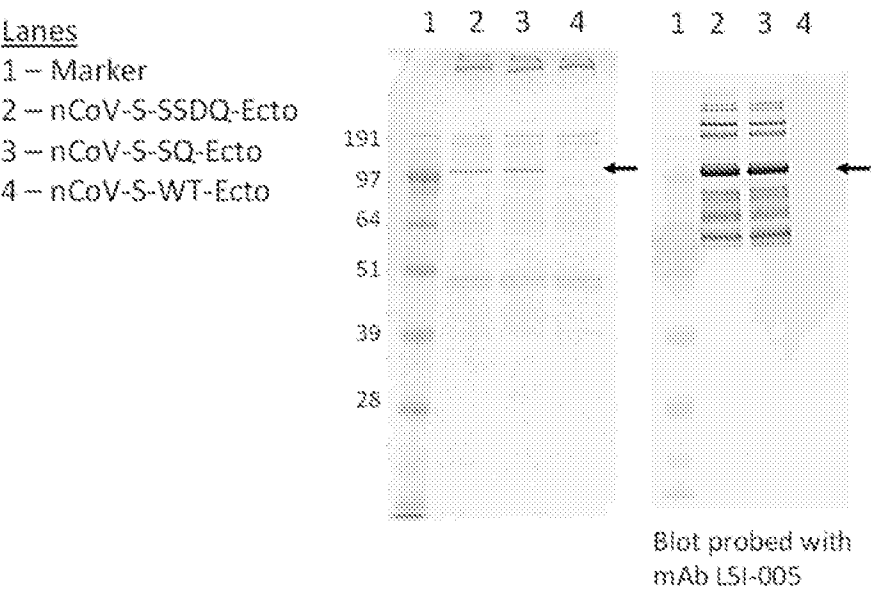


FIGURE 7

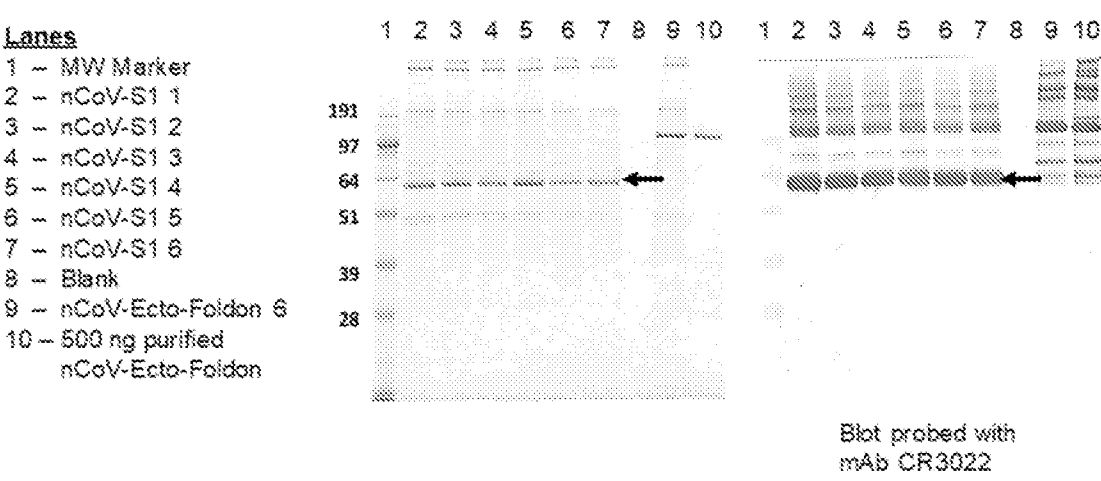


FIGURE 8

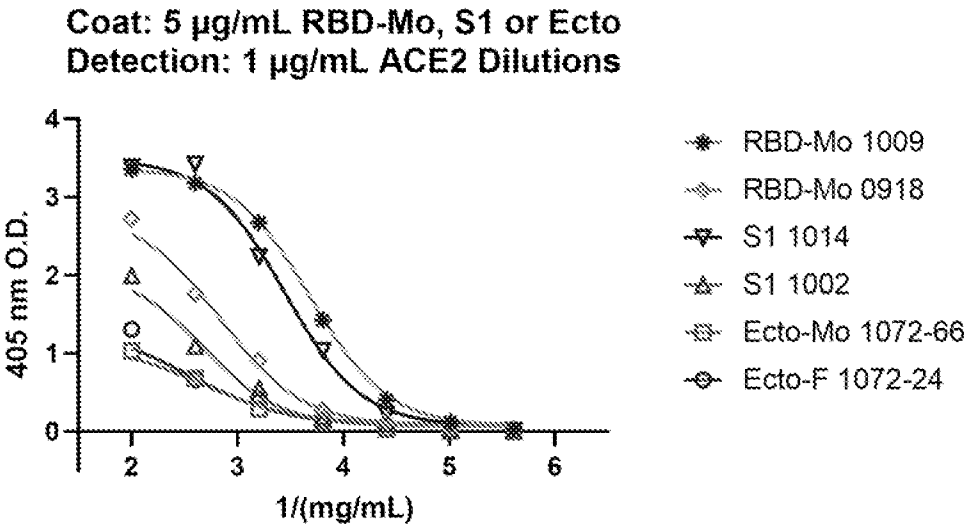


FIGURE 9

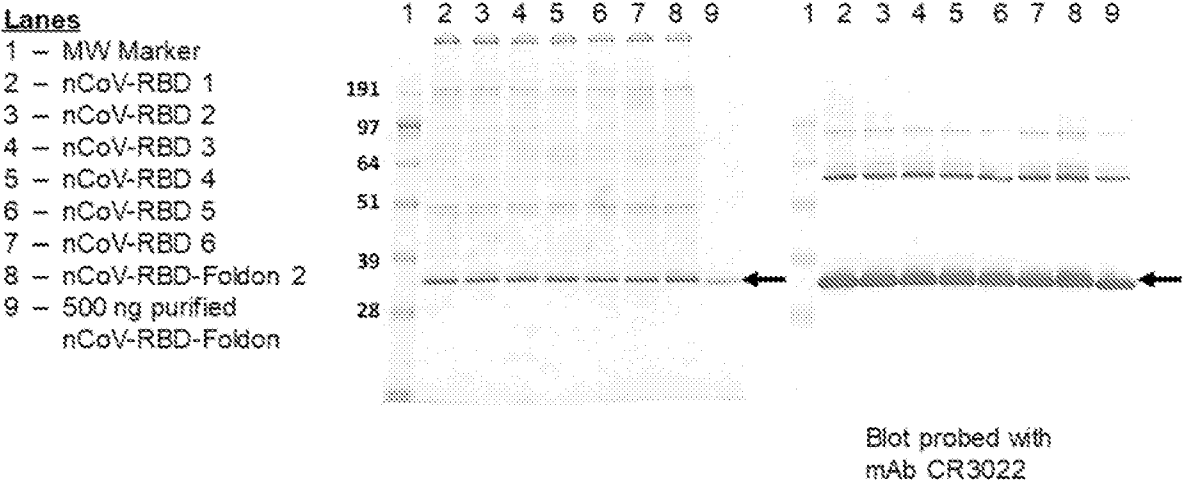


FIGURE 10

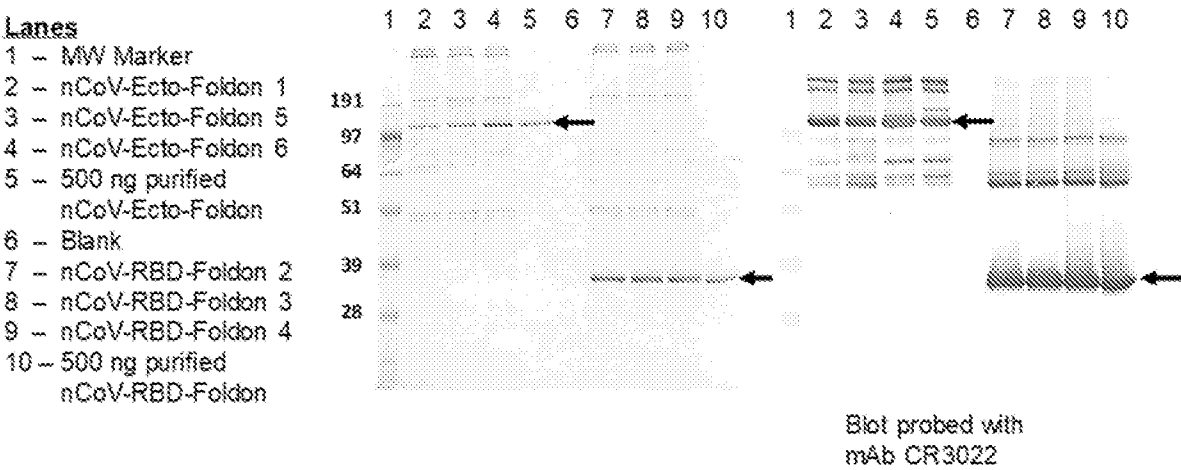


FIGURE 11

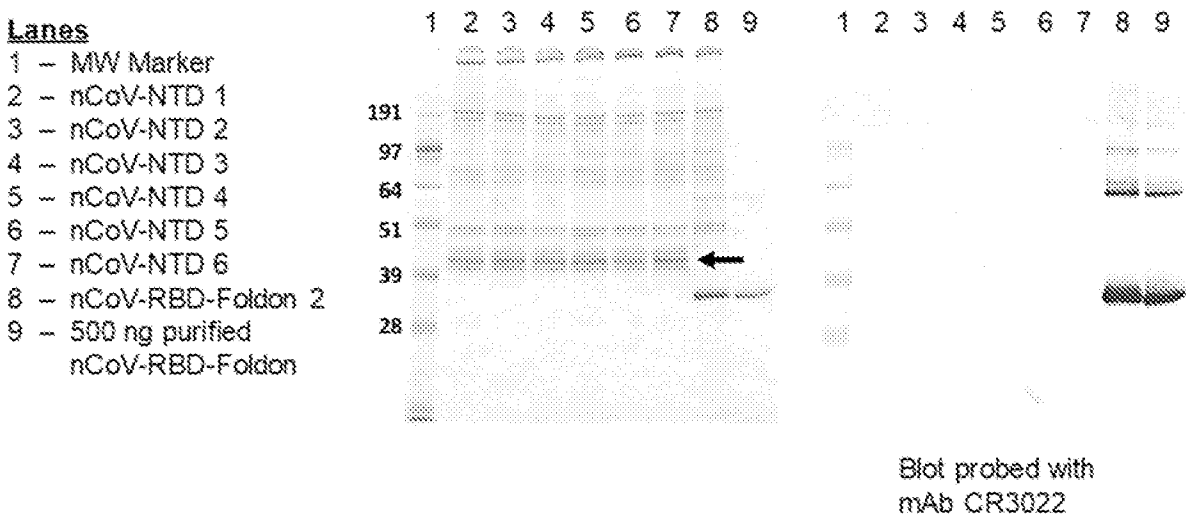


FIGURE 12

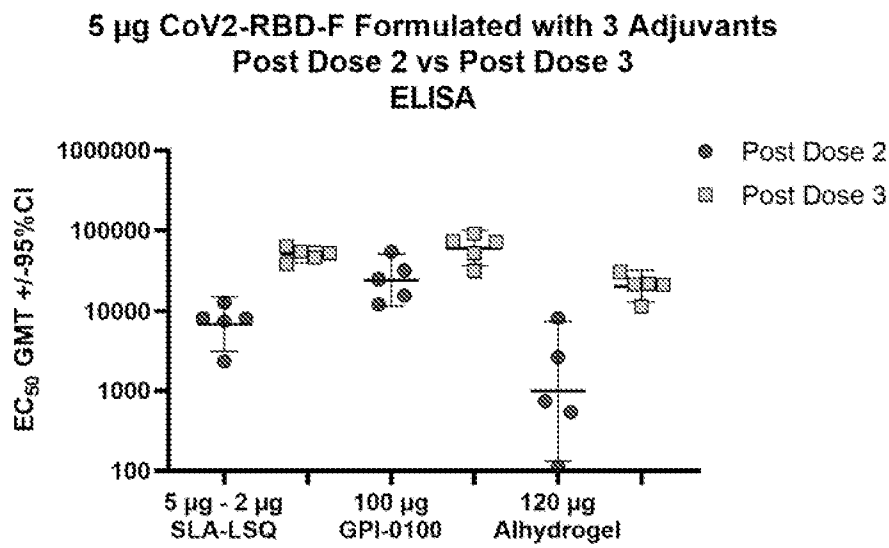


FIGURE 13

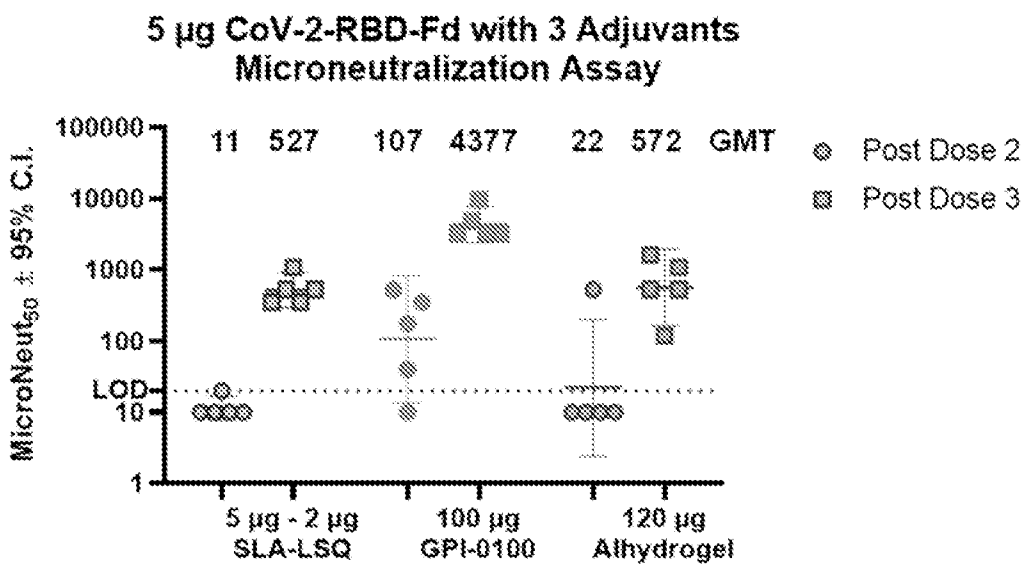


FIGURE 14

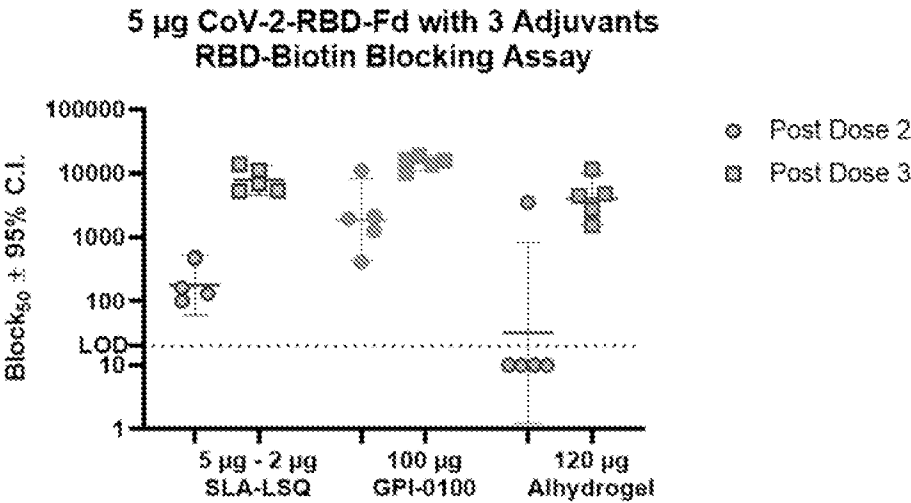


FIGURE 15A

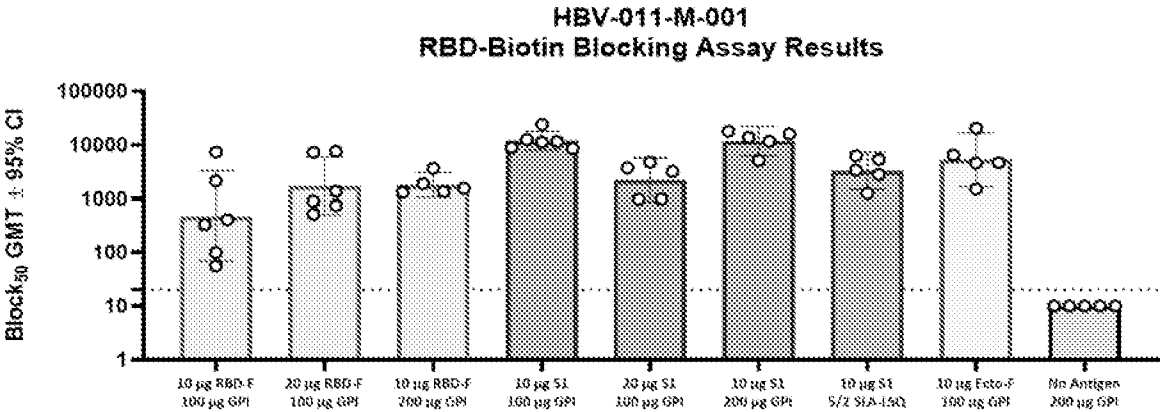


FIGURE 15B

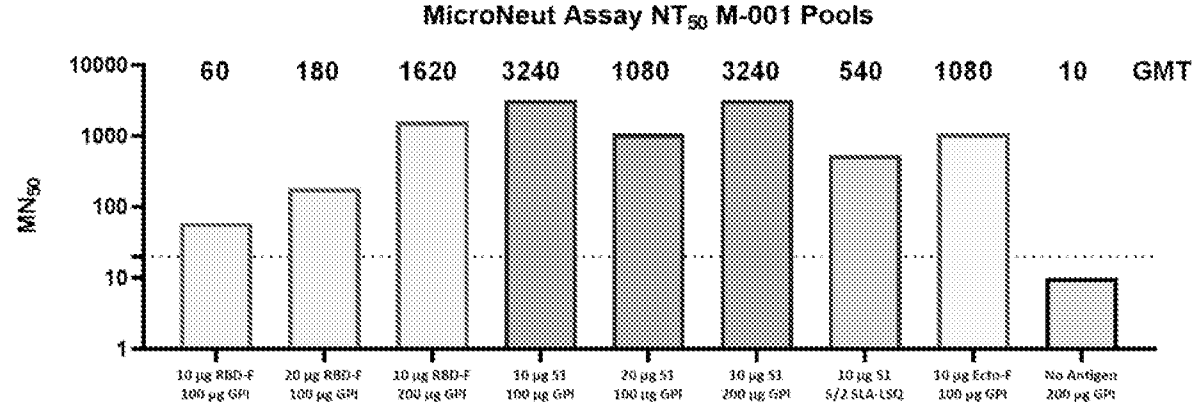


FIGURE 16A

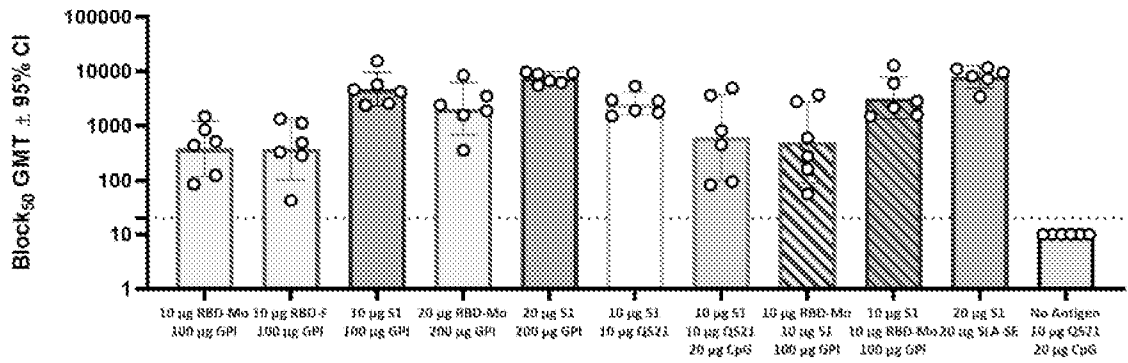


FIGURE 16B

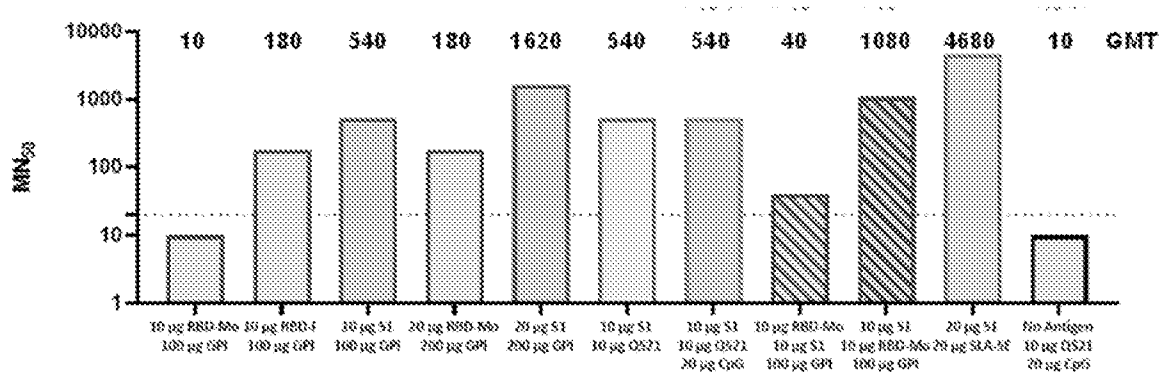
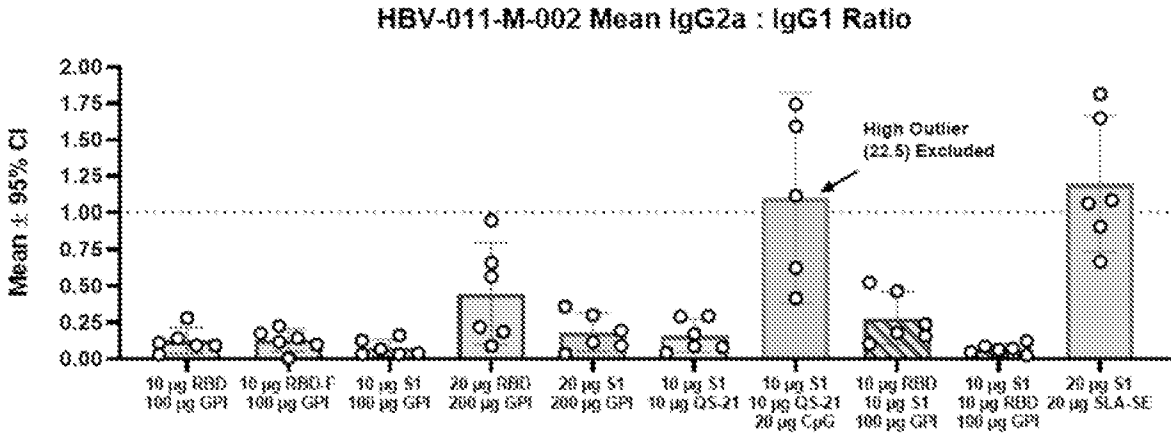


FIGURE 17



16/17

FIGURE 18

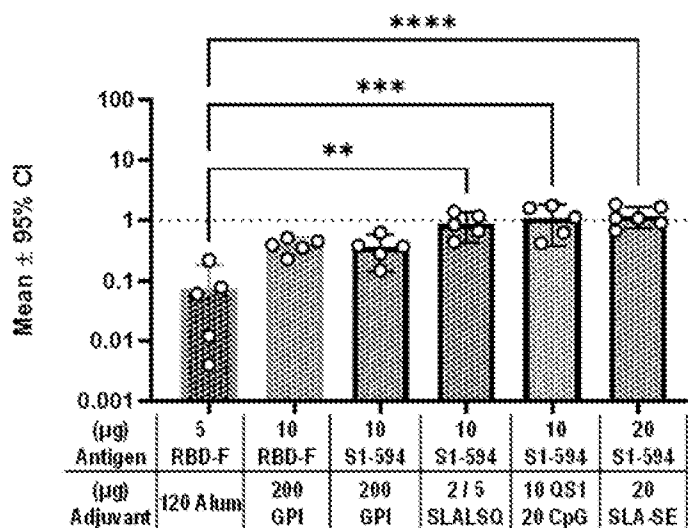


FIGURE 19A

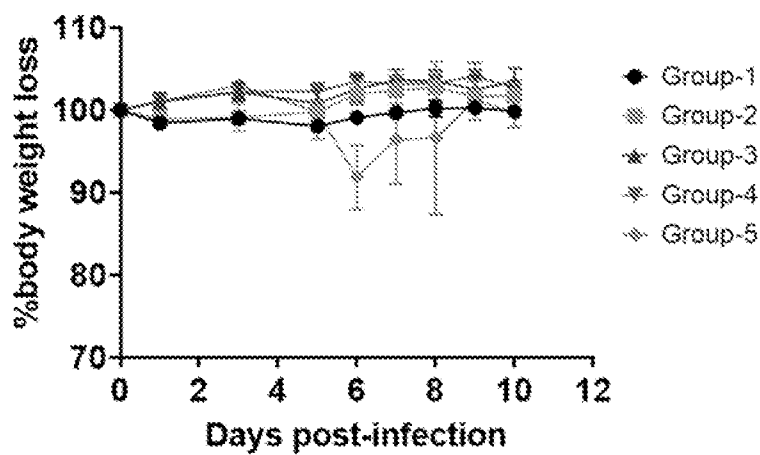


FIGURE 19B

Group	Antigen	Adjuvant	Pre-chall MN ₅₀ titer	Survival
1	20 μg nCoV-RBD-Mo	200 μg GPI-0100	1050	6/6 (100%)*
2	20 μg nCoV-RBD-Mo	10 μg QS21 / 20 μg CpG	441	6/6 (100%)*
3	20 μg nCoV-S1	10 μg QS21	981	6/6 (100%)*
4	20 μg nCoV-S1	10 μg QS21 / 20 μg CpG	1236	6/6 (100%)*
5	none	10 μg QS21 / 20 μg CpG	10	2/6 (33%)

17/17

FIGURE 20A

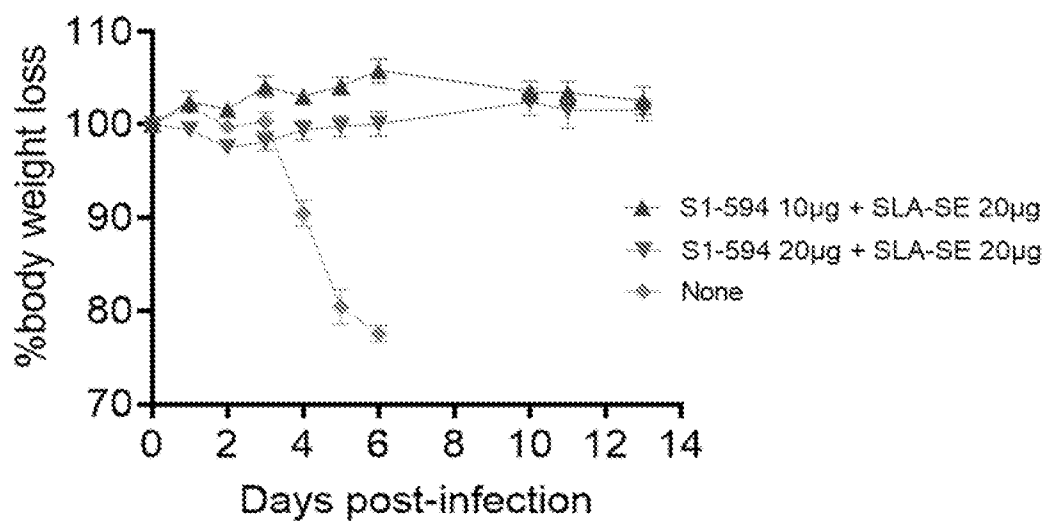


FIGURE 20B

Group	Antigen	Adjuvant	Survival	Pre-chall MN ₅₀ titer	Post-chall MN ₅₀ titer	Fold Increase
1	10 µg nCoV-S1	20 µg SLA-SE	6/6 (100%)*	2572	3968	1.5
2	20 µg nCoV-S1	20 µg SLA-SE	6/6 (100%)*	981	2887	2.9
5	none	20 µg SLA-SE	0/6 (0%)	10	N/A	N/A