

## Modified Viral Structural Protein

### FIELD OF INVENTION

5 [0001] The present disclosure relates to modified viral structural protein. The present invention also relates to virus-like particles (VLPs) comprising modified viral structural protein and methods of producing the VLPs in a host or host cells.

### BACKGROUND

10 [0002] Coronaviruses (CoVs) are the largest group of viruses belonging to the Nidovirales order, which includes Coronaviridae, Arteriviridae, Mesoniviridae, and Roniviridae families. The Coronavirinae comprise one of two subfamilies in the Coronaviridae family, with the other being the Torovirinae. The Coronavirinae are further subdivided into four genera, the alpha, beta, gamma, and delta coronaviruses. Members of alpha coronavirus and beta coronavirus are found exclusively in mammals. 15 The alphacoronavirus genus includes two human virus species, HCoV-229E and HCoV-NL63. Important animal alphacoronaviruses are transmissible gastroenteritis virus of pigs and feline infectious peritonitis virus.

[0003] Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, also known as 20 2019-nCoV and HCoV-19) is a novel lineage B betacoronavirus (Beta-CoV) and causes coronavirus disease 2019 (COVID-19), a respiratory illness with high mortality and morbidity resulting in major public health impacts worldwide. Outbreaks of SARS-CoV-2, such as the pandemic starting in 2020, are a paramount challenge for healthcare systems due to the incubation period and transmissibility of the virus. Treatments for COVID-19 are urgently needed, but long-term management of SARS-CoV-2 outbreaks 25 will require an effective vaccine.

[0004] Coronavirus virions are spherical with diameters of approximately 118–140 nm as depicted in recent studies by cryo-electron tomography and cryo-electron microscopy.

[0005] The most prominent structural feature of coronaviruses is the club-shaped spike projections emanating from the surface of the virion. Coronavirus particles consist of a

helical nucleocapsid structure, formed by the association between nucleocapsid (N) phosphoproteins and the viral genomic RNA, surrounded by a lipid bilayer where three or four types of structural proteins are inserted: the spike (S), the membrane (M), and the envelope (E) proteins and, for some coronaviruses only, the hemagglutinin-esterase (HE) protein (Masters PS. The molecular biology of coronaviruses. Adv Virus Res. 2006; 66:193–292.)

[0006] The membrane (M) protein is the most abundant structural protein in the virion. It is a small (~25–30 kDa) protein with three transmembrane domains and is thought to give the virion its shape. The envelope (E) protein is a short, integral membrane protein of 76–109 amino acids, ranging from 8.4 to 12 kDa in size. The primary and secondary structure reveals that E has a short, hydrophilic amino terminus consisting of 7–12 amino acids, followed by a large hydrophobic transmembrane domain (TMD) of 25 amino acids, and ends with a long, hydrophilic carboxyl terminus, which comprises the majority of the protein. The E protein is involved in several aspects of the virus' life cycle, such as assembly, budding, envelope formation, and pathogenesis.

[0007] The spike (S) protein is a glycoprotein that is required for the recognition of host receptors for many coronaviruses as well as the fusion of viral and host cell membranes for viral entry into cells (Belouzard et al., Viruses 2012 Jun;4(6):1011-33). As the primary glycoprotein on the surface of the viral envelope, S proteins of Coronaviridae are a major target of neutralizing antibodies elicited by natural infection, including SARS-CoV-2 infection, and are key antigens targeted in experimental vaccine candidates.

[0008] SARS-CoV-2 S protein, like S protein of other coronaviruses, is initially synthesized as a precursor protein. Individual precursor S protein forms a homotrimer and undergoes glycosylation within the Golgi compartment as well as processing to remove the signal peptide. The S protein requires a two-step, protease-mediated activation to facilitate membrane fusion. SARS-CoV-2 S protein is distinguished by a RRAR furin cleavage site at the S1/S2 junction that is presumably processed in the Golgi compartment to yield two separate polypeptides: the S1 and S2 polypeptide (or subunit), which remain non-covalently bound as S1/S2 protomers within the homotrimer in the prefusion conformation (Walls et al. Cell 2020 181(2) p281-292; Li et al. eLife 2019; 8:

e51230.). Furin cleavage at the S1/S2 junction and cleavage at the S2' site, upstream of the fusion peptide, occurs during viral entry at the cell surface or in endosomes and can be mediated by several proteases.

5 [0009] This trimer is held in the prefusion conformation prior to binding to target receptors on a host cell via receptor binding domain (RBD) epitopes. Receptor binding destabilizes the prefusion trimer, resulting in shedding of the S1 subunit and transition of the S2 subunit to a stable post-fusion conformation through fusion of the virus to the cell membrane (Wrapp et al. *Science*, 13 Mar 2020, Vol. 367, Issue 6483, pp. 1260-1263). Neutralizing antibodies from individuals infected with SARS-CoV-2 have been shown to  
10 target the RBD of the S1 subunit of the S protein (Premkumar, L., 2020 *Science Immunology* 11 Jun 2020: Vol. 5, Issue 48).

[0010] Stabilization of the S protein ectodomain in the prefusion conformation tends to increase the recombinant expression yield, possibly by preventing triggering or misfolding that results from a tendency to adopt the more stable post-fusion structure  
15 (Hsieh et al. *Science* 2020, 369 p.1501-1505).

[0011] Mutations to the S protein ectodomain have been shown to facilitate stabilization of the prefusion conformation. WO 2018/081318 and its companion publication by Pallesen, J. et al. (PNAS August 29, 2017 114 (35)) disclose double proline substitutions at or near a junction between a heptad repeat 1 (HR1) and a central helix that stabilize the  
20 S ectodomain trimer of MERS-CoV spike protein in a prefusion conformation and substitutions to prevent protease cleavage at a S1/S2 cleavage site and the S2' cleavage site of the S ectodomain. SARS-CoV-2 S protein stabilized with double proline substitutions at homologous amino acid residues have been used to determine high-resolution structures by cryo-EM (Wrapp et al *Science* 2020 367, 1260–1263; Walls et al.  
25 *Cell* 2020, 181, 281–292). Further, disruption of the furin recognition site is thought to retain S protein in a prefusion conformation (Wrapp et al *Science* 2020 367, 1260–1263). However, even with these substitutions, the SARS-CoV-2 S protein ectodomain remains unstable and difficult to produce reliably in mammalian cells, hindering development of effective and high-yield subunit vaccines (Hsieh et al. *Science* 2020, 369 p.1501-1505).

[0012] Hsieh et al. (Science 2020, 369 p.1501-1505) designed and expressed in mammalian cells over 100 structure-guided SARS-CoV-2 spike protein variants based upon previously determined cryo-EM structure. The variants were biochemically, biophysically and structurally characterized to identify substitutions that lead to an increase in yield and stability. Hsieh et al. reports multiple prolines, disulfide bonds, salt bridges, and cavity-filling substitutions that increase expression and/or stability of the spike relative to the double proline substitutions. The best identified variant, HexaPro, has six beneficial proline substitutions leading to 10-fold higher expression than its parental construct and is able to withstand heat stress, storage at room temperature, and multiple freeze-thaws.

[0013] The S2 subunit can be divided into three domains: a large ectodomain, a transmembrane domain (TM) and a cytoplasmic tail (CT). The cytoplasmic tail of the S protein has previously been shown to be required for assembly. Two distinct retention signals may be found in the CT of Coronaviridae: i) an endoplasmic reticulum retrieval signal (ERRS) and/or ii) a tyrosine-dependent localization signal (YxxI or YxxF motif). The ERRS comprises the dibasic KxHxx motif which binds to the coatamer complex I (COPI). The motif is required for the localization of the SARS S protein to the ERGIC/Golgi region when coexpressed with SARS membrane (M) protein, and localization can be disrupted by mutating the KxHxx motif (McBride et al. J. Virol. Feb 2007, 81 (5) 2418-2428). S proteins containing an ERRS are recruited into COPI vesicles and retrieved from the Golgi to the endoplasmic reticulum (ER) in retrograde. The repeated cycling of S proteins between the ER and the Golgi leads to S protein intracellular retention. S protein of Alphacoronavirus and Betacoronavirus both comprise an ERRS (Ujike et al. Journal of General Virology (2016), 97, 1853–1864).

[0014] S protein of Betacoronavirus, such as S protein of MERS-CoV, SARS-CoV and SARS-CoV 2, possess only an ERRS and cannot be retained intracellularly, resulting in the release of S protein into the plasma membrane. Mutant SARS-CoV S protein lacking the ERRS is transported to the plasma membrane, while native S protein, when co-expressed with M protein, interacts with the M protein near the budding site, leading to S protein intracellular retention, suggesting that the ERRS of SARS-CoV contributes to S protein accumulation specifically in the post-medial Golgi compartment by interaction

with M protein, leading to S protein incorporation into VLPs (Ujike et al. Journal of General Virology (2016), 97, 1853–1864). Removal of the ERRS has recently been found to facilitate incorporation of SARS-CoV-2 S protein into lentiviral pseudovirions (Ou et al., 2020 Nature Communications volume 11, Article number: 1620).

5 [0015] Yu et al. (2020 Science) constructed a set of prototype DNA vaccines expressing six variants of the SARS-CoV-2 S protein with various deletions of the cytoplasmic tail, and transmembrane domain, which were assessed for their immunogenicity and protective efficacy against SARS-CoV-2 viral challenge in rhesus macaques. While the soluble fragments of the SARS-CoV-2 S protein ectodomain elicited reduced levels of  
10 sgRNA (indicative of viral replication), optimal protection was achieved with the full-length S protein immunogen.

[0016] Broer et al. (2006 J. Virol. p. 1302–1310) studied the roles of the transmembrane and cytoplasmic domains of the S protein in the infectivity and membrane fusion activity of SARS-CoV, using a SARS-CoV S-pseudotyped retrovirus (SARSpp). SARSpp, in  
15 which the cytoplasmic domain of S was replaced by the cytoplasmic domain derived from vesicular stomatitis virus G protein (VSV-G), were infectious, up to 40% of wild type. In contrast, SARSpp containing both the TMD and the cytoplasmic domain of VSV-G, were severely impaired in infectivity (<5%). This shows that the TMD of S may be involved in the entry process of SARS-CoV.

20 [0017] Vaccination provides protection against disease by inducing a subject to mount an immune response to a likely agent prior to infection. Conventionally, this has been accomplished through the use of live attenuated or whole inactivated forms of the infectious agents as immunogens. To avoid the danger of using a whole virus (such as killed or attenuated viruses) as a vaccine, viral proteins or subunits, or recombinant  
25 versions thereof, have been pursued as vaccines. A major obstacle to employing viral proteins, either native or recombinant, as vaccine agents is ensuring that the conformation of the protein mimics the antigens in their natural environment. Suitable adjuvants and, in the case of peptides, carrier proteins, may be used to boost the immune response. In addition, viral proteins or subunits as vaccines may elicit primarily humoral responses

and thus fail to evoke lasting immunity. Subunit vaccines may be ineffective for diseases in which whole inactivated virus can be demonstrated to provide superior protection.

[0018] Virus-like particles (VLPs) may be used in immunogenic compositions to express viral proteins in a preferred conformation with improved antigen presentation to the immune system. VLPs closely resemble mature virions, but they do not contain viral genomic material, and they are non-replicative which makes them safe for administration as a vaccine. In addition, VLPs can be engineered to express viral glycoproteins on the surface of the VLP, which is their native physiological configuration. Since VLPs resemble intact virions and are multivalent particulate structures, VLPs may be more effective in inducing neutralizing antibodies to the glycoprotein than soluble envelope protein antigens.

[0019] A variety of expression systems have been utilized to produce VLPs, including mammalian cell lines, bacteria, insect cell lines, yeast and plant cells. VLPs for over thirty different viruses have been generated in insect and mammalian systems for vaccine purposes (Noad, R. and Roy, P., 2003, Trends Microbiol 11: 438-44). VLPs have also been produced in plants (see WO2009/076778; WO2009/009876; WO 2009/076778; WO 2010/003225; WO 2010/003235; WO2010/006452; WO2011/03522; WO 2010/148511; WO2014153674, and WO2012/083445).

[0020] VLPs have been produced with native surface proteins from Severe acute respiratory syndrome coronavirus (SARS-CoV or SARS-CoV-1), including S protein, M protein, E protein in insect and mammalian cells (Liu et al., 2008, J Virol., p. 11318–11330). SARS-CoV-2 virus like particles (VLPs) have also been assembled by co-expressing viral surface proteins S, M, and E in mammalian cells (Xu et al. Front. Bioeng. Biotechnol., 30 July 2020). Studies have further shown that the M protein is indispensable for virus-like particle (VLP) formation (Siu et al. Journal of Virology (2008) 82:11318–11330, Huang et al. Journal of Virology (2004) 78:12557–12565). In mammalian cells, expression of membrane protein (M) and small envelope protein (E) are essential for efficient formation and release of SARS-CoV-2 VLPs (Xu et al. Front. Bioeng. Biotechnol., 30 July 2020)). Nevertheless, the minimal requirement for assembly of SARS-CoV VLPs is still controversial. Y. Huang et al. (Journal of Virology (2004)

78:12557–12565) described formation of VLPs in transfected human cells that only required co-expression of the M and N viral proteins, whereas Siu et al. (Journal of Virology (2008)82:11318–11330) showed that both E and N proteins must be co-expressed with M protein for the efficient production and release of SARS-CoV VLPs in transfected mammalian cells.

[0021] WO2012/083445 discloses the production of SARS CoV S protein in plants, wherein the transmembrane domain and the cytosolic tail domain (TM/CT) of the S protein were replaced with TM/CT from an influenza HA protein.

[0022] A few groups have proposed immunization with SARS-CoV VLPs as an effective vaccine strategy. VLPs produced in insect cells or chimeric MHV/SARS-CoV VLPs produced in mammalian cells were used in these studies (Lokugamage et al. Vaccine 2008 Feb 6;26(6):797-808, Lu et al. 2007 Immunology 122496-5024).

[0023] However, effective scale-up and manufacture of SARS-CoV-2 VLPs at the quantity required to meet the need of widespread vaccination of the global population, requires efficient viral structural protein and VLP production.

#### SUMMARY OF THE INVENTION

[0024] The present invention relates to modified viral structural proteins. The present invention also relates to virus-like particles (VLPs) comprising modified viral structural protein and methods of producing the VLPs in a host or host cells. More specifically, the invention relates to modified coronavirus S proteins. The present invention also relates to virus-like particles (VLPs) comprising modified S proteins and methods of producing the VLPs in a host or host cells.

[0025] In one aspect it is provided a modified coronavirus S-protein comprising, in series,

- an ectodomain derived from a coronavirus S-protein,
- a transmembrane and cytosolic tail domain (TMCT), wherein the TMCT is a chimeric TMCT, comprising:

- a transmembrane domain (TM), wherein the TM or a portion of the TM is derived from a coronavirus S-protein, and
- a cytosolic tail (CT), wherein the CT or a portion of the CT is derived from an influenza hemagglutinin (HA) protein.

5 [0026] The modified S-protein as described herein may form trimers. Accordingly it is also provided a trimer comprising modified coronavirus S-protein as described herewith.

[0027] In a further aspect, a virus like particle (VLP) comprising the modified S-protein or trimers comprising the modified S-protein as described above is provided.

10 Accordingly, the VLP comprises a modified coronavirus S-protein or trimer that comprise the modified S-protein, the modified S-protein comprising

- an ectodomain derived from a coronavirus S-protein,
- a transmembrane and cytosolic tail domain (TMCT), wherein the TMCT is a chimeric TMCT, comprising:

15 - a transmembrane domain (TM), wherein the TM or a portion of the TM is derived from a coronavirus S-protein and

- a cytosolic tail (CT), wherein the CT or a portion of the CT is derived from an influenza hemagglutinin (HA) protein.

[0028] The VLP may further comprise plant lipids.

20 [0029] The TM may be directly fused to the CT. The TM may be derived from the coronavirus S-protein TM and the CT may be derived from the influenza HA protein CT. Furthermore, the TM may be a chimeric TM comprising a N terminal sequence derived from the coronavirus S-protein TM and a C terminal sequence derived from the influenza HA protein TM. The chimeric TM may comprise a N terminal sequence derived from the coronavirus S-protein TM comprising at least 20 amino acids corresponding to amino acids 1-20 of SEQ ID NO: 18 or SEQ ID NO: 169, or at least 21 amino acids

25 corresponding to amino acids 1-21 of SEQ ID NO: 118 or 164, or at least 22 amino acids corresponding to amino acids 1-22 of SEQ ID NO: 123 and one or more than one amino acid from the C-terminal end of the influenza HA protein TM. The one or more than one

amino acid from the C-terminal end of the influenza HA protein TM may be selected from AGL or conserved substitution of AGL, MAGL or conserved substitution of MAGL. The chimeric TM may comprise amino acids corresponding to amino acids of 1-20 of SEQ ID NO: 18.

5 [0030] The CT may be chimeric CT comprising a N terminal sequence derived from the coronavirus S-protein CT and a C terminal sequence derived from the influenza HA protein CT. The chimeric CT may comprise a C terminal sequence derived from influenza HA protein CT comprising at least 11 amino acids corresponding to amino acids 27-37 of SEQ ID NO: 18, 126, 127, 128, 129, 130 or 131 and one or more than one  
10 amino acid from the N-terminal end of the coronavirus S-protein CT. The one or more than one amino acid from the N-terminal end of the coronavirus S-protein CT may be selected from C or a conserved substitution of C, CC or a conserved substitution of CC, or CCM or a a conserved substitution of CCM. The chimeric CT may comprise amino acids corresponding to amino acids of 27-37 of SEQ ID NO: 18, 126, 128, 129, 130 or  
15 131; or amino acids 27-36 of SEQ ID NO: 127. In one aspect the chimeric TMCT may comprise a chimeric TM comprising amino acids corresponding to amino acids 1-20 of SEQ ID NO: 18 or SEQ ID NO: 169, or to amino acids 1-21 of SEQ ID NO: 118 or SEQ ID NO: 164, or amino acids 1-22 of SEQ ID NO: 123, a chimeric CT comprising amino acids corresponding to amino acids 27-37 of SEQ ID NO: 18, 126, 127, 128, 129, 130 or  
20 131, or a combination thereof.

[0031] The CT or portion of the CT may comprise from 80% to 100% identity with the sequence of SEQ ID NO: 15, or with amino acids 35-50 of SEQ ID NO 6, 8, 7, 9, 10, 12, 13 or 14, or with amino acids 34-49 of SEQ ID NO 11, or with amino acids 553-568 of SEQ ID NO:3 or with amino acids 22-37 of SEQ ID NO:18, or with amino acids 21-40 of  
25 SEQ ID NO: 19, or with amino acids 21-39 of SEQ ID NO: 37, or with amino acids 25-36 of SEQ ID NO: 38 or with amino acids 24-34 of SEQ ID NO: 39, or amino acids 22-37 of SEQ ID NO: 126, 128, 129, 130 or 131; or amino acids 22-36 of SEQ ID NO: 127. The TM or portion of the TM may comprises from 80% to 100% identity with the sequence of SEQ ID NO: 132 or 133.

[0032] The TMCT may comprise a sequence having about 80% to about 100% identity with the sequence of SEQ ID NO: 18, 19, 37, 38, 39, 64, 126, 127, 128, 129, 130, 131, 118, 119, 120, 123, 124, 125, 134, 135 164, 165, 166, 169, 170, 171, 172 or 173.

5 [0033] The modified S protein may comprises an S1 subunit and an S2 subunit, wherein the S2 subunit comprises the chimeric TMCT.

[0034] The modified S-protein may be produced as a precursor protein, the precursor protein comprising the modified S-protein and a signal peptide. The precursor protein comprising the modified S-protein and a signal peptide may comprise from 80% to 100% identity with amino acids 1-1234 of SEQ ID 1, or with amino acids 1-1234 of SEQ ID  
10 NO: 5, amino acids 1-1219 of SEQ ID NO: 21 or with amino acids 1-1243 of SEQ ID NO: 30 and wherein the amino acid sequence of the CT comprises from 80% to 100% identity with the sequence of SEQ ID NO: 15, or with amino acids 35-50 of SEQ ID NO 6, 8, 7, 9, 10, 12, 13 or 14, or with amino acids 34-49 of SEQ ID NO 11, or with amino acids 553-568 of SEQ ID NO:3.

15 [0035] The signal peptide may be native or non-native to the S-protein. The non-native signal peptide may be derived from the signal peptide of protein disulfide-isomerase (PDI). The modified S-protein may further comprise plant specific N-glycans.

[0036] The CT or portion of the CT in the modified S-protein may be derived from an influenza hemagglutinin (HA) protein that is derived from influenza type B or influenza  
20 subtype H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, or H16. The influenza hemagglutinin (HA) protein may be derived from influenza type B or influenza subtype H1, H3, H5, H6, H7 or H9.

[0037] The ectodomain of the modified S-protein may be derived from SARS-CoV-2, SARS-CoV-1, MERS-CoV, OC43-CoV or 229E-CoV, the TM or the portion of the TM  
25 may be derived from SARS-CoV-2, SARS-CoV-1, MERS-CoV, OC43-CoV or 229E-CoV, or both the ectodomain and the TM or the portion of the TM may be derived from SARS-CoV-2, SARS-CoV-1, MERS-CoV, OC43-CoV or 229E-CoV.

[0038] In a further aspect, the modified S-protein may comprise one or more than one amino acid substitution when compared to a wild-type coronavirus amino acid sequence. The one or more than one substitution may maintain the S-protein in a pre-fusion state.

5 [0039] The one or more than one amino acid substitution may comprise i) substitutions that restricts the processing at a cleavage site between S1 and S2 subunit, ii) substitution of one or more than one amino acid to one or more than one proline, or iii) substitutions that restrict the processing at the cleavage site between the S1 and the S2 subunit and substitution of one or more than one amino acid to one or more than one proline.

10 [0040] The one or more than one substitution may maintain the S-protein in a pre-fusion state or produces are higher yield of the modified S-protein when expressed in a host or host cell, when compared to the yield of a corresponding S-protein without the one or more than one substitutions expressed in the host or host cell.

15 [0041] The one or more than one amino acid substitution may correspond to amino acids at positions 667, 668, 670, 802, 923, 927, 971, 972 or a combination thereof, when compared to reference amino acid sequence of SEQ ID NO: 2.

20 [0042] In one aspect, the one or more than one amino acid substitution correspond to amino acids at positions 971 and 972, when compared to reference amino acid sequence of SEQ ID NO: 2. In another aspect, the one or more than one amino acid substitution correspond to amino acids at positions 802, 927, 971 and 972, when compared to reference amino acid sequence of SEQ ID NO: 2. Furthermore, the modified S-protein may comprise one or more than one amino acid substitution corresponding to amino acids at positions 667, 668, 670, or a combination thereof, when compared to reference amino acid sequence of SEQ ID NO: 2. Accordingly, the modified S-protein may comprise substitutions that correspond to amino acids at positions 667, 668 and 670,  
25 when compared to reference amino acid sequence of SEQ ID NO: 2.

[0043] In one aspect, the one or more than one substitution may correspond to amino acids at positions 667, 668, 670, 971 and 972, when compared to reference amino acid sequence of SEQ ID NO: 2. The substitution of the amino acid corresponding to the amino acid at position 667 of SEQ ID NO: 2 may be to glycine or a conserved

substitution of glycine, the substitution of the amino acid corresponding to position 668 of SEQ ID NO: 2 may be to a serine or a conserved substitution of serine, the substitution of the amino acid corresponding to position 670 of SEQ ID NO: 2 may be to a serine or a conserved substitution of serine, the substitution of the amino acid corresponding to the amino acid at position 971 of SEQ ID NO: 2 may be to a proline or a conserved substitution of proline and the substitution of the amino acid corresponding to the amino acid at position 972 of SEQ ID NO: 2 may be to a proline or a conserved substitution of proline. The modified S-protein as described above may further comprises an amino acid substitution corresponding to amino acid at position 923, when compared to reference amino acid sequence of SEQ ID NO: 2.

[0044] In another aspect the one or more than one amino acid substitution may correspond to amino acids at positions 667, 668, 670, 802, 927, 971 and 972, when compared to reference amino acid sequence of SEQ ID NO: 2. The substitution of the amino acid corresponding to the amino acid at position 667 of SEQ ID NO: 2 may be to glycine or a conserved substitution of glycine, the substitution of the amino acid corresponding to position 668 of SEQ ID NO: 2 may be to a serine or a conserved substitution of serine, the substitution of the amino acid corresponding to position 670 of SEQ ID NO: 2 may be to a serine or a conserved substitution of serine, the substitution of the amino acid corresponding to the amino acid at positions 802 of SEQ ID NO: 2 may be to a proline or a conserved substitution of proline, the substitution of the amino acid corresponding to the amino acid at positions 927 of SEQ ID NO: 2 may be to a proline or a conserved substitution of proline, the substitution of the amino acid corresponding to the amino acid at positions 971 of SEQ ID NO: 2 may be to a proline or a conserved substitution of proline and the substitution of the amino acid corresponding to the amino acid at positions 972 of SEQ ID NO: 2 may be to a proline or a conserved substitution of proline.

[0045] In another aspect the modified S-protein as described above may further comprises an amino acid substitution corresponding to amino acid at position 923, when compared to reference amino acid sequence of SEQ ID NO: 2. The substitution in the modified S-protein of the amino acid corresponding to the amino acid at position 667 of

SEQ ID NO: 2 may be to glycine or a conserved substitution of glycine, the substitution of the amino acid corresponding to position 668 of SEQ ID NO: 2 may be to a serine or a conserved substitution of serine, the substitution of the amino acid corresponding to position 670 of SEQ ID NO: 2 may be to a serine or a conserved substitution of serine, the substitution of the amino acid corresponding to the amino acid at positions 802, 927, 971 and 972 of SEQ ID NO: 2 may be to a proline or a conserved substitution of proline and the substitution of the amino acid corresponding to position 923 of SEQ ID NO: 2 may be to phenylalanine or a conserved substitution of phenylalanine.

[0046] The modified the S-protein may comprise from 80% to 100% identity with amino acids of SEQ ID NO: 5, 21, 30, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 95, 96, 97, 108, 109, 110, 144, 145, 146, 155, 156 or 157, or with amino acids 24-1259 of SEQ ID NO: 47 amino acids 25-1259 of SEQ ID NO: 48, amino acids 25-1259 of SEQ ID NO: 49, amino acids 25-1259 of SEQ ID NO: 50, amino acids 25-1259 of SEQ ID NO: 51, amino acids 25-1259 of SEQ ID NO: 52, amino acids 25-1259 of SEQ ID NO: 53, amino acids 25-1259 of SEQ ID NO: 54, amino acids 25-1259 of SEQ ID NO: 55, amino acids 25-1259 of SEQ ID NO: 56, amino acids 25-1259 of SEQ ID NO: 57, amino acids 25-1259 of SEQ ID NO: 58, amino acids 25-1262 of SEQ ID NO: 59, amino acids 25-1261 of SEQ ID NO: 60, amino acids 25-1258 of SEQ ID NO: 61, or amino acids 25-1256 of SEQ ID NO: 62, amino acids 25-1243 of SEQ ID NO: 95, amino acids 25-1240 of SEQ ID NO: 96, amino acids 25-1243 of SEQ ID NO: 97, amino acids 25-1341 of SEQ ID NO: 108, amino acids 25-1338 of SEQ ID NO: 109, amino acids 25-1341 of SEQ ID NO: 110, amino acids 25-1351 of SEQ ID NO: 144, amino acids 25-1348 of SEQ ID NO: 145, amino acids 25-1351 of SEQ ID NO: 146, amino acids 25-1159 of SEQ ID NO: 155, amino acids 25-1156 of SEQ ID NO: 156, or amino acids 25-1159 of SEQ ID NO: 157.

[0047] In another aspect, it is provided a nucleic acid that comprises a nucleotide sequence that encodes the modified S-protein as described above.

[0048] In a further aspect a composition comprising an effective dose of the modified S-protein, the trimer comprising the modified S-protein or VLP comprising the modified S-protein as described above and a pharmaceutically acceptable carrier, adjuvant, vehicle or

excipient is provided. In yet another aspect, it is provided a vaccine for inducing an immune response. The vaccine comprises an effective dose of the modified S protein, the trimer comprising the modified S-protein or VLP comprising the modified S-protein as described above as described above.

5 [0049] The composition further comprises a pharmaceutically acceptable carrier, adjuvant, vehicle or excipient. In a further aspect, it is provided a vaccine for inducing an immune response. The vaccine comprises an effective dose of the VLP comprising a modified coronavirus as described above. The vaccine may be a multivalent vaccine, comprising a mixture of VLP.

10 [0050] In yet another aspect a (non-human) host or host cell comprising the modified S-protein, trimer or VLP as described above is provided. In yet another aspect a host or host cell comprising the VLP as described above is provided. In another aspect it is provided a composition comprising an effective dose of the VLP comprising the modified S-protein as described above is provided.

15 [0051] In yet another aspect, a S-protein trimer is provided. The trimer comprises modified coronavirus S-protein, the modified S-protein comprising

- an ectodomain derived from a coronavirus S-protein,
- a transmembrane and cytosolic tail domain (TMCT), wherein the TMCT is a chimeric TMCT, comprising:

20 - a transmembrane domain (TM), wherein the TM or a portion of the TM is derived from a coronavirus S-protein, and

- a cytosolic tail (CT), wherein the CT or a portion of the CT is derived from an influenza hemagglutinin (HA) protein. The modified S-protein in the trimer may comprise one or more than one amino acid substitution when compared to a wild-type coronavirus amino acid sequence, as described above. In a further aspect a composition comprising an effective dose of the trimer as described above and a pharmaceutically acceptable carrier, adjuvant, vehicle or excipient is provided. In another aspect, virus like particle (VLP) comprising the trimer as described above are also provided. The VLP may further comprise plant lipids. In another aspect it is provided a composition comprising

25

an effective dose of the VLP comprising the trimer as described above and a pharmaceutically acceptable carrier, adjuvant, vehicle or excipient is provided. In a further aspect, it is provided a vaccine for inducing an immune response. The vaccine comprises an effective dose of the trimer as described above. In a further aspect, it is provided a vaccine for inducing an immune response. The vaccine comprises an effective dose of the VLP comprising the trimer as described above. The vaccine may be a multivalent vaccine, comprising a mixture of VLP. In yet another aspect a non-human host or host cell comprising the trimer or the VLP comprising the trimer as described above is provided.

[0052] In another aspect, it is provided a method for inducing immunity to a Coronavirus infection in a subject, the method comprising administering the composition or vaccines as described above. The composition or vaccine may be administered once to the subject or the composition or vaccine may be administered multiple times to the subject. The composition or vaccine may be administered as an initial dose and one or more than one subsequent doses may be administered between 1 day and 6 month from the administration of the initial dose. The subsequent dose may be administered after 21 days from the administration of the initial dose.

[0053] In another aspect an antibody or antibody fragment prepared using the composition or vaccine as described above are provided.

[0054] In yet another aspect, it is provided A) a method of producing a virus like particle (VLP) in a (non-human) host or host cell comprising:

a) introducing into a non-human host or host cell the nucleic acid comprising a nucleotide sequence that encodes the modified S-protein as described above; or providing the non-human host or host cell comprising the nucleic acid comprising a nucleotide sequence that encodes the modified S-protein as described above, and

b) incubating the non-human host or host cell under conditions that permit the expression of the nucleic acid, thereby producing the VLP.

[0055] In a further step c) the non-human host or host cell may be harvested.

[0056] In a further aspect, it is provided B) a method of increasing yield of production of a Coronavirus S-protein in a (non-human) host or host cell comprising:

- 5 a) introducing the nucleic acid comprising a nucleotide sequence that encodes the modified S-protein as described above into the non-human host or host cell; or providing a non-human host or host cell comprising the nucleic acid comprising a nucleotide sequence that encodes the modified S-protein as described above; and
- 10 b) incubating the non-human host or host cell under conditions that permit expression of the modified S-protein encoded by the nucleic acid, thereby producing modified S-protein at a higher yield compared to a host or host cell expressing the unmodified S-protein under similar or identical conditions.

[0057] In a further step c) the non-human host or host cell may be harvested.

[0058] In yet another aspect, it is provided C) a method of increasing yield of production of virus like particle (VLP) in a (non-human) host or host cell comprising:

- 15 a) introducing into the non-human host or host cell, a nucleic acid encoding a modified coronavirus S-protein comprising
- an ectodomain derived from a coronavirus S-protein,
  - a transmembrane and cytosolic tail domain (TMCT), wherein the TMCT is a chimeric TMCT, comprising:
- 20 - a transmembrane domain (TM), wherein the TM or a portion of the TM is derived from a coronavirus S-protein and
- a cytosolic tail (CT), wherein the CT or a portion of the CT is derived from an influenza hemagglutinin (HA) protein;
- 25 providing a non-human host or host cell comprising the nucleic acid encoding the modified S-protein comprising

- an ectodomain derived from a coronavirus S-protein,
- a transmembrane and cytosolic tail domain (TMCT), wherein the TMCT is a chimeric TMCT, comprising:

5

- a transmembrane domain (TM), wherein the TM or a portion of the TM is derived from a coronavirus S-protein and

- a cytosolic tail (CT), wherein the CT or a portion of the CT is derived from an influenza hemagglutinin (HA) protein; and

10

- b) incubating the non-human host or host cell under conditions that permit expression of the modified S-protein encoded by the nucleic acid, thereby producing VLP comprising modified S-protein at a higher yield compared to the yield of VLP in a host of host cell that expresses unmodified S protein under similar or identical conditions.

[0059] In a further aspect, it is provided D) a method of producing a virus like particle (VLP) in a (non-human) host or host cell comprising:

15

- a) introducing into a non-human host or host cell, a nucleic acid encoding a modified coronavirus S-protein comprising

- an ectodomain derived from a coronavirus S-protein,

- a transmembrane and cytosolic tail domain (TMCT), wherein the TMCT is a chimeric TMCT, comprising:

20

- a transmembrane domain (TM), wherein the TM or a portion of the TM is derived from a coronavirus S-protein and

- a cytosolic tail (CT), wherein the CT or a portion of the CT is derived from an influenza hemagglutinin (HA) protein; or

25

- providing a non-human host or host cell comprising the nucleic acid encoding the modified S-protein comprising

- an ectodomain derived from a coronavirus S-protein,
- a transmembrane and cytosolic tail domain (TMCT), wherein the TMCT is a chimeric TMCT, comprising:

5

- a transmembrane domain (TM), wherein the TM or a portion of the TM is derived from a coronavirus S-protein and

- a cytosolic tail (CT), wherein the CT or a portion of the CT is derived from an influenza hemagglutinin (HA) protein; and

b) incubating the non-human host or host cell under conditions that permit the expression of the nucleic acid, thereby producing the VLP.

10

[0060] In a further aspect the VLP of method A), B), C) or D) may further be extracted and purified from the host or host cell. The host or host cell may comprise a plant, a plant cell, a fungi, a fungi cell, an insect, an insect cell, an animal or an animal cell. The host or host cell of method A), B), C) or D) may be a plant, portion of a plant or plant cell.

15

[0061] In another aspect it is provided a VLP produced by the method of A), B), C) or D).

[0062] Furthermore, in yet another aspect it is provided a composition comprising an adjuvant and virus-like particles (VLP), the VLP comprising modified coronavirus S-protein, the modified S-protein comprising

- an ectodomain derived from a coronavirus S-protein,

20

- a transmembrane and cytosolic tail domain (TMCT), wherein the TMCT is a chimeric TMCT, comprising:

- a transmembrane domain (TM), wherein the TM or a portion of the TM is derived from a coronavirus S-protein and

25

- a cytosolic tail (CT), wherein the CT or a portion of the CT is derived from an influenza hemagglutinin (HA) protein and wherein the modified S-

protein further comprises substitutions at positions 667, 668, 670, 971 and 972 when compared to reference amino acid sequence of SEQ ID NO: 2.

[0063] In yet another aspect it is provided a composition comprising virus-like particle (VLP), the VLP comprising modified coronavirus S-protein, the modified S protein comprising

- an ectodomain derived from a coronavirus S-protein,
- a transmembrane and cytosolic tail domain (TMCT), wherein the TMCT is a chimeric TMCT, comprising:

- a transmembrane domain (TM), wherein the TM or a portion of the TM is derived from a coronavirus S-protein and

- a cytosolic tail (CT), wherein the CT or a portion of the CT is derived from an influenza hemagglutinin (HA) protein; and wherein the modified S-protein comprises a glycine substitution at position 667, a serine substitution at position 668, a serine substitution at position 670, a proline substitution at position 971 and a proline substitution at position 972, the position corresponding to reference amino acid sequence of SEQ ID NO: 2. The influenza hemagglutinin (HA) protein may be derived from influenza type B or influenza subtype H1, H3, H5, H6, H7 or H9. The composition may further comprise an adjuvant.

[0064] In a further aspect it is provided a composition comprising virus-like particle (VLP), the VLP comprising modified coronavirus S-protein, the modified S protein comprising the sequence of SEQ ID NO: 21. The composition may further comprise an adjuvant.

[0065] This summary of the invention does not necessarily describe all features of the invention.

## **BRIEF DESCRIPTION OF THE DRAWINGS**

[0066] These and other features of the invention will become more apparent from the following description in which reference is made to the appended drawings wherein:

[0067] **FIGURE 1** shows a schematic representation of Coronavirus S protein and the location of S1/S2 (aa 685/686) and S2' cleavage sites. SP: signal peptide (aa 1-15); NTD: N-terminal domain (aa 16-306); RBD: receptor-binding domain (aa 335-527); FP: fusion peptide (aa 816-833); HR1: heptad repeat 1 (aa 908-991); HR2: heptad repeat 2 (aa 1166-1207); TM: transmembrane domain (aa 1214-1234); CT: cytoplasmic tail (aa 1235-1273). The residue numbers (aa) of each region correspond to their positions in the S protein of SARS-CoV-2 (2019-nCoV).

[0068] **FIGURE 2** shows an alignment of amino acid sequences from exemplary influenza strains. The C-terminal region of the ectodomain, the transmembrane domain (TM), and the cytoplasmic tail domain (CT) of hemagglutinin (HA) are shown for H1 A/California/07/2009 (SEQ ID NO: 6), H2 A/Singapore/1/1957 HA (SEQ ID NO:7), H3 A/Minnesota/41/2019 HA (SEQ ID NO:8), H5 A/Indonesia/5/05 HA (SEQ ID NO:9), H6 A/Teal/Hong Kong/W312/97 HA (SEQ ID NO:10), H7 A/Guangdong/17SF003/2016 HA (SEQ ID NO:11), H9 A/Hong Kong/1073/99 HA (SEQ ID NO:12), and B/Washington/02/2019 HA (SEQ ID NO:13). The consensus sequence for these sequences is also shown (SEQ ID NO: 14).

[0069] **FIGURE 3A** shows quantified fold-change difference in SARS-CoV-2 S protein accumulation in plants expressing: SARS-CoV-2 S protein with a native (wild-type) transmembrane domain and cytoplasmic tail (wtTMCT) under the control of the following 5'UTRs: nbMT78 (construct 8586), nbCSY65 (construct 8589) and nbHEL40 (construct 8591); a modified SARS-CoV-2 protein wherein the native (wild-type) transmembrane domain and cytoplasmic tail (wtTMCT) has been replaced with the TMCT of influenza hemagglutinin (HA) of influenza H5 A/Indonesia/5/05 (H5iTMCT) under the control of nbMT78 (construct 8592), nbCSY65 (construct 8595) and nbHEL40 (construct 8597); and a modified SARS-CoV-2 protein wherein the native (wild-type) cytoplasmic tail (wtCT) has been replaced with the CT of influenza hemagglutinin (HA) of influenza H5 A/Indonesia/5/05 (H5iCT) under the control of nbMT78 (construct 8610), nbCSY65 (construct 8611) and nbHEL40 (construct 8671). The SARS-CoV-2 S protein sequences (referred to as nCOV S (GSAS-2P)) have the following substitutions: R667G, R668S, R670S, K971P and V972P with respect to the reference sequence of SEQ

ID NO: 2. The results have been normalized to the SARS-CoV-2 S protein accumulation from construct 8591, which is set as 1. **FIGURE 3B** shows protein separation of clarified crude extract on a non-reducing SDS-PAGE gel. The following modified S proteins were expressed in plants: Lane 1: S protein with wild-type transmembrane and cytosolic tail domain (wt TMCT) under the control of nbMT78; lane 2: S protein with wild-type transmembrane and cytosolic tail domain (wt TMCT) under the control of nbCSY65; lane 3: S protein with wild-type transmembrane and cytosolic tail domain (wt TMCT) under the control of nbHEL40; lane 4: modified S protein with a SARS-CoV-2 ectodomain, and a transmembrane and cytosolic tail domain from hemagglutinin (HA) of H5 A/Indonesia/5/05 (H5i TMCT) under the control of nbMT78; lane 5: modified S protein with a SARS-CoV-2 ectodomain and a transmembrane and cytosolic tail domain from hemagglutinin (HA) of H5 A/Indonesia/5/05 (H5i TMCT) under the control of nbCSY65; lane 6: modified S protein with a SARS-CoV-2 ectodomain and a transmembrane and cytosolic tail domain from hemagglutinin (HA) of H5 A/Indonesia/5/05 (H5i TMCT) under the control of nbHEL40; lane 7: modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane domain and cytosolic tail domain (CT) from hemagglutinin (HA) of influenza H5 A/Indonesia/5/05 (H5i CT) under the control of nbMT78; lane 8: modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane domain and cytosolic tail domain (CT) from hemagglutinin (HA) of influenza H5 A/Indonesia/5/05 (H5i CT) under the control of nbCSY65; lane 9: modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane domain and cytosolic tail domain (CT) from hemagglutinin (HA) of influenza H5 A/Indonesia/5/05 (H5i CT) under the control of nbHEL40. The modified S protein has a molecular weight of about 150 kDa and is indicated by an arrow. **FIGURE 3C** shows a Western blot analysis of the same series of lysates depicted in Figure 3B and the lanes correspond to the lanes as described in Figure 3B. The top panel shows detection with an anti-SARS-CoV-2 S1 antibody (40150-R007). The bottom panel shows detection with an anti-SARS-CoV-2 S2 antibody (NB100-56578). The monomer of the SARS-CoV-2 protein (comprising the S1 and S2 subunit) has a molecular weight of about 150 kDa (non-reducing).

[0070] **FIGURE 4A** shows quantified fold-change difference in SARS-CoV-2 S protein accumulation in plants expressing: a modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane, and cytosolic tail domain from hemagglutinin (HA) of H5 A/Indonesia/5/05 (H5 Indo); a modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane, and cytosolic tail domain from hemagglutinin (HA) of H1 A/California/7/2009 (H1 California); a modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane, and cytosolic tail domain from hemagglutinin (HA) of H3 A/Minnesota/41/2019 (H3 Minnesota); a modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane, and cytosolic tail domain from hemagglutinin (HA) of H6 A/Teal/Hong Kong/W312/97 (H6 Hong Kong); a modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane, and cytosolic tail domain from hemagglutinin (HA) of H7 A/Guangdong/17SF003/2016 (H7 Guangdong); a modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane, and cytosolic tail domain from hemagglutinin (HA) of H9 A/Hong Kong/1073/99 (H9 Hong Kong); and a modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane, and cytosolic tail domain from hemagglutinin (HA) of B/Washington/02/2019 (B Washington). The results have been normalized to the SARS-CoV-2 S protein accumulation from construct 8671 encoding a modified SARS-CoV-2 protein wherein the native (wild-type) cytoplasmic tail (wtCT) has been replaced with the CT of influenza hemagglutinin (HA) of influenza H5 A/Indonesia/5/05 (H5iCT) under the control of nbHEL40 (H5 Indo), which is set as 1. **FIGURE 4B** shows a Western blot analysis of crude lysate from plants expressing the modified S proteins from the following constructs: lane 2, a modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane, and cytosolic tail domain from hemagglutinin (HA) of H5 A/Indonesia/5/05 (H5 Indo); lane 3, a modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane, and cytosolic tail domain from hemagglutinin (HA) of H1 A/California/07/2009 (H1 Calif); lane 4, a modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane, and cytosolic tail domain from hemagglutinin (HA) of H3 A/Minnesota/41/2019 (H3 Minn); lane 5, a modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane, and cytosolic tail domain from hemagglutinin (HA) of H6 A/Teal/Hong

Kong/W312/97 (H6 HK); lane 6, a modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane, and cytosolic tail domain from hemagglutinin (HA) of H7 A/Guangdong/17SF003/2016 (H7 Guan); lane 7, a modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane, and cytosolic tail domain from hemagglutinin (HA) of H9 A/Hong Kong/1073/99 (H9 HK); lane 8, a modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane, and cytosolic tail domain from hemagglutinin (HA) of B/Washington/02/2019 (B Wash). Lane 1 is crude lysate from a plant treated with a mock Agroinfiltration. Sino 400150-R007 antibody was used for detection of the S1 subunit of the SARS-COV-2 S protein. The monomer of the SARS-CoV-2 protein (comprising the S1 and S2 subunit) has a molecular weight of about 150 kDa (non-reducing). **FIGURE 4C** shows a Western blot analysis of the same series of lysates depicted in Figure 4B, except with NB100-56578 antibody detecting the S2 subunit of the SARS-COV-2 S protein.

[0071] **FIGURE 5A** shows the amino acid sequence of the C-terminal region of the native SARS-CoV-2 S protein (wtTM/wtCT), the C-terminal region of influenza H5 hemagglutinin (HA) (H5iTM/H5iCT), the C-terminal region of modified SARS-CoV-2 S protein with wild-type transmembrane domain (TM) and influenza H5 HA cytosolic tail (CT) domain (wtTM/H5iCT), and the C-terminal regions of four alternative versions of modified S protein (wtTM/H5iCT V1-V4) with variable margin between the SARS-CoV-2 transmembrane (TM) domain and H5 A/Indonesia/5/05 HA cytosolic tail (CT) domain. The TM domain from Coronavirus S-protein is underlined and the CT domain derived from influenza HA is shown in bold. **FIGURE 5B** shows quantified fold-change difference in SARS-CoV-2 S protein accumulation in plants expressing each of the four variant modified S proteins with a chimeric transmembrane and cytosolic tail domain (TMCT), as depicted in Figure 5A (wtTM/H5iCT, V1-V4), relative to modified SARS-CoV-2 S protein accumulation in plants expressing modified SARS-CoV-2 S protein having a chimeric TMCT with a wild-type transmembrane domain (TM) and influenza H5 HA cytosolic tail (CT) domain (wtTM/H5iCT) which is set as 1.

[0072] **FIGURE 6A** shows an electron micrograph of virus like particles (VLP) comprising SARS-COV-2 S protein with wild-type transmembrane and cytosolic tail domain (wtTMCT; construct 8591) **FIGURE 6B** shows an electron micrograph of virus

like particles (VLP) comprising a modified S protein with a SARS-CoV-2 ectodomain and an influenza H5 hemagglutinin transmembrane domain and cytosolic tail domain (H5i TMCT; construct 8597). **FIGURE 6C** shows an electron micrograph of virus like particles (VLP) comprising a modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane domain and an influenza H5 hemagglutinin cytosolic tail domain (H5i CT; construct 8671). **FIGURE 6D** shows an electron micrograph of virus like particles (VLP) comprising an alternative version of modified S protein (H5i CT V1; construct 8980) having a SARS-CoV-2 ectodomain and a chimeric transmembrane and cytosolic tail domain (TMCT). **FIGURE 6E** shows an electron micrograph of virus like particles (VLP) comprising an alternative version of modified S protein (H5i CT V2; construct 8981) having a SARS-CoV-2 ectodomain and a chimeric transmembrane and cytosolic tail domain (TMCT). **FIGURE 6F** shows an electron micrograph of virus like particles (VLP) comprising an alternative version of modified S protein (H5i CT V3; construct 8982) with a SARS-CoV-2 ectodomain having a SARS-CoV-2 ectodomain and a chimeric transmembrane and cytosolic tail domain (TMCT). **FIGURE 6G** shows an electron micrograph of virus like particles (VLP) comprising an alternative version of modified S protein (H5i CT V4; construct 8983) having a SARS-CoV-2 ectodomain and a chimeric transmembrane and cytosolic tail domain (TMCT). **FIGURE 6H** shows an electron micrograph of virus like particles (VLP) comprising a modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane domain and an influenza H1 hemagglutinin cytosolic tail domain (H1 CT; construct 7390). **FIGURE 6I** shows an electron micrograph of virus like particles (VLP) comprising a modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane domain and an influenza H3 hemagglutinin cytosolic tail domain (H3 CT; construct 7391). **FIGURE 6J** shows an electron micrograph of virus like particles (VLP) comprising a modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane domain and an influenza H6 hemagglutinin cytosolic tail domain (H6 CT; construct 7392). **FIGURE 6K** shows an electron micrograph of virus like particles (VLP) comprising a modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane domain and an influenza H7 hemagglutinin cytosolic tail domain (H7 CT; construct 7393). **FIGURE 6L** shows an electron micrograph of virus like particles (VLP) comprising a modified S protein with a

SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane domain and an influenza H9 hemagglutinin cytosolic tail domain (H9 CT; construct 7394). **FIGURE 6M** shows an electron micrograph of virus like particles (VLP) comprising a modified S protein with a SARS-CoV-2 ectodomain, a SARS-CoV-2 transmembrane domain and an influenza HA B hemagglutinin cytosolic tail domain (HA B CT; construct 7395).

[0073] **FIGURE 7A** shows a schematic representation of acceptor vector 8501. **FIGURE 7B** shows a schematic representation of acceptor vector 8500. **FIGURE 7C** shows a schematic representation of acceptor vector 8716.

[0074] **FIGURE 8A** shows a schematic representation of vector 8586. **FIGURE 8B** shows a schematic representation of vector 8589. **FIGURE 8C** shows a schematic representation of vector 8591.

[0075] **FIGURE 9A** shows a schematic representation of vector 8592. **FIGURE 9B** shows a schematic representation of vector 8595. **FIGURE 9C** shows a schematic representation of vector 8597.

[0076] **FIGURE 10A** shows a schematic representation of vector 8610. **FIGURE 10B** shows a schematic representation of vector 8611. **FIGURE 10C** shows a schematic representation of vector 8671.

[0077] **FIGURE 11A** shows quantified fold-change of accumulation in plants expressing modified SARS-CoV-2 S protein (wtTM/H5iCT) with additional substitutions. The modified SARS-CoV-2 S proteins have the following substitutions: “GSAS-2P”: R667G, R668S, R670S, K971P and V972P; “GSAS-4P”: R667G, R668S, R670S, K971P, V972P, F802P and A927P; and “GSAS-6P”: R667G, R668S, R670S, K971P, V972P, F802P, A877P, A884P and A927P (with respect to reference sequence of SEQ ID NO: 2). The results have been normalized to the accumulation of modified SARS-CoV-2 (wtTM/H5iCT) and GSAS+2P substitutions, which is set as 1. **FIGURE 11B** shows quantified fold-change of accumulation in plants expressing modified SARS-CoV-2 S protein (wtTM/H5iCT) with each of the GSAS-2P, GSAS-4P, and GSAS-6P substitutions as described for Figure 11A as compared to the quantified fold change of accumulation wherein each modified SARS-CoV-2 S protein further incorporates a

L923F substitution. The results have been normalized to the accumulation of modified SARS-CoV-2 (wtTM/H5iCT) and GSAS+2P substitutions, which is set as 1.

[0078] **FIGURE 12A** shows a schematic representation of vector 8980. **FIGURE 12B** shows a schematic representation of vector 8981. **FIGURE 12C** shows a schematic representation of vector 8982. **FIGURE 12D** shows a schematic representation of vector 8983.

[0079] **FIGURE 13A** shows a schematic representation of vector 7390. **FIGURE 13B** shows a schematic representation of vector 7391. **FIGURE 13C** shows a schematic representation of vector 7392. **FIGURE 13D** shows a schematic representation of vector 7393. **FIGURE 13E** shows a schematic representation of vector 7394. **FIGURE 13F** shows a schematic representation of vector 7395.

[0080] **FIGURE 14A** shows a schematic representation of vector 8953. **FIGURE 14B** shows a schematic representation of vector 8940.

[0081] **FIGURE 15A** shows a schematic representation of vector 8933. **FIGURE 15B** shows a schematic representation of vector 8960. **FIGURE 15C** shows a schematic representation of vector 8947.

[0082] **FIGURE 16A** shows a Western blot analysis of crude lysate from plants expressing the modified S proteins from the following constructs: lane 1, a modified S protein with a SARS-CoV-1 ectodomain, transmembrane, and cytosolic tail domain ("wtTMCT", construct 9231); lane 2, a modified S protein with an ectodomain from SARS-CoV-1, and a transmembrane and cytosolic tail domain (TMCT) from hemagglutinin (HA) of H5 A/Indonesia/5/05 ("H5iTMCT", construct 9232); lane 3, a modified S protein with an ectodomain and transmembrane domain from SARS-CoV-1 and a cytosolic tail domain from hemagglutinin (HA) of H5 A/Indonesia/5/05 (H5 Indo) ("H5iCT", construct 9233); lane 4, a modified S protein with an ectodomain and transmembrane domain from SARS-CoV-1 and a cytosolic tail domain from hemagglutinin (HA) of H5 A/Indonesia/5/05 (H5 Indo) with a variable margin between the SARS-CoV-1 transmembrane (TM) domain and H5 A/Indonesia/5/05 HA cytosolic

tail (CT) domain (“H5iCT(V4)”, construct 9234); lane 5, a modified S protein with an ectodomain and transmembrane domain from a SARS-CoV-1 and a cytosolic tail domain from hemagglutinin (HA) of H1 A/California/7/2009 (“H1cCT”, construct 9235). The primary antibody used for detection was SARS-CoV Spike S1 Subunit Antibody from Sino Biologicals (40150-MM08, 1/5000). The secondary antibody used for detection was Goat anti-Mouse from JIR (115-035-146, 1/10 000). The modified S protein has a molecular weight of about 150 kDa.

[0083] **FIGURE 16B, 16C and 16D** shows Western blot analysis of fractions F5 (30%), F6 (30%), F7 (25%), F8 (25%), F9 (25%), F10 (15%) and F11 (15%) from a discontinuous iodixanol density gradient. Accumulation of protein in these fractions is indicative for the formation of higher molecular weight structures i.e. VLP formation. For Western blots from fractions of crude lysate, the primary antibody used for detection was SARS-CoV Spike S1 subunit antibody from Sino Biologicals, 40150-MM08 (1/5000) and the secondary antibody used for detection was Goat anti-Mouse, JIR, 115-035-146 (1/10000). **FIGURE 16B:** Crude lysate from plants expressing SARS-CoV-1 S protein (with 2P+R667A substitution), with a native TMCT domain (wtTMCT, construct 9231) were analyzed. **FIGURE 16C:** Crude lysate from plants expressing modified SARS-CoV-1 S protein (with 2P+R667A substitution) having a TMCT from H5 A/Indonesia/5/05 HA (H5iTMCT, construct 9232). **FIGURE 16D:** Crude lysate from plants expressing modified SARS-CoV-1 S protein (with 2P+R667A substitution) having a cytoplasmic tail from H5 A/Indonesia/5/05 HA (H5iCT, construct 9233).

[0084] **FIGURE 17A** shows an electron micrograph of virus like particles (VLP) comprising SARS-COV-1 S protein (with 2P+R667A substitution) with native TMCT domain (wtTMCT, construct 9231). **FIGURE 17B** shows an electron micrograph of virus like particles (VLP) comprising modified SARS-CoV-1 S protein (with 2P+R667A substitution) having a TMCT from H5 A/Indonesia/5/05 HA (H5iTMCT, construct 9232). **FIGURE 17C** shows an electron micrograph of virus like particles (VLP) comprising modified SARS-CoV-1 S protein (with 2P+R667A substitution) having a cytoplasmic tail from H5 A/Indonesia/5/05 HA (H5iCT, construct 9233).

[0085] **FIGURE 18A** shows a schematic representation of vector 9231. **FIGURE 18B** shows a schematic representation of vector 9232. **FIGURE 18C** shows a schematic representation of vector 9233. **FIGURE 18D** shows a schematic representation of vector 9234. **FIGURE 18E** shows a schematic representation of vector 9235.

5 [0086] **FIGURE 19A** shows a Western blot analysis of crude lysate from plants expressing the modified S proteins from the following constructs: lane 1, a modified S protein with a MERS-CoV ectodomain, transmembrane, and cytosolic tail domain (“wtTMCT”, construct 9246); lane 2, a modified S protein with an ectodomain from MERS-CoV, and a transmembrane and cytosolic tail domain (TMCT) from  
10 hemagglutinin (HA) of H5 A/Indonesia/5/05 (“H5iTMCT”, construct 9247); lane 3, a modified S protein with an ectodomain and transmembrane domain from MERS-CoV and a cytosolic tail domain from hemagglutinin (HA) of H5 A/Indonesia/5/05 (H5 Indo) (“H5iCT”, construct 9249); lane 4, a modified S protein with an ectodomain and transmembrane domain from MERS-CoV and a cytosolic tail domain from  
15 hemagglutinin (HA) of H5 A/Indonesia/5/05 (H5 Indo) with a variable margin between the MERS-CoV transmembrane (TM) domain and H5 A/Indonesia/5/05 HA cytosolic tail (CT) domain (“H5iCT(V4)”, construct 9250); lane 5, a modified S protein with an ectodomain and transmembrane domain from a MERS-CoV and a cytosolic tail domain from hemagglutinin (HA) of H1 A/California/7/2009 (“H1cCT”, construct 9251). The  
20 primary antibody used for detection was MERS-CoV spike protein S1 antibody (N-terminal) from Sino Biological, (100208-RP02, 1/5000). The secondary antibody used for detection was Goat anti-Mouse from JIR (115-035-144, 1/10 000). The modified S protein has a molecular weight of about 175 kDa. **FIGURE 19B** shows an electron micrograph of virus like particles (VLP) comprising MERS-COV S protein (with ASVG+2P substitution) with native TMCT domain (wtTMCT, construct 9246).  
25 **FIGURE 19C** shows an electron micrograph of virus like particles (VLP) comprising modified MERS-CoV S protein (with ASVG+2P substitution) having a TMCT from H5 A/Indonesia/5/05 HA (H5iTMCT, construct 9247). **FIGURE 19D** shows an electron micrograph of virus like particles (VLP) comprising modified MERS-CoV S protein (with ASVG+2P substitution) having a cytoplasmic tail from H5 A/Indonesia/5/05 HA (H5iCT, construct 9249). **FIGURE 19E** shows an electron micrograph of virus like  
30

particles (VLP) comprising modified MERS-CoV S protein (with ASVG+2P substitution) having a cytoplasmic tail from H5 A/Indonesia/5/05 HA with a variable margin between the MERS-CoV transmembrane (TM) domain and H5 A/Indonesia/5/05 HA cytosolic tail (CT) domain (H5iCT (V4), construct 9250). **FIGURE 19F** shows an electron micrograph of virus like particles (VLP) comprising modified MERS-CoV S protein (with ASVG+2P substitution) having a cytoplasmic tail from H1 A/California/7/2009 HA (H1cCT, construct 9251).

[0087] **FIGURE 20A** shows a schematic representation of vector 9246. **FIGURE 20B** shows a schematic representation of vector 9247. **FIGURE 20C** shows a schematic representation of vector 9249. **FIGURE 20D** shows a schematic representation of vector 9250. **FIGURE 20E** shows a schematic representation of vector 9251.

[0088] **FIGURE 21** shows a schematic representation of acceptor vector 7147.

[0089] **FIGURE 22** shows an alignment of the native SARS-CoV-2, SARS-CoV-1, and MERS-CoV S protein sequences with the native signal peptide removed (SEQ ID NO: 2, 114, and 115). Residues corresponding to the RRAR furin cleavage site (667-670) and each of F802P, A877P, A884P, A927P, K971P, V972P in native SARS-CoV-2 S protein without signal peptide (SEQ ID NO: 2) are boxed along with homologous residues from native SARS-CoV-1 S protein without signal peptide (SEQ ID NO: 114) and native MERS S protein (SEQ ID NO: 115).

[0090] **FIGURE 23A** shows a Western blot analysis of crude lysate from plants expressing the modified S proteins from the following constructs: lane 3, a modified S protein with a OC43-CoV ectodomain, transmembrane, and cytosolic tail domain ("wtTMCT", construct 9269); lane 4, a modified S protein with an ectodomain from OC43-CoV, and a transmembrane and cytosolic tail domain (TMCT) from hemagglutinin (HA) of H5 A/Indonesia/5/05 ("H5iTMCT", construct 9270); lane 5, a modified S protein with an ectodomain and transmembrane domain from OC43-CoV and a cytosolic tail domain from hemagglutinin (HA) of H5 A/Indonesia/5/05 (H5 Indo) ("H5iCT", construct 9272); lane 6, a modified S protein with an ectodomain and transmembrane domain from OC43-CoV and a cytosolic tail domain from hemagglutinin (HA) of H5 A/Indonesia/5/05

(H5 Indo) with a variable margin between the OC43-CoV transmembrane (TM) domain and H5 A/Indonesia/5/05 HA cytosolic tail (CT) domain ("H5iCT (V4)", construct 9273); lane 7, a modified S protein with an ectodomain and transmembrane domain from a OC43-CoV and a cytosolic tail domain from hemagglutinin (HA) of H1 A/California/7/2009 ("H1cCT", construct 9274). The primary antibody used for detection was anti-coronavirus OC43 spike protein from Antibodies-online (ABIN2754654, 1/1000. The secondary antibody used for detection was Goat anti-Rabbit from JIR (111-035-144, 1/10 000). The modified S protein has a molecular weight of about 150 kDa. **FIGURE 23B** shows an electron micrograph of virus like particles (VLP) comprising modified OC43-CoV S protein (with GGSGS+2P substitution) having a TMCT from H5 A/Indonesia/5/05 HA (H5iTMCT, construct 9270). **FIGURE 23C** shows an electron micrograph of virus like particles (VLP) comprising modified OC43-CoV S protein (with GGSGS+2P substitution) having a cytoplasmic tail from H5 A/Indonesia/5/05 HA (H5iCT, construct 9272). **FIGURE 23D** shows an electron micrograph of virus like particles (VLP) comprising modified OC43-CoV S protein (with GGSGS+2P substitution) having a cytoplasmic tail from H5 A/Indonesia/5/05 HA with a variable margin between the OC43-CoV transmembrane (TM) domain and H5 A/Indonesia/5/05 HA cytosolic tail (CT) domain (H5iCT (V4), construct 9273). **FIGURE 23E** shows an electron micrograph of virus like particles (VLP) comprising modified OC43-CoV S protein (with GGSGS+2P substitution) having a cytoplasmic tail from H1 A/California/7/2009 HA (H1cCT, construct 9274).

[0091] **FIGURE 24A** shows a schematic representation of vector 9269. **FIGURE 24B** shows a schematic representation of vector 9270. **FIGURE 24C** shows a schematic representation of vector 9272. **FIGURE 24D** shows a schematic representation of vector 9273. **FIGURE 24E** shows a schematic representation of vector 9274.

[0092] **FIGURE 25A** shows an electron micrograph of virus like particles (VLP) comprising 229E-CoV S protein (with R567A+2P substitution) with native TMCT domain (wtTMCT, construct 9310). **FIGURE 25B** shows an electron micrograph of virus like particles (VLP) comprising modified 229E-CoV S protein (with R567A+2P substitution) having a TMCT from H5 A/Indonesia/5/05 HA (H5iTMCT, construct

9311). **FIGURE 25C** shows an electron micrograph of virus like particles (VLP) comprising modified 229E-CoV S protein (with R567A+2P substitution) having a cytoplasmic tail from H5 A/Indonesia/5/05 HA (H5iCT, construct 9312). **FIGURE 25D** shows an electron micrograph of virus like particles (VLP) comprising modified 229E-CoV S protein (with R567A+2P substitution) having a cytoplasmic tail from H5 A/Indonesia/5/05 HA with a variable margin between the 229E-CoV transmembrane (TM) domain and H5 A/Indonesia/5/05 HA cytosolic tail (CT) domain (H5iCT (V4), construct 9313). **FIGURE 25E** shows an electron micrograph of virus like particles (VLP) comprising modified 229E-CoV S protein (with R567A+2P substitution) having a cytoplasmic tail from H1 A/California/7/2009 HA (H1cCT, construct 9314).

[0093] **FIGURE 26A** shows a schematic representation of vector 9310. **FIGURE 26B** shows a schematic representation of vector 9311. **FIGURE 26C** shows a schematic representation of vector 9312. **FIGURE 26D** shows a schematic representation of vector 9313. **FIGURE 26E** shows a schematic representation of vector 9314.

#### **DETAILED DESCRIPTION**

[0094] The following description is of a preferred embodiment.

[0095] As used herein, the terms “comprising,” “having,” “including” and “containing,” and grammatical variations thereof, are inclusive or open-ended and do not exclude additional, un-recited elements and/or method steps. The term “consisting essentially of” when used herein in connection with a use or method, denotes that additional elements and/or method steps may be present, but that these additions do not materially affect the manner in which the recited method or use functions. The term “consisting of” when used herein in connection with a use or method, excludes the presence of additional elements and/or method steps. A use or method described herein as comprising certain elements and/or steps may also, in certain embodiments, consist essentially of those elements and/or steps, and in other embodiments consist of those elements and/or steps, whether or not these embodiments are specifically referred to. In addition, the use of the singular includes the plural, and “or” means “and/or” unless otherwise stated. The term “plurality” as used herein means more than one, for example, two or more, three or more,

four or more, and the like. Unless otherwise defined herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art. As used herein, the term “about” refers to an approximately  $\pm 10\%$  variation from a given value. It is to be understood that such a variation is always included in any given value provided herein, whether or not it is specifically referred to. The use of the word “a” or “an” when used herein in conjunction with the term “comprising” may mean “one,” but it is also consistent with the meaning of “one or more,” “at least one” and “one or more than one.”

[0096] The present description relates to modified viral structural protein and their production in a host or host cell. The modified viral structural protein comprises in series, an ectodomain, a transmembrane domain (TM) or portion of a TM, and a cytosolic tail (CT) domain or portion of a CT, wherein the ectodomain and the TM or portion of the TM are derived from a Coronaviridae, and the CT or portion of the CT is derived from an influenza hemagglutinin (HA) protein.

[0097] The modified viral structural protein may be a modified Coronavirus structural protein, wherein the cytosolic tail domain or portion of the cytosolic tail domain has been replaced with the cytosolic tail domain or portion of the cytosolic tail domain of an influenza hemagglutinin (HA) protein. For example, the modified viral structural protein may be a modified Coronavirus spike or surface (S) protein, wherein the cytosolic tail domain or portion of the cytosolic tail domain of the S protein has been replaced with the cytosolic tail domain or portion of the cytosolic tail domain of an influenza hemagglutinin (HA) protein.

[0098] The present disclosure provides modified viral structural protein, wherein the ectodomain and the transmembrane domain of the modified viral structural protein may be derived from the ectodomain and the transmembrane domain of a Coronavirus S protein and wherein the cytosolic tail domain is derived from the cytosolic tail domain of an influenza hemagglutinin (HA) protein.

[0099] The modified S-protein may be a chimeric modified S-protein or a chimeric S-protein. By “chimeric S-protein”, it is meant a protein or polypeptide that comprises

amino acid sequences and/or protein domains or portions of protein domains from two or more than two sources that are fused as a single polypeptide. For example but not limited to, the ectodomain and the transmembrane domain (TM) or portion of the TM of the chimeric S-protein may be derived from a first viral structural protein, for example a Coronavirus S protein, and the cytoplasmic tail (CT) or portion of the CT may be derived from a second viral structural protein, for example the CT may be derived from influenza HA. Furthermore, the ectodomain may be derived from a first viral structural protein for example a first Coronavirus S protein, the TM or portion of the TM may be derived from a second viral structural protein, for example a second Coronavirus S protein and the CT or portion of the CT may be derived from a third viral structural protein, for example the CT may be derived from influenza HA. Accordingly, the modified S-protein or chimeric S-protein may comprise a chimeric transmembrane and cytosolic tail domain (TMCT).

[00100] The modified coronavirus S-protein may comprise, in series,

- an ectodomain derived from a coronavirus S-protein,
- a transmembrane and cytosolic tail domain (TMCT), wherein the TMCT may be a chimeric TMCT, that may comprise:

- a transmembrane domain (TM), wherein the TM or a portion of the TM may be derived from a coronavirus S-protein and

- a cytosolic tail (CT), wherein the CT or a portion of the CT is derived from an influenza hemagglutinin (HA) protein.

[00101] The TM or portion of the TM may directly be fused or joined to the CT or portion of the CT or the TM or portion of the TM may be fused or joined to the CT or portion of the CT by an intervening peptide sequence.

[00102] Furthermore, the TM may be a chimeric TM that may comprise a N terminal sequence derived from the coronavirus S-protein TM and a C terminal sequence derived from the influenza HA protein TM. The CT may be a chimeric CT that may comprise a N terminal sequence derived from the coronavirus S-protein CT and a C terminal sequence derived from the influenza HA protein CT.

[00103] Accordingly, the chimeric TMCT, may comprise a native coronavirus S-protein TM, a chimeric coronavirus S-protein/influenza HA TM, a native influenza HA CT, a chimeric influenza HA/coronavirus S-protein CT or a combination thereof. The chimeric coronavirus S-protein/influenza HA TM comprises sequences from the TM of coronavirus S-protein and sequences from the TM of influenza HA. Similarly, the chimeric influenza HA/coronavirus S-protein CT comprises sequences from the CT of influenza HA and sequences from the CT of coronavirus S-protein.

[00104] A “chimeric transmembrane and cytosolic tail domain” or “chimeric TMCT” refers to a TMCT that is not native to the coronavirus S-protein TMCT. The chimeric TMCT comprises sequences that are not found together in nature. Thus the TMCT may comprise sequences that are heterologous to the ectodomain of the coronavirus S-protein. The term "heterologous" refers to a sequence or domain originating from different biological or synthetic sources. For example, the chimeric TMCT may comprise a TM or portion of a TM that is derived from the same coronavirus S-protein as the ectodomain, i.e. the TM may be homologous to the ectodomain of the S-protein or the TM or portion of the TM may be derived from a different viral TM, for example a TM from a different coronavirus S-protein as the ectodomain, i.e. the TM may be heterologous to the ectodomain of the S-protein. The CT or portion of the CT may be derived from a CT that is heterologous to the ectodomain, the TM, or both the ectodomain and the TM of the modified S-protein.

[00105] The coronavirus S protein, the modified S protein or the ectodomain and the transmembrane domain or portion of the transmembrane domain of the modified coronavirus S protein may be derived from any member of the Coronaviridae family of viruses. For example the coronavirus S-protein, the modified S-protein or the ectodomain and the transmembrane domain of the modified coronavirus S-protein may for example be derived from a Coronavirus, such as an Alphacoronavirus (Alpha-CoV), a Betacoronavirus (Beta-CoV), a Gammacoronavirus (Gamma-CoV) or a Deltacoronavirus (Delta-CoV). For example, the Coronavirus may be an Alphacoronavirus (Alpha-CoV) or a Betacoronavirus (Beta-CoV). The Alphacoronavirus may be a Duvinacovirus, such as for example HCoV-229E (229E-CoV), or may be a Setracovirus, such as for example HCoV-NL63. In a preferred embodiment, the Coronavirus is a Betacoronavirus (Beta-

CoV). The Betacoronavirus may be a lineage A Betacoronavirus, such as for example HCoV-OC43 (OC43-CoV) or HCoV-HKU1 (HKU1-CoV), a lineage B Betacoronavirus, such as for example SARS-CoV (also referred to as SARS-CoV-1) or SARS-CoV-2 and variants thereof or a lineage C Betacoronavirus, such as for example MERS-CoV.

5 [00106] The coronavirus S-protein, the modified S-protein or the ectodomain and the transmembrane domain or portion of the transmembrane domain of the modified coronavirus S-protein may further be derived from variants of the SARS-CoV-2 lineage, including but not limited to the B.1.1.7 strain ("Alpha" variant) (20I/501Y.V1, MW531680.1), the B.1.351 strain ("Beta" variant) (20H/501Y.V2), the P.1 strain  
10 ("Gamma" variant) (20J/501Y.V3), the B.1.617.2 strain ("Delta" variant), the B.1.525 strain, the B.1.429 strain (the "ETA" variant) or other variants of strains comprising mutations that arise naturally in the coronavirus S protein, or naturally occurring recombinant strains thereof.

[00107] In one embodiment the ectodomain and the transmembrane domain or  
15 portion of the transmembrane domain of the modified viral structural protein are derived from the spike protein (S) of a Coronavirus of the SARS-CoV-2 lineage (also referred to as SARS-CoV-2 variants). In other embodiments the ectodomain and the transmembrane domain or portion of the transmembrane domain of the modified viral structural protein are derived from the spike protein (S) of SARS-CoV-1, MERS-CoV, OC43-CoV or  
20 229E-CoV or variants thereof.

[00108] With reference to modified viral structural protein, the term "modified" as used herein may refer to the replacement of the cytoplasmic tail domain (CT) or portion of the CT in a structural protein from Coronaviridae with the CT or portion of the CT of a heterologous virus. For example a modified viral structural protein may be a Coronavirus  
25 S protein wherein the CT or portion of the CT of the S protein has been replaced with the CT or portion of the CT of influenza hemagglutinin (HA).

[00109] Therefore the modified viral structural protein may be a modified coronavirus spike (S) protein comprising a transmembrane domain (TM) or portion of a TM, and a cytosolic tail (CT) or portion of a CT, wherein the CT or portion of the CT  
30 may be derived from an influenza hemagglutinin (HA) protein and wherein the TM or

portion of the TM is heterologous to the CT or portion of the CT. Furthermore, the modified S protein comprises a transmembrane domain (TM) or portion of the TM, and a cytosolic tail (CT) or portion of the CT, wherein the CT or portion of the CT may be derived from an influenza hemagglutinin (HA) protein and wherein the CT or portion of the CT is heterologous to the TM or portion of the TM.

[00110] Therefore, in one aspect it is provided a modified coronavirus spike (S) protein comprising a transmembrane domain (TM) or portion of a TM, and a cytosolic tail (CT) or portion of a CT, wherein the CT or portion of the CT is derived from an influenza hemagglutinin (HA) protein and wherein the TM or portion of the TM is heterologous to the CT or portion of the CT. The modified coronavirus spike (S) protein is also referred to as modified S protein.

[00111] The cytoplasmic tail domain may also be referred to as “cytoplasmic tail”, “cytosolic tail”, “cytosolic tail domain”, “CT”, “CTD”, “cytoplasmic domain”, “cytoplasm domain”, “CP”, “CPD” or “C-terminal domain” and similar expressions. The cytoplasmic tail domain may also encompass portions of the cytoplasmic tail domain.

[00112] It has been found that the modified viral structural protein such as a modified S protein as disclosed herewith has improved characteristics as compared to the wild-type or unmodified viral structural protein (for example the S-protein). Examples of improved characteristics of the modified viral structural protein such as the modified S protein include but are not limited to: increased yield of the modified viral structural protein when expressed in a host or host cell as compared to the wild-type or unmodified viral structural protein; improved integrity, stability, or both integrity and stability, of the viral structural protein when expressed in a host or host cell as compared to the wild-type or unmodified viral structural protein; improved integrity, stability, or both integrity and stability, of virus like particles (VLPs) that are comprised of the modified viral structural protein as compared to the integrity, stability or integrity and stability of VLPs comprising to viral structural protein that does not comprise the modification as described herewith; increased yield of VLPs comprising modified viral structural protein when expressed in host cells as compared to the yield of VLPs that do not comprise the

modified viral structural protein that are expressed in same or substantially similar host cells.

[00113] Furthermore, methods of producing virus like particle (VLP) comprising modified viral structural protein such as the modified S protein in a host or host cell are also described. It has been observed that when VLPs are produced that comprise a modified viral structural protein such as the modified S protein wherein the native or wild-type CT has been replaced with a CT of influenza HA as described herein, the yield of VLP production in a host is increased compared to the yield of VLP that comprise viral structural protein that either i) comprise the native CT or ii) comprise a modified viral structural protein wherein the transmembrane domain (TM) and the CT have been replaced with the TM and the CT of an influenza HA.

[00114] The transmembrane domain may also be referred to as “TM” or “TMD”. The transmembrane and cytoplasmic tail domain may be referred to as TMCT or TM/CT.

[00115] Figure 3A shows that when a modified S protein (e.g. modified SARS-CoV-2 S-protein) was expressed in plants, the yield or protein accumulation (expressed as fold-change) of the modified S protein was increased approximately 2 fold when the native transmembrane and cytoplasmic tail (TMCT) was replaced with a TMCT from influenza HA (constructs 8592, 8595, and 8597) compared to the yield or protein accumulation of S protein with native TMCT (constructs 8586, 8589, and 8591). Furthermore, when a modified S protein (e.g. modified SARS-CoV-2 S-protein) wherein only the cytoplasmic tail (CT) was replaced with the CT of influenza HA (constructs 8610, 8611, and 8671) was expressed in plants, the protein accumulation of the modified S protein with the CT of influenza HA (expressed as fold-change), further increased between approximately 1.74 to 2.14 times, as compared to accumulation of modified S protein wherein the TMCT had been replaced with the TMCT of influenza HA. Correspondingly, the protein accumulation of the modified S protein with the CT of influenza HA increased between approximately 3.57 to 4.40 times, as compared to accumulation of S protein with the native transmembrane and cytoplasmic tail (wtTMCT).

[00116] Figure 3B shows that higher protein accumulation was observed for modified S protein (modified SARS-CoV-2 S-protein) with a cytoplasmic tail from influenza HA (H5i CT) when compared to protein accumulation of S protein with a wild-type TMCT (wt TMCT) or a modified S protein with the TMCT of influenza HA (H5i TMCT) from crude plant extract. Modified S protein with a cytoplasmic tail from influenza HA (H5i CT) is visible by Coomassie blue staining alone. The bands for modified S protein with a cytoplasmic tail from influenza HA (H5i CT) are more pronounced and thicker compared to the band of S protein with a wild-type TMCT (wt TMCT) or modified S protein with the TMCT of influenza HA (H5i TMCT) – see bands at about 150 kDa marked as S protein. Thickness of bands correspond to the amount of protein present, indicating that more protein accumulated for the H5i CT S protein. This higher protein accumulation was observed irrespective of the expression enhancer that was used.

[00117] As further discussed in more detail below, similar results were obtained, wherein the modified S-protein comprises a SARS-CoV-1 S protein with a cytoplasmic tail from influenza HA (see Figures 16A) or a MERS CoV S protein with a cytoplasmic tail from influenza HA (see Figure 19A).

[00118] Figure 3C shows S protein (SARS-CoV-2 S protein) accumulation by Western blot analysis of crude plant extract. When a modified S protein with a cytoplasmic tail from influenza HA (H5i CT) was expressed in plants, higher accumulation of modified S protein was observed compared to S protein with a wild-type TMCT (wt TMCT) and a modified S protein wherein both the transmembrane domain and the cytoplasmic tail (TMCT) domain have been replaced with the TMCT from influenza HA (H5i TMCT). The Western blot analysis in Figure 3C further shows that the SARS CoV-2 S-protein comprises both an S1 domain/subunit (top panel, detection with anti-SARS-CoV-2 S1 antibody) and an S2 domain/subunit (bottom panel, detection with an anti-SARS-CoV-2 S2 antibody) and has a molecular weight of about 150 kDa.

[00119] The present description provides a modified viral structural protein, wherein the modified viral structural protein may be a modified Coronavirus Spike or

Surface Protein (S protein). The modified S protein comprising, in series, an ectodomain, a transmembrane domain (TM) or portion of a TM, and a cytosolic tail (CT) domain or portion of a CT, wherein the ectodomain and the transmembrane domain are derived from Coronavirus, and the CT or portion of the CT is derived from a CT of influenza hemagglutinin (HA) protein. The ectodomain and the transmembrane domain or portion of the TM may be derived from the same Coronavirus. Therefore, the ectodomain and the transmembrane domain or portion of the TM of the modified structural protein are homologues (i.e. not heterologous) to each other, whereas the CT or portion of the CT is heterologous to the ectodomain and the transmembrane domain.

[00120] Furthermore, the transmembrane domain (TM) or portion of the TM of the modified S protein may be derived from a different Coronavirus than the ectodomain. Therefore, the TM or portion of the TM in the modified S protein may be heterologous (not homologous) to both the ectodomain and the CT domain or portion of the CT of the modified S protein. Similarly, the ectodomain may be heterologous (not homologous) to the TM or portion of the TM and the CT domain or portion of the CT of the modified S protein. For example, the ectodomain of the modified S protein may be derived from a first Coronavirus, the TM or portion of the TM may be derived from a second Coronavirus and the CT or portion of the CT may be derived from an influenza HA. The first Coronavirus and the second Coronavirus may belong to different Coronavirus families, sub-groups, types, subtypes, lineages or strains. The first Coronavirus and second Coronavirus may therefore be heterologous to each other and also each heterologous to the virus family from which the CT or portion of the CT is derived.

[00121] For example, the first Coronavirus from which the S protein ectodomain is derived, may be from any Coronavirus such for example an Alphacoronavirus (Alpha-CoV) or a Betacoronavirus (Beta-CoV). A non-limiting example of the first coronavirus from which the ectodomain of the S protein may be derived is a Duvinacovirus, such for example HCoV-229E, a Setracovirus, such for example HCoV-NL63. a lineage A Betacoronavirus, such for example HCoV-OC43 or HCoV-HKU1, a lineage B Betacoronavirus, such for example SARS-CoV or SARS-CoV 2 or a lineage C Betacoronavirus such for example MERS-CoV. The second Coronavirus, from which the

TM is derived, may belong to a different Coronavirus family, sub-group, type, subtype, lineage or strain than the first Coronavirus from which the ectodomain is derived. For example the second Coronavirus from which the S protein TM is derived, may be from any Coronavirus such for example an Alphacoronavirus (Alpha-CoV) or Betacoronavirus (Beta-CoV), as long as the second Coronavirus is heterologous to the first Coronavirus. A non-limiting example of the second coronavirus from which the TM of the S protein may be derived is a Duvinacovirus, such for example HCoV-229E (also referred to as 229E-CoV), a Setracovirus, such for example HCoV-NL63 (NL63-CoV), a lineage A Betacoronavirus, such for example HCoV-OC43 (also referred to as OC43-CoV) or HCoV-HKU1 (HKU1-CoV), a lineage B Betacoronavirus, such for example SARS-CoV (also referred to as SARS-CoV 1) or SARS-CoV 2 or a lineage C Betacoronavirus such for example MERS-CoV (also simply referred to as "MERS").

[00122] The domains in a Coronavirus S protein, such as the SARS-CoV-1 S-protein, SARS-CoV-2 S-protein, MERS CoV S-protein, OC43-CoV S-protein, or 229E-CoV S-protein, may readily be identified by methods known within the art. For example, domains such as transmembrane domains, may be identified by determining the degree of hydrophobicity of an amino acid sequence of the protein, for example using a transmembrane prediction program (e.g. Expert Protein Analysis System; ExPASy.org, operated by the Swiss Institute of Bioinformatics; or the Dense Alignment Surface Method, Cserzo M., et al. 1997, Prot. Eng. vol. 10, no. 6, 673-676; Lolkema J.S. 1998, FEMS Microbiol Rev. 22, no 4, 305-322), by determining the hydropathy profile of the amino acid sequence of the protein (e.g. Kyte-Doolittle Hydropathy Profile), by determining the three-dimensional protein structure and identifying the structure that is thermodynamically stable in a membrane (e.g. a single alpha helix, a stable complex of several transmembrane alpha helices, a transmembrane beta barrel, a beta-helix, or any other structure that is thermodynamically stable in a membrane).

[00123] Furthermore, domains within a Coronavirus S protein may be determined by comparison to known protein sequences for example by sequence alignment. Methods of alignment of sequences for comparison are well-known in the art and as further described below.

[00124] Domains and domain organization of Coronavirus S protein are well known and have been described. All Coronavirus spike proteins (S protein) share the same organization in two subunits or domains: a N-terminal subunit (or domain) named S1 that is responsible for receptor binding and a C-terminal S2 subunit (or domain) responsible for virus attachment, membrane fusion and virus entry.

[00125] Figure 1A shows a schematic representation of the Coronavirus S protein with its subunits and domains and the location of the S1/S2 and S2' cleavage sites. The S1 subunit is distal to the virus membrane and contains the receptor-binding domain (RBD) that mediates virus attachment to its host receptor. The S2 subunit contains fusion protein machinery, such as the fusion peptide, two heptad-repeat sequences (HR1 and HR2), a central helix typical of fusion glycoproteins and a transmembrane domain, and the cytosolic tail domain (see for example Kirchdoerfer et al. Nature 2016 Mar 3;531(7592):118-2, which is herein incorporated by reference).

[00126] The transmembrane domain (TM) and the cytoplasmic tail domain (CT) are positioned at the C-terminal end of the S2 subunit. While these domains are conserved in all coronaviruses (see Figure 1A and Corver et al. 2009, Virol J. 2009; 6: 230, which is herein incorporated by reference), different references, groups and authors have referred to different amino acid numbering with respect to these domains.

[00127] For example, amino acids (aa): 1214–1234 may be assigned to the TM and aa 1235-1273 may be assigned to the CT in the S protein of SARS-CoV-2 (see for example UniProtKB - P0DTC2 (SPIKE\_SARS2)). When aligning the sequence of SARS-CoV-2 (SEQ ID NO. 1) with the sequence of SARS-CoV-1 of Kirchdoerfer et al. (Nature 2016 Mar 3;531(7592):118-2) the SARS-CoV-2 TM corresponds to amino acids 1214-1236 and the SARS-CoV-2 CT corresponds to amino acids: 1237-1273.

[00128] For the purpose of this disclosure, the TM and CT of the native (unmodified) S protein corresponds to the following amino acids when aligned to a Coronavirus S protein reference sequence (SEQ ID NO: 1): TM: amino acids 1214-1234 and CT: amino acids: 1235-1273.

[00129] When 15 amino acids comprising the signal peptide (SP) are removed from the S protein, the TM corresponds to amino acids 1199-1219 of reference sequence SEQ ID NO: 2 and the CT corresponds to amino acids 1220-1258 of SEQ ID NO:2. (see also Table 1 for reference sequences and numbering).

5 [00130] The TM of Coronavirus S-protein has a highly conserved N-terminal aromatic rich stretch, followed by a hydrophobic sequence (see Figure 22 and Corver et al. Virology Journal volume 6, 230 (2009)). The consensus sequence of Coronavirus S-protein TM domain is:

WYXWLGFIAGLXAXXX{X}VXXXL (SEQ ID NO: 132), (wherein {X} may be absent).

10 [00131] For example, the Coronavirus S-protein TM domain consensus sequence may be:

WY[I/V]WLGFIAGL[V/I]A[L/I][A/V][L/M]{X}V[F/T][F/I]XL (SEQ ID NO: 133), (wherein {X} may be C or absent).

15 [00132] **Table 1.** Non-limiting examples of positions of TM and CT domains in modified S protein and corresponding amino acid positions in reference sequences.

| S Protein                                                                             | Transmembrane Domain (TM) | Cytoplasmic Tail Domain (CT) |
|---------------------------------------------------------------------------------------|---------------------------|------------------------------|
| <b>Modified S Protein<sup>1</sup></b><br><b>[SARS-CoV-2 H5iCT]</b><br>(SEQ ID NO: 21) | 1199-1219                 | 1220-1235                    |
| <b>SARS-CoV-2<sup>2</sup></b><br>(SEQ ID NO: 1)                                       | 1214-1234                 | 1235-1273                    |
| <b>SARS-CoV-2<sup>3</sup></b><br>(SEQ ID NO: 2)                                       | 1199-1219                 | 1220-1258                    |
| <b>SARS-CoV</b><br><b>(SARS-CoV-1)<sup>2</sup></b><br>(SEQ ID NO: 112)                | 1196-1216                 | 1217-1255                    |
| <b>SARS-CoV</b><br><b>(SARS-CoV-1)<sup>3</sup></b><br>(SEQ ID NO: 114)                | 1183-1203                 | 1204-1242                    |
| <b>MERS</b><br><b>(MERS-CoV)<sup>2</sup></b><br>(SEQ ID NO: 113)                      | 1297-1318                 | 1319-1353                    |
| <b>MERS</b><br><b>(MERS-CoV)<sup>3</sup></b><br>(SEQ ID NO: 115)                      | 1280-1301                 | 1302-1336                    |
| <b>OC43-CoV<sup>2</sup></b><br>(SEQ ID NO: 158)                                       | 1233-1325                 | 1326-1360                    |
| <b>OC43-CoV<sup>3</sup></b><br>(SEQ ID NO: 160)                                       | 1291-1311                 | 1312-1346                    |
| <b>229E-CoV<sup>2</sup></b><br>(SEQ ID NO: 159)                                       | 1116-1135                 | 1136-1173                    |

| S Protein                                 | Transmembrane Domain (TM) | Cytoplasmic Tail Domain (CT) |
|-------------------------------------------|---------------------------|------------------------------|
| 229E-CoV <sup>3</sup><br>(SEQ ID NO: 161) | 1100-1119                 | 1120-1157                    |

<sup>1</sup> numbering excludes signal peptide, CT is derived from influenza HA

<sup>2</sup> numbering includes signal peptide, CT is native

<sup>3</sup> numbering excludes signal peptide, CT is native

5 [00133] While there are differences in the numbering of the residues assigned to the TM and CT domain, a person of skill in the art will be able to determine the borders or boundaries of these domains in a Coronavirus S-protein by using known methods as for example described below.

10 [00134] In the modified Coronavirus S protein, the heterologous CT or portion of the CT that may be derived from influenza HA may be directly fused to the C-terminal end of the TM or portion of the TM of the Coronavirus S protein, or the heterologous CT or portion of the CT may be fused to the C-terminal end of the TM or portion of the TM of the Coronavirus S protein with an intervening peptide sequence (also referred to as a linker or linker sequence). Accordingly, the modified S-protein may comprise a  
15 intervening peptide, wherein the intervening peptide sequence fuses the CT or portion of the CT to the C-terminal end of the TM or portion of the TM.

[00135] The heterologous CT, portion of the CT or the intervening peptide sequence with the heterologous CT may be fused to an amino acid in the C-terminal portion of the TM domain (for example within 4 amino acids of the C-terminus of the  
20 TM domain as defined in Table 1 with reference to SEQ ID NO: 1, 2, 21, 114, 115, 160 or 161) or the N-terminal portion of the CT domain (for example within 4 amino acids of the N-terminus of the CT domain as defined in Table 1 with reference to SEQ ID NO: 1, 2, 21, 114, 115, 160 or 161).

25 [00136] For example, the Coronavirus TM may end at an amino acid residue that corresponds to any one of amino acids 1230-1238 of SEQ ID NO: 1. Accordingly, the C-terminal end of the Coronavirus TM may be an amino acid that corresponds to any one of amino acids 1230-1238 of SEQ ID NO: 1. In one example the Coronavirus TM may end at an amino acid residue that corresponds to amino acid 1230 in SEQ ID NO: 1. In another example, the TM may end at an amino acid residue that corresponds to amino

acid 1231 in SEQ ID NO: 1. In a further example, the TM may end at an amino acid residue that corresponds to amino acid 1232 in SEQ ID NO: 1. In another example, the TM may end at an amino acid residue that corresponds to amino acid 1233 in SEQ ID NO: 1. In a further example, the TM may end at an amino acid residue that corresponds to amino acid 1234 in SEQ ID NO: 1. In another example, the TM may end at an amino acid residue that corresponds to amino acid 1235 in SEQ ID NO: 1. In another example, the TM may end at an amino acid residue that corresponds to amino acid 1236 in SEQ ID NO: 1. In another example, the TM may end at an amino acid residue that corresponds to amino acid 1237 in SEQ ID NO: 1. In another example, the TM may end at an amino acid residue that corresponds to amino acid 1238 in SEQ ID NO: 1. In a preferred embodiment the TM may end at an amino acid residue that corresponds to amino acid 1234 in SEQ ID NO: 1.

[00137] In another example, the Coronavirus TM or portion of the TM may end at an amino acid residue that corresponds to any one of amino acids 1215-1219 of SEQ ID NO: 2 or 21. Accordingly, the C-terminal end of the Coronavirus TM or portion of the TM may be an amino acid that corresponds to any one of amino acids 1215-1224 of SEQ ID NO: 2 or 21. In one example the Coronavirus TM or portion of the TM may end at an amino acid residue that corresponds to amino acid 1215 in SEQ ID NO: 2 or 21. In another example, the TM or portion of the TM may end at an amino acid residue that corresponds to amino acid 1216 in SEQ ID NO: 2 or 21. In a further example, the TM or portion of the TM may end at an amino acid residue that corresponds to amino acid 1217 in SEQ ID NO: 2 or 21. In another example, the TM or portion of the TM may end at an amino acid residue that corresponds to amino acid 1218 in SEQ ID NO: 2 or 21. In another example, the TM or portion of the TM may end at an amino acid residue that corresponds to amino acid 1219 in SEQ ID NO: 2 or 21.

[00138] The intervening peptide sequence that may fuse the heterologous CT to the C-terminal end of the TM or portion of the TM from the Coronavirus S protein may have a length from 0-10 amino acids. Accordingly, the intervening peptide sequence may have a length of 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acids. The intervening peptide sequence may be derived from a Coronavirus protein, for example the intervening peptide sequence may be derived from the C-terminal end of the TM from a Coronavirus

S protein or from the N-terminal end of the CT of a Coronavirus S protein or both. The intervening peptide sequence may further be derived from an influenza HA protein, for example the intervening peptide sequence may be derived from the C-terminal end of the TM from influenza HA protein or from the N-terminal end of the CT of influenza HA protein, or both. Furthermore, the intervening peptide sequence may be heterologous to the Coronavirus and /or the HA portion of the modified S protein or the intervening peptide sequence may be an artificial sequence.

[00139] Non-limiting examples of sequences of the TM/CT domain (also referred to as chimeric TMCT) of the modified S protein are as shown below. The sequence of the TM domain from Coronavirus S-protein is underlined and the CT domain derived from influenza HA is shown in bold. Sequences in italic and bold are sequences derived from the TM of influenza HA. Sequences in italic and underlined are sequences derived from the CT of Coronavirus S-protein.

#### *SARS-CoV-2*

- WYIWLGFIAGLIAIVMTIML**SLWMC****SN****SLQCRICI** (SEQ ID NO: 18) (wtTM/H5iCT)
- WYIWLGFIAGLIAIVMTIM**MAGLSLWMC****SN****SLQCRICI** (SEQ ID NO: 19) (wtTM/H5iCT V1)
- WYIWLGFIAGLIAIVMTIM**AGLSLWMC****SN****SLQCRICI** (SEQ ID NO: 37) (wtTM/H5iCT V2)
- WYIWLGFIAGLIAIVMTIML**CC****MC****SN****SLQCRICI** (SEQ ID NO: 38) (wtTM/H5iCT V3)
- WYIWLGFIAGLIAIVMTIML**CC****SN****SLQCRICI** (SEQ ID NO: 39) (wtTM/H5iCT V4)
- WYIWLGFIAGLIAIVMTIML**S****FWMC****SN****SLQCRICI** (SEQ ID NO: 126) (wtTM/H1iCT)
- WYIWLGFIAGLIAIVMTIML**MWACQKGNIRCNICI** (SEQ ID NO: 127) (wtTM/H3iCT)
- WYIWLGFIAGLIAIVMTIML**GLWMC****SN****SMQCRICI** (SEQ ID NO: 128) (wtTM/H6iCT)
- WYIWLGFIAGLIAIVMTIML**VFICVKNGNMRCTICI** (SEQ ID NO: 129) (wtTM/H7iCT)
- WYIWLGFIAGLIAIVMTIML**LF****WAM****SN****SCRCNICI** (SEQ ID NO: 130) (wtTM/H9iCT)
- WYIWLGFIAGLIAIVMTIML**VVYMVSRDNVSCSICL** (SEQ ID NO: 131) (wtTM/BiCT)

#### *SARS-CoV-1*

- WYVWLGFIAGLIAIVMTIILL**SLWM** **CS****SN****SLQCRICI** (SEQ ID NO: 118) (wtTM/H5iCT)
- WYVWLGFIAGLIAIVMTIILL**CC****SN****SLQCRICI** (SEQ ID NO: 119) (wtTM/H5iCT V4)

- WYVWLGFIAGLIAIVMTILL**SFWMCSNGSLQCRICI** (SEQ ID NO: 120) (wtTM/H1cCT)

#### *MERS-CoV*

- WYIWLGFIAGLVALALCVFFIL**SLWMCSNGSLQCRICI** (SEQ ID NO: 123) (wtTM/H5iCT)

- WYIWLGFIAGLVALALCVFFILCC**SNGSLQCRICI** (SEQ ID NO: 124) (wtTM/H5iCT V4)

5 - WYIWLGFIAGLVALALCVFFIL**SFWMCSNGSLQCRICI** (SEQ ID NO: 125) (wtTM/H1cCT)

#### *OC43-CoV*

- WYVWLLICLAGVAMLVLLFFI**SLWMCSNGSLQCRICI** (SEQ ID NO: 164) (wtTM/H5iCT)

- WYVWLLICLAGVAMLVLLFFI**CCSNGSLQCRICI** (SEQ ID NO: 165) (wtTM/H5iCT V4)

- WYVWLLICLAGVAMLVLLFFI**SFWMCSNGSLQCRICI** (SEQ ID NO: 166) (wtTM/H1cCT)

10 *229E-CoV*

- WWVWLCISVVLIFVVSMLLL**SLWMCSNGSLQCRICI** (SEQ ID NO: 169) (wtTM/H5iCT)

- WWVWLCISVVLIFVVSMLLL**CCSNGSLQCRICI** (SEQ ID NO: 170) (wtTM/H5iCT V4)

- WWVWLCISVVLIFVVSMLLL**SFWMCSNGSLQCRICI** (SEQ ID NO: 171) (wtTM/H1cCT)

15 [00140] The modified coronavirus S-protein may comprise a chimeric TMCT. For example, the chimeric TMCT may comprise N-terminal sequences derived from coronavirus S-protein and C-terminal sequence derived from influenza HA protein as indicated in Table 1B. The TM may comprise the sequences as indicated in the column labeled as “S-protein TM sequences” and the CT may comprise the sequences as indicated in the column labeled as “HA protein CT sequence”. The CT and TM may be joined by the sequences as indicated in columns labeled as “S-protein CT sequence” and/or “HA protein TM sequence” (also referred to as intervening sequences or linker, as further described below).

20 [00141] **Table 1B:** Non-limiting examples of TM and CT sequences in modified S protein. The amino acid positions within the reference sequences are indicated.

| SEQ ID NO: | S-protein TM sequence | S-protein CT sequence | HA protein TM sequence | HA protein CT sequence |
|------------|-----------------------|-----------------------|------------------------|------------------------|
| 18         | 1-21                  | -                     | -                      | 22-37                  |
| 19         | 1-20                  | -                     | 21-24                  | 25-40                  |

|     |      |       |       |       |
|-----|------|-------|-------|-------|
| 37  | 1-20 | -     | 21-23 | 24-39 |
| 38  | 1-21 | 22-24 | -     | 25-36 |
| 39  | 1-21 | 22-23 | -     | 24-34 |
| 126 | 1-21 |       |       | 22-37 |
| 127 | 1-21 |       |       | 22-36 |
| 128 | 1-21 |       |       | 22-37 |
| 129 | 1-21 |       |       | 22-37 |
| 130 | 1-21 |       |       | 22-37 |
| 131 | 1-21 |       |       | 22-37 |
| 118 | 1-21 |       |       | 22-37 |
| 119 | 1-21 | 22-23 | -     | 24-34 |
| 120 | 1-21 |       |       | 22-37 |
| 123 | 1-22 |       |       | 23-38 |
| 124 | 1-22 | 23-24 | -     | 25-35 |
| 125 | 1-22 |       |       | 23-38 |
| 164 | 1-21 |       |       | 22-37 |
| 165 | 1-21 | 22-23 | -     | 24-34 |
| 166 | 1-21 |       |       | 22-37 |
| 169 | 1-20 |       |       | 21-36 |
| 170 | 1-20 | 21-22 | -     | 23-33 |
| 171 | 1-20 |       |       | 21-36 |

[00142] For example, the N-terminal sequence derived from coronavirus S-protein TM may comprise at least the following:

5                    –at least 19 amino acids corresponding to amino acids 1-19 of SEQ ID NO: 18, 19, 37, 38, 39, 118, 119, 123, 124, 164, 165, 169 or 170;

                    at least 20 amino acids corresponding to amino acids 1-20 of SEQ ID NO: 18, 19, 37, 38, 39, 118, 119, 123, 124, 164, 165, 169 or 170;

                    at least 21 amino acids corresponding to amino acids 1-21 of SEQ ID NO: 18, 19, 37, 38, 39, 118, 119, 123, 124, 164, 165, 169 or 170;

10                   at least 22 amino acids corresponding to amino acids 1-22 of SEQ ID NO: 18, 19, 37, 38, 39, 118, 119, 123, 124, 164, 165, 169 or 170;

                    at least 23 amino acids corresponding to amino acids 1-23 of SEQ ID NO: 18, 19, 37, 38, 39, 118, 119, 123, 124, 164, 165, 169 or 170;

at least 24 amino acids corresponding to amino acids 1-24 of SEQ ID NO: 18, 19, 37, 38, 39, 118, 119, 123, 124, 164, 165, 169 or 170.

5 [00143] The N-terminal sequence derived from the coronavirus S-protein TM may comprise at least 20 amino acids corresponding to amino acids 1-20 of SEQ ID NO: 18 or 169, or at least 21 amino acids corresponding to amino acids 1-21 of SEQ ID NO: 118 or 164, or at least 22 amino acids corresponding to amino acids 1-22 of SEQ ID NO: 123 and one or more than one amino acid from the C-terminal end of the influenza HA protein TM. The N-terminal sequence derived from the coronavirus S-protein TM may comprise at least 20 amino acids corresponding to amino acids 1-20 of SEQ ID NO: 18 or 10 169, or at least 21 amino acids corresponding to amino acids 1-21 of SEQ ID NO: 118 or 164, or at least 22 amino acids corresponding to amino acids 1-22 of SEQ ID NO: 123 and one or more than one amino acid from the C-terminal end of the influenza HA protein TM. The one or more than one amino acid from the C-terminal end of the influenza HA protein TM may comprise 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acids. For 15 example the one or more than one amino acid may be 2, 3 or 4 amino acids. The one or more than one amino acid from the C-terminal end of the influenza HA protein TM may be A, C, G, L, S, M, W or conserved substitution of A, C, G, L, S, M, W, or a combination thereof. In one example the one or more than one amino acid from the C-terminal end of the influenza HA protein TM may be selected from AG or conserved substitution of AG, AGL or conserved substitution of AGL, MAGL or conserved substitution of MAGL. 20

[00144] The modified coronavirus S-protein may also comprise a chimeric CT comprising a N-terminal sequence derived from the coronavirus S-protein CT and a C-terminal sequence derived from the influenza HA protein CT.

25 [00145] The N-terminal sequence derived from the coronavirus S-protein CT may comprise one or more than one amino acid. The one or more than one amino acid from the N terminal end of the coronavirus S-protein CT may comprise 0, 1, 2, 3, 4 or 5 amino acids. For example the one or more than one amino acid may be 1, 2 or 3 amino acids. The one or more than one amino acid from the N-terminal end of the coronavirus S-protein CT may be C or M or conserved substitutions of C or M. In one example the one 30

or more than one amino acids from the N-terminal end of the coronavirus S-protein may be selected from C or a conserved substitution of C, CC or a conserved substitution of CC, or CCM or a or a conserved substitution of CCM.

5 [00146] The C-terminal sequence derived from the influenza HA protein CT may comprise at least 11 amino acids corresponding to amino acids 27-37 of SEQ ID NO: 18. The N-terminal sequence derived from the influenza HA protein CT may further comprise at least 12 amino acids corresponding to amino acids 26-37 of SEQ ID NO: 18, at least 13 amino acids corresponding to amino acids 25-37 of SEQ ID NO: 18, at least 14 amino acids corresponding to amino acids 24-37 of SEQ ID NO: 18, at least 15 amino acids corresponding to amino acids 23-37 of SEQ ID NO: 18, or at least 16 amino acids corresponding to amino acids 22-37 of SEQ ID NO: 18.

[00147] In another example, the C-terminal sequence derived from the influenza HA protein CT may comprise at least the following:

- 15 - at least 11 amino acids corresponding to amino acids 27-37 of SEQ ID NO: 126, 127, 128, 129, 130 or 131;
- at least 12 amino acids corresponding to amino acids 26-37 of SEQ ID NO: 126, 127, 128, 129, 130 or 131;
- at least 13 amino acids corresponding to amino acids 25-37 of SEQ ID NO: 126, 127, 128, 129, 130 or 131;
- 20 - at least 14 amino acids corresponding to amino acids 24-37 of SEQ ID NO: 126, 127, 128, 129, 130 or 131;
- at least 15 amino acids corresponding to amino acids 23-37 of SEQ ID NO: 126, 127, 128, 129, 130 or 131; or
- 25 - at least 16 amino acids corresponding to amino acids 22-37 of SEQ ID NO: 126, 127, 128, 129, 130 or 131.

[00148] For example the CT may comprise the sequences as indicated in Table 1B (HA protein CT sequence). For example the CT may comprise amino acids 22-37 of SEQ ID NO: 18, 126, 128, 129, 130, 131, 118, 120, 164 or 166; or amino acids 25-40 of SEQ ID NO: 19; or amino acids 24-39 of SEQ ID NO: 37; or amino acids 25-36 of SEQ ID

NO: 38; or amino acids 24-34 of SEQ ID NO: 39 or 119; or amino acids 22-36 of SEQ ID NO: 127; or amino acids 22-37 of SEQ ID NO: 118 or 164; or amino acids 23-38 of SEQ ID NO: 123 or 125; or amino acids 25-35 of SEQ ID NO: 124; or amino acids 24-34 of SEQ ID NO: 165; or amino acids 21-36 of SEQ ID NO: 169; or amino acids 23-33 of SEQ ID NO: 170; or amino acids of SEQ ID NO: 21-36.

[00149] The influenza CT or portion of the CT may be fused or joined to the TM or portion of the TM of the S-protein with an intervening peptide sequence. For example, the intervening peptide sequence may be derived from the influenza CT, the S-protein TM or a combination thereof or the intervening peptide sequence may be an artificial sequence. The intervening peptide sequence may be of varying length. For example, the intervening peptide sequence may be 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acid long, preferably the intervening peptide sequence is between 2 and 8 amino acids long. In one example the intervening peptide sequence is 2 amino acid long and may for example comprise the sequence LC. In another example the intervening peptide sequence is 4 amino acids long and may for example comprise the sequence LCCM. In another example the intervening peptide sequence may be 5 amino acids long and may for example comprise the sequence LSLWM. In another example the intervening peptide sequence may be 7 amino acids long and may for example comprise the sequence AGLSLWM. In a further example the intervening peptide sequence may be 8 amino acids long and may for example comprise the sequence MAGLSLWM.

[00150] For example the TMCT of the modified S-protein may comprise the following sequence or a sequence that has 90-100%, or any amount therebetween sequence identity, or sequence similarity to:

-WYIWLGFIAGLIAIVMTIM - (X)<sub>n</sub> - CSNGSXXCXICI (SEQ ID NO: 64),  
 -WYVWLGFIAGLIAIVMTIL - (X)<sub>n</sub> - CSNGSXXCXICI (SEQ ID NO: 134), or  
 -WYIWLGFIAGLVALALCVFFIL - (X)<sub>n</sub> - CSNGSXXCXICI (SEQ ID NO: 135),  
 -WYVWLLICLAGVAMLVLLFFI - (X)<sub>n</sub> - CSNGSXXCXICI (SEQ ID NO: 172),

-WWWVWLCISVVLIFFVVSMLLL - (X)<sub>n</sub> - CSNGSXXCXICI (SEQ ID NO: 173),

wherein (X)<sub>n</sub> is the intervening peptide sequence, wherein the intervening peptide sequence may have a length from 0 to n amino acid residues, wherein n may be any length from 0-10, for example 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acids and wherein X may comprise any amino acid, for example X may be A, C, G, F, L, S, M, W or conserved substitution of A, C, G, F, L, S, M, W, or a combination thereof.

Non-limiting intervening peptide sequences (X)<sub>n</sub> may include the following:

- (X)<sub>0</sub>, the intervening peptide sequence is absent;
- (X)<sub>1</sub>, comprising for example the sequence L or C, or conserved substitution of L or C;
- (X)<sub>2</sub>, comprising for example the sequence LC, or conserved substitution of LC;
- (X)<sub>3</sub>, comprising for example the sequence LCC, or conserved substitution of LCC;
- (X)<sub>4</sub>, comprising for example the sequence LCCM, or conserved substitutions of LCCM, SLWM or a conserved substitution of SLWM, SFWM or a conserved substitution of SFWM;
- (X)<sub>5</sub>, comprising for example the sequence LSLWM, or conserved substitutions of LSLWM, LSLWM or a conserved substitution of LSLWM, LSFWM or a conserved substitution of LSFWM;
- (X)<sub>6</sub>, comprising for example the sequence GLSLWM, or conserved substitutions of GLSLWM;
- (X)<sub>7</sub>, comprising for example the sequence AGLSLWM, or conserved substitutions of AGLSLWM, or
- (X)<sub>8</sub>, comprising for example the sequence MAGLSLWM, or conserved substitutions of MAGLSLWM.

[00151] Figure 5B shows the fold change of protein accumulation of modified S protein with alternative versions of the C-terminal region with variable margin (intervening peptide sequence) between the SARS-COV-2 transmembrane (TM) domain and H5 A/Indonesia/5/05 HA cytosolic tail (CT) domain, wtTM/H5iCT V1 (SEQ ID NO:

19, product of construct 8980), wtTM/H5iCT V2 (SEQ ID NO: 37, product of construct 8981), wtTM/H5iCT V3 (SEQ ID NO: 38, product of construct #8982) and wtTM/H5iCT V4 (SEQ ID NO: 39, product of construct 8983), when expressed in plants, compared to the protein accumulation of a reference modified S protein wherein the cytoplasmic tail of Coronavirus S-protein has been replaced with the Coronavirus S-protein of H5 A/Indonesia/5/05 HA wtTM/H5iCT (SEQ ID NO: 18, product of construct 8671). All tested modified S proteins showed protein accumulation, with no statistically significant differences between the alternative versions and the wtTM/H5iCT reference control.

[00152] Similarly, a modified S protein comprising a SARS-CoV-1 S protein with a wtTM/H5iCT V4 version of the TMCT (Figure 16A) or a MERS S protein with a wtTM/H5iCT V4 version of the TMCT (Figure 19A), when expressed in plants, showed increased protein accumulation compared to protein accumulation of the wild type S proteins (wtTMCT) or S proteins wherein the TMCT has been replaced with the TMCT of H5 A/Indonesia/5/05 HA (H5iTMCT). Furthermore, a OC43 CoV S-protein with a wtTM/H5iCT V4 version of the TMCT when expressed in plants, showed increased protein accumulation compared to protein accumulation of the OC43 CoV S-protein with wild type TMCT (wtTMCT) (Figure 23 A).

[00153] Accordingly, the modified S protein may comprise a TM and CT domain (TM/CT), wherein the CT or a portion of the CT is fused to the C-terminal end of the TM or portion of the TM via an intervening peptide sequence, wherein the intervening peptide sequence comprises the sequence  $X_n$ .

[00154] Furthermore, the modified S protein may comprise a TM and CT domain (TM/CT) comprising a sequence having about 70, 75, 80, 85, 87, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100% or any amount therebetween sequence identity, or sequence similarity, with the sequence of SEQ ID NO: 18, 19, 37, 38, 39, 64, 126, 127, 128, 129, 130, 131, 118, 119, 120, 123, 124, 125, 134, 135, 164, 165, 166, 169, 170, 171, 172 or 173.

[00155] The modified S protein may comprise a CT or portion of the CT comprising a sequence having about 70, 75, 80, 85, 87, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100% or any amount therebetween sequence identity, or sequence similarity, with

amino acids 22-37 of SEQ ID NO:18, amino acids 21-40 of SEQ ID NO: 19, amino acids 21-39 of SEQ ID NO: 37, amino acids 25-36 of SEQ ID NO: 38 or amino acids 24-34 of SEQ ID NO: 39, amino acids 22-37 of SEQ ID NO:126, amino acids 22-36 of SEQ ID NO:127, amino acids 22-37 of SEQ ID NO:128, amino acids 22-37 of SEQ ID NO:129, amino acids 22-37 of SEQ ID NO:130, or amino acids 22-37 of SEQ ID NO:131.

[00156] The modified S protein may comprise a TM or portion of the TM comprising a sequence having about 70, 75, 80, 85, 87, 90, 91, 92, 93 94, 95, 96, 97, 98, 99, 100% or any amount therebetween sequence identity, or sequence similarity, with amino acids 1-20 of SEQ ID NO:18, amino acids 1-20 of SEQ ID NO: 19, amino acids 1-20 of SEQ ID NO: 37, amino acids 1-24 of SEQ ID NO: 38 , amino acids 1-23 of SEQ ID NO: 39, amino acids 1-21 of SEQ ID NO: 118, amino acids 1-23 of SEQ ID NO: 119, amino acids 1-22 of SEQ ID NO: 123, amino acids 1-24 of SEQ ID NO: 124, amino acids 1-21 of SEQ ID NO: 164, amino acids 1-23 of SEQ ID NO: 165, amino acids 1-20 of SEQ ID NO: 169, or amino acids 1-22 of SEQ ID NO: 170. Furthermore, the modified S protein as described herewith may comprise a TM or portion of TM that comprises from 80% to 100% identity with the sequence of SEQ ID NO: 132 or 133.

[00157] Furthermore, the modified the S-protein may comprise from 70% to 100% sequence identity, or sequence similarity, with the sequence of SEQ ID NO: 5, 59, 60, 61, 62, 95, 96, 97, 108, 109 or 110, for example the modified S protein may comprise a sequence having about 70, 75, 80, 85, 87, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100% or any amount therebetween sequence identity, or sequence similarity, with the sequence of SEQ ID NO: 5, 59, 60, 61, 62, 95, 96, 97, 108, 109 or 110.

[00158] The cytoplasmic tail domain (CT) or portion of the CT of a viral structural protein such as for example a Coronavirus S protein, may be replaced with the CT or portion of the CT from an influenza hemagglutinin (HA) as described below, the resulting protein is referred to as modified viral structural protein. Accordingly, a coronavirus S protein wherein the native CT or portion of the native CT has been replaced with the CT or portion of the CT from HA may be referred to as modified coronavirus S-protein or modified S-protein. As further described above the HA CT or

portion of the HA CT may either be directly fused to the N-terminal end of the Coronavirus TM domain or may be fused to the N-terminal end of the Coronavirus TM or portion of the TM via a intervening peptide sequence. Therefore, the HA CT or a portion of a HA CT may be fused to the C-terminal end of the S-protein TM or portion of the S-protein TM via an intervening peptide sequence.

[00159] Influenza “hemagglutinin” or “HA” is a homotrimeric membrane type I glycoprotein, generally comprising a signal peptide, an HA1 domain, and an HA2 domain comprising a membrane-spanning anchor site at the C-terminus and a small cytoplasmic tail (see for example Figure 1C and Figure 2). The amino acid sequences of HA from various influenza strains are well known within the art. Furthermore, amino acid sequences and nucleotide sequences encoding HA are well known and are available—see, for example, the BioDefence Public Health base (Influenza Virus; see URL: [biohealthbase.org](http://biohealthbase.org)) or National Center for Biotechnology Information (see URL: [ncbi.nlm.nih.gov](http://ncbi.nlm.nih.gov)), both of which are incorporated herein by reference. Exemplary amino acid sequences of HA cytoplasmic tail domains from different influenza strains are shown in Figure 2.

[00160] While different references and groups assign different length to the CT of HA, it has been shown that the N-terminal sequence of the CT is conserved among HA from different influenza subtypes and strains and that at least five residues have sequence identity for at least 10 of 13 HA subtypes (Simpson and Lamb 1992, Journal of Virology, 790-803). Figure 2 shows an alignment of amino acid sequences from exemplary influenza strains and conserved sequences in the N-terminal part of the HA protein. The consensus sequence of influenza cytoplasmic tail (CT) domain is:

XXWMCSNGSXXCXICI (SEQ ID NO: 15) (see also Fig. 2, C-terminal end of SEQ ID NO: 14)

[00161] CT sequences that correspond to the HA cytoplasmic tail domain consensus sequence may be fused to the C-terminal end of the TM of Coronavirus S protein either directly or via an intervening peptide sequence (linker sequence) as discussed above.

[00162] Furthermore, amino acid residues located in N-terminal or C-terminal from the native influenza HA TM/CT boundary may also be included in the CT sequence that is fused either directly or via an intervening peptide sequence to the TM or a portion of the TM of the modified Coronavirus S protein.

5 [00163] Therefore the sequence of the CT or a portion of the CT may for example start at an amino acid residue that corresponds to any one of amino acids 30-40 of SEQ ID NO: 14. Accordingly, the N-terminal end of the CT sequence may be an amino acid that corresponds to any one of amino acids 30-40 of SEQ ID NOs: 6, 7, 8, 9, 10, 11, 12, 13 or 14. In one example the CT sequence may start at an amino acid residue that  
10 corresponds to amino acid 30 in SEQ ID NOs: 6, 7, 8, 9, 10, 11, 12, 13 or 14. In another example, the CT sequence may start at an amino acid residue that corresponds to amino acid 31 of SEQ ID NOs: 6, 7, 8, 9, 10, 11, 12, 13 or 14. In a further example, the CT sequence may start at an amino acid residue that corresponds to amino acid 32 of SEQ ID NOs: 6, 7, 8, 9, 10, 11, 12, 13 or 14. In another example, the CT sequence may start at an  
15 amino acid residue that corresponds to amino acid 33 of SEQ ID NOs: 6, 7, 8, 9, 10, 11, 12, 13 or 14. In a further example, the CT sequence may start at an amino acid residue that corresponds to amino acid 34 of SEQ ID NOs: 6, 7, 8, 9, 10, 11, 12, 13 or 14. In another example, the CT sequence may start at an amino acid residue that corresponds to amino acid 35 of SEQ ID NOs: 6, 7, 8, 9, 10, 11, 12, 13 or 14. In a further example, the  
20 CT sequence may start at an amino acid residue that corresponds to amino acid 36 of SEQ ID NOs: 6-13 or 14. In another example, the CT sequence may start at an amino acid residue that corresponds to amino acid 37 of SEQ ID NOs: 6, 7, 8, 9, 10, 11, 12, 13 or 14. In a further example, the CT sequence may start at an amino acid residue that corresponds to amino acid 38 of SEQ ID NOs: 6-13 or 14. In another example, the CT  
25 sequence may start at an amino acid residue that corresponds to amino acid 39 of SEQ ID NOs: 6, 7, 8, 9, 10, 11, 12, 13 or 14. In a further example, the CT sequence may start at an amino acid residue that corresponds to amino acid 40 of SEQ ID NOs: 6, 7, 8, 9, 10, 11, 12, 13 or 14.

30 [00164] The cytoplasmic tail (CT) or portion of the CT of the modified S protein may be derived from a CT or portion of the CT of hemagglutinin (HA) of any one

influenza type, subtype or strain. For example the CT may be derived from an HA from influenza type A or influenza type B. For example the CT may be derived from an HA of influenza subtype H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, or H16. The CT may for example be derived from a HA of subtype H1, H2, H3, H5, H6, H7 or H9. Furthermore, the CT or portion of the CT may be derived from an HA of influenza type B. The type B influenza may be from the lineage B/Yamagata or B/Victoria.

[00165] For example, the CT or portion of the CT of the modified S protein may be derived from a CT of hemagglutinin (HA) influenza H1, H3, H5, H6, H7, H9 or B strain. Non limiting examples of influenza stains from which the HA CT might be derived are influenza H1 California/7/2009, H2 A/Singapore/1/1957, H3 A/Minnesota/41/2019, H5 A/Indonesia/5/05, H6 A/Teal/Hong Kong/W312/97, H7 A/Guangdong/17SF003/2016, H9 A/Hong Kong/1073/99 or B/Washington/02/2019. Non limiting examples of amino acid sequences of the HA CT are shown in Figure 2.

[00166] As shown in Figure 4A, when the native cytoplasmic tail (CT) of SARS-CoV-2 S protein was replaced with the CT from influenza HA H1 California/7/2009 (H1 Calif), H3 A/Minnesota/41/2019 (H3 Minn), H6 A/Teal/Hong Kong/W312/97 (H6 HK), H7 A/Guangdong/17SF003/2016 (H7 Guan), H9 A/Hong Kong/1073/99 (H9 HK) or B/Washington/02/2019 (B Wash), similar fold change in protein accumulation were observed for these modified SARS-CoV-2 S protein, when compared to SARS-CoV-2 S with a CT from H5 A/Indonesia/5/05 (H5 Indo). Western blot analysis confirmed these observations (see Figures 4B and 4C).

[00167] Similar results were obtained, when the native cytoplasmic tail (CT) of SARS-CoV-1 S protein, the native CT of MERS S protein, or the native CT of OC43 CoV S protein was replaced with the CT from influenza HA H1 California/7/2009 (H1cCT) (see Figures 16A, 19A, and 23A).

[00168] Accordingly, the cytoplasmic tail domain (CT) or portion of the CT may have about 70, 75, 80, 85, 86, 87, 88, 89, 90, 91, 92, 93 94, 95, 96, 97, 98, 99, 100% or any amount therebetween sequence identity, or sequence similarity, with the sequence of

5 SEQ ID NO: 15, or with amino acids 30-50 of SEQ ID NO 6, 7, 8, 9, 10, 12, 13, 14, or  
with amino acids 31-50 of SEQ ID NO 6, 7, 8, 9, 10, 12, 13, 14, or with amino acids 32-  
50 of SEQ ID NO 6, 7, 8, 9, 10, 12, 13, 14, or with amino acids 33-50 of SEQ ID NO 6,  
7, 8, 9, 10, 12, 13, 14, or with amino acids 34-50 of SEQ ID NO 6, 7, 8, 9, 10, 12, 13, 14,  
10 or with amino acids 35-50 of SEQ ID NO 6, 7, 8, 9, 10, 12, 13, 14, or with amino acids  
36-50 of SEQ ID NO 6, 7, 8, 9, 10, 12, 13, 14, or with amino acids 37-50 of SEQ ID NO  
6, 7, 8, 9, 10, 12, 13, 14, or with amino acids 38-50 of SEQ ID NO 6, 7, 8, 9, 10, 12, 13,  
14, or with amino acids 39-50 of SEQ ID NO 6, 7, 8, 9, 10, 12, 13, 14, or with amino  
acids 40-50 of SEQ ID NO 6, 7, 8, 9, 10, 12, 13, 14, or with amino acids 31-49 of SEQ  
10 ID NO 11, or with amino acids 32-49 of SEQ ID NO 11, or with amino acids 33-49 of  
SEQ ID NO 11, or with amino acids 34-49 of SEQ ID NO 11, or with amino acids 35-49  
of SEQ ID NO 11, or with amino acids 36-49 of SEQ ID NO 11, or with amino acids 37-  
49 of SEQ ID NO 11, or with amino acids 38-49 of SEQ ID NO 11, or with amino acids  
39-49 of SEQ ID NO 11, or with amino acids 548-568 of SEQ ID NO:3, or with amino  
15 acids 549-568 of SEQ ID NO:3, or with amino acids 550-568 of SEQ ID NO:3, or with  
amino acids 551-568 of SEQ ID NO:3, or with amino acids 552-568 of SEQ ID NO:3, or  
with amino acids 553-568 of SEQ ID NO:3, or with amino acids 554-568 of SEQ ID  
NO:3, or with amino acids 555-568 of SEQ ID NO:3, or with amino acids 556-568 of  
SEQ ID NO:3, or with amino acids 557-568 of SEQ ID NO:3, or with amino acids 558-  
20 568 of SEQ ID NO:3.

[00169] Furthermore, the modified S-protein may comprise from 70% to 100%  
sequence identity, or sequence similarity, with the sequence of SEQ ID NO: 5, 53, 54, 55,  
56, 57 or 58, for example the modified S protein may comprise a sequence having about  
70, 75, 80, 85, 87, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100% or any amount  
25 therebetween sequence identity, or sequence similarity, with the sequence of SEQ ID  
NO: 5, 53, 54, 55, 56, 57 or 58 or with amino acids 25-1259 of SEQ ID NO: 53, amino  
acids 25-1259 of SEQ ID NO: 54, amino acids 25-1259 of SEQ ID NO: 55, amino acids  
25-1259 of SEQ ID NO: 56, amino acids 25-1259 of SEQ ID NO: 57 or amino acids 25-  
1259 of SEQ ID NO: 58.

30 [00170] In further embodiments, the modified S protein ectodomain and/or  
transmembrane domain may be obtained from a coronavirus S protein other than SARS-

CoV-2 S protein, for example from SARS-CoV-1 S protein, MERS-CoV S protein, OC43-CoV S protein, 229E-CoV S protein and the like.

[00171] As shown in Figure 16A, higher protein accumulation was observed for modified SARS-CoV-1 S protein with a CT from influenza H5 HA (H5iCT), a CT from influenza H1 HA (H1cCT) or a CT from influenza H5 HA with a variable margin between the SARS-CoV-1 TM and the H5 HA CT ("H5iCT(V4)), when compared to protein accumulation of S protein with a wild-type TMCT (wt TMCT) or a modified S protein with the TMCT of influenza H5 HA (H5i TMCT) from crude plant extract. The modified S-proteins assembled into high molecular weight structures (see Figures 16B-16D), which were confirmed to be VLPs (see Figures 17A-17C). Although the amount of protein accumulation for SARS-CoV-1 protein with native TMCT was below detection limits of the Western Blot analysis when presented on the same gel with SARS-CoV-1 S-proteins with modified TMCT and/or CT (see Figure 16A), signals could be detected by Western Blot analysis when only SARS-CoV-1 protein with native TMCT was present on a gel (see Figure 16B) and VLPs were observed by electron microscopy (see Figure 17B).

[00172] Similar results were obtained for modified MERS-CoV S protein (see Figure 19A). Higher protein accumulation was observed for modified MERS-CoV S protein with a CT from influenza H5 HA (H5iCT), a CT from influenza H1 HA (H1cCT) or a CT from influenza H5 HA with a variable margin between the MERS-CoV TM and the H5 HA CT ("H5iCT(V4)), when compared to protein accumulation of S protein with a wild-type TMCT (wt TMCT) or a modified S protein with the TMCT of influenza H5 HA (H5i TMCT) (see protein band at approx. 175 kDa). The highest accumulation was observed for the modified MERS-CoV with an influenza H1 HA CT (H1cCT). The smaller band observed at approx. 100 kDa is most likely a proteolytic cleavage fragment of the S-protein. Without wishing to be bound by theory, it is believed that the replacement of the native CT with an influenza HA CT stabilizes the MERS S-protein and reduces cleavage of the S-protein.

[00173] As further shown in Figure 23A, low protein yields were observed in plants expressing OC43 CoV S-protein with a native OC43 CoV S-protein TMCT.

However, when the native OC43 CoV S-protein TMCT was replaced with a TMCT from influenza H5 HA (H5iTMCT), a CT from influenza H5 HA (H5iCT), a CT from influenza H1 HA (H1cCT) or a CT from influenza H5 HA with a variable margin between the OC43-CoV TM and the H5 HA CT ("H5iCT(V4)) higher protein accumulations were observed compared to the OC43 CoV S-protein with native TMCT (see bands at about 150 kDa. The larger band shown in the gel is believed to be a protein trimer). Similar results were observed with modified 229E-CoV S-protein (data not shown).

[00174] Furthermore, MERS-CoV S-protein, OC43-CoV S-protein, and 229E-CoV S-protein with a TMCT from influenza H5 HA (H5iTMCT), a CT from influenza H5 HA (H5iCT), or a CT from influenza H1 HA were observed to form VLPs as shown in Figures 19B-19F, 23B-23E, and 25A-25E.

[00175] The present disclosure therefore provides a "modified viral structural protein", a "viral structural fusion protein" or a "chimeric viral structural protein", wherein the ectodomain and the transmembrane domain (TM) of the viral structural protein or a portion of the TM are derived from a Coronavirus and the cytosolic tail (CT) or a portion of the CT is derived from an influenza protein. For example, the ectodomain and the transmembrane domain may be derived from a Coronavirus Spike (S) protein and the cytosolic tail (CT) or a portion of the CT may be derived from influenza HA protein. Modified S protein may comprise, in series i) an ectodomain derived from a coronavirus S-protein (comprising the S1 subunit and the FP, HR1 and HR2 domains of the S2 subunit), ii) a Coronavirus transmembrane domain (TM) or a portion of a Coronavirus TM and iii) an influenza HA cytoplasmic tail domain (CT) or a portion of a HA CT. Therefore, in the modified S protein, the CT or portion of the CT is heterologous to the TM and the ectodomain. Similarly, the TM (and the ectodomain) of the modified S protein are heterologous to the CT. The ectodomain and the transmembrane domain (TM) may be derived from the same Coronavirus (i.e. the ectodomain and the TM may be homologous to each other) or the ectodomain may be derived from a first Coronavirus and the TM may be derived from a second Coronavirus (i.e. the ectodomain and the TM are heterologous to each other).

[00176] By “chimeric protein”, or “chimeric polypeptide”, also referred to as a “fusion protein”, it is meant a protein or polypeptide that comprises amino acid sequences from two or more than two sources, for example but not limited to an ectodomain and a transmembrane domain derived from a first viral structural protein for example derived from Coronavirus S protein and a cytoplasmic tail (CT) derived from a second viral structural protein for example a CT from influenza HA, that are fused as a single polypeptide.

[00177] The modified coronavirus S-protein may comprise a transmembrane and cytosolic tail domain (TMCT), wherein the TMCT is a chimeric TMCT. The chimeric TMCT may comprise a transmembrane domain (TM), wherein the TM or a portion of the TM is derived from a coronavirus S-protein and a cytosolic tail (CT), wherein the CT or a portion of the CT is derived from an influenza hemagglutinin (HA) protein. Furthermore the chimeric TMCT may comprise a native coronavirus S-protein TM, a chimeric coronavirus S-protein/influenza HA TM, a native influenza HA CT, a chimeric influenza HA/coronavirus S-protein CT or a combination thereof. For example, the modified coronavirus S-protein may comprise a chimeric TMCT with a native influenza HA CT and a chimeric TM, wherein the chimeric TM comprises a N-terminal sequence which is derived from the TM of the coronavirus S-protein and a C-terminal sequence which is derived from the TM of influenza HA protein. In another example the modified coronavirus S-protein may comprise a chimeric TMCT with a native coronavirus S-protein TM and a chimeric CT, wherein the chimeric CT comprises a N-terminal sequence derived from the coronavirus S-protein and a C-terminal sequence derived from the influenza HA protein. In a further example, the modified coronavirus S-protein may comprise a chimeric TMCT with a chimeric TM, wherein the chimeric TM comprises a N-terminal sequence which is derived from the TM of the coronavirus S-protein and a C-terminal sequence which is derived from the TM of influenza HA protein and a chimeric CT, wherein the chimeric CT comprises a N-terminal sequence derived from the coronavirus S-protein and a C-terminal sequence derived from the influenza HA protein.

[00178] When referring to a modified S-protein or modified coronavirus spike (S)-protein in the present disclosure, it is meant a modified coronavirus spike (S)-protein comprising a transmembrane domain (TM) or portion of a S-protein TM, and a cytosolic

tail (CT) or a portion of a CT, wherein the CT is derived from an influenza hemagglutinin (HA) protein and wherein the TM is heterologous to the CT.

[00179] The modified the S-protein may comprise from 70% to 100% sequence identity, or sequence similarity, with the sequence of SEQ ID NO: 5, 21, 30, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 95, 96, 97, 108, 109, 110, 144, 145, 146, 155, 156 or 157, for example the modified S protein may comprise a sequence having about 70, 75, 80, 85, 87, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100% or any amount therebetween sequence identity, or sequence similarity, with the sequence of SEQ ID NO: 5, 21, 30, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 95, 96, 97, 108, 109, 110, 144, 145, 146, 155, 156 or 157, or with amino acids 25-1259 of SEQ ID NO: 47, amino acids 25-1259 of SEQ ID NO: 48, amino acids 25-1259 of SEQ ID NO: 49, amino acids 25-1259 of SEQ ID NO: 50, amino acids 25-1259 of SEQ ID NO: 51, amino acids 25-1259 of SEQ ID NO: 52, amino acids 25-1259 of SEQ ID NO: 53, amino acids 25-1259 of SEQ ID NO: 54, amino acids 25-1259 of SEQ ID NO: 55, amino acids 25-1259 of SEQ ID NO: 56, amino acids 25-1259 of SEQ ID NO: 57, amino acids 25-1259 of SEQ ID NO: 58, amino acids 25-1262 of SEQ ID NO: 59, amino acids 25-1261 of SEQ ID NO: 60, amino acids 25-1258 of SEQ ID NO: 61, amino acids 25-1256 of SEQ ID NO: 62, amino acids 25-1243 of SEQ ID NO: 95, amino acids 25-1240 of SEQ ID NO: 96, amino acids 25-1243 of SEQ ID NO: 97, amino acids 25-1341 of SEQ ID NO: 108, amino acids 25-1338 of SEQ ID NO: 109, amino acids 25-1341 of SEQ ID NO: 110, amino acids 25-1351 of SEQ ID NO: 144, amino acids 25- 1348 of SEQ ID NO: 145, amino acids 25-1351 of SEQ ID NO: 146, amino acids 25- 1159 of SEQ ID NO: 155, amino acids 25-1156 of SEQ ID NO: 156, or amino acids 25- 1159 of SEQ ID NO: 157.

[00180] The modified S-protein may further be produced or synthesized as modified S-protein precursor (also referred to as precursor S-protein), wherein the S-protein precursor comprises the modified S-protein and a signal peptide, wherein the signal peptide is native to Coronavirus (i.e. homologues to the ectodomain) or the signal peptide might be non-native or heterologous to the ectodomain. In a non-limiting example, the native signal peptide may be replaced with the signal peptide from protein disulfide isomerase (PDI).

[00181] The modified S-protein precursor may comprise a signal peptide that is non-native or heterologous to the ectodomain. The non-native signal peptide may replace the entire native signal peptide or may replace a portion of the native signal peptide of the Coronavirus S protein. Furthermore, the non-native or heterologous signal peptide may be directly fused to the N-terminus of the modified S protein or the non-native or heterologous signal peptide may be fused to the N-terminus of the modified S protein with an intervening peptide sequence.

[00182] A signal peptide (also referred to as signal sequence, targeting signal, localization signal, localization sequence, transit peptide, leader sequence or leader peptide) is a short peptide present at the N-terminus of the majority of newly synthesized proteins that are destined toward the secretory pathway. The signal peptide is responsible for targeting proteins to the endomembrane system, including the endoplasmic reticulum and the Golgi apparatus, where it is co-translationally removed by a signal peptidase located within the ER lumen and the mature proteins are generated. Since experimental methods for identification of targeting sequences are time-consuming and laborious, different computational approaches predicting targeting signals were developed, and are well known within the art. Signal peptides generally have low sequence similarity, but share some characteristic features. For predicting the signal sequence and its cleavage site, many prediction methods have been developed which take these characteristic features into account, such for example SignalP (Bendtsen et al., J Mol Biol. 2004 Jul 16;340(4):783-95.; Petersen et al., Nature Methods volume 8, pages785–786(2011), Signal-CF (Chou and Shen, Biochem Biophys Res Commun. 2007 Jun 8;357(3):633-40), and Signal-BLAST (Frank and Sippl, Bioinformatics, 2008 Oct 1;24(19):2172-6), which are herewith incorporated by reference.

[00183] Using the SignalP prediction program, a signal peptide cleavage site for the SARS-CoV-2 S protein is predicted between position 15 and 16 of the sequence corresponding to the sequence of SEQ ID NO:1. However, a signal peptide cleavage site for the SARS-CoV-2 S protein may be predicted or occur between other consecutive positions of the sequence corresponding to the sequence of SEQ ID NO:1. For example, a signal peptide cleavage site for the SARS-CoV-2 S protein may also be predicted or may

occur between position 13 and 14 of the sequence corresponding to the sequence of SEQ ID NO:1.

[00184] The N-terminal region of the native SARS-CoV-2 S protein (including the native signal peptide sequence) is shown below:

5 MFVFLVLLPLVSSQCVNLTTTRTQLPPAYTNS (SEQ ID NO: 63)

[00185] A predicted signal peptide sequence (SP) is underlined. The sequence shaded in grey corresponds to the sequence depicted in Table 2. The first amino acid residue of the mature SARS-CoV-2 S protein may be Valine (V) with its position designated as 1 (+1), which corresponds to V16 of the precursor S protein (native SARS-CoV-2 S protein with the native signal peptide). The first amino acid residue of the mature SARS-CoV-2 S protein may be at other residues of SEQ ID NO:1 or SEQ ID NO: 63 as indicated in Table 2. For example, the first amino acid residue of the mature SARS-CoV-2 S protein may be Glutamine (Q) with its position designated as 14 (-2).

**Table 2:** Portion of the SARS-CoV-2 S-protein sequence surrounding a signal peptide (SP) cleavage site. Numbering of residues is either from the N-terminus that includes the native signal peptide (# with SP) or from a predicted cleavage site at position V1 (# without SP), which is equivalent to V16 in the sequence that includes the signal peptide (precursor S protein).

|               | <b>S</b> | <b>S</b> | <b>Q</b> | <b>C</b> | <b>V</b> | <b>N</b> | <b>L</b> | <b>T</b> | <b>T</b> | <b>R</b> | <b>T</b> |
|---------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| #_with_SP     | 12       | 13       | 14       | 15       | 16       | 17       | 18       | 19       | 20       | 21       | 22       |
| #_without_SP* | -4       | -3       | -2       | -1       | 1        | 2        | 3        | 4        | 5        | 6        | 7        |

\*mature protein

[00186] Signal peptides or peptide sequences for directing localization of an expressed protein or polypeptide to the apoplast include, but are not limited to, a native (with respect to the protein) signal or leader sequence, or a heterologous signal sequence, for example but not limited to, a rice amylase signal peptide (McCormick 1999, Proc Natl Acad Sci USA 96:703-708) or a protein disulfide isomerase signal peptide (PDI). Therefore, as described herein, the modified S protein may be produced as precursor

protein comprising a modified S-protein and a heterologous amino acid signal peptide sequence. For example, the modified S protein precursor may comprise the signal peptide from Protein disulphide isomerase (PDI SP; nucleotides 32-103 of Accession No. Z11499).

5 [00187] The present disclosure therefore also provides for a modified S protein precursor comprising a modified S-protein and a native, or a non-native signal peptide, and nucleic acids encoding such protein.

[00188] The modified viral structural protein may be a modified S protein, wherein the modified S protein is a monomeric or single chain modified S protein. The  
10 monomeric or single chain modified S protein may include an S1 domain (subunit) and an S2 domain (subunit), wherein the S2 domain (subunit) has been modified to replace the native CT of the S protein with the CT of influenza HA protein and wherein the modified S protein is a single contiguous polypeptide chain. Monomeric or single chain modified S protein may trimerize to form a trimer, referred to as a trimeric modified S  
15 protein. A trimer is a macromolecular complex formed by three, usually non-covalently bound proteins.

[00189] The S protein is cleaved at a conserved activation cleavage site into 2 polypeptide chains, the S1 subunit and S2 subunit, which remain associated as S1/S2 protomers within the homotrimer. Without wishing to be bound by theory, the cleavage  
20 of the S protein into subunits may be important for virus infectivity, but it may not be essential for the trimerization of the protein.

[00190] The modified S protein may further comprise one or more than one substitution, replacement or mutation. For example, the modified S protein may comprise one or more than one substitution, replacement or mutation in the ectodomain to increase  
25 expression, yield, stability or to increase expression, yield and stability of the modified S protein in a suitable expression system.

[00191] For example the modified S protein, may comprise substitutions or mutations to the S1/S2 and/or S2' protease cleavage sites to prevent protease cleavage at

these sites. Therefore, when produced in a host or host cells, the modified S protein is not cleaved into separate S1 and S2 subunits or polypeptide chains.

[00192] The modified viral structural protein, such as the modified S protein, may further assemble into trimers of modified viral structural protein. It is therefore further provided a Coronavirus protein trimer comprising the modified S protein as described herein. The trimer may comprise single chain modified S protein wherein the single chain modified S protein comprises an S1 subunit and an S2 subunit, wherein the CT of the S2 subunit has been replaced with the CT of influenza hemagglutinin (HA).

[00193] The trimer may further be stabilized in a prefusion conformation. The modified viral structural protein, such as the modified S protein, therefore may further comprise one or more than one substitution, replacement or mutation to inhibit a conformational change in the S protein from the prefusion conformation to the post-fusion conformation, and thereby stabilizing the S protein or S protein trimer in the prefusion conformation.

[00194] By “amino acid substitution” or “substitution” it is meant the replacement of an amino acid in the amino acid sequence of a protein with a different amino acid. The terms amino acid, amino acid residue or residue are used interchangeably in the disclosure. One or more amino acids may be replaced with or substituted with one or more amino acids that are different than the original or wild-type amino acid at this position, without changing the overall length of the amino acid sequence of the protein.

[00195] For example, the modified viral structural protein, such as the modified S protein may be stabilized by proline substitutions, substitutions allowing the formation of disulfide bonds and salt bridges, and/or cavity-filling substitutions.

[00196] Hsieh et al. (Science 2020, 369 p. 1501-1505 which is incorporated herein by reference) designed and expressed a variety of SARS-CoV-2 spike protein variants in mammalian cells. An S protein variant with six proline substitutions, referred to as HexaPro, expressed 9.8X higher than S protein compared to variant that only had a double proline substitutions, had ~5°C increase in T<sub>m</sub>, and retained the trimeric prefusion conformation in mammalian cell lines. The HexaPro variant is considered the best variant by Hsieh et al.

[00197] In the current disclosure, the highest yields were observed with combinations of four proline substitutions corresponding to positions 802, 927, 971 and 972 (“4P”) of SEQ ID NO: 2 and an additional single amino acid substitution at position 923. Furthermore, higher yields were also observed with combinations of six proline substitutions corresponding to positions 802, 877, 884, 927, 971 and 972 (“6P”) and an additional single amino acid substitution at position 923.

[00198] As provided herewith, the modified S protein may further comprise one or more than one substitution, replacement or mutation to increase stability, yield or stability and yield of the modified protein in a host or cost cell, such for example in a plant or plant cells.

[00199] The modified S protein as described herein may comprise one or more than one mutation, modification, or substitution in its amino acid sequence at any one or more amino acid that corresponds to an amino acid within a reference sequence as described below.

[00200] By “correspond to an amino acid”, “corresponding to an amino acid” “or “corresponding to the sequence” and the like, it is meant that an amino acid (or nucleotide) corresponds to an amino acids (or nucleotide) in a sequence alignment with a reference Coronavirus sequence as described below. The corresponding amino acid positions in Coronavirus sequence may be determined by alignment to known sequences of Coronavirus S protein. Methods of alignment of sequences for comparison are well-known in the art and are further described below. Examples of corresponding amino acids are shown in Table 3.

[00201] **Table 3:** Positions of corresponding amino acid/residue position in Coronavirus S-proteins. (Reference sequences are indicated).

|                                                   | Amino acid position in S-protein (with ref. to indicated sequence) |     |     |     |     |     |     |      |      |
|---------------------------------------------------|--------------------------------------------------------------------|-----|-----|-----|-----|-----|-----|------|------|
| <b>SARS-CoV-2<sup>1</sup></b><br>(SEQ ID NO: 2)   | 667                                                                | 668 | 670 | 802 | 877 | 884 | 927 | 971  | 972  |
| <b>SARS-CoV-2<sup>2</sup></b><br>(SEQ ID NO: 1)   | 682                                                                | 683 | 685 | 817 | 892 | 899 | 942 | 986  | 987  |
| <b>SARS-CoV-1<sup>1</sup></b><br>(SEQ ID NO: 114) | 651                                                                | 652 | 654 | 786 | 861 | 868 | 911 | 955  | 956  |
| <b>SARS-CoV-1<sup>2</sup></b><br>(SEQ ID NO: 112) | 664                                                                | 665 | 667 | 799 | 874 | 881 | 924 | 968  | 969  |
| <b>MERS-CoV<sup>1</sup></b>                       | 730                                                                | 731 | 733 | 872 | 949 | 956 | 999 | 1043 | 1044 |

|                                                 | Amino acid position in S-protein (with ref. to indicated sequence) |     |     |     |     |     |      |      |      |
|-------------------------------------------------|--------------------------------------------------------------------|-----|-----|-----|-----|-----|------|------|------|
| (SEQ ID NO: 115)                                |                                                                    |     |     |     |     |     |      |      |      |
| <b>MERS-CoV<sup>2</sup></b><br>(SEQ ID NO: 113) | 747                                                                | 748 | 750 | 889 | 966 | 973 | 1016 | 1060 | 1061 |
| <b>OC43-CoV<sup>1</sup></b><br>(SEQ ID NO: 160) | 748                                                                | 749 | 751 | 898 | N/A | 976 | 1019 | 1063 | 1064 |
| <b>OC43-CoV<sup>2</sup></b><br>(SEQ ID NO: 158) | 762                                                                | 763 | 765 | 912 | N/A | 990 | 1033 | 1077 | 1078 |
| <b>229E-CoV<sup>1</sup></b><br>(SEQ ID NO: 161) | 551                                                                | 552 | 554 | 675 | N/A | 754 | 811  | 855  | 856  |
| <b>229E-CoV<sup>2</sup></b><br>(SEQ ID NO: 159) | 567                                                                | 568 | 570 | 691 | N/A | 770 | 827  | 871  | 872  |

1 numbering excludes signal peptide (SP)

2 numbering includes signal peptide (SP)

5 [00202] For example, the modified S protein may have one or more than one (for example two consecutive) proline substitutions at or near the boundary between a HR1 domain and a central helix domain that stabilize the S ectodomain trimer in the prefusion conformation, as described for example in WO 2018/081318, which is herein incorporated by reference. Furthermore, the one or more than one substitution may restrict and/or may prevent the processing or cleavage at the cleavage site between the S1 and the S2 subunit.

15 [00203] The modified S protein may comprise one or more than one substitution at a position as indicated in Table 3. For example the modified S protein may comprise one or more than one substitution at a position that corresponds to position 667, 668, 670, 802, 877, 884, 923, 927, 971, 972, or a combination thereof in reference sequence of SEQ ID NO: 2 (SARS-CoV-2). Corresponding positions in S-proteins of SARS-CoV-1, MERS-CoV, OC43-CoV and 229E-CoV are indicated in Table 3. Corresponding amino acid positions in S-protein from other Coronavirus may be determined by methods known within the art.

GSAS - 2P (971 and 972)

20 [00204] For example, the modified S protein may have one or more than one substitution at one or more than one amino acid corresponding to amino acid at positions 667, 668, 670, 971 or 972 of amino acid sequence of SEQ ID NO: 2.

[00205] In one aspect, the modified S protein may comprise a substitution, modification or mutation, corresponding to positions 667, 668, 670 or a combination thereof (numbering in accordance with SEQ ID NO: 2). For example, the amino acid corresponding to position 667 may be substituted for glycine (G) or a conserved substitution of glycine (G), the amino acid corresponding to position 668 may be substituted for serine (S) or a conserved substitution of serine (S), and the amino acid corresponding to position 670 may be substituted for serine (S) or a conserved substitution of serine (S).

[00206] The modified S protein may further comprise a substitution, modification or mutation, corresponding to positions 971, 972 or at positions 971 and 972 (numbering in accordance with SEQ ID NO: 2). For example, the amino acid corresponding to position 971 and/or 972 may be substituted for proline (P) or a conserved substitution of proline (P).

[00207] The modified S protein may comprise one or more than one substitution wherein the one or more than one substitutions comprise or consist of one or more than one substitution of an amino acid corresponding to amino acid at positions 667, 668, 670, 971, 972 of SEQ ID NO: 2. The modified S protein with one or more than one substitutions may be stabilized in a prefusion conformation. Furthermore, the modified S protein may form trimer that are stabilized in a prefusion conformation.

[00208] For example, the modified S protein may comprise the following substitutions (numbering in accordance with SEQ ID NO: 2): R667G, R668S, R670S (herein referred to as "GSAS"). The modified S protein may also have the following substitutions (numbering in accordance with SEQ ID NO: 2): K971P and V972P (herein referred to as "2P"). Furthermore the modified S protein may have the following substitutions (numbering in accordance with SEQ ID NO: 2): R667G, R668S, R670S, K971P and V972P (herein referred to as "GSAS-2P").

[00209] For example the modified S protein may have an amino acid sequence that has about 70, 75, 80, 85, 87, 90, 91, 92, 93 94, 95, 96, 97, 98, 99, 100% or any amount therebetween, sequence identity, or sequence similarity, with the amino acid sequence of SEQ ID NO: 47 sequence or with amino acids 25-1259 of SEQ ID NO: 47, wherein the

amino acid sequence has glycine (G) or a conserved substitution of glycine (G) at position 667, serine (S) or a conserved substitution of serine (S) at position 668, serine (S) or a conserved substitution of serine (S) at position 670, proline (P) or a conserved substitution of proline (P) at positions 971 and 972, wherein the modified S protein, when expressed, forms VLP.

[00210] In another example, the modified S protein may have one or more than one substitution at one or more than one amino acid corresponding to amino acid at positions 654, 955 or 956 of amino acid sequence of SEQ ID NO: 114 or at positions 730, 733, 1043 or 1044 of amino acid sequence of SEQ ID NO: 115.

[00211] For example, the modified S protein may comprise the following substitutions: R654A (numbering in accordance with SEQ ID NO: 114) or R730A and/or R733G (numbering in accordance with SEQ ID NO: 115). The modified S protein may also have the following substitutions: K955P and/or V956P (numbering in accordance with SEQ ID NO: 114) or V1043P and/or L1044P (numbering in accordance with SEQ ID NO: 115). Furthermore the modified S protein may have the following substitutions: R654A, K955P and V956P (numbering in accordance with SEQ ID NO: 114) or R730A, R733G, V1043P, L1044P (numbering in accordance with SEQ ID NO: 115).

GSAS-4P (802, 927, 971 and 972)

[00212] The modified S protein may further have substitution at amino acids corresponding to amino acid at positions 667, 668, and 670 and further one or more than one substitution at one or more than one residue corresponding to positions 802, 927, 971 and 972 (numbering in accordance with SEQ ID NO: 2). For example, the amino acid corresponding to positions 802, 927, 971 and 972 may be substituted for proline (P) or a conserved substitution of proline (P).

[00213] As shown in Figure 11A, modified S protein having the “GSAS” modifications and the following modifications: F802P, A927P, K971P, V972P (referred to as “GSAS-4P”, expressed from construct 8953) showed an increase of 2.47-fold increase in yield of modified S protein when compared to the yield of the “GSAS-2P” S protein (expressed from construct 8671).

[00214] Accordingly, the modified S protein may comprise one or more than one substitution wherein the one or more than one substitution comprise or consist of one or more than one substitution of an amino acid corresponding to amino acid at positions 667, 668, 670, 802, 927, 971 and 972 of SEQ ID NO: 2.

5 [00215] For example the modified S protein may have an amino acid sequence that has about 70, 75, 80, 85, 87, 90, 91, 92, 93 94, 95, 96, 97, 98, 99, 100% or any amount therebetween, sequence identity, or sequence similarity, with the amino acid sequence of SEQ ID NO: 48 or with amino acids 25-1259 of SEQ ID NO: 48, wherein the amino acid  
10 sequence has glycine (G) or a conserved substitution of glycine (G) at position 667, serine (S) or a conserved substitution of serine (S) at position 668, serine (S) or a conserved substitution of serine (S) at position 670, proline (P) or a conserved substitution of proline (P) at positions 802, 927, 971 and 972, wherein the modified S protein, when expressed, forms VLP.

15 [00216] In another example, the modified S protein may have one or more than one substitution at one or more than one amino acid corresponding to amino acid at positions 654, 786, 911, 955 or 956 of amino acid sequence of SEQ ID NO: 114 or at positions 730, 733, 872, 999, 1043 or 1044 of amino acid sequence of SEQ ID NO: 115.

[00217] For example, the modified S protein may comprise the following substitutions: R654A (numbering in accordance with SEQ ID NO: 114) or R730A and/or  
20 R733G (numbering in accordance with SEQ ID NO: 115). The modified S protein may also have the following substitutions: F786P, S911P, K955P and/or V956P (numbering in accordance with SEQ ID NO: 114) or A872P, N999P, V1043P and/or L1044P (numbering in accordance with SEQ ID NO: 115). Furthermore the modified S protein may have the following substitutions: R654A, F786P, S911P, K955P and V956P  
25 (numbering in accordance with SEQ ID NO: 114) or R730A, R733G, A872P, N999P, V1043P, L1044P (numbering in accordance with SEQ ID NO: 115).

GSAS-6P (802, 877, 884, 927, 971 and 972)

[00218] The modified S protein may further have substitution at amino acids corresponding to amino acid at positions 667, 668, and 670 and further one or more than  
30 one substitution at one or more than one residue corresponding to positions 802, 877,

884, 927, 971, and 972 (numbering in accordance with SEQ ID NO: 2). For example, the amino acid corresponding to position 802, 877, 884, 927, 971, and 972 may be substituted for proline (P) or a conserved substitution of proline (P) (numbering in accordance with SEQ ID NO: 2).

5 [00219] As shown in Figure 11A, modified S protein having the “GSAS” modifications and the following modifications: F802P, A877P, A884P, A927P, K971P, V972P (referred to as “GSAS-6P”, expressed from construct 8940) showed an increase of 2.11-fold increase in yield of S protein when compared to the yield of the “GSAS-2P” S protein (expressed from construct 8671).

10 [00220] Accordingly, the modified S protein may comprise one or more than one substitution wherein the one or more than one substitution comprise or consist of one or more than one substitution of an amino acid corresponding to amino acid at positions 667, 668, 670, 802, 877, 884, 927, 971 and 972 of SEQ ID NO: 2.

15 [00221] For example the modified S protein may have an amino acid sequence that has about 70, 75, 80, 85, 87, 90, 91, 92, 93 94, 95, 96, 97, 98, 99, 100% or any amount therebetween, sequence identity, or sequence similarity, with the amino acid sequence of SEQ ID NO: 49 or with amino acids 25-1259 of SEQ ID NO: 48, wherein the amino acid sequence has glycine (G) or a conserved substitution of glycine (G) at position 667, serine (S) or a conserved substitution of serine (S) at position 668, serine (S) or a  
20 conserved substitution of serine (S) at position 670, proline (P) or a conserved substitution of proline (P) at position 802, 802, 877, 884, 927, 971 and 972, wherein the modified S protein when expressed forms VLP.

[00222] In another example, the modified S protein may have one or more than one substitution at one or more than one amino acid corresponding to amino acid at  
25 positions 654, 786, 861, 868, 911, 955 or 956 of amino acid sequence of SEQ ID NO: 114 or at positions 730, 733, 872, 949, 956, 999, 1043 or 1044 of amino acid sequence of SEQ ID NO: 115.

[00223] For example, the modified S protein may comprise the following substitutions: R654A (numbering in accordance with SEQ ID NO: 114) or R730A and/or  
30 R733G (numbering in accordance with SEQ ID NO: 115). The modified S protein may

also have the following substitutions: F786P, A861P, A868P, S911P, K955P and/or V956P (numbering in accordance with SEQ ID NO: 114) or A872P, S949P, A956P, N999P, V1043P and/or L1044P (numbering in accordance with SEQ ID NO: 115).

Furthermore the modified S protein may have the following substitutions: R654A, F786P, A861P, A868P, S911P, K955P and V956P (numbering in accordance with SEQ ID NO: 114) or R730A, R733G, A872P, S949P, A956P, N999P, V1043P and L1044P (numbering in accordance with SEQ ID NO: 115).

Substitution at position 923

[00224] The modified S protein as described herewith may further comprise a substitution, modification, or mutation, corresponding to position 923 (numbering in accordance with SEQ ID NO: 2). For example the amino acid corresponding to position 923 may be substituted for phenylalanine (F) or a conserved substitution of phenylalanine (F).

[00225] As shown in Figure 11B, modified S protein having the “GSAS-2P” modifications and a L923F substitution (expressed from construct 8933) showed an increase of 1.36-fold yield of S protein when compared to the yield of the “GSAS-2P” S protein without the L923F substitution (expressed from construct 8671). The modified S protein having the “GSAS-4P” modifications and a L923F substitution (expressed from construct 8960) showed an increase of 2.88-fold in yield of the modified S protein when compared to the yield of the “GSAS-2P” S protein without the L923F substitution (expressed from construct 8671). The modified S protein having the “GSAS-6P” modifications and a L923F substitution (expressed from construct 8947) showed an increase of 2.47-fold in yield when compared to the yield of the “GSAS-2P” S protein without the L923F substitution (expressed from construct 8671).

[00226] Accordingly, the modified S protein may comprise one or more than one substitution wherein the one or more than one substitution comprises or consists of one or more than one substitution of an amino acid corresponding to amino acids at positions 667, 668, 670, 927, 971, 972, 802, 877, 884, 923 or a combination thereof of SEQ ID NO: 2. For example the modified S-protein may comprise one or more than one substitution wherein the one or more than one substitution comprises or consists of one or

more than one substitution of an amino acid corresponding to amino acids at positions 667, 668, 670, 971, 972, 923, or a combination thereof of SEQ ID NO: 2 (GSAS-2P-923), 667, 668, 670, 927, 971, 972, 802 923, or a combination thereof of SEQ ID NO: 2 (GSAS-4P-923) or 667, 668, 670, 927, 971, 972, 802, 877, 884, 923 or a combination thereof of SEQ ID NO: 2 of SEQ ID NO: 2 (GSAS-6P-923).

[00227] For example, the modified S protein may comprise one or more than one substitution wherein the one or more than one substitution comprises or consists of one or more than one substitution of an amino acid corresponding to amino acids at positions

- 667, 668, 670, 971, 972 and 923 of SEQ ID NO: 2 (GSAS-2P-923),
- 667, 668, 670, 927, 971, 972, 802 and 923 of SEQ ID NO: 2 (GSAS-4P-923) or
- 667, 668, 670, 927, 971, 972, 802, 877, 884 and 923 of SEQ ID NO: 2 (GSAS-6P-923).

[00228] For example, the modified S protein may have an amino acid sequence that has about 70, 75, 80, 85, 87, 90, 91, 92, 93 94, 95, 96, 97, 98, 99, 100% or any amount therebetween, sequence identity, or sequence similarity, with the amino acid sequence of SEQ ID NO: 50 or amino acids 25-1259 of SEQ ID NO: 50, wherein the amino acid sequence has glycine (G) or a conserved substitution of glycine (G) at position 667, serine (S) or a conserved substitution of serine (S) at positions 668 and 670, proline (P) or a conserved substitution of proline (P) at positions 971 and 972, and phenylalanine (F) or a conserved substitution of phenylalanine (F) at position 923; SEQ ID NO: 51 or amino acids 25-1259 of SEQ ID NO: 51, wherein the amino acid sequence has glycine (G) or a conserved substitution of glycine (G) at position 667, serine (S) or a conserved substitution of serine (S) at positions 668 and 670, proline (P) or a conserved substitution of proline (P) at positions 927, 971, 972 and 802, and phenylalanine (F) or a conserved substitution of phenylalanine (F) at position 923; or SEQ ID NO: 52 or amino acids 25-1259 of SEQ ID NO: 52, wherein the amino acid sequence has glycine (G) or a conserved substitution of glycine (G) at position 667, serine (S) or a conserved substitution of serine (S) at positions 668 and 670, proline (P) or a conserved substitution of proline (P) at positions 927, 971, 972, 802, 877 and 884, and phenylalanine (F) or a conserved substitution of phenylalanine (F) at position 923, wherein the modified S protein when expressed forms VLP.

[00229] Accordingly, it is provided a modified coronavirus S-protein which may comprise:

1. a chimeric transmembrane and cytoplasmic tail domain (TMCT);

5 2. one or more than one substitution corresponding to amino acid at positions 667, 668, and/or 670 (numbering in accordance with SEQ ID NO: 2) when compared to a corresponding wildtype coronavirus S-protein;

3. one or more than one substitution corresponding to amino acid at positions 802, 877, 884, 927, 971, and/or 972 (numbering in accordance with SEQ ID NO: 2) when compared to a corresponding wildtype coronavirus S-protein;

10 4. a substitution corresponding to position 923 (numbering in accordance with SEQ ID NO: 2) when compared to a corresponding wildtype coronavirus S-protein;

5. or a combination of the modifications and/or substitutions as described under 1-4.

[00230] As used herein, the term “conserved substitution” or “conservative substitution” and grammatical variations thereof, refers to an amino acid that is different from an reference amino acid (substitution), but is in the same class of amino acid as the described substitution or described residue (i.e., a nonpolar residue replacing a nonpolar residue, an aromatic residue replacing an aromatic residue, a polar-uncharged residue replacing a polar-uncharged residue, a charged residue replacing a charged residue).  
15 Further information about conservative substitutions can be found, for instance, in Sahin-Toth et al. (Protein ScL, 3:240-247, 1994), Hochuli et al (Bio/Technology, 6:1321-1325, 1988) Henikoff S, and Henikoff JG (Proc. Natl. Acad. Sci. USA 89: 10915-10919, 1992) and in widely used textbooks of genetics and molecular biology.

[00231] The modified viral structural protein may further be glycosylated.  
25 Coronavirus S protein, Coronavirus M protein and Coronavirus E protein are glycosylated and both N-linked glycosylation and O-linked glycosylation occur.

[00232] The modified viral structural protein may comprise glycosylation pattern that are unique to the host or host cell in which the modified viral structural protein is expressed. For example, when expressed in plants or plant cells, the modified viral

structural protein may comprise plant-specific N-glycans. Therefore, it is also provided modified viral structural protein having plant specific N-glycans.

[00233] As described herein, the cytosolic tail domain (CT) of the modified viral structural protein may be replaced with the CT from influenza hemagglutinin (HA). The ectodomain and the transmembrane domain (TM) of the viral structural protein as described above are fused to an influenza HA cytosolic tail domain (CT) such that the CT is heterologous with respect to the ectodomain and the transmembrane domain of the viral structural protein, such as the S protein. The modified S protein may self-assemble into virus-like particles (VLPs).

[00234] The present description therefore further relates to virus-like particles (VLPs). More specifically, the present description is directed to VLPs comprising modified viral structural proteins such as modified S-protein, and methods of producing VLPs with modified viral structural proteins such as modified S-protein in a host or host cell. The VLPs comprise a modified viral structural protein such as modified S-protein as described herewith.

[00235] As shown in Figure 6C, 17A, 17B, and 17C, modified viral structural protein as exemplified by a modified S protein (modified SARS-CoV-2 or modified SARS-CoV-1 S protein), wherein the native or wild-type CT has been replaced by a CT from influenza HA protein self-assemble into VLPs when expressed in plants. The VLPs are similar to VLPs produced with a S protein with native TM/CT sequence (see Figures 6A and 17A) or modified S protein with H5 influenza TM/CT sequence (see Figures 6B and 17B) in the same plant expression system.

[00236] Furthermore, as shown in Figures 6D, 6E, 6F and 6G, modified S protein with variable margin or boundaries (intervening peptide sequence) between the TM and influenza CT domain also self-assemble into VLPs when expressed in plants.

[00237] In addition, as shown in Figures 6H, 6I, 6J, 6K, 6L and 6M, modified S protein, wherein the native or wild-type CT has been replaced by a CT from influenza HA protein from H1, H3, H6, H7, H9 and B influenza, respectively, also self-assemble into VLPs when expressed in plants.

[00238] Furthermore, as shown in Figures 19B-19F, 23B-23E, and 25A-25E, modified S-protein derived from MERS-CoV, OC43-CoV, and 229E-CoV, wherein the modified S-protein has a TMCT from influenza H5 HA (H5iTMCT), a CT from influenza H5 HA (H5iCT), or a CT from influenza H1 HA also formed VLPs.

5 [00239] The term virus-like particle" (VLP), or "virus-like particles" or "VLPs" refers to virus-like structures that are generally morphologically and antigenically similar to virions produced in an infection, but lack genetic information sufficient to replicate and thus are non-infectious. VLPs are structures that self-assemble and comprise one or more structural proteins such as for example modified viral structural proteins, for  
10 example but not limited to a modified S protein. Therefore, the VLP may comprise modified S protein. The VLP may further comprise viral structural proteins, wherein the viral structural proteins consist of modified S protein. Therefore, in some embodiments the VLP may lack or be free of the Coronavirus M protein and/or Coronavirus E protein. In some embodiments the VLPs produced from the modified viral structural protein as  
15 described herewith, therefore do not comprise a Coronavirus M protein, a Coronavirus E protein or Coronavirus M protein and Coronavirus E protein. Furthermore, in some embodiment the VLP do not comprise structural or non-structural proteins from viruses that are heterologous to *Coronaviridae* or influenza virus, for example the VLP do not comprise structural and non-structural protein from viruses that are not from  
20 *Coronaviridae*.

[00240] In another embodiment the VLP may comprise Coronavirus E protein, Coronavirus M protein and modified Coronavirus S protein. In another embodiment the VLP may comprise Coronavirus E protein and modified Coronavirus S protein. In another embodiment the VLP may comprise Coronavirus M protein and modified  
25 Coronavirus S protein. Furthermore, the VLP may comprise Coronavirus E protein, modified Coronavirus M protein and modified Coronavirus S protein. The VLP may further comprise modified Coronavirus E protein, modified Coronavirus M protein and modified Coronavirus S protein. In another embodiment the VLP may comprise modified Coronavirus E protein and modified Coronavirus S protein. In another embodiment the  
30 VLP may comprise modified Coronavirus M protein and modified Coronavirus S protein.

[00241] VLPs may be produced in suitable host or host cells including plants and plant cells. Following extraction from the host or host cell and upon isolation and further purification under suitable conditions, VLPs may be recovered as intact structures.

5 [00242] The VLPs may be purified or extracted using any suitable method for example chemical or biochemical extraction. VLPs are relatively sensitive to desiccation, heat, pH, surfactants and detergents. Therefore it may be useful to use methods that maximize yields, minimize contamination of the VLP fraction with cellular proteins, maintain the integrity of the proteins, or VLPs, and, where required, the associated lipid envelope or membrane, methods of loosening the cell wall to release the proteins, or  
10 VLP. Minimizing or eliminating the use of detergent or surfactants such for example SDS or Triton<sup>TM</sup> X-100 may be beneficial for improving the yield of VLP extraction. VLPs may be then assessed for structure and size by, for example, electron microscopy (see Figure 4B), or by size exclusion chromatography.

15 [00243] For enveloped viruses, such as Coronavirus, it may be advantageous for a lipid layer or membrane to be retained by the virus. The composition of the lipid may vary with the system (e.g. a plant-produced enveloped virus would include plant lipids or phytosterols in the envelope), and may contribute to an improved immune response.

[00244] Therefore, the VLPs that are produced in a host or host cell, may comprise lipids from the plasma membrane of the host or host cell. For example VLPs produced in  
20 plants may contain lipids of plant origin ("plant lipids"), VLPs produced in insect cells may comprise lipids from the plasma membrane of insect cells (generally referred to as "insect lipids"), and VLPs produced in mammalian cells may comprise lipids from the plasma membrane of mammalian cells (generally referred to as "mammalian lipids").

25 [00245] The plant lipids or plant-derived lipids may be in the form of a lipid bilayer, and may further comprise an envelope surrounding the VLP. The plant-derived lipids may comprise lipid components of the plasma membrane of the plant where the VLP is produced, including phospholipids, tri-, di- and monoglycerides, as well as fat-soluble sterol or metabolites comprising sterols. Examples include phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylinositol, phosphatidylserine,  
30 glycosphingolipids, phytosterols or a combination thereof. Examples of phytosterols

include campesterol, stigmasterol, ergosterol, brassicasterol, delta-7-stigmasterol, delta-7-avenasterol, daunosterol, sitosterol, 24-methylcholesterol, cholesterol or beta-sitosterol.

As one of skill in the art would understand, the lipid composition of the plasma membrane of a cell may vary with the culture or growth conditions of the cell or organism, or species, from which the cell is obtained. Generally, beta-sitosterol is the most abundant phytosterol.

[00246] Without wishing to be bound by theory, plant-made VLPs comprising plant derived lipids, may induce a stronger immune reaction than VLPs made in other manufacturing systems and the immune reaction induced by these plant-made VLPs may be stronger when compared to the immune reaction induced by live or attenuated whole virus vaccines.

[00247] Furthermore, in addition to the potential adjuvant effect of the presence of plant lipids, the ability of plant N-glycans to facilitate the capture of glycoprotein antigens by antigen presenting cells, may be advantageous of the production of VLPs in plants.

[00248] The VLP produced within a plant may comprise a modified viral structural protein comprising plant-specific N-glycans. Therefore, this disclosure also provides for a VLP comprising modified viral structural protein having plant specific N-glycans. Furthermore, it is provided VLP comprising plant lipids and modified viral structural protein having plant specific N-glycans.

[00249] Methods of producing virus like particle (VLP) comprising modified structural protein in a host or host cell are also provided. Furthermore, methods of increasing yield of production of virus like particle (VLP) comprising modified structural protein in a host or host cell are also provided. The methods comprise the introduction of a nucleic acid comprising a sequence that encodes a modified structural protein into the host or host cell, and incubating the host or host cell under conditions that permit the expression of the nucleic acid, thereby producing the VLP. The modified viral structural protein may be produced at a higher yield compared to a host or host cell expressing the unmodified viral structural protein.

[00250] For example, as shown in Figure 3A, yields of VLPs expressed in plants may be increased when the cytoplasmic tail (CT) of a viral structural protein is replaced with the CT of influenza HA to produce a modified viral structural protein, such for example a modified S protein. As further shown in Figures 11A and 11B, when the modified S protein further comprises one or more than one substitution wherein the one or more than one substitution comprise or consist of one or more than one substitution of an amino acid corresponding to amino acid at positions 667, 668, 670, 802, 923, 927, 971 and/or 972 of SEQ ID NO: 2, yield of VLPs comprising the modified S protein when expressed in plants, may be further increased.

[00251] The yield of the modified viral structural protein (such as modified S protein) or the yield of a VLP comprising modified viral structural protein produced in a host or host cell, such for example a plant or plant cells, may be increased by 1.1- 10 fold, or any amount therebetween when compared to the yield of a corresponding unmodified viral structural protein or the yield of VLP that comprises the corresponding unmodified viral structural protein. For example the yield of the modified viral structural protein (such as modified S protein) or the yield of a VLP (comprising the modified viral structural protein) in a host or host cell may be increased by 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 4, 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.6, 5.7, 5.8, 5.9, 6, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, 7, 7.1, 7.2, 7.3, 7.4, 7.5, 7.6, 7.7, 7.8, 7.9, 8, 8.1, 8.2, 8.3, 8.4, 8.5, 8.6, 8.7, 8.8, 8.9, 9, 9.1, 9.2, 9.3, 9.4, 9.5, 9.6, 9.7, 9.8, 9.9, 10 fold or any amount therebetween, compared to the yield of a corresponding unmodified viral structural protein or the yield of a VLP wherein the VLP comprises a corresponding unmodified viral structural protein, when produced in a host or host cell under identical conditions.

[00252] The modified viral structural protein described herewith includes modified S proteins with amino acid sequences that have about 70, 75, 80, 85, 87, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100% or any amount therebetween, sequence identity, or sequence similarity, with the amino acid sequence of SEQ ID NO: 1, 2, 5, 21, 30, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 95, 96, 97, 108, 109, 110, 112, 113, 114, 115, 144, 145, 146, 155, 156, 157, 158, 159, 160, or 161, or with amino acids 25-1259 of SEQ ID NO: 47, amino acids 25-1259 of SEQ ID NO: 48, amino acids 25-1259 of SEQ

5 ID NO: 49, amino acids 25-1259 of SEQ ID NO: 50, amino acids 25-1259 of SEQ ID NO: 51, amino acids 25-1259 of SEQ ID NO: 52, amino acids 25-1259 of SEQ ID NO: 53, amino acids 25-1259 of SEQ ID NO: 54, amino acids 25-1259 of SEQ ID NO: 55, amino acids 25-1259 of SEQ ID NO: 56, amino acids 25-1259 of SEQ ID NO: 57, amino acids 25-1259 of SEQ ID NO: 58, amino acids 25-1262 of SEQ ID NO: 59, amino acids 25-1261 of SEQ ID NO: 60, amino acids 25-1258 of SEQ ID NO: 61, amino acids 25-1256 of SEQ ID NO: 62, amino acids 25-1243 of SEQ ID NO: 95, amino acids 25-1240 of SEQ ID NO: 96, amino acids 25-1243 of SEQ ID NO: 97, amino acids 25-1341 of SEQ ID NO: 108, amino acids 25-1338 of SEQ ID NO: 109, amino acids 25-1341 of SEQ ID NO: 110, amino acids 25-1351 of SEQ ID NO: 144, amino acids 25-1348 of SEQ ID NO: 145, amino acids 25-1351 of SEQ ID NO: 146, amino acids 25-1159 of SEQ ID NO: 155, amino acids 25-1156 of SEQ ID NO: 156, or amino acids 25-1159 of SEQ ID NO: 157, and wherein modified S proteins when expressed in a host or host cell form VLP. The amino acid sequence of the ectodomain and the transmembrane domain of the modified S proteins has about 70, 75, 80, 85, 87, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100% or any amount therebetween sequence identity, or sequence similarity, with amino acids 1-1234 of SEQ ID NO: 1, with amino acids 1-1219 of SEQ ID NO: 2, with amino acids 1-1234 of SEQ ID NO: 5, with amino acids 1-1219 of SEQ ID NO: 21, with amino acids 1-1243 of SEQ ID NO: 30, with the amino acids 25-1243 of SEQ ID NO: 47, with the amino acids 25-1243 of SEQ ID NO: 48, with the amino acids 25-1243 of SEQ ID NO: 49, with the amino acids 25-1243 of SEQ ID NO: 50, with the amino acids 25-1243 of SEQ ID NO: 51, with the amino acids 25-1243 of SEQ ID NO: 52, with the amino acids 25-1243 of SEQ ID NO: 53, with the amino acids 25-1243 of SEQ ID NO: 54, with the amino acids 25-1243 of SEQ ID NO: 55, with the amino acids 25-1243 of SEQ ID NO: 56, with the amino acids 25-1243 of SEQ ID NO: 57, with the amino acids 25-1243 of SEQ ID NO: 58, with the amino acids 25-1242 of SEQ ID NO: 59, with the amino acids 25-1242 of SEQ ID NO: 60, with the amino acids 25-1246 of SEQ ID NO: 61, or with the amino acids 25-1245 of SEQ ID NO: 62, amino acids 25-1227 of SEQ ID NO: 95, amino acids 25-1227 of SEQ ID NO: 96, amino acids 25-1227 of SEQ ID NO: 97, amino acids 25-1325 of SEQ ID NO: 108, amino acids 25-1325 of SEQ ID NO: 109, amino acids 25-1325 of SEQ ID NO: 110, amino acids 25-1335 of SEQ ID NO: 144,

amino acids 25- 1335 of SEQ ID NO: 145, amino acids 25-1335 of SEQ ID NO: 146, amino acids 25- 1143 of SEQ ID NO: 155, amino acids 25-1143 of SEQ ID NO: 156, or amino acids 25- 1143 of SEQ ID NO: 157, and the amino acid sequence of the cytoplasmic tail domain (CT) of the modified S protein has about 70, 75, 80, 85, 87, 90, 91, 92, 93 94, 95, 96, 97, 98, 99, 100% or any amount therebetween sequence identity, or sequence similarity, with the sequence of SEQ ID NO: 15, or with amino acids 35-50 of SEQ ID NO 6, 8, 7, 9, 10, 12, 13, 14, or with amino acids 34-49 of SEQ ID NO 11, or with amino acids 553-568 of SEQ ID NO:3 and wherein modified S proteins when expressed in a host or host cell form VLP.

[00253] Furthermore, the modified viral structural protein may be encoded by a nucleotide sequence that has about 70, 75, 80, 85, 87, 90, 91, 92, 93 94, 95, 96, 97, 98, 99, 100% or any amount therebetween, sequence identity, or sequence similarity, with the nucleotide sequence according to SEQ ID NO: 22, 26, 29, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 90, 91, 92, 95, 96, 97, 103, 104, 105, 139, 140, 141, 150, 151, or 152 and wherein the nucleotide sequence encodes modified S proteins that when expressed in a host or host cell form VLP.

[00254] It is further provided nucleotide sequence encoding a modified S proteins with amino acid sequences that have about 70, 75, 80, 85, 87, 90, 91, 92, 93 94, 95, 96, 97, 98, 99, 100% or any amount therebetween, sequence identity, or sequence similarity, with the amino acid sequence of SEQ ID NO: 1, 2, 5, 21, 30, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 95, 96, 97, 108, 109, 110, 144, 145, 146, 155, 156 or 157, and wherein modified S proteins when expressed in a host or host cell form VLP. The nucleotide sequence may encode an amino acid sequence of the ectodomain and the transmembrane domain of the modified S proteins that has about 70, 75, 80, 85, 87, 90, 91, 92, 93 94, 95, 96, 97, 98, 99, 100% or any amount therebetween sequence identity, or sequence similarity, with amino acids 1-1234 of SEQ ID NO:1, with amino acids 1-1219 of SEQ ID NO: 2, with amino acids 1-1234 of SEQ ID NO: 5, with amino acids 1-1219 of SEQ ID NO: 21 or with amino acids 1-1243 of SEQ ID NO: 30, with the amino acids 25-1243 of SEQ ID NO: 47, with the amino acids 25-1243 of SEQ ID NO: 48, with the amino acids 25-1243 of SEQ ID NO: 49, with the amino acids 25-1243 of SEQ ID NO: 50, with the amino acids 25-1243 of SEQ ID NO: 51, with the amino acids 25-1243 of

SEQ ID NO: 52, with the amino acids 25-1243 of SEQ ID NO: 53, with the amino acids  
 25-1243 of SEQ ID NO: 54, with the amino acids 25-1243 of SEQ ID NO: 55, with the  
 amino acids 25-1243 of SEQ ID NO: 56, with the amino acids 25-1243 of SEQ ID NO:  
 57, with the amino acids 25-1243 of SEQ ID NO: 58, with the amino acids 25-1242 of  
 5 SEQ ID NO: 59, with the amino acids 25-1242 of SEQ ID NO: 60, with the amino acids  
 25-1246 of SEQ ID NO: 61, with the amino acids 25-1245 of SEQ ID NO: 62, amino  
 acids 25-1227 of SEQ ID NO: 95, amino acids 25-1227 of SEQ ID NO: 96, amino acids  
 25-1227 of SEQ ID NO: 97, amino acids 25-1325 of SEQ ID NO: 108, amino acids 25-  
 1325 of SEQ ID NO: 109, amino acids 25-1325 of SEQ ID NO: 110, amino acids 25-  
 10 1335 of SEQ ID NO: 144, amino acids 25- 1335 of SEQ ID NO: 145, amino acids 25-  
 1335 of SEQ ID NO: 146, amino acids 25- 1143 of SEQ ID NO: 155, amino acids 25-  
 1143 of SEQ ID NO: 156, or amino acids 25- 1143 of SEQ ID NO: 157, and the amino  
 acid sequence of the cytoplasmic tail domain of the modified S protein has about 70, 75,  
 80, 85, 87, 90, 91, 92, 93 94, 95, 96, 97, 98, 99, 100% or any amount therebetween  
 15 sequence identity, or sequence similarity, with the sequence of SEQ ID NO: 15, or with  
 amino acids 35-50 of SEQ ID NO 6, 8, 7, 9, 10, 12, 13, 14, or with amino acids 34-49 of  
 SEQ ID NO 11, or with amino acids 553-568 of SEQ ID NO:3 and wherein modified S  
 proteins when expressed in a host or host cell form VLP.

[00255] It is further provided a nucleotide sequence encoding a modified S proteins  
 20 with amino acid sequences that have about 70, 75, 80, 85, 87, 90, 91, 92, 93 94, 95, 96,  
 97, 98, 99, 100% or any amount therebetween, sequence identity, or sequence similarity,  
 with the amino acid sequence of SEQ ID NO: 5, 21, 30, or 47-62, or with amino acids 24-  
 1259 of SEQ ID NO: 47 amino acids 25-1259 of SEQ ID NO: 48, amino acids 25-1259  
 of SEQ ID NO: 49, amino acids 25-1259 of SEQ ID NO: 50, amino acids 25-1259 of  
 25 SEQ ID NO: 51, amino acids 25-1259 of SEQ ID NO: 52, amino acids 25-1259 of SEQ  
 ID NO: 53, amino acids 25-1259 of SEQ ID NO: 54, amino acids 25-1259 of SEQ ID  
 NO: 55, amino acids 25-1259 of SEQ ID NO: 56, amino acids 25-1259 of SEQ ID NO:  
 57, amino acids 25-1259 of SEQ ID NO: 58, amino acids 25-1262 of SEQ ID NO: 59,  
 amino acids 25-1261 of SEQ ID NO: 60, amino acids 25-1258 of SEQ ID NO: 61, or  
 30 amino acids 25-1256 of SEQ ID NO: 62, amino acids 25-1243 of SEQ ID NO: 95, amino  
 acids 25-1240 of SEQ ID NO: 96, amino acids 25-1243 of SEQ ID NO: 97, amino acids

25-1341 of SEQ ID NO: 108, amino acids 25-1338 of SEQ ID NO: 109, amino acids 25-1341 of SEQ ID NO: 110, amino acids 25-1351 of SEQ ID NO: 144, amino acids 25-1348 of SEQ ID NO: 145, amino acids 25-1351 of SEQ ID NO: 146, amino acids 25-1159 of SEQ ID NO: 155, amino acids 25-1156 of SEQ ID NO: 156, or amino acids 25-1159 of SEQ ID NO: 157, and wherein modified S proteins when expressed in a host or host cell form VLP.

[00256] The terms “percent similarity”, “sequence similarity”, “percent identity”, or “sequence identity”, when referring to a particular sequence, are used for example as set forth in the University of Wisconsin GCG software program, or by manual alignment and visual inspection (see, e.g., Current Protocols in Molecular Biology, Ausubel et al., eds. 1995 supplement). Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, using for example the algorithm of Smith & Waterman, (1981, Adv. Appl. Math. 2:482), by the alignment algorithm of Needleman & Wunsch, (1970, J. Mol. Biol. 48:443), by the search for similarity method of Pearson & Lipman, (1988, Proc. Natl. Acad. Sci. USA 85:2444), by computerized implementations of these algorithms (for example: GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, Wis.).

[00257] An example of an algorithm suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al., (1977, Nuc. Acids Res. 25:3389-3402) and Altschul et al., (1990, J. Mol. Biol. 215:403-410), respectively. BLAST and BLAST 2.0 are used, with the parameters described herein, to determine percent sequence identity for the nucleic acids and proteins of the disclosure. For example the BLASTN program (for nucleotide sequences) may use as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program may use as defaults a word length of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, 1989, Proc. Natl. Acad. Sci. USA 89:10915) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both

strands. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (see URL: [ncbi.nlm.nih.gov/](http://ncbi.nlm.nih.gov/)).

[00258] A nucleic acid sequence or nucleotide sequence referred to in the present disclosure, may be “substantially homologous”, “substantially similar” or “substantially identical” to a sequence, or a compliment of the sequence if the nucleic acid sequence or nucleotide sequence hybridise to one or more than one nucleotide sequence or a compliment of the nucleic acid sequence or nucleotide sequence as defined herein under stringent hybridisation conditions. Sequences are “substantially homologous” “substantially similar” “substantially identical” when at least about 70%, or between 70 to 100%, or any amount therebetween, for example 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100%, or any amount therebetween, of the nucleotides match over a defined length of the nucleotide sequence providing that such homologous sequences exhibit one or more than one of the properties of the sequence, or the encoded product as described herein.

[00259] Many organisms display a bias for use of particular codons to code for insertion of a particular amino acid in a growing peptide chain. Codon preference or codon bias, differences in codon usage between organisms, is afforded by degeneracy of the genetic code, and is well documented among many organisms. Codon bias often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, inter alia, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization. The process of optimizing the nucleotide sequence coding for a heterologously expressed protein may be an important step for improving expression yields. The optimization requirements may include steps to improve the ability of the host to produce the foreign protein.

[00260] There are different codon-optimization techniques known in the art for improving, the translational kinetics of translationally inefficient protein coding regions. These techniques mainly rely on identifying the codon usage for a certain host organism.

If a certain gene or sequence should be expressed in this organism, the coding sequence of such genes and sequences will then be modified such that one will replace codons of the sequence of interest by more frequently used codons of the host organism.

5 [00261] “Codon optimization” is defined as modifying a nucleic acid sequence for enhanced expression in a host or host cell of interest by replacing at least one, more than one, or a significant number, of codons of the native sequence with codons that may be more frequently or most frequently used in the genes of another organism or species. Various species exhibit particular bias for certain codons of a particular amino acid.

10 [00262] The present disclosure includes synthetic polynucleotide sequences that have been codon optimized for example the sequences have been optimized for human codon usage or plant codon usage. The codon optimized polynucleotide sequences may then be expressed in the host for example plants. More specifically the sequences optimized for human codon usage or plant codon usage may be expressed in plants. Without wishing to be bound by theory, it is believed that the sequences optimized for human codon  
15 increases the guanine-cytosine content (GC content) of the sequence and improves expression yields when plants are used as host.

[00263] The term “construct”, “vector” or “expression vector”, as used herein, refers to a recombinant nucleic acid for transferring exogenous nucleotide sequences (for example a nucleotide sequences encoding the modified viral structural protein as described  
20 herewith) into host cells (e.g. plant cells) and directing expression of the exogenous nucleic acid sequences in the host cells. “Expression cassette” refers to a nucleic acid comprising a nucleotide sequence of interest under the control of, and operably (or operatively) linked to, an appropriate promoter or other regulatory elements for transcription of the nucleic acid of interest in a host cell. As one of skill in the art would  
25 appreciate, the expression cassette may comprise a termination (terminator) sequence that is any sequence that is active the host cell (e.g. plant host). For example in plants, the termination sequence may be derived from the RNA-2 genome segment of a bipartite RNA virus, e.g. a comovirus, the termination sequence may be a NOS terminator, or terminator sequence may be obtained from the 3’UTR of the alfalfa plastocyanin gene.

[00264] The nucleic acid comprising a nucleotide sequence encoding a modified viral structural protein, as described herein may further comprise sequences that enhance expression of the viral structural protein in the host, portion of the host or host cell. Sequences that enhance expression may include, a 5' UTR enhancer element, or a plant-derived expression enhancer, in operative association with the nucleic acid encoding the modified viral structural protein. The sequence encoding the modified viral structural protein may also be optimized to increase expression by for example optimizing for human codon usage, increased GC content, or a combination thereof.

[00265] By “regulatory region” “regulatory element” or “promoter” it is meant a portion of nucleic acid typically, but not always, upstream of the protein coding region of a gene, which may be comprised of either DNA or RNA, or both DNA and RNA. When a regulatory region is active, and in operative association, or operatively linked, with a nucleotide sequence of interest, this may result in expression of the nucleotide sequence of interest. A regulatory element may be capable of mediating organ specificity, or controlling developmental or temporal gene activation. A “regulatory region” includes promoter elements, core promoter elements exhibiting a basal promoter activity, elements that are inducible in response to an external stimulus, elements that mediate promoter activity such as negative regulatory elements or transcriptional enhancers. “Regulatory region”, as used herein, also includes elements that are active following transcription, for example, regulatory elements that modulate gene expression such as translational and transcriptional enhancers, translational and transcriptional repressors, upstream activating sequences, and mRNA instability determinants. Several of these latter elements may be located proximal to the coding region.

[00266] In the context of this disclosure, the term “regulatory element” or “regulatory region” typically refers to a sequence of DNA, usually, but not always, upstream (5') to the coding sequence of a structural gene, which controls the expression of the coding region by providing the recognition for RNA polymerase and/or other factors required for transcription to start at a particular site. However, it is to be understood that other nucleotide sequences, located within introns, or 3' of the sequence may also contribute to the regulation of expression of a coding region of interest. An example of a regulatory

element that provides for the recognition for RNA polymerase or other transcriptional factors to ensure initiation at a particular site is a promoter element. Most, but not all, eukaryotic promoter elements contain a TATA box, a conserved nucleic acid sequence comprised of adenosine and thymidine nucleotide base pairs usually situated approximately 25 base pairs upstream of a transcriptional start site. A promoter element may comprise a basal promoter element, responsible for the initiation of transcription, as well as other regulatory elements that modify gene expression.

[00267] There are several types of regulatory regions, including those that are developmentally regulated, inducible or constitutive. A regulatory region that is developmentally regulated, or controls the differential expression of a gene under its control, is activated within certain organs or tissues of an organ at specific times during the development of that organ or tissue. However, some regulatory regions that are developmentally regulated may preferentially be active within certain organs or tissues at specific developmental stages, they may also be active in a developmentally regulated manner, or at a basal level in other organs or tissues within the plant as well. Examples of tissue-specific regulatory regions, for example see-specific a regulatory region, include the napin promoter, and the cruciferin promoter (Rask et al., 1998, J. Plant Physiol. 152: 595-599; Bilodeau et al., 1994, Plant Cell 14: 125-130). An example of a leaf-specific promoter includes the plastocyanin promoter (see US 7,125,978, which is incorporated herein by reference).

[00268] An inducible regulatory region is one that is capable of directly or indirectly activating transcription of one or more DNA sequences or genes in response to an inducer. In the absence of an inducer the DNA sequences or genes will not be transcribed. Typically the protein factor that binds specifically to an inducible regulatory region to activate transcription may be present in an inactive form, which is then directly or indirectly converted to the active form by the inducer. However, the protein factor may also be absent. The inducer can be a chemical agent such as a protein, metabolite, growth regulator, herbicide or phenolic compound or a physiological stress imposed directly by heat, cold, salt, or toxic elements or indirectly through the action of a pathogen or disease agent such as a virus. A plant cell containing an inducible regulatory

region may be exposed to an inducer by externally applying the inducer to the cell or plant such as by spraying, watering, heating or similar methods. Inducible regulatory elements may be derived from either plant or non-plant genes (e.g. Gatz, C. and Lenk, I.R.P., 1998, Trends Plant Sci. 3, 352-358). Examples, of potential inducible promoters include, but not limited to, tetracycline-inducible promoter (Gatz, C., 1997, Ann. Rev. Plant Physiol. Plant Mol. Biol. 48, 89-108), steroid inducible promoter (Aoyama, T. and Chua, N.H., 1997, Plant J. 2, 397-404) and ethanol-inducible promoter (Salter, M.G., et al, 1998, Plant Journal 16, 127-132; Caddick, M.X., et al, 1998, Nature Biotech. 16, 177-180) cytokinin inducible IB6 and CKI1 genes (Brandstatter, I. and Kieber, J.J., 1998, Plant Cell 10, 1009-1019; Kakimoto, T., 1996, Science 274, 982-985) and the auxin inducible element, DR5 (Ulmasov, T., et al., 1997, Plant Cell 9, 1963-1971).

[00269] A constitutive regulatory region directs the expression of a gene throughout the various parts of a plant and continuously throughout plant development. Examples of known constitutive regulatory elements include promoters associated with the CaMV 35S transcript. (p35S; Odell et al., 1985, Nature, 313: 810-812; which is incorporated herein by reference), the rice actin 1 (Zhang et al, 1991, Plant Cell, 3: 1155-1165), actin 2 (An *et al.*, 1996, *Plant J.*, 10: 107-121), or tms 2 (U.S. 5,428,147), and triosephosphate isomerase 1 (Xu et. al., 1994, Plant Physiol. 106: 459-467) genes, the maize ubiquitin 1 gene (Cornejo et al, 1993, Plant Mol. Biol. 29: 637-646), the *Arabidopsis* ubiquitin 1 and 6 genes (Holtorf et al, 1995, Plant Mol. Biol. 29: 637-646), the tobacco translational initiation factor 4A gene (Mandel et al, 1995 Plant Mol. Biol. 29: 995-1004), the Cassava Vein Mosaic Virus promoter, pCAS, (Verdaguer et al., 1996); the promoter of the small subunit of ribulose biphosphate carboxylase, pRbcS: (Outchkourov et al., 2003), the pUbi (for monocots and dicots ).

[00270] The term "constitutive" as used herein does not necessarily indicate that a nucleotide sequence under control of the constitutive regulatory region is expressed at the same level in all cell types, but that the sequence is expressed in a wide range of cell types even though variation in abundance is often observed.

[00271] One or more of the genetic constructs of the present disclosure may also include further enhancers, either translation or transcription enhancers, as may be required. Enhancers may be located 5' or 3' to the sequence being transcribed. Enhancer regions are well known to persons skilled in the art, and may include an ATG initiation codon, adjacent sequences or the like. The initiation codon, if present, may be in phase with the reading frame ("in frame") of the coding sequence to provide for correct translation of the transcribed sequence.

[00272] The term "5'UTR" or "5' untranslated region", "5' leader sequence" or "5' UTR enhancer element" refers to regions of an mRNA that are not translated. The 5'UTR typically begins at the transcription start site and ends just before the translation initiation site or start codon of the coding region. The 5' UTR may modulate the stability and/or translation of an mRNA transcript.

[00273] The term "plant-derived expression enhancer", as used herein, refers to a nucleotide sequence obtained from a plant, the nucleotide sequence encoding a 5'UTR. Examples of a plant derived expression enhancer are described in US Provisional Patent Application No.62/643,053 (Filed March 14, 2018) and International Application No. PCT/CA2019/050319 (Filed March 14, 2019); which are incorporated herein by reference) or in Diamos A.G. et al. (2016, Front Plt Sci. 7:1-15; which is incorporated herein by reference). The plant-derived expression enhancer may be selected from nbEPI42, nbSNS46, nbCSY65, nbHEL40, nbSEP44, nbMT78, nbATL75, nbDJ46, nbCHP79, nbEN42, atHSP69, atGRP62, atPK65, atRP46, nb30S72, nbGT61, nbPV55, nbPPI43, nbPM64 and nbH2A86 as described in US 62/643,053 and PCT/CA2019/050319. The plant derived expression enhancer may be used within a plant expression system comprising a regulatory region that is operatively linked with the plant-derived expression enhancer sequence and a nucleotide sequence of interest, for example a nucleotide sequence encoding a modified S protein.

[00274] Stability and/or translation efficiency of an RNA may further be improved by the inclusion of a 3' untranslated region (3'UTR). The one or more genetic constructs of the present description may therefore further comprise a 3' UTR.

[00275] A 3' untranslated region may contain a polyadenylation signal and any other regulatory signals capable of effecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by effecting the addition of polyadenylic acid tracks to the 3' end of the mRNA precursor. Polyadenylation signals are commonly recognized by the presence of homology to the canonical form 5' AATAAA-3' although variations are not uncommon. Non-limiting examples of suitable 3' regions are the 3' transcribed non-translated regions containing a polyadenylation signal of *Agrobacterium* tumor inducing (Ti) plasmid genes, such as the nopaline synthase (Nos gene) and plant genes such as the soybean storage protein genes, the small subunit of the ribulose-1, 5-bisphosphate carboxylase gene (ssRUBISCO; US 4,962,028; which is incorporated herein by reference), the promoter used in regulating plastocyanin expression, described in US 7,125,978 (which is incorporated herein by reference), 3' UTR derived from a Arracacha virus B isolate gene (AvB) (SEQ ID NO: 40), 3'UTR derived from Beet necrotic yellow vein virus (trBNYVV) (SEQ ID NO: 41), 3'UTR derived from Southern bean mosaic virus (SBMV) (SEQ ID NO: 42), 3'UTR derived from Turnip ringspot virus (TuRSV) (SEQ ID NO: 43), 3' UTR derived from Cowpea Mosaic Virus (CPMV) (SEQ ID NO: 44), 3'UTR derived from Broad bean true mosaic virus (BBTMV) (SEQ ID NO: 45) or 3'UTR derived from Ourmia melon virus (trOUMV) (SEQ ID NO: 46). The 3'UTR might be used in conjunction with 5'UTR derived from heterologous sequences to modulate expression levels.

[00276] It is therefore provided a "construct", "vector", "expression vector" or "expression cassette" that comprises a nucleic acid comprising a nucleotide sequence of interest (such as a modified viral structural protein) under the control of, and operably (or operatively) linked to a 3'UTR. Furthermore, the nucleic acid may comprise a 3'UTR operably (or operatively) linked to a nucleotide sequence of interest (such as a modified viral structural protein).

[00277] The modified viral structural protein may be targeted to any intracellular or extracellular space, organelle or tissue of a host of host cell such as plant or plant cell as desired. In order to localize the expressed protein to a particular location, the nucleic acid encoding the protein may be linked to a nucleic acid sequence encoding a signal peptide or leader sequence. A signal peptide may alternately be referred to as a transit

peptide, signal sequence, leader sequence, targeting signal, localization signal, localization sequence, transit peptide, or leader peptide.

5 [00278] The one or more than one modified genetic constructs of the present description may be expressed in any suitable host or host cell that is transformed by the nucleic acids, or nucleotide sequence, or constructs, or vectors of the present disclosure. The host or host cell may be from any source including plants, fungi, bacteria, insect and animals for example mammals. Therefore the host or host cell may be selected from a plant or plant cell, a fungi or a fungi cell, a bacteria or bacteria cell, an insect or an insect cell, and animal or an animal cell. The mammal or animal may not be a human. In a  
10 preferred embodiment the host or host cell is a plant, portion of a plant or plant cell.

[00279] The term “plant”, “portion of a plant”, “plant portion”, “plant matter”, “plant biomass”, “plant material”, plant extract”, or “plant leaves”, as used herein, may comprise an entire plant, tissue, cells, or any fraction thereof, intracellular plant components, extracellular plant components, liquid or solid extracts of plants, or a  
15 combination thereof, that are capable of providing the transcriptional, translational, and post-translational machinery for expression of one or more than one nucleic acids described herein, and/or from which an expressed protein or VLP may be extracted and purified. Plants may include, but are not limited to, herbaceous plants. The herbaceous plants may be annuals, biennials or perennials plants. Plants may further include, but are  
20 not limited to agricultural crops including for example canola, Brassica spp., maize, Nicotiana spp., (tobacco) for example, *Nicotiana benthamiana*, *Nicotiana rustica*, *Nicotiana tabacum*, *Nicotiana alata*, *Arabidopsis thaliana*, alfalfa, potato, sweet potato (*Ipomoea batatas*), ginseng, pea, oat, rice, soybean, wheat, barley, sunflower, cotton, corn, rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), safflower  
25 (*Carthamus tinctorius*).

[00280] The term “plant portion”, as used herein, refers to any part of the plant including but not limited to leaves, stem, root, flowers, fruits, a plant cell obtained from leaves, stem, root, flowers, fruits, a plant extract obtained from leaves, stem, root, flowers, fruits, or a combination thereof. In one embodiment the plant portion refers to  
30 the areal portion of a plant such as for example leaves, stem, flowers and fruits. The term

“plant extract”, as used herein, refers to a plant-derived product that is obtained following treating a plant, a portion of a plant, a plant cell, or a combination thereof, physically (for example by freezing followed by extraction in a suitable buffer), mechanically (for example by grinding or homogenizing the plant or portion of the plant followed by extraction in a suitable buffer), enzymatically (for example using cell wall degrading enzymes), chemically (for example using one or more chelators or buffers), or a combination thereof. A plant extract may be further processed to remove undesired plant components for example cell wall debris. A plant extract may be obtained to assist in the recovery of one or more components from the plant, portion of the plant or plant cell, for example a protein (including protein complexes, protein suprastructures and/or VLPs), a nucleic acid, a lipid, a carbohydrate, or a combination thereof from the plant, portion of the plant, or plant cell. If the plant extract comprises proteins, then it may be referred to as a protein extract. A protein extract may be a crude plant extract, a partially purified plant or protein extract, or a purified product, that comprises one or more proteins, protein complexes such for example protein trimers, protein suprastructures, and/or VLPs, from the plant tissue. If desired a protein extract, or a plant extract, may be partially purified using techniques known to one of skill in the art, for example, the extract may be subjected to salt or pH precipitation, centrifugation, gradient density centrifugation, filtration, chromatography, for example, size exclusion chromatography, ion exchange chromatography, affinity chromatography, or a combination thereof. A protein extract may also be purified, using techniques that are known to one of skill in the art.

[00281] The constructs of the present disclosure can be introduced into plant cells using Ti plasmids, Ri plasmids, plant virus vectors, direct DNA transformation, micro-injection, electroporation, etc. For reviews of such techniques see for example Weissbach and Weissbach, *Methods for Plant Molecular Biology*, Academy Press, New York VIII, pp. 421-463 (1988); Geierman and Corey, *Plant Molecular Biology*, 2d Ed. (1988); and Miki and Iyer, *Fundamentals of Gene Transfer in Plants*. In *Plant Metabolism*, 2d Ed. DT. Dennis, DH Turpin, DD Lefebvre, DB Layzell (eds), Addison Wesley, Langmans Ltd. London, pp. 561-579 (1997). Other methods include direct DNA uptake, the use of liposomes, electroporation, for example using protoplasts, micro-

injection, microprojectiles or whiskers, and vacuum infiltration. See, for example, Bilang, et al. (Gene 100: 247-250 (1991), Scheid et al. (Mol. Gen. Genet. 228: 104-112, 1991), Guerche et al. (Plant Science 52: 111-116, 1987), Neuhauser et al. (Theor. Appl Genet. 75: 30-36, 1987), Klein et al., Nature 327: 70-73 (1987); Howell et al. (Science 208: 1265, 1980), Horsch et al. (Science 227: 1229-1231, 1985), DeBlock et al., Plant Physiology 91: 694-701, 1989), Methods for Plant Molecular Biology (Weissbach and Weissbach, eds., Academic Press Inc., 1988), Methods in Plant Molecular Biology (Schuler and Zielinski, eds., Academic Press Inc., 1989), Liu and Lomonosoff (J Virol Meth, 105:343-348, 2002,), U.S. Pat. Nos. 4,945,050; 5,036,006; and 5,100,792, U.S. patent application Ser. Nos. 08/438,666, filed May 10, 1995, and 07/951,715, filed Sep. 25, 1992, (all of which are hereby incorporated by reference).

[00282] As described below, transient expression methods may be used to express the constructs of the present disclosure (see Liu and Lomonosoff, 2002, Journal of Virological Methods, 105:343-348; which is incorporated herein by reference).

Alternatively, a vacuum-based transient expression method, as described by Kapila et al., 1997, which is incorporated herein by reference) may be used. These methods may include, for example, but are not limited to, a method of Agro-inoculation or Agroinfiltration, syringe infiltration, however, other transient methods may also be used as noted above. With Agro-inoculation, Agroinfiltration, or syringe infiltration, a mixture of *Agrobacteria* comprising the desired nucleic acid enter the intercellular spaces of a tissue, for example the leaves, aerial portion of the plant (including stem, leaves and flower), other portion of the plant (stem, root, flower), or the whole plant. After crossing the epidermis the *Agrobacteria* infect and transfer t-DNA copies into the cells. The t-DNA is episomally transcribed and the mRNA translated, leading to the production of the protein of interest in infected cells, however, the passage of t-DNA inside the nucleus is transient.

[00283] To aid in identification of transformed plant cells, the constructs of this disclosure may be further manipulated to include plant selectable markers. Useful selectable markers include enzymes that provide for resistance to chemicals such as an antibiotic for example, gentamycin, hygromycin, kanamycin, or herbicides such as

phosphinothrycin, glyphosate, chlorosulfuron, and the like. Similarly, enzymes providing for production of a compound identifiable by colour change such as GUS (beta-glucuronidase), or luminescence, such as luciferase or GFP, may be used.

[00284] Also considered part of this disclosure are transgenic plants, plant cells  
5 or seeds containing the gene construct of the present disclosure that may be used as a platform plant suitable for transient protein expression described herein. Methods of regenerating whole plants from plant cells are also known in the art (for example see Guerineau and Mullineaux (1993, Plant transformation and expression vectors. In: Plant Molecular Biology Labfax (Croy RRD ed) Oxford, BIOS Scientific Publishers, pp 121-  
10 148). In general, transformed plant cells are cultured in an appropriate medium, which may contain selective agents such as antibiotics, where selectable markers are used to facilitate identification of transformed plant cells. Once callus forms, shoot formation can be encouraged by employing the appropriate plant hormones in accordance with known methods and the shoots transferred to rooting medium for regeneration of plants.  
15 The plants may then be used to establish repetitive generations, either from seeds or using vegetative propagation techniques. Transgenic plants can also be generated without using tissue culture. Methods for stable transformation, and regeneration of these organisms are established in the art and known to one of skill in the art. Available techniques are reviewed in Vasil et al. (Cell Culture and Somatic Cell Genetics of Plants, Vol I, II and III, Laboratory Procedures and Their Applications, Academic Press, 1984),  
20 and Weissbach and Weissbach (Methods for Plant Molecular Biology, Academic Press, 1989). The method of obtaining transformed and regenerated plants is not critical to the present disclosure.

[00285] If plants, plant portions or plant cells are to be transformed or co-transformed  
25 by two or more nucleic acid constructs, the nucleic acid construct may be introduced into the *Agrobacterium* in a single transfection event so that the nucleic acids are pooled, and the bacterial cells transfected. Alternatively, the constructs may be introduced serially. In this case, a first construct is introduced into the *Agrobacterium* as described, the cells are grown under selective conditions (e.g. in the presence of an antibiotic) where only the  
30 singly transformed bacteria can grow. Following this first selection step, a second

nucleic acid construct is introduced into the *Agrobacterium* as described, and the cells are grown under double-selective conditions, where only the double-transformed bacteria can grow. The double-transformed bacteria may then be used to transform a plant, portion of the plant or plant cell as described herein, or may be subjected to a further transformation step to accommodate a third nucleic acid construct.

[00286] Alternatively, if plants, plant portions, or plant cells are to be transformed or co-transformed by two or more nucleic acid constructs, the nucleic acid construct may be introduced into the plant by co-infiltrating a mixture of *Agrobacterium* cells with the plant, plant portion, or plant cell, each *Agrobacterium* cell may comprise one or more constructs to be introduced within the plant. In order to vary the relative expression levels within the plant, plant portion or plant cell, of a nucleotide sequence of interest within a construct, during the step of infiltration, the concentration of the various *Agrobacteria* populations comprising the desired constructs may be varied.

[00287] The modified viral surface protein or VLP comprising modified viral surface protein as described herewith, may be used to elicit an immune response in a subject.

[00288] An “immune response” generally refers to a response of the adaptive immune system of a subject. The adaptive immune system generally comprises a humoral response, and a cell-mediated response. The humoral response is the aspect of immunity that is mediated by secreted antibodies, produced in the cells of the B lymphocyte lineage (B cell). Secreted antibodies bind to antigens on the surfaces of invading microbes (such as viruses or bacteria), which flags them for destruction. Humoral immunity is used generally to refer to antibody production and the processes that accompany it, as well as the effector functions of antibodies, including Th2 cell activation and cytokine production, memory cell generation, opsonin promotion of phagocytosis, pathogen elimination and the like. The terms “modulate” or “modulation” or the like refer to an increase or decrease in a particular response or parameter, as determined by any of several assays generally known or used, some of which are exemplified herein.

[00289] A cell-mediated response is an immune response that does not involve antibodies but rather involves the activation of macrophages, natural killer cells (NK),

antigen-specific cytotoxic T-lymphocytes, and the release of various cytokines in response to an antigen. Cell-mediated immunity is used generally to refer to some Th cell activation, Tc cell activation and T-cell mediated responses. Cell mediated immunity may be of particular importance in responding to viral infections.

5 [00290] For example, the induction of antigen specific CD8 positive T lymphocytes may be measured using an ELISPOT assay; stimulation of CD4 positive T-lymphocytes may be measured using a proliferation assay. Anti-Coronavirus antibody titers may be quantified using an ELISA assay; isotypes of antigen-specific or cross-reactive antibodies may also be measured using anti-isotype antibodies (e.g. anti-IgG, IgA, IgE or IgM).  
10 Methods and techniques for performing such assays are well-known in the art.

[00291] Cytokine presence or levels may also be quantified. For example a T-helper cell response (Th1/Th2) will be characterized by the measurement of IFN- $\gamma$  and IL-4 secreting cells using by ELISA (e.g. BD Biosciences OptEIA kits). Peripheral blood mononuclear cells (PBMC) or splenocytes obtained from a subject may be cultured, and  
15 the supernatant analyzed. T lymphocytes may also be quantified by fluorescence-activated cell sorting (FACS), using marker specific fluorescent labels and methods as are known in the art.

[00292] A microneutralization assay may also be conducted to characterize an immune response in a subject, see for example the methods of Rowe *et al.*, 1973. Virus  
20 neutralization titers may be quantified in a number of ways, including: enumeration of lysis plaques (plaque assay) following crystal violet fixation/coloration of cells; microscopic observation of cell lysis in *in vitro* culture; and 2) ELISA and spectrophotometric detection of Coronavirus.

[00293] The term “epitope” or “epitopes”, as used herein, refers to a structural part of  
25 an antigen to which an antibody specifically binds.

[00294] A method of producing an antibody or antibody fragment is provided, the method comprises administering the modified viral structural protein, a trimer or trimeric modified viral structural protein or VLP comprising the modified viral structural protein

as described herewith to a subject, or a host animal, thereby producing the antibody or the antibody fragment. Antibodies or the antibody fragments produced by the method are also provided.

5 [00295] The present disclosure therefore also provides the use of a viral structural protein or VLP comprising the modified viral structural protein, as described herein, for inducing immunity to a Coronavirus infection in a subject. Also disclosed herein is an antibody or antibody fragment, prepared by administering the modified viral structural protein or VLP comprising the modified viral structural protein, to a subject or a host animal.

10 [00296] Further provided is a composition comprising an effective dose of modified viral structural protein or VLP comprising the modified viral structural protein, as described herein, and a pharmaceutically acceptable carrier, adjuvant, vehicle, or excipient, for inducing an immune response in a subject. Also provided is a vaccine for inducing an immune response against Coronavirus in a subject, wherein the vaccine  
15 comprises an effective dose of the modified viral structural protein or VLP comprising the modified viral structural protein.

[00297] Further provided is a composition that may comprise a mixture of VLPs provided that at least one of the VLPs within the composition comprises modified coronavirus S protein as described herein. For example, each coronavirus S protein  
20 including one or more than one modified S protein, from each of one or more than one Coronavirus family, sub-group, type, subtype, lineage or strain may be expressed and the corresponding VLPs purified. Virus like particles obtained from two or more than two Coronavirus families, sub-groups, types, subtypes, lineages or strains (for example, two, three, four, five, six, seven, eight, nine, 10 or more Coronavirus families, sub-groups,  
25 types, subtypes, lineages or strains) may be combined as desired to produce a mixture of VLPs, provided that one or more than one VLP in the mixture of VLPs comprises a modified S protein as described herein. The VLPs may be combined or produced in a desired ratio, for example about equivalent ratios, or may be combined in such a manner that one Coronavirus family, sub-group, type, subtype, lineage or strain comprises the  
30 majority of the VLPs in the composition. It is further provided a composition of VLPs

comprising one or more than one modified S protein with ectodomain and/or TM or portion of a TM derived from each of one or more than one Coronavirus family, sub-group, type, subtype, lineage or strain, such that a mixture of different modified S protein as provided for in this disclosure may be present in any individual VLP of the composition.

[00298] The composition or vaccine may comprise VLP comprising the modified viral structural protein, such as the modified S protein from one type of Coronavirus family, sub-group, type, subtype, lineage or strain, or the composition or vaccine may comprise multiple VLP types, wherein each VLP type comprises modified S protein, wherein the modified S proteins in the same VLP are derived from one type of Coronavirus family, sub-group, type, subtype, lineage or strain i.e. the composition or vaccine may comprise a mixture of different Coronavirus VLP, wherein each VLP may comprise a modified S protein from the same Coronavirus family, sub-group, type, subtype, lineage or strain. For example the composition or vaccine may comprise a first VLP comprising a first modified S protein from a first Coronavirus family, sub-group, type, subtype, lineage or strain and a second VLP comprising a second modified S protein from a second Coronavirus family, sub-group, type, subtype, lineage or strain. Furthermore the composition may also comprise a third VLP comprising a third modified S protein from a third Coronavirus family, sub-group, type, subtype, lineage or strain and/or the composition or vaccine may comprise a fourth VLP comprising a fourth modified S protein from a fourth Coronavirus family, sub-group, type, subtype, lineage or strain. Accordingly, the description also provides compositions or vaccines that are monovalent (univalent), or multivalent (polyvalent). The monovalent composition or vaccine may immunize a subject against a single type of Coronavirus strain, whereas the multivalent composition or vaccine may immunize a subject against more than one Coronavirus strain. For example, the composition or vaccine may be a bivalent composition or vaccine, which upon administration, may immunize a subject against two different types of Coronavirus families, sub-groups, types, subtypes, lineages or strains. Furthermore, the composition or vaccine may be a trivalent composition, or the vaccine or composition may be a tetravalent or quadrivalent composition or vaccine.

[00299] Furthermore, the multivalent composition may comprise VLP comprising one or more than one modified S proteins with different HA cytoplasmic tails. For example, the multivalent composition may comprise a VLP or plurality of VLPs comprising two or more modified S proteins, each comprising a S protein ectodomain, a S protein transmembrane domain, and a cytoplasmic tail derived from HA from an influenza H1, H3, H5, H6, H7, H9 or B strain. Non-limiting examples of influenza strains are for example H1 California/7/2009, H3 A/Minnesota/41/2019, H5 A/Indonesia/5/05, H6 A/Teal/Hong Kong/W312/97, H7 A/Guangdong/17SF003/2016, H9 A/Hong Kong/1073/99 or B/Washington/02/2019.

[00300] The multivalent composition or vaccine with multiple type VLPs may further comprise a pharmaceutically acceptable carrier, adjuvant, vehicle, or excipient, for inducing an immune response in a subject.

[00301] Adjuvant systems to enhance a subject's immune response to a vaccine antigen are well known and may be used in conjunction with the vaccine or pharmaceutical composition as described herewith. There are many types of adjuvants that may be used. Common adjuvants for human use are aluminum hydroxide, aluminum phosphate and calcium phosphate. There are also a number of adjuvants based on oil emulsions (oil in water or water in oil emulsions such as Freund's incomplete adjuvant (FIA), Montanide™, Adjuvant 65, and Lipovant™), products from bacterial (or their synthetic derivatives), endotoxins, fatty acids, paraffinic, or vegetable oils, cholesterol, and aliphatic amines or natural organic compounds such for example squalene. Non-limiting adjuvants that might be used include for example oil-in water emulsions of squalene oil (for example MF-59 or AS03), adjuvant composed of the synthetic TLR4 agonist glucopyranosyl lipid A (GLA) integrated into stable emulsion (SE) (GLA-SE) or CpG 1018 a toll-like receptor (TLR9) agonist adjuvant.

[00302] Therefore the vaccine or pharmaceutical composition may comprise one or more than one adjuvant. For example the vaccine or pharmaceutical composition may comprise aluminum hydroxide, aluminum phosphate, calcium phosphate, an oil in water or water in oil emulsions, an emulsion comprising squalene (for example MF-59 or AS03), an emulsion comprising GLA-SE, or CpG 1018 adjuvant.

[00303] The pharmaceutical compositions, vaccines or formulations of the present description may be manufactured in a manner that is itself known, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or tableting processes.

5 [00304] The pharmaceutical compositions, vaccines or formulations may be produced by mixing or premixing of any constituent components before administration, for example by manual or mechanically-aided mixing of two or more vaccine suspensions, pharmaceutically acceptable carriers, adjuvants, vehicles, or excipients as a step performed before the final formulation, vaccine, or pharmaceutical composition is  
10 administered.

[00305] The pharmaceutical compositions, vaccines or formulations may be administered to a subject orally, intradermally, intranasally, intramuscularly, intraperitoneally, intravenously, or subcutaneously.

15 [00306] Injectables can be prepared in conventional forms, either as liquid solutions or suspensions, solid forms suitable for solution or suspension in liquid prior to injection, or as emulsions. Suitable excipients are, for example, water, saline, dextrose, mannitol, lactose, lecithin, albumin, sodium glutamate, cysteine hydrochloride, and the like. In addition, if desired, the injectable pharmaceutical compositions may contain  
20 minor amounts of nontoxic auxiliary substances, such as wetting agents, pH buffering agents, and the like. Physiologically compatible buffers include, but are not limited to, Hanks's solution, Ringer's solution, or physiological saline buffer. If desired, absorption enhancing preparations (for example, liposomes), may be utilized.

25 [00307] The composition or vaccine may be administered to a subject once (single dose). Furthermore, the vaccine or composition may be administered to a subject multiple times (multi-dose). Therefore the composition, formulation, or vaccine may be administered to a subject in a single dose to illicit an immune response or the composition, formulation, or vaccine may be administered multiple time (multi dosages). For example a dose of the composition or vaccine may be administered 2, 3, 4 or 5 times. Accordingly, the composition or vaccine may be administered to a subject in an initial  
30 dose and one or more than one doses may subsequently be administered to the subject.

5

Administration of the doses may be separated in time from each other. For example after the administration of an initial dose, one or more than one subsequent dose may be administered 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, 1 week, 2 weeks, 3 weeks, 1 month, 2 months, 3 months, 4 months, 5 months or 6 months or any time in between from the administration of the initial dose. Furthermore, the composition or vaccine may be administered annually. For example the composition or vaccine may be administered as a seasonal vaccine.

[00308] The disclosure further provides the following sequences.

**Table 4:** SEQ ID NOs and Description of Sequences

| SEQ ID NO: | Description of Sequence                                                             |
|------------|-------------------------------------------------------------------------------------|
| 1          | Native SARS-CoV-2 S protein wtTM/CT AA (P0DTC2)                                     |
| 2          | Native SARS-CoV-2 S protein wtTM/CT AA (P0DTC2) without signal peptide (SP)         |
| 3          | H5 A/Indonesia/5/05 Hemagglutinin (HA) AA (A5A5L7)                                  |
| 4          | H5 A/Indonesia/5/05 Hemagglutinin (HA) viral cDNA (EF541394.1)                      |
| 5          | Modified SARS-CoV-2 with H5 A/Indonesia/5/05 (H5i) Hemagglutinin CT AA              |
| 6          | H1 California/7/2009 Hemagglutinin TM/CT region AA                                  |
| 7          | H2 A/Singapore/1/1957 HA Hemagglutinin TM/CT region AA                              |
| 8          | H3 A/Minnesota/41/2019 Hemagglutinin TM/CT region AA                                |
| 9          | H5 A/Indonesia/5/05 Hemagglutinin TM/CT region AA                                   |
| 10         | H6 A/Teal/Hong Kong/W312/97 Hemagglutinin TM/CT region AA                           |
| 11         | H7 A/Guangdong/17SF003/2016 Hemagglutinin TM/CT region AA                           |
| 12         | H9 A/Hong Kong/1073/99 Hemagglutinin TM/CT region AA                                |
| 13         | B/Washington/02/2019 Hemagglutinin TM/CT region AA                                  |
| 14         | Consensus Sequence of C-Terminal Region of Influenza Hemagglutinin TM/CT region     |
| 15         | Consensus Sequence of CT Domain of Influenza Hemagglutinin                          |
| 16         | TM/CT Region of Native SARS-CoV-2 S protein (wtTM/CT)                               |
| 17         | TM/CT of H5 A/Indonesia/5/05 Hemagglutinin (H5iTM/CT)                               |
| 18         | TM/CT Region of Modified SARS-CoV-2 S protein with H5i Hemagglutinin CT             |
| 19         | TM/CT Region of Modified S protein with H5i Hemagglutinin CT (Variation 1)          |
| 20         | (no SP) SARS-CoV-2 S protein GSAS+PP wtTM/CT AA                                     |
| 21         | (no SP) Modified SARS-CoV-2 S protein GSAS+PP with H5i Hemagglutinin CT AA          |
| 22         | PDI-SARS-CoV-2 S protein GSAS+PP wtTM/CT-DNA                                        |
| 23         | PDI-SARS-CoV-2 S protein GSAS+PP wtTM/CT-AA (product of construct 8591)             |
| 24         | IF(PDI)-CoV(opt2).c                                                                 |
| 25         | IF(AVB)-CoV(opt2).r                                                                 |
| 26         | PDI-Modified SARS-CoV-2 S protein GSAS+PP H5iTM/CT-DNA                              |
| 27         | PDI-Modified SARS-CoV-2 S protein GSAS+PP H5iTM/CT-AA (product of construct 8597)   |
| 28         | IF(Avb)-H5I.r                                                                       |
| 29         | PDI-Modified SARS-CoV-2 S protein GSAS+PP wtTM/H5iCT-DNA                            |
| 30         | PDI-Modified SARS-CoV-2 S protein GSAS+PP wtTM/H5iCT-AA (product of construct 8671) |
| 31         | Cloning vector 8501 from left to right T-DNA                                        |

| SEQ ID NO: | Description of Sequence                                                                        |
|------------|------------------------------------------------------------------------------------------------|
| 32         | Construct 8586 from 2X35S(+C) prom to NOS term                                                 |
| 33         | Cloning vector 8500 from left to right T-DNA]                                                  |
| 34         | Construct 8589 from 2X35S(+C) prom to NOS term                                                 |
| 35         | Cloning vector 8716 from left to right T-DNA                                                   |
| 36         | Construct 8591 from 2X35S(+C) prom to NOS term                                                 |
| 37         | C-Terminal Region of Modified S protein with H5i Hemagglutinin CT (alternative 2)              |
| 38         | C-Terminal Region of Modified S protein with H5i Hemagglutinin CT (alternative 3)              |
| 39         | C-Terminal Region of Modified S protein with H5i Hemagglutinin CT (alternative 4)              |
| 40         | 3'UTR AvB (Arracacha Virus B Isolate)                                                          |
| 41         | 3'UTR trBNYVV (Beet necrotic yellow vein virus)                                                |
| 42         | 3' UTR SBMV (Southern bean mosaic virus)                                                       |
| 43         | 3' UTR TuRSV (Turnip ringspot virus)                                                           |
| 44         | 3' UTR CPMV (Cowpea Mosaic virus)                                                              |
| 45         | 3' BBTMV (Broad bean true mosaic virus)                                                        |
| 46         | 3' UTR trOUMV (Ourmia melon virus)                                                             |
|            | <b>Modified S Protein with Substitutions</b>                                                   |
| 47         | PDI-S Protein GSAS-2P (product of construct 8671)                                              |
| 48         | PDI-S Protein GSAS-4P (product of construct 8953)                                              |
| 49         | PDI-S Protein GSAS-6P (product of construct 8940)                                              |
| 50         | PDI-S Protein GSAS-2P-923 (product of construct 8933)                                          |
| 51         | PDI-S Protein GSAS-4P-923 (product of construct 8960)                                          |
| 52         | PDI-S Protein GSAS-6P-923 (product of construct 8947)                                          |
|            | <b>Modified S with CT from different HA strains</b>                                            |
| 53         | PDI-S-protein+H1 Cal (product of construct 7390)                                               |
| 54         | PDI-S-protein+H3 Minn (product of construct 7391)                                              |
| 55         | PDI-S-protein+H6 HK (product of construct 7392)                                                |
| 56         | PDI-S-protein+H7 Guangdong (product of construct 7393)                                         |
| 57         | PDI-S-protein+H9 HK (product of construct 7394)                                                |
| 58         | PDI-S-protein+ B/ Wash (product of construct 7395)                                             |
|            | <b>Modified S Protein C-terminal variations</b>                                                |
| 59         | PDI-Modified S protein with H5i Hemagglutinin CT (V1) (product of construct 8980)              |
| 60         | PDI-Modified S protein with H5i Hemagglutinin CT (V2) (product of construct 8981)              |
| 61         | PDI-Modified S protein with H5i Hemagglutinin CT (V3) (product of construct 8982)              |
| 62         | PDI-Modified S protein with H5i Hemagglutinin CT (V4) (product of construct 8983)              |
|            | <b>N-terminal region</b>                                                                       |
| 63         | N-terminal region of native SARS-CoV-2 protein (including native signal peptide)               |
| 64         | TM/CT Region of Modified SARS-CoV-2 S protein with intervening peptide sequence X <sub>n</sub> |
|            | <b>Modified S Protein with Substitutions (DNA)</b>                                             |
| 65         | PDI-S Protein GSAS+4P-DNA                                                                      |
| 66         | PDI-S Protein GSAS+6P-DNA                                                                      |
| 67         | PDI-S Protein GSAS+2P+L923F-DNA                                                                |
| 68         | PDI-S Protein GSAS+4P+L923F-DNA                                                                |
| 69         | PDI-S Protein GSAS+6P+L923F-DNA                                                                |
|            | <b>Modified S Protein C-terminal variations (DNA)</b>                                          |
| 70         | PDI-Modified S protein with H5i Hemagglutinin CT (V1) DNA                                      |
| 71         | PDI-Modified S protein with H5i Hemagglutinin CT (V2) DNA                                      |
| 72         | PDI-Modified S protein with H5i Hemagglutinin CT (V3) DNA                                      |

| SEQ ID NO: | Description of Sequence                                               |
|------------|-----------------------------------------------------------------------|
| 73         | PDI-Modified S protein with H5i Hemagglutinin CT (V4) DNA             |
|            | <b>Modified S with CT from different HA strains (DNA) and Primers</b> |
| 74         | PDI-S-protein+H1 Cal DNA                                              |
| 75         | PDI-S-protein+H3 Minn DNA                                             |
| 76         | PDI-S-protein+H6 HK DNA                                               |
| 77         | PDI-S-protein+H7 Guangdong DNA                                        |
| 78         | PDI-S-protein+H9 HK DNA                                               |
| 79         | PDI-S-protein+ B/ Wash DNA                                            |
| 80         | IF-H1HawaiiCT.r                                                       |
| 81         | IF-H3MinnesotaCT.r                                                    |
| 82         | IF-HongKongCT.r                                                       |
| 83         | IF-GuangdongCT.r                                                      |
| 84         | IF-H9HKCT.r                                                           |
| 85         | IF-BWashCT.r                                                          |
|            | <b>Modified SARS-COV-1 S Protein (DNA) and Primers</b>                |
| 86         | IF(nbHEL40)-PDI.c                                                     |
| 87         | IF(AvB+wtCT).r                                                        |
| 88         | PDI-SARS-COV-1 wtTMCT-DNA                                             |
| 89         | PDI-SARS-COV-1 H5iTMCT-DNA                                            |
| 90         | PDI-SARS-COV-1 H5iCT-DNA                                              |
| 91         | PDI-SARS-COV-1 H5iCT(V4)-DNA                                          |
| 92         | PDI-SARS-COV-1 H1cCT-DNA                                              |
|            | <b>Modified SARS-COV-1 S Protein (AA)</b>                             |
| 93         | PDI-SARS-COV-1 wtTMCT-AA                                              |
| 94         | PDI-SARS-COV-1 H5iTMCT-AA                                             |
| 95         | PDI-SARS-COV-1 H5iCT-AA                                               |
| 96         | PDI-SARS-COV-1 H5iCT(V4)-AA                                           |
| 97         | PDI-SARS-COV-1 H1cCT-AA                                               |
|            | <b>Modified MERS S Protein (DNA) and Primers</b>                      |
| 98         | IF(AvB+wtCT-MERS).r                                                   |
| 99         | IF(H1cCT-wtTM).r                                                      |
| 100        | IF(H5iTMCT).r                                                         |
| 101        | PDI-MERS-wtTMCT-DNA                                                   |
| 102        | PDI-MERS-H5iTMCT-DNA                                                  |
| 103        | PDI-MERS-H5iCT-DNA                                                    |
| 104        | PDI-MERS-H5iCT(V4)-DNA                                                |
| 105        | PDI-MERS-H1cCT-DNA                                                    |
|            | <b>Modified SARS-COV-1 S Protein (AA)</b>                             |
| 106        | PDI-MERS-wtTMCT-AA                                                    |
| 107        | PDI-MERS-H5iTMCT-AA                                                   |
| 108        | PDI-MERS-H5iCT-AA                                                     |
| 109        | PDI-MERS-H5iCT(V4)-AA                                                 |
| 110        | PDI-MERS-H1cCT-AA                                                     |
|            | <b>Other Sequences</b>                                                |
| 111        | Cloning vector 7147 from left to right T-DNA                          |
| 112        | Native SARS-CoV-1 S protein wtTM/CT AA (P59594)                       |
| 113        | Native MERS S protein wtTM/CT AA (AFY13307)                           |

| SEQ ID NO:                                           | Description of Sequence                                                                        |
|------------------------------------------------------|------------------------------------------------------------------------------------------------|
| 114                                                  | Native SARS-CoV-1 S protein wtTM/CT AA (P59594) without signal peptide                         |
| 115                                                  | Native MERS S protein wtTM/CT AA (AFY13307) without signal peptide                             |
| 116                                                  | TMCT region of modified PDI-SARS-COV-1 wtTMCT-AA                                               |
| 117                                                  | TMCT region of modified PDI-SARS-COV-1 H5iTMCT-AA                                              |
| 118                                                  | TMCT region of modified PDI-SARS-COV-1 H5iCT-AA                                                |
| 119                                                  | TMCT region of modified PDI-SARS-COV-1 H5iCT(V4)-AA                                            |
| 120                                                  | TMCT region of modified PDI-SARS-COV-1 H1cCT-AA                                                |
| 121                                                  | TMCT region of modified PDI-MERS-wtTMCT-AA                                                     |
| 122                                                  | TMCT region of modified PDI-MERS-H5iTMCT-AA                                                    |
| 123                                                  | TMCT region of modified PDI-MERS-H5iCT-AA                                                      |
| 124                                                  | TMCT region of modified PDI-MERS-H5iCT(V4)-AA                                                  |
| 125                                                  | TMCT region of modified PDI-MERS-H1cCT-AA                                                      |
| 126                                                  | TMCT region of modified PDI-S-protein+H1 Cal                                                   |
| 127                                                  | TMCT region of modified PDI-S-protein+H3 Minn                                                  |
| 128                                                  | TMCT region of modified PDI-S-protein+H6 HK                                                    |
| 129                                                  | TMCT region of modified PDI-S-protein+H7 Guangdong                                             |
| 130                                                  | TMCT region of modified PDI-S-protein+H9 HK                                                    |
| 131                                                  | TMCT region of modified PDI-S-protein+ B/ Wash                                                 |
| 132                                                  | Consensus Sequence of TM Domain of Coronavirus S-protein                                       |
| 133                                                  | Consensus Sequence of TM Domain of Coronavirus S-protein                                       |
| 134                                                  | TM/CT Region of Modified SARS-CoV-1 S protein with intervening peptide sequence X <sub>n</sub> |
| 135                                                  | TM/CT Region of Modified MERS S protein with intervening peptide sequence X <sub>n</sub>       |
| <b>Modified OC43-CoV S Protein (DNA) and Primers</b> |                                                                                                |
| 136                                                  | IF(AvB+wtCT-OC43).r                                                                            |
| 137                                                  | PDI-OC43-wtTMCT-DNA                                                                            |
| 138                                                  | PDI-OC43-H5iTMCT-DNA                                                                           |
| 139                                                  | PDI-OC43-H5iCT-DNA                                                                             |
| 140                                                  | PDI-OC43-H5iCT(V4)-DNA                                                                         |
| 141                                                  | PDI-OC43-H1cCT-DNA                                                                             |
| <b>Modified OC43-CoV S Protein (AA)</b>              |                                                                                                |
| 142                                                  | PDI-OC43-wtTMCT-AA                                                                             |
| 143                                                  | PDI-OC43-H5iTMCT-AA                                                                            |
| 144                                                  | PDI-OC43-H5iCT-AA                                                                              |
| 145                                                  | PDI-OC43-H5iCT(V4)-AA                                                                          |
| 146                                                  | PDI-OC43-H1cCT-AA                                                                              |
| <b>Modified 229E-CoV S Protein (DNA) and Primers</b> |                                                                                                |
| 147                                                  | IF(CoV229EwtCT).r                                                                              |
| 148                                                  | PDI-229E -wtTMCT-DNA                                                                           |
| 149                                                  | PDI-229E -H5iTMCT-DNA                                                                          |
| 150                                                  | PDI-229E -H5iCT-DNA                                                                            |
| 151                                                  | PDI-229E -H5iCT(V4)-DNA                                                                        |
| 152                                                  | PDI-229E -H1cCT-DNA                                                                            |
| <b>Modified 229E-CoV S Protein (AA)</b>              |                                                                                                |
| 153                                                  | PDI-229E -wtTMCT-AA                                                                            |
| 154                                                  | PDI-229E -H5iTMCT-AA                                                                           |
| 155                                                  | PDI-229E -H5iCT-AA                                                                             |
| 156                                                  | PDI-229E -H5iCT(V4)-AA                                                                         |

| SEQ ID NO: | Description of Sequence                                                                      |
|------------|----------------------------------------------------------------------------------------------|
| 157        | PDI-229E-H1cCT-AA                                                                            |
|            | <b>Other Sequences</b>                                                                       |
| 158        | Native OC43-CoV S protein wtTM/CT AA (AVR40344)                                              |
| 159        | Native 229E S protein wtTM/CT AA (P15423)                                                    |
| 160        | Native OC43-CoV S protein wtTM/CT AA (AVR40344) without signal peptide                       |
| 161        | Native 229E S protein wtTM/CT AA (P15423) without signal peptide                             |
| 162        | TMCT region of modified PDI-OC43-COV wtTMCT-AA                                               |
| 163        | TMCT region of modified PDI-OC43-COV H5iTMCT-AA                                              |
| 164        | TMCT region of modified PDI-OC43-COV H5iCT-AA                                                |
| 165        | TMCT region of modified PDI-OC43-COV H5iCT(V4)-AA                                            |
| 166        | TMCT region of modified PDI-OC43-COV H1cCT-AA                                                |
| 167        | TMCT region of modified PDI-229E-wtTMCT-AA                                                   |
| 168        | TMCT region of modified PDI-229E-H5iTMCT-AA                                                  |
| 169        | TMCT region of modified PDI-229E-H5iCT-AA                                                    |
| 170        | TMCT region of modified PDI-229E-H5iCT(V4)-AA                                                |
| 171        | TMCT region of modified PDI-229E-H1cCT-AA                                                    |
| 172        | TM/CT Region of Modified OC43-CoV S protein with intervening peptide sequence X <sub>n</sub> |
| 173        | TM/CT Region of Modified 229E-CoV S protein with intervening peptide sequence X <sub>n</sub> |

[00309] The present invention will be further illustrated in the following examples.

### Examples

#### Example 1: Preparation of Modified Structural Protein Constructs

5 [00310] The SARS-CoV-2 S protein constructs were produced using techniques well known within the art. For example SARS-COV-2 Spike Protein with wtTMCT (Constructs number 8586, 8589, 8591, see Figure 8A-8C) were cloned as described below. Constructs for SARS-COV-2 Spike Protein with H5iTMCT (Constructs number 8592, 8595, 8597, see Figure 9A-9C) and constructs for SARS-COV-2 Spike Protein  
10 with H5iCT (Constructs number 8610, 8611, 8671, see Figure 10A-10C) were obtained using similar techniques and sequences primers, templates and products are described in Table 5.

#### *SARS-COV-2 Spike Protein with wtTMCT (Constructs number 8586, 8589, 8591)*

15 [00311] A sequence encoding mature SARS-CoV-2 Spike (S) protein 2 (SEQ ID NO: 23) with GSAS+K971P+V972P ectodomain mutations and with native transmembrane domain and native cytoplasmic tail (wtTMCT) from SARS-CoV-2, fused

to alfalfa PDI secretion signal peptide (PDISP) was cloned into three different expression systems using the following PCR-based method. A fragment containing the SARS-COV-2 Spike protein (wtTMCT) coding sequence was amplified using primers IF(PDI)-CoV(opt2).c (SEQ ID NO: 24) and IF(AVB)-CoV(opt2).r (SEQ ID NO: 25), using PDISP-SARS-COV-2 Spike Protein with wtTMCT gene sequence (SEQ ID NO: 22) as template. The PCR product was cloned into three different expression systems using In-Fusion cloning system (Clontech, Mountain View, CA).

[00312] For the first expression system, construct number 8501 (Figure 7A), was digested with AatII and StuI restriction enzymes and the linearized plasmid was used for the first In-Fusion assembly reaction. Construct number 8501 is an acceptor plasmid intended for “In Fusion” cloning of genes of interest in a 2X35S(+C)/nbMT78/PDI/AvB/NOS-based expression cassette. This acceptor plasmid also incorporates a gene construct for the co-expression of the TBSV P19 suppressor of silencing under the alfalfa Plastocyanin gene promoter and terminator. The backbone is a pCAMBIA binary plasmid and the sequence from left to right t-DNA borders is presented in SEQ ID NO: 31. The resulting construct was given number 8586 (SEQ ID NO: 32). The amino acid sequence of mature spike protein from SARS-COV-2 fused to alfalfa PDI secretion signal peptide (PDISP) is presented in SEQ ID NO: 23. A representation of plasmid 8586 is presented in Figure 8A.

[00313] For the second expression system, construct number 8500 (Figure 7B), was also digested with AatII and StuI restriction enzymes and the linearized plasmid was used for the second In-Fusion assembly reaction. Construct number 8500 is an acceptor plasmid intended for “In Fusion” cloning of genes of interest in a 2X35S(+C)/nbCSY65/PDI/AvB/NOS-based expression cassette. This acceptor plasmid also incorporates a gene construct for the co-expression of the TBSV P19 suppressor of silencing under the alfalfa Plastocyanin gene promoter and terminator. The backbone is a pCAMBIA binary plasmid and the sequence from left to right t-DNA borders is presented in SEQ ID NO: 33. The resulting construct was given number 8589 (SEQ ID NO: 34). The amino acid sequence of mature spike protein from SARS-COV-2 fused to

alfalfa PDI secretion signal peptide (PDISP) is presented in SEQ ID NO: 23. A representation of plasmid 8589 is presented in Figure 8B.

[00314] For the third expression system, construct number 8716 (Figure 7C), was also digested with AatII and StuI restriction enzymes and the linearized plasmid was used for the third In-Fusion assembly reaction. Construct number 8716 is an acceptor plasmid intended for “In Fusion” cloning of genes of interest in a 2X35S(+C)/nbHEL40/PDI/AvB/NOS based expression cassette. This acceptor plasmid also incorporates a gene construct for the co-expression of the TBSV P19 suppressor of silencing under the alfalfa Plastocyanin gene promoter and terminator. The backbone is a pCambia binary plasmid and the sequence from left to right t-DNA borders is presented in SEQ ID NO: 35. The resulting construct was given number 8591 (SEQ ID NO: 36). The amino acid sequence of mature spike protein from SARS-COV-2 fused to alfalfa PDI secretion signal peptide (PDISP) is presented in SEQ ID NO: 23. A representation of plasmid 8591 is presented in Figure 8C.

*SARS-COV-2 Spike Protein with H5iTMCT (Constructs number 8592, 8595, 8597)*

[00315] A sequence encoding mature Spike (S) protein from SARS-CoV-2 (SEQ ID NO: 27) with GSAS+K971P+V972P ectodomain mutations, and with transmembrane domain and cytoplasmic tail from H5 A/Indonesia/5/05 HA (H5iTMCT), was fused to alfalfa PDI secretion signal peptide (PDISP) and cloned into the same three expression systems described above by a similar PCR-based method (see table 5 for primers and Example 3 for sequences used). Construct number 8592 (Figure 9A) was derived from acceptor construct 8501, construct number 8595 (Figure 9B) was derived from acceptor construct 8500 and construct number 8597 (Figure 9C) was derived from acceptor construct 8716 using similar techniques as described above and the primers, templates and products is provided in Table 5 below.

*SARS-COV-2 Spike Protein with H5iCT (Constructs number 8610, 8611, 8671)*

[00316] A sequence encoding mature Spike (S) protein from SARS-CoV-2 (SEQ ID NO: 30) with GSAS+K971P+V972P ectodomain mutations, and cytoplasmic tail from

H5 A/Indonesia/5/05 HA (H5iCT), was fused to alfalfa PDI secretion signal peptide (PDISP) and cloned into the same three expression systems as described above by a similar PCR-based method (see table 5 for primers and Example 3 for sequences used). Construct number 8610 (Figure 10A) was derived from acceptor construct 8501, construct number 8611 (Figure 10B) was derived from acceptor construct 8500 and construct number 8671 (Figure 10C) was derived from acceptor construct 8716 using similar techniques as described above and the primers, templates and products is provided in Table 5 below.

*SARS-COV-2 Spike Protein with Alternative TM CT Fusion Sequences (Constructs number 8980, 8981, 8982, 8983)*

[00317] A sequence encoding mature Spike (S) protein from SARS-CoV-2 with GSAS+K971P+V972P ectodomain mutations, and cytoplasmic tail from H5 A/Indonesia/5/05 HA (H5iCT), as depicted in SEQ ID NO: 19, was fused to alfalfa PDI secretion signal peptide (PDISP) and cloned into the same expression system as described for construct 8671, yielding construct 8980 (Figure 12A). Similar constructs were created for a sequence encoding mature Spike (S) protein from SARS-CoV-2 with GSAS+K971P+V972P ectodomain mutations, and cytoplasmic tail from H5 A/Indonesia/5/05 HA (H5iCT), as depicted in SEQ ID NO: 37 (construct 8981, Figure 12B), SEQ ID NO: 38 (construct 8982, Figure 12C), and SEQ ID NO: 39 (construct 8983, Figure 12D).

*SARS-COV-2 Spike Protein with CT from other HA Strains (Constructs number 7390, 7391, 7392, 7393, 7394, and 7395)*

[00318] A sequence encoding mature Spike (S) protein from SARS-CoV-2 with GSAS+K971P+V972P ectodomain mutations, and cytoplasmic tail from H1 A/California/7/2009 HA (H1CT), was fused to alfalfa PDI secretion signal peptide (PDISP) and cloned into the same expression system as described for construct 8671 above by a similar PCR-based method (see table 5 for primers and Example 3 for sequences used). The resulting construct 7390 thus encodes a modified S protein comprising a H1 A/California/7/2009 HA cytoplasmic tail (H1CT) (Figure 13A). Similar

constructs were created for H3 A/Minnesota/41/2019 (Construct 7391, H3 CT) (Figure 13B), H6 A/Teal/Hong Kong/W312/97 (Construct 7392, H6 CT) (Figure 13C), H7 A/Guangdong/17SF003/2016 (Construct 7393, H7 CT) (Figure 13D), H9 A/Hong Kong/1073/99 (Construct 7394, H9h CT) (Figure 13E) or B/Washington/02/2019 (Construct 7395, HA B CT) (Figure 13F).

*SARS-COV-2 Spike Protein with Substitutions (Constructs numbers 8933, 8960, 8947)*

[00319] Modified SARS-CoV-2 S protein constructs comprising combinations of mutations in the S protein, such as R667G, R668S, R670S, F802P, A877P, A884P, A927P, K971P, V972P, and L923F were produced using techniques well known within the art and basically as described above. The constructs have the following substitutions: Construct 8933: R667G, R668S, R670S, K971P, V972P and L923F (“GSAS-2P-923”); construct 8960: R667G, R668S, R670S, F802P, A927P, K971P, V972P and L923F (“GSAS-4P-923”) and construct 8947: R667G, R668S, R670S, F802P, A877P, A884P, A927P, K971P, V972P and L923F (“GSAS-6P-923”).

*SARS-COV-1 Spike Protein with wtTMCT and modified TMCT (Constructs number 9231, 9232, 9233, 9234, 9235)*

[00320] A sequence encoding mature SARS-CoV-1 Spike (S) protein (SEQ ID NO: 88) with R654A+K955P+V956P ectodomain mutations and with native transmembrane domain and native cytoplasmic tail (wtTMCT) from SARS-CoV-1, fused to alfalfa PDI secretion signal peptide (PDISP) was cloned into the following expression system by a PCR-based method. A fragment containing the PDISP-SARS-COV-1 Spike protein (wtTMCT) coding sequence was amplified using primers IF(nbHEL40)-PDI.c (SEQ ID NO: 86) and IF(AvB+wtCT).r (SEQ ID NO: 87), using PDISP-SARS-COV-1 Spike Protein with wtTMCT gene sequence (SEQ ID NO: 88) as template. The PCR product was cloned into the following expression system using In-Fusion cloning system (Clontech, Mountain View, CA).

[00321] Construct number 7147 (Figure 21) was digested with AatII and StuI restriction enzymes and the linearized plasmid was used for the first In-Fusion assembly reaction. Construct number 7147 is an acceptor plasmid intended for “In Fusion” cloning

of genes of interest in a 2X35S(+C)/nbHEL40/AvB/NOS-based expression cassette. This acceptor plasmid also incorporates a gene construct for the co-expression of the TBSV P19 suppressor of silencing under the alfalfa Plastocyanin gene promoter and terminator. The backbone is a pCAMBIA binary plasmid and the sequence from left to right t-DNA borders is presented in SEQ ID NO: 111. The resulting construct was given number 9231. The amino acid sequence of mature spike protein from SARS-COV-1 fused to alfalfa PDI secretion signal peptide (PDISP) is presented in SEQ ID NO: 93. A representation of plasmid 9231 is presented in Figure 18A.

[00322] A sequence encoding mature Spike (S) protein from SARS-CoV-1 with R654A+ K955P+V956P ectodomain mutations, and either i) transmembrane domain and cytoplasmic tail from H5 A/Indonesia/5/05 HA (H5iTMCT), ii) cytoplasmic tail from H5 A/Indonesia/5/05 HA (H5iCT and variation H5iCT(V4)), or iii) cytoplasmic tail from H1 A/California/7/2009 HA (H1cCT), were fused to alfalfa PDI secretion signal peptide (PDISP) and cloned into the same expression system as described for construct 9231 above by a similar PCR-based method (see table 5 for primers and Example 3 for sequences used). The resulting constructs 9232, 9233, 9234, 9235 thus encode a modified S protein comprising a H5 A/Indonesia/5/05 TMCT (H5iTMCT) (Figure 18B, SEQ ID NO: 94), a modified SARS-COV-1 S protein comprising a H5 A/Indonesia/5/05 CT (H5iCT) (Figure 18C, SEQ ID NO: 95), a modified S protein comprising a H5 A/Indonesia/5/05 CT variant (H5iCT(V4)) (Figure 18D, SEQ ID NO: 96), or a modified S protein comprising a H1 A/California/7/2009 CT (H1cCT) (Figure 18E, SEQ ID NO: 97.)

*MERS-CoV Spike Protein with wtTMCT and modified TMCT (Constructs number 9246, 9247, 9249, 9250, 9251)*

[00323] A sequence encoding mature MERS-CoV Spike (S) protein (SEQ ID NO: 101) with R730A+R733G+ V1043P+L1044P ectodomain mutations and with native transmembrane domain and native cytoplasmic tail (wtTMCT) from MERS-CoV, fused to alfalfa PDI secretion signal peptide (PDISP) was cloned into the following expression system by a PCR-based method. A fragment containing the PDISP-MERS-COV Spike

protein (wtTMCT) coding sequence was amplified using primers IF(nbHEL40)-PDI.c (SEQ ID NO: 86) and IF(AvB+wtCT-MERS).r (SEQ ID NO: 98), using PDISP-MERS-COV Spike Protein with wtTMCT gene sequence (SEQ ID NO: 101) as template. The PCR product was cloned into the following expression system using In-Fusion cloning system (Clontech, Mountain View, CA).

[00324] Construct number 7147 (Figure 21) was digested with AatII and StuI restriction enzymes and the linearized plasmid was used for the first In-Fusion assembly reaction. Construct number 7147 is an acceptor plasmid intended for “In Fusion” cloning of genes of interest in a 2X35S(+C)/nbHEL40/AvB/NOS-based expression cassette. This acceptor plasmid also incorporates a gene construct for the co-expression of the TBSV P19 suppressor of silencing under the alfalfa Plastocyanin gene promoter and terminator. The backbone is a pCAMBIA binary plasmid and the sequence from left to right t-DNA borders is presented in SEQ ID NO: 111. The resulting construct was given number 9246. The amino acid sequence of mature spike protein from MERS-COV fused to alfalfa PDI secretion signal peptide (PDISP) is presented in SEQ ID NO: 106. A representation of plasmid 9246 is presented in Figure 20A.

[00325] A sequence encoding mature Spike (S) protein from MERS-CoV with R730A+R733G+ V1043P+L1044P ectodomain mutations, and either i) transmembrane domain and cytoplasmic tail from H5 A/Indonesia/5/05 HA (H5iTMCT), ii) cytoplasmic tail from H5 A/Indonesia/5/05 HA (H5iCT and variation H5iCT(V4)), or iii) cytoplasmic tail from H1 A/California/7/2009 HA (H1cCT), were fused to alfalfa PDI secretion signal peptide (PDISP) and cloned into the same expression system as described for construct 9246 above by a similar PCR-based method (see table 5 for primers and Example 3 for sequences used). The resulting constructs 9247, 9249, 9250, 9251 thus encode a modified MERS-COV S protein comprising a H5 A/Indonesia/5/05 TMCT (H5iTMCT) (Figure 20B, SEQ ID NO: 107), a modified S protein comprising a H5 A/Indonesia/5/05 CT (H5iCT) (Figure 20C, SEQ ID NO: 108), a modified S protein comprising a H5 A/Indonesia/5/05 CT variant (H5iCT(V4)) (Figure 20D, SEQ ID NO: 109), or a modified S protein comprising a H1 A/California/7/2009 CT (H1cCT) (Figure 20E, SEQ ID NO: 110.)

*OC43-CoV Spike Protein with wtTMCT and modified TMCT (Constructs number 9269, 9270, 9272, 9273 and 9274)*

[00326] A sequence encoding mature OC43-CoV Spike (S) protein (SEQ ID NO: 137) with R761G+R762G+R764G+R765S+A1077P+L1078P ectodomain mutations and with native transmembrane domain and native cytoplasmic tail (wtTMCT) from OC43-CoV, fused to alfalfa PDI secretion signal peptide (PDISP) was cloned into the following expression system by a PCR-based method. A fragment containing the PDISP- OC43-COV Spike protein (wtTMCT) coding sequence was amplified using primers IF(nbHEL40)-PDI.c (SEQ ID NO: 86) and IF(AvB+wtCT-OC43).r (SEQ ID NO: 136), using PDISP-OC43-COV Spike Protein with wtTMCT gene sequence (SEQ ID NO: 137) as template. The PCR product was cloned into the following expression system using In-Fusion cloning system (Clontech, Mountain View, CA).

[00327] Construct number 7147 (Figure 21) was digested with AatII and StuI restriction enzymes and the linearized plasmid was used for the first In-Fusion assembly reaction. Construct number 7147 is an acceptor plasmid intended for “In Fusion” cloning of genes of interest in a 2X35S(+C)/nbHEL40/AvB/NOS-based expression cassette. This acceptor plasmid also incorporates a gene construct for the co-expression of the TBSV P19 suppressor of silencing under the alfalfa Plastocyanin gene promoter and terminator. The backbone is a pCambia binary plasmid and the sequence from left to right t-DNA borders is presented in SEQ ID NO: 111. The resulting construct was given number 9269. The amino acid sequence of mature spike protein from OC43-COV fused to alfalfa PDI secretion signal peptide (PDISP) is presented in SEQ ID NO: 142. A representation of plasmid 9269 is presented in Figure 24A.

[00328] A sequence encoding mature Spike (S) protein from OC43-CoV with R761G+R762G+R764G+R765S+A1077P+L1078P ectodomain mutations, and either i) transmembrane domain and cytoplasmic tail from H5 A/Indonesia/5/05 HA (H5iTMCT), ii) cytoplasmic tail from H5 A/Indonesia/5/05 HA (H5iCT and variation H5iCT(V4)), or iii) cytoplasmic tail from H1 A/California/7/2009 HA (H1cCT), were fused to alfalfa PDI secretion signal peptide (PDISP) and cloned into the same expression system as described

for construct 9269 above by a similar PCR-based method (see table 5 for primers and Example 3 for sequences used). The resulting constructs 9270, 9272, 9273 and 9274 thus encode a modified OC43-COV S protein comprising a H5 A/Indonesia/5/05 TMCT (H5iTMCT) (Figure 24B, SEQ ID NO: 143), a modified S protein comprising a H5 A/Indonesia/5/05 CT (H5iCT) (Figure 24C, SEQ ID NO: 144), a modified S protein comprising a H5 A/Indonesia/5/05 CT variant (H5iCT(V4)) (Figure 24D, SEQ ID NO: 145), or a modified S protein comprising a H1 A/California/7/2009 CT (H1cCT) (Figure 24E, SEQ ID NO: 146).

*229E-CoV Spike Protein with wtTMCT and modified TMCT (Constructs number 9310, 9311, 9312, 9313 and 9314)*

[00329] A sequence encoding mature 229E-CoV Spike (S) protein (SEQ ID NO: 148) with R567A+T871P+I872P ectodomain mutations and with native transmembrane domain and native cytoplasmic tail (wtTMCT) from 229E -CoV, fused to alfalfa PDI secretion signal peptide (PDISP) was cloned into the following expression system by a PCR-based method. A fragment containing the PDISP-229E -COV Spike protein (wtTMCT) coding sequence was amplified using primers IF(nbHEL40)-PDI.c (SEQ ID NO: 86) and IF(CoV229EwtCT).r (SEQ ID NO: 147), using PDISP-OC43-COV Spike Protein with wtTMCT gene sequence (SEQ ID NO: 148) as template. The PCR product was cloned into the following expression system using In-Fusion cloning system (Clontech, Mountain View, CA).

[00330] Construct number 7147 (Figure 21) was digested with AatII and StuI restriction enzymes and the linearized plasmid was used for the first In-Fusion assembly reaction. Construct number 7147 is an acceptor plasmid intended for “In Fusion” cloning of genes of interest in a 2X35S(+C)/nbHEL40/AvB/NOS-based expression cassette. This acceptor plasmid also incorporates a gene construct for the co-expression of the TBSV P19 suppressor of silencing under the alfalfa Plastocyanin gene promoter and terminator. The backbone is a pCAMBIA binary plasmid and the sequence from left to right t-DNA borders is presented in SEQ ID NO: 111. The resulting construct was given number 9310. The amino acid sequence of mature spike protein from 229E -CoV fused to alfalfa PDI

secretion signal peptide (PDISP) is presented in SEQ ID NO: 153. A representation of plasmid 9310 is presented in Figure 26A.

[00331] A sequence encoding mature Spike (S) protein from 229E -CoV with R567A+T871P+I872P ectodomain mutations, and either i) transmembrane domain and cytoplasmic tail from H5 A/Indonesia/5/05 HA (H5iTMCT), ii) cytoplasmic tail from H5 A/Indonesia/5/05 HA (H5iCT and variation H5iCT(V4)), or iii) cytoplasmic tail from H1 A/California/7/2009 HA (H1cCT), were fused to alfalfa PDI secretion signal peptide (PDISP) and cloned into the same expression system as described for construct 9310 above by a similar PCR-based method (see table 5 for primers and Example 3 for sequences used). The resulting constructs 9311, 9312, 9313 and 9314 thus encode a modified 229E -COV S protein comprising a H5 A/Indonesia/5/05 TMCT (H5iTMCT) (Figure 26B, SEQ ID NO: 154), a modified S protein comprising a H5 A/Indonesia/5/05 CT (H5iCT) (Figure 26C, SEQ ID NO: 155), a modified S protein comprising a H5 A/Indonesia/5/05 CT variant (H5iCT(V4)) (Figure 26D, SEQ ID NO: 156), or a modified S protein comprising a H1 A/California/7/2009 CT (H1cCT) (Figure 26E, SEQ ID NO: 157).

Table 5. List of Construct Numbers, Primer and Templates

| Construct # | 5' UTR  | Gene of interest (GOI) | TMCT Region     | Primer 1 (forward primer of fragment 1) | Primer 2 (reverse primer of fragment 1) | Template for first PCR / Resulting gene | Resulting protein | Acceptor plasmid |
|-------------|---------|------------------------|-----------------|-----------------------------------------|-----------------------------------------|-----------------------------------------|-------------------|------------------|
| 8586        | nbMT78  | SARS-CoV-2 S           | wt TMCT         | SEQ ID NO: 24                           | SEQ ID NO: 25                           | SEQ ID NO: 22                           | SEQ ID NO: 23     | SEQ ID NO: 31    |
| 8589        | nbCSY65 | SARS-CoV-2 S           | wt TMCT         | SEQ ID NO: 24                           | SEQ ID NO: 25                           | SEQ ID NO: 22                           | SEQ ID NO: 23     | SEQ ID NO: 33    |
| 8591        | nbHEL40 | SARS-CoV-2 S           | wt TMCT         | SEQ ID NO: 24                           | SEQ ID NO: 25                           | SEQ ID NO: 22                           | SEQ ID NO: 23     | SEQ ID NO: 35    |
| 8592        | nbMT78  | SARS-CoV-2 S           | H5iTMCT         | SEQ ID NO: 24                           | SEQ ID NO: 28                           | SEQ ID NO: 26                           | SEQ ID NO: 27     | SEQ ID NO: 31    |
| 8595        | nbCSY65 | SARS-CoV-2 S           | H5iTMCT         | SEQ ID NO: 24                           | SEQ ID NO: 28                           | SEQ ID NO: 26                           | SEQ ID NO: 27     | SEQ ID NO: 33    |
| 8597        | nbHEL40 | SARS-CoV-2 S           | H5iTMCT         | SEQ ID NO: 24                           | SEQ ID NO: 28                           | SEQ ID NO: 26                           | SEQ ID NO: 27     | SEQ ID NO: 35    |
| 8610        | nbMT78  | SARS-CoV-2 S           | wtTM/H5iCT      | SEQ ID NO: 24                           | SEQ ID NO: 28                           | SEQ ID NO: 29                           | SEQ ID NO: 30     | SEQ ID NO: 31    |
| 8611        | nbCSY65 | SARS-CoV-2 S           | wtTM/H5iCT      | SEQ ID NO: 24                           | SEQ ID NO: 28                           | SEQ ID NO: 29                           | SEQ ID NO: 30     | SEQ ID NO: 33    |
| 8671        | nbHEL40 | SARS-CoV-2 S           | wtTM/H5iCT      | SEQ ID NO: 24                           | SEQ ID NO: 28                           | SEQ ID NO: 29                           | SEQ ID NO: 30     | SEQ ID NO: 35    |
| 8953        | nbHEL40 | SARS-CoV-2 S           | wtTM/H5iCT      | SEQ ID NO: 24                           | SEQ ID NO: 28                           | SEQ ID NO: 65                           | SEQ ID NO: 48     | SEQ ID NO: 35    |
| 8940        | nbHEL40 | SARS-CoV-2 S           | wtTM/H5iCT      | SEQ ID NO: 24                           | SEQ ID NO: 28                           | SEQ ID NO: 66                           | SEQ ID NO: 49     | SEQ ID NO: 35    |
| 8933        | nbHEL40 | SARS-CoV-2 S           | wtTM/H5iCT      | SEQ ID NO: 24                           | SEQ ID NO: 28                           | SEQ ID NO: 67                           | SEQ ID NO: 50     | SEQ ID NO: 35    |
| 8960        | nbHEL40 | SARS-CoV-2 S           | wtTM/H5iCT      | SEQ ID NO: 24                           | SEQ ID NO: 28                           | SEQ ID NO: 68                           | SEQ ID NO: 51     | SEQ ID NO: 35    |
| 8947        | nbHEL40 | SARS-CoV-2 S           | wtTM/H5iCT      | SEQ ID NO: 24                           | SEQ ID NO: 28                           | SEQ ID NO: 69                           | SEQ ID NO: 52     | SEQ ID NO: 35    |
| 8980        | nbHEL40 | SARS-CoV-2 S           | wtTM/H5iCT (V1) | SEQ ID NO: 24                           | SEQ ID NO: 28                           | SEQ ID NO: 70                           | SEQ ID NO: 59     | SEQ ID NO: 35    |
| 8981        | nbHEL40 | SARS-CoV-2 S           | wtTM/H5iCT (V2) | SEQ ID NO: 24                           | SEQ ID NO: 28                           | SEQ ID NO: 71                           | SEQ ID NO: 60     | SEQ ID NO: 35    |
| 8982        | nbHEL40 | SARS-CoV-2 S           | wtTM/H5iCT (V3) | SEQ ID NO: 24                           | SEQ ID NO: 28                           | SEQ ID NO: 72                           | SEQ ID NO: 61     | SEQ ID NO: 35    |
| 8983        | nbHEL40 | SARS-CoV-2 S           | wtTM/H5iCT (V4) | SEQ ID NO: 24                           | SEQ ID NO: 28                           | SEQ ID NO: 73                           | SEQ ID NO: 62     | SEQ ID NO: 35    |
| 7390        | nbHEL40 | SARS-CoV-2 S           | wtTM/H1cCT      | SEQ ID NO: 24                           | SEQ ID NO: 80                           | SEQ ID NO: 74                           | SEQ ID NO: 53     | SEQ ID NO: 35    |
| 7391        | nbHEL40 | SARS-CoV-2 S           | wtTM/H3mCT      | SEQ ID NO: 24                           | SEQ ID NO: 81                           | SEQ ID NO: 75                           | SEQ ID NO: 54     | SEQ ID NO: 35    |
| 7392        | nbHEL40 | SARS-CoV-2 S           | wtTM/H6hkCT     | SEQ ID NO: 24                           | SEQ ID NO: 82                           | SEQ ID NO: 76                           | SEQ ID NO: 55     | SEQ ID NO: 35    |
| 7393        | nbHEL40 | SARS-CoV-2 S           | wtTM/H7gCT      | SEQ ID NO: 24                           | SEQ ID NO: 83                           | SEQ ID NO: 77                           | SEQ ID NO: 56     | SEQ ID NO: 35    |
| 7394        | nbHEL40 | SARS-CoV-2 S           | wtTM/H9hkCT     | SEQ ID NO: 24                           | SEQ ID NO: 84                           | SEQ ID NO: 78                           | SEQ ID NO: 57     | SEQ ID NO: 35    |
| 7395        | nbHEL40 | SARS-CoV-2 S           | wtTM/BwCT       | SEQ ID NO: 24                           | SEQ ID NO: 85                           | SEQ ID NO: 79                           | SEQ ID NO: 58     | SEQ ID NO: 35    |
| 9231        | nbHEL40 | SARS-COV-1 S           | wtTMCT          | SEQ ID NO: 86                           | SEQ ID NO: 87                           | SEQ ID NO: 88                           | SEQ ID NO: 93     | SEQ ID NO: 111   |
| 9232        | nbHEL40 | SARS-COV-1 S           | H5iTMCT         | SEQ ID NO: 86                           | SEQ ID NO: 28                           | SEQ ID NO: 89                           | SEQ ID NO: 94     | SEQ ID NO: 111   |
| 9233        | nbHEL40 | SARS-COV-1 S           | wtTM/H5iCT      | SEQ ID NO: 86                           | SEQ ID NO: 28                           | SEQ ID NO: 90                           | SEQ ID NO: 95     | SEQ ID NO: 111   |

| Construct # | 5' UTR  | Gene of interest (GOI) | TMCT Region     | Primer 1 (forward primer of fragment 1) | Primer 2 (reverse primer of fragment 1) | Template for first PCR / Resulting gene | Resulting protein | Acceptor plasmid |
|-------------|---------|------------------------|-----------------|-----------------------------------------|-----------------------------------------|-----------------------------------------|-------------------|------------------|
| 9234        | nbHEL40 | SARS-COV-1 S           | wtTM/H5ICT (V4) | SEQ ID NO: 86                           | SEQ ID NO: 28                           | SEQ ID NO: 91                           | SEQ ID NO: 96     | SEQ ID NO: 111   |
| 9235        | nbHEL40 | SARS-COV-1 S           | wtTM/H1cCT      | SEQ ID NO: 86                           | SEQ ID NO: 80                           | SEQ ID NO: 92                           | SEQ ID NO: 97     | SEQ ID NO: 111   |
| 9246        | nbHEL40 | MERS S                 | wtTMCT          | SEQ ID NO: 86                           | SEQ ID NO: 98                           | SEQ ID NO: 101                          | SEQ ID NO: 106    | SEQ ID NO: 111   |
| 9247        | nbHEL40 | MERS S                 | H5iTMCT         | SEQ ID NO: 86                           | SEQ ID NO: 100                          | SEQ ID NO: 102                          | SEQ ID NO: 107    | SEQ ID NO: 111   |
| 9249        | nbHEL40 | MERS S                 | wtTM/H5ICT      | SEQ ID NO: 86                           | SEQ ID NO: 100                          | SEQ ID NO: 103                          | SEQ ID NO: 108    | SEQ ID NO: 111   |
| 9250        | nbHEL40 | MERS S                 | wtTM/H5ICT (V4) | SEQ ID NO: 86                           | SEQ ID NO: 100                          | SEQ ID NO: 104                          | SEQ ID NO: 109    | SEQ ID NO: 111   |
| 9251        | nbHEL40 | MERS S                 | wtTM/H1cCT      | SEQ ID NO: 86                           | SEQ ID NO: 99                           | SEQ ID NO: 105                          | SEQ ID NO: 110    | SEQ ID NO: 111   |
| 9269        | nbHEL40 | OC43 S                 | wtTMCT          | SEQ ID NO: 86                           | SEQ ID NO: 136                          | SEQ ID NO: 137                          | SEQ ID NO: 142    | SEQ ID NO: 111   |
| 9270        | nbHEL40 | OC43 S                 | H5iTMCT         | SEQ ID NO: 86                           | SEQ ID NO: 100                          | SEQ ID NO: 138                          | SEQ ID NO: 143    | SEQ ID NO: 111   |
| 9272        | nbHEL40 | OC43 S                 | wtTM/H5ICT      | SEQ ID NO: 86                           | SEQ ID NO: 100                          | SEQ ID NO: 139                          | SEQ ID NO: 144    | SEQ ID NO: 111   |
| 9273        | nbHEL40 | OC43 S                 | wtTM/H5ICT (V4) | SEQ ID NO: 86                           | SEQ ID NO: 100                          | SEQ ID NO: 140                          | SEQ ID NO: 145    | SEQ ID NO: 111   |
| 9274        | nbHEL40 | OC43 S                 | wtTM/H1cCT      | SEQ ID NO: 86                           | SEQ ID NO: 99                           | SEQ ID NO: 141                          | SEQ ID NO: 146    | SEQ ID NO: 111   |
| 9310        | nbHEL40 | 229E S                 | wtTMCT          | SEQ ID NO: 86                           | SEQ ID NO: 147                          | SEQ ID NO: 148                          | SEQ ID NO: 153    | SEQ ID NO: 111   |
| 9311        | nbHEL40 | 229E S                 | H5iTMCT         | SEQ ID NO: 86                           | SEQ ID NO: 100                          | SEQ ID NO: 149                          | SEQ ID NO: 154    | SEQ ID NO: 111   |
| 9312        | nbHEL40 | 229E S                 | wtTM/H5ICT      | SEQ ID NO: 86                           | SEQ ID NO: 100                          | SEQ ID NO: 150                          | SEQ ID NO: 155    | SEQ ID NO: 111   |
| 9313        | nbHEL40 | 229E S                 | wtTM/H5ICT (V4) | SEQ ID NO: 86                           | SEQ ID NO: 100                          | SEQ ID NO: 151                          | SEQ ID NO: 156    | SEQ ID NO: 111   |
| 9314        | nbHEL40 | 229E S                 | wtTM/H1cCT      | SEQ ID NO: 86                           | SEQ ID NO: 99                           | SEQ ID NO: 152                          | SEQ ID NO: 157    | SEQ ID NO: 111   |

## Example 2: Methods

### *Agrobacterium tumefaciens* Transfection

[00332] *Agrobacterium tumefaciens* strain AGL1 was transfected by electroporation with the SARS-CoV-2 modified S protein expression vectors using the methods described by D'Aoust *et al.*, 2008 (*Plant Biotech. J.* 6:930-40). Transfected *Agrobacterium* were grown in YEB medium supplemented with 10 mM 2-(N-morpholino)ethanesulfonic acid (MES), 20  $\mu$ M acetosyringone, 50  $\mu$ g/ml kanamycin and 25  $\mu$ g/ml of carbenicillin pH5.6 to an OD<sub>600</sub> between 0.6 and 1.6. *Agrobacterium* suspensions were centrifuged before use and resuspended in infiltration medium (10 mM MgCl<sub>2</sub> and 10 mM MES pH 5.6).

### Preparation of Plant Biomass, Inoculum and Agroinfiltration

[00333] *N. benthamiana* plants were grown from seeds in flats filled with a commercial peat moss substrate. The plants were allowed to grow in the greenhouse under a 16/8 photoperiod and a temperature regime of 25°C day/20°C night. Three weeks after seeding, individual plantlets were picked out, transplanted in pots and left to grow in the greenhouse for three additional weeks under the same environmental conditions.

[00334] *Agrobacteria* transfected with each expression vector were grown in a YEB medium supplemented with 10 mM 2-(N-morpholino)ethanesulfonic acid (MES), 20  $\mu$ M acetosyringone, 50  $\mu$ g/ml kanamycin and 25  $\mu$ g/ml of carbenicillin pH 5.6 until they reached an OD<sub>600</sub> between 0.6 and 1.6. *Agrobacterium* suspensions were centrifuged before use and resuspended in infiltration medium (10 mM MgCl<sub>2</sub> and 10 mM MES pH 5.6) and stored overnight at 4° C. On the day of infiltration, culture batches were diluted in 2.5 culture volumes and allowed to warm before use. Whole plants of *N. benthamiana* were placed upside down in the bacterial suspension in an air-tight stainless steel tank under a vacuum of 20-40 Torr for 2-min. Plants were returned to the greenhouse for a 6 or 9 day incubation period until harvest.

Leaf Harvest and Total Protein and VLP Extraction

[00335] Following incubation, the aerial part of plants was harvested, frozen at  $-80^{\circ}\text{C}$  and crushed into pieces. Total soluble proteins were extracted by mechanically homogenizing (Polytron) each sample of frozen-crushed plant material in two volumes of cold 50 mM Tris buffer at pH 8.0 + 500 mM NaCl, 0.4  $\mu\text{g/ml}$  metabisulfite and 1 mM phenylmethanesulfonyl fluoride. After homogenization, the slurries were centrifuged at 10,000 g for 10 min at  $4^{\circ}\text{C}$  and these clarified crude extracts (supernatant) kept for analysis.

[00336] The total protein content of clarified crude extracts was determined by the Bradford assay (Bio-Rad, Hercules, California) using bovine serum albumin as the reference standard. Proteins were separated by SDS-PAGE under reducing conditions using Criterion™ TGX Stain-Free™ precast gels (Bio-Rad Laboratories, Hercules, CA). Proteins were visualized by staining the gels with Coomassie Brilliant Blue. Alternatively, proteins were visualized with Gel Doc™ EZ imaging system (Bio-Rad Laboratories, Hercules, CA) and electrotransferred onto polyvinylidene difluoride (PVDF) membranes (Roche Diagnostics Corporation, Indianapolis, Indiana) for immunodetection. Prior to immunoblotting, the membranes were blocked with 5% skim milk and 0.1% Tween-20 in Tris-buffered saline (TBS-T) for 16-18 h at  $4^{\circ}\text{C}$ .

[00337] For VLP purification, proteins were extracted from frozen biomass by mechanical extraction using a blender with two volumes of extraction buffer (50 mM Tris buffer at pH 7.0 + 500 mM NaCl) and pH was lowered to 6.1 using 0.5M citric acid. The slurry was filtered through a large pore nylon filter to remove large debris and centrifuged 5000 g for 5 min at  $4^{\circ}\text{C}$ . The supernatant was collected and centrifuged again at 5000 g for 30 min ( $4^{\circ}\text{C}$ ) to remove additional debris and passed through clarification filters. The supernatant was then loaded on a discontinuous iodixanol density gradient. Analytical density gradient centrifugation was performed as follows: 38 mL tubes containing discontinuous iodixanol density gradient in Tris buffer (3 ml at 35%, 3 ml at 30%, 3 ml at 25%, 3 ml at 15% and 5 ml at 10% of iodixanol) were prepared and overlaid with 22 ml of the extracts containing the virus-like particles. The gradients were centrifuged at 120 000 g for 2 hours ( $4^{\circ}\text{C}$ ). After centrifugation, 1 mL fractions were

collected from the bottom to the top and fractions were analyzed by SDS-PAGE combined with protein staining or Western blot. Fractions 6 to 9 were pooled and buffer-exchanged using Amicon centrifugation device. Protein content is determined by Bradford assay.

5        Protein Analysis and Immunoblotting

[00338]        Immunoblotting was performed with a first incubation with a primary mAb, (anti-S1, Sino Biological, cat#40150-R007 or anti-S2, Novus biological, cat#NB100-56578) diluted in 2% skim milk in TBS-Tween 20 0.1%. Peroxydase-conjugated goat anti-rabbit (Jackson Immunoresearch, cat #115-035-144) was used as  
10        secondary antibody for chemiluminescence detection in 2% skim milk in TBS-Tween 20 0.1% Immunoreactive complexes were detected by chemiluminescence using luminol as the substrate (Roche Diagnostics Corporation). Horseradish peroxidase-enzyme conjugation of human IgG antibody was carried out by using the EZ-Link Plus® Activated Peroxidase conjugation kit (Pierce, Rockford, Ill.).

15        [00339]        *In planta* yields were assessed on clarified crude extracts and analyzed using a capillary-based electrophoresis method (Protein Simple, BioTechne) technology and a WES analysis system. In brief, soluble proteins from crude extracts were separated by molecular weight in a capillary and fixed to the matrix. A standard curve using purified VLPs is used to determine S protein quantity and Anti-S2 antibody (Novus  
20        biological, cat#NB100-56578) is used for detection according to the manufacturer instructions. Yields are then normalized using a comparator construct which is set to 1.

[00340]        The primary antibody for detection of SARS-CoV S protein was SARS-CoV Spike S1 subunit antibody from Sino Biologicals, 40150-MM08 (1/5000) and the secondary antibody used for detection was Goat anti-Mouse, JIR, 115-035-146  
25        (1/10000). The primary antibody for detection of MERS CoV S protein was MERS-CoV spike protein S1 antibody (N-terminal) from Sino Biological, (100208-RP02, 1/5000). The secondary antibody used for detection was Goat anti-Mouse from JIR (115-035-144, 1/10 000). The primary antibody used for detection was anti-coronavirus OC43 spike

protein from Antibodies-online (ABIN2754654, 1/1000. The secondary antibody used for detection was Goat anti-Rabbit from JIR (111-035-144, 1/10 000).

### Electron Microscopy

[00341] To determine whether expressed S protein assembled into VLPs, transmission electron microscopy (TEM) of immuno-trapped particles was performed on purified VLPs. Glow discharged carbon/copper grids (10s, 0.3 mbar) were placed on 20  $\mu$ L of purified VLPs (100  $\mu$ g/mL) for 5 min and then washed 4 times with sterile distilled water. The grids were floated on 20  $\mu$ L of 2 % uranyl acetate for 1 min, excess solution is then removed by touching a moist filter paper and allowed to dry for 24h on a filter paper before viewing under a TEM (Tecnai Microscope).

### Example 3: Sequences

[00342] The following sequences were use in the examples described above.

Native SARS-CoV-2 S protein wtTM/CT AA (P0DTC2) (SEQ ID NO:1)

MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWF  
HAIHVSNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSTQSLIIVNNATNVVIKV  
CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMDLEGKQGNFKNLREFVFK  
NIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLTFGDSSSGWTA  
GAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSTFVEKGIYQTSNFRVQPTES  
IVRFPNITNLCPFGEVFNATRFASVYAWNKRKISNCVADYSVLNSASFSTFKCYGVSPTKLNLDL  
CFNTVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPPDFTGCVIAWNSNNLDSKVGGNYNYLYRL  
FRKSNLKPFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHA  
PATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEIL  
DITPCSFSGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTRVYSTGNSVFQTRAGCL  
IGAETHVNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNIAIP  
TNFTISVTTEILPVSMTKTSVDCTMYICGDSTECNLLLQYGSFCTQLNRALTGIAVEQDKNTQE  
VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIA  
ARDLICAQKFNGLTVLPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIG  
VTQNVLYENQKLIANQFNSAIGKIQDSLSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAISS  
VLNDILSRDLKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRV  
DFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNTTAPAICHGDKAHFPREGVVFVSNGTHWFVT  
QRNFYEPQIITDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDIS  
GINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLC  
CMTSCCCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

Native SARS-CoV-2 S protein wtTM/CT AA (P0DTC2) without signal peptide (SP) (SEQ ID NO:2)

VNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSNGTKRFD  
NPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSTQSLIIVNNATNVVIKVCEFQFCNDPFLGVYY

5 HKNKNSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPI  
 NLVRDLPPQGSALPLVDLPIGINITRFQTLALHRSYLTGPDSSSGWTAGAAAYVGYLQPRTF  
 LLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGE  
 VFNATRFASVYAWNRRKRISNCVADYSVLVNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGD  
 10 EVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDIST  
 EIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHAPATVCGPKKSTNLVK  
 NKCVMFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLLEILDITPCSFGGVSVITP  
 GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAHEVNNSECDIP  
 IGAGICASYQTQTNSPRRARSVASQSI IAYTMSLGAENSVAYSNNIAIPTNFTISVTTEILPVS  
 15 MTKTSVDCTMYICGDSTECNLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIK  
 DFGGFNFSQILPDPSPKSKRSFIEDLLFNKVTLDADAGFIKQYGDCLGDIAARDLICAQKFNGLTV  
 LPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIAN  
 QFNSAIGKIQDSLSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRDKVEAE  
 VQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLSFPQS  
 20 APHGVVFLHVTYVPAQEKNFTTAPAI CHDGAHFPREGVFVSNGTHWFTVQRNFYEPQIITTDNT  
 FVSGNCDVVI GIVNNTVYDPLQPELDSFKEELDKY FKNHTSPDVLGDISGINASVUNI QKEIDR  
 LNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMTIMLCCMTSCCSCLKGCCSC  
 GSCCKFDEDDSEFVLKGVKLHYT

#### H5 A/Indonesia/5/05 Hemagglutinin (HA) AA (A5A5L7) (SEQ ID NO:3)

20 MEKIVLLLLAIVSLVKSQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKP  
 LILRDCSVAGWLLGNPMCDDEFINPEWSYIVEKANPTNDLCYPGSFNDYEELKHLLSRINHFEKI  
 QIIPKSSWSDEASSGVSSACPYLGSPSFFRNVVWLIKKNSTYPTIKKSYNNTNQEDLLVLWGIH  
 HPNDAAEQTRLYQNPTTYISIGTSTLNQRLVPKIATR SKVNGQSGRMEFFWTILKPNDAINFESN  
 25 GNFIAPYAYKIVKKGDSAIMKSELEYGNCNTKCQTPMGAINSSMPFHNHPLTIGECPKYVKS  
 RLVLATGLRNSPQRESRRKKRGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQSGSYAADKESTQKA  
 IDGVTNKVNSIIDKMNTQFEAVGREFNLERRIENLNKMKMEDGFLDVWTYNAELLVLMENERTLD  
 FHDSNVKNLYDKVRLQLRDNAKELGNCCFEFYHKCDNECMESIRNGTYNYPQYSEEARLKREEIS  
 GVKLESIGTYQILSIYSTVASSLALAIMMAGLSLWMCNGLQCRICI

#### H5 A/Indonesia/5/05 Hemagglutinin (HA) viral cDNA (EF541394.1) (SEQ ID NO:4)

30 CTGTAAAAAATGGAGAAAATAGTGCTTCTTCTTGCAATAGTCAGTCTTGTTAAAAAGTGATCAGATT  
 TGCATTGGTTACCATGCAAACAATTCAACAGAGCAGGTTGACACAATCATGGAAAAGAACGTTAC  
 TGTTACACATGCCCAAGACATACTGGAAAAGACACACAACGGGAAGCTCTGCGATCTAGATGGAG  
 TGAAGCCTCTAATTTTAAGAGATTGTAGTGTAGCTGGATGGCTCCTCGGGAACCCAATGTGTGAC  
 35 GAATTCATCAATGTACCGGAATGGTCTTACATAGTGGAGAAGGCCAATCCAACCAATGACCTCTG  
 TTACCCAGGGAGTTTCAACGACTATGAAGAACTGAAACACCTATTGAGCAGAATAAACCATTTTG  
 AGAAAATTCAAATCATCCCCAAAAGTTCTTGGTCCGATCATGAAGCCTCATCAGGAGTGAGCTCA  
 GCATGTCCATACCTGGGAAGTCCCTCCTTTTTTAGAAATGTGGTATGGCTTATCAAAAAGAAGAG  
 TACATACCCAACAATAAAGAAAAGCTACAATAATACCAACCAAGAAGATCTTTTGGTACTGTGGG  
 GAATTCACCATCCTAATGATGCGGCAGAGCAGACAAGGCTATATCAAAACCCAACCACCTATATT  
 40 TCCATTGGGACATCAACACTAAACCAGAGATTGGTACCAAAAATAGCTACTAGATCCAAAGTAAA  
 CGGGCAAAGTGGAAGGATGGAGTTCTTCTGGACAATTTTAAAACCTAATGATGCAATCAACTTCG  
 AGAGTAATGGAATTTTCATTGCTCCAGAATATGCATACAAAATTGTCAAGAAAGGGGACTCAGCA  
 ATTATGAAAAGTGAATTGGAATATGGTAAGTCAACACCAAGTGTCAAACCTCAATGGGGGCGAT  
 45 AAACCTAGTATGCCATTCCACAACATACACCTCTCACCATCGGGGAATGCCCCAAATATGTGA  
 AATCAAACAGATTAGTCCTTGCAACAGGGCTCAGAAATAGCCCTCAAAGAGAGAGCAGAAGAAAA  
 AAGAGAGGACTATTTGGAGCTATAGCAGGTTTATAGAGGGAGGATGGCAGGGAATGGTAGATGG  
 TTGGTATGGGTACCACCATAGCAATGAGCAGGGGAGTGGGTACGCTGCAGACAAAGAATCCACTC  
 AAAAGGCAATAGATGGAGTCACCAATAAGGTCAACTCAATCATTGACAAAATGAACACTCAGTTT

5 GAGGCCGTTGGAAGGGAATTTAATAACTTAGAAAGGAGAATAGAGAATTTAAACAAGAAGATGGA  
 AGACGGGTTTCTAGATGTCTGGACTTATAATGCCGAACCTCTGGTTCTCATGGAAAATGAGAGAA  
 CTCTAGACTTTTCATGACTCAAATGTTAAGAACCTCTACGACAAGGTCCGACTACAGCTTAGGGAT  
 AATGCAAAGGAGCTGGGTAACGGTTGTTTCGAGTTCATCACAAATGTGATAATGAATGTATGGA  
 AAGTATAAGAAACGGAACGTACAACATATCCGCAGTATTCAGAAGAAGCAAGATTAAAAAGAGAGG  
 AAATAAGTGGGGTAAATTTGGAATCAATAGGAACCTACCAAATACTGTCAATTTATTCAACAGTG  
 GCGAGTTCCTAGCACTGGCAATCATGATGGCTGGTCTATCTTTATGGATGTGCTCCAATGGATC  
 GTTACAATGCAGAATTTGCATTTAAATTTGTGAGTTCAG

**Modified SARS-CoV-2 with H5 A/Indonesia/5/05 Hemagglutinin CT AA (SEQ ID NO:5)**

10 MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWF  
 HAIHVSNGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWI FGTTLDSTQSLLI VNNATNVVIKV  
 CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLLMDLEGKQGNFKNLREFVFK  
 NIDGYFKIYSKHTPINLVRDL PQGFSALEPLVDLP IGINITRFQTLALHRSYLT PGDSSSGWTA  
 15 GAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTEVKGIYQTSNFRVQPTES  
 IVRFPNITNLCPFGEVFENATRFASVYAWNRRKISNCVADYSVLYNSASFSTFKCYGVSPTKLNDL  
 CFTNVYADSFVIRGDEV RQIAPGQTGKIADYNYKLPDDFTGCVI AWNSNNLDSKVGNNYNYLYRL  
 FRKSNLKPFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVL SFELLHA  
 PATVCGPKKSTNLVKNKCVNFENGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEIL  
 DITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTRVYSTGNSNVFQTRAGCL  
 20 IGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSI IAYTMSLGAENSVAYSNNSIAIP  
 TNFTISVTTEILPVSMTKTSVDCTMYICGDSTEC SNLLLQYGSFCTQLNRALTGI AVEQDKNTQE  
 VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVT LADAGFIKQYGDCLGDIA  
 ARDLICAQKFNGLT VLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAA LQIPFAMQMAYRFNGIG  
 VTQNVLYENQKLIANQFNSAIGKIQDSLSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAISS  
 25 VLNDILSRLDKVEAEVQIDRLITGR LQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRV  
 DFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICH DGKAHFPREGV FVSNGTHWFVT  
 QRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKY FKNHTSPDVDLGDIS  
 GINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLS  
 LWMCSNGSLQCRICI

30 **H1 A/California/7/2009 Hemagglutinin TM/CT AA (SEQ ID NO:6)**

IDGVKLESTRIYQILAIYSTVASSLVLVSLGAISFWMCSNGSLQCRICI

**H2 A/Singapore/1/1957 Hemagglutinin TM/CT AA (SEQ ID NO:7)**

IKGVKLSSMGVYQILAIYATVAGSLSLAIMMAGISFWMCSNGSLQCRICI

**H3 A/Minnesota/41/2019 Hemagglutinin TM/CT AA (SEQ ID NO:8)**

35 IKGVELKSGYKD WILWISFAISCFLLCVALLGFIMWACQKGNIRCNICI

**H5 A/Indonesia/5/05 Hemagglutinin TM/CT AA (SEQ ID NO:9)**

ISGVKLESIGTYQILSIYSTVASSLALAIMMAGLSLWMCSNGSLQCRICI

**H6 A/Teal/Hong Kong/W312/97 Hemagglutinin TM/CT AA (SEQ ID NO:10)**

IESVKLENLGVYQILAIYSTVSSSLVLVGLIMAMGLWMCSNGSMQCRICI

H7 A/Guangdong/17SF003/2016 Hemagglutinin TM/CT AA (SEQ ID NO:11)

IDPVKLSSGYKDVILWFSFGASCFILLAIIVMGLVFICVKNGNMRCTICI

H9 A/Hong Kong/1073/99 Hemagglutinin TM/CT AA (SEQ ID NO:12)

IEGVKLESEGTYKILTIYSTVASSLVLAMGFAAFLFWAMSNGSCRCNICI

5 B/Washington/02/2019 Hemagglutinin TM/CT AA (SEQ ID NO:13)

AASLNDDGLDNHTILLYYSTAASSLAVTLMIAIFVVYMVSRDNVSCSICL

Consensus Sequence of C-Terminal Region of Influenza Hemagglutinin (SEQ ID NO:14)

IXGVKLXSXGXIXILXIYSTVASSLXLXXXXXXXXXXWMCSNGSXXCXIC I

Consensus Sequence of CT Domain of Influenza Hemagglutinin (SEQ ID NO:15)

10 XXWMCSNGSXXCXIC I

C-Terminal Region of Native SARS-CoV-2 S protein wtTM/CT (SEQ ID NO:16)

WYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCCLKGCCSCGSCCKFDEDDSEPVLKGVK  
LHYT

C-Terminal Region of H5 A/Indonesia/5/05 Hemagglutinin (SEQ ID NO:17)

15 ISGVKLESIGTYQILSIYSTVASSLALAIMMAGLSLWMCSNGSLQCRIC I

C-Terminal Region of Modified SARS-CoV-2 S protein with H5i Hemagglutinin CT (SEQ ID  
NO:18)

WYIWLGFIAGLIAIVMVTIMLSLWMCSNGSLQCRIC I

20 C-Terminal Region of Modified SARS-CoV-2 S protein with H5i Hemagglutinin CT, Variation 1  
(SEQ ID NO:19)

WYIWLGFIAGLIAIVMVTIMMAGLSLWMCSNGSLQCRIC I

(no SP) SARS-CoV-2 S protein GSAS+PP wtTM/CT AA (SEQ ID NO:20)

25 VNLTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFD  
NPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSTQSLIVNNATNVVIKVCEFCNDPFLGVYY  
HKNNKSWMESEFRVYSSANNCTFEYVSQPFLLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPI  
NLVRDLPGGSALEPLVDLPIGINITRFQTLALHRSYLT PGDSSSGWTAGAAAYVGYLQPRTE  
LLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGE  
VFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNLDLCFTNVYADSFVIRGD

EVRQIAPGQTGKIADYNYKL PDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDIST  
 EIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHAPATVCGPKKSTNLVK  
 NKC VNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTL EILDITPCSFGGVSVITP  
 GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGA EHVNNSYEC DIP  
 5 IGAGICASYQTQTNSPGSASSVASQSIIAYTMSLGAENSVAYSNN SIAIPTNFTISVTTEILPVS  
 MTKTSVDCTMYICGDSTEC SNLLQLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIK  
 DFGGFNFSQILPDPSKPSKRSFIEDLLFNKVT LADAGFIKQYGDCLGDIAARDL ICAQKFNGLTV  
 LPPLLTDEMI AQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIAN  
 10 QFNSAIGKIQDSL SSTASALGKLQDVVNQNAQALNTLVKQLSSNFGA ISSVLNDILSR LDPPEAE  
 VQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQS  
 APHGVVFLHVTVVPAQEKNFTTAPAICHGDKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT  
 FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD VDLGDISGINASVVNIQKEIDR  
 LNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMTIMLCCMTSCC SCLKGCCSC  
 GSCCKFDEDDSEPV LKGVKLHYT

15 (no SP) Modified SARS-CoV-2 S protein GSAS+PP with H5i Hemagglutinin CT AA (SEQ ID NO:21)

VNLTRTQQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFD  
 NPVLPFNDGVYFASTEKSNIIRGWIFGTTLD SKTQSLIIVNNATNVVIKCFEQFCNDPFLGVYY  
 HKNNKSWMESEFRVYSSANNCTFEYVSQPF LMDLEGKQGNFKNLREFVFNIDGYFKIYSKHTPI  
 20 NLVRDLPPQGSAL EPLVDLPIGINITRFQTL LALHRSYLT PGDSSSGWTAGAAAAYVGYLQPRTF  
 LLKYNENGTTITDAVDCALDPLSE TKTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGE  
 VFNATRFASVYAWNKRKISNCVADYSVLNSASFSTFKCYGVSPTKLN DLCTNVYADSFVIRGD  
 EVRQIAPGQTGKIADYNYKL PDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDIST  
 25 EIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHAPATVCGPKKSTNLVK  
 NKC VNFNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTL EILDITPCSFGGVSVITP  
 GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGA EHVNNSYEC DIP  
 IGAGICASYQTQTNSPGSASSVASQSIIAYTMSLGAENSVAYSNN SIAIPTNFTISVTTEILPVS  
 MTKTSVDCTMYICGDSTEC SNLLQLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIK  
 30 DFGGFNFSQILPDPSKPSKRSFIEDLLFNKVT LADAGFIKQYGDCLGDIAARDL ICAQKFNGLTV  
 LPPLLTDEMI AQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIAN  
 QFNSAIGKIQDSL SSTASALGKLQDVVNQNAQALNTLVKQLSSNFGA ISSVLNDILSR LDPPEAE  
 VQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQS  
 APHGVVFLHVTVVPAQEKNFTTAPAICHGDKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT  
 35 FVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD VDLGDISGINASVVNIQKEIDR  
 LNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMTIMLSLWMC SNGSLQCRICI

PDI-SARS-CoV-2 S protein GSAS+PP wtTM/CT-DNA (SEQ ID NO:22)

ATGGCGAAAAACGTTGCGATTTTCGGCTTATTGTTTCTCTTCTTGTGTGGTTCCCTTCTCAGAT  
 CTTTCGCGTGAATCTTACGACGCGAACACAGTTACCACCCGCATATACAAATAGCTTCACTCGGG  
 GTGTTTATTACCCCGACAAAGTGTT CAGGTCCTCCGTGCTCCACTCAACACAGGACCTCTTTCTT  
 40 CCTTTCTTTTCTAACGTGACATGGTTTCATGCCATTCATGTATCCGGCACTAACGGTACTAAGAG  
 GTTCGATAATCCTGTGCTCCCTTTCAATGACGGCGTTTACTTTGCAAGCACAGAGAAGAGTAACA  
 TCATCCGAGGTTGGATCTTTGGCACTACCCTCGATTCAAAGACGCAGAGCCTCCTCATTGTGAAC  
 AATGCCACTAACGTGGTGATCAAAGTTTGCAGAGTTTCAGTTCTGCAATGACCCTTTCTTGGGGGT  
 GTACTATCATAAGAACAACAAGTCTTGGATGGAATCTGAATCCGCGTCTATAGCAGCGCCAACA  
 45 ACTGCACCTTTGAATACGTGTCCAGCCCTTCCTTATGGACCTGGAGGGAAAGCAGGGAAACTTT  
 AAGAACTCTGAGAGAGTTCGTGTTTAAAAATATCGACGGCTATTTTAAGATCTATTCTAAGCACAC  
 GCCTATTAATCTCGTGCGGATCTTCCACAAGGCTTCAGCGCCCTGGAACCACTCGTGGACCTCC  
 CAATTGGTATCAACATCACTAGATTT CAGACTCTGCTTGCCCTCCACCGATCCTATCTGACACCC

5 GGAGACTCCTCTAGCGGCTGGACTGCCGGCGCTGCCGCTTATTACGTTGGTTATCTTCAGCCACG  
CACGTTCTGCTGAAGTATAACGAGAATGGTACTATTACCGATGCCGTGGATTGTGCCCTTGACC  
CCCTGTCCGAAACTAAGTGCACACTCAAGTCATTCACTGTGGAAAAAGGAATCTACCAGACAAGC  
AATTTTCGGGTCCAGCCTACTGAGAGCATTGTGCGCTTTCCTAACATCACAAATCTTTGCCCTT  
CGGAGAGGTTTTCAATGCTACACGGTTTGCCTCCGTGTATGCCGTGGAACCGCAAGAGAATTTCCA  
ATTGCGTGGCCGATTACTCCGTGCTCTACAATAGTGCAAGCTTTAGCACCTTTAAGTGCTATGGC  
GTATCCCCTACTAAGCTTAACGACTTGTGTTTCACAAACGTGTATGCCGACTCCTTTGTGATACG  
GGGCGACGAAGTTAGACAGATAGCACCAGGACAGACGGGAAAGATAGCTGACTACAACTATAAGC  
10 TTCTTGATGACTTCACTGGCTGCGTTATCGCGTGGAATTTCTAACAACCTGGACTCAAAAGTCGGC  
GGCAACTATAACTATCTCTATCGGCTGTTCCGCAAGAGTAACCTTAAGCCCTTTGAGAGAGATAT  
AAGCACTGAAATCTACCAGGCTGGCAGTACGCCCTGTAATGGCGTGGAAGGCTTTAATTGTTATT  
TTCCACTGCAATCCTATGGTTTTTCAGCCAACCAATGGCGTGGGCTACCAACCATAACCGCTCGTG  
GTGCTCTCCTTTGAACTGCTCCACGCTCCCGCGACTGTCTGCGGCCCCAAGAAGTCCACGAACCT  
15 TGTGAAGAATAAGTGCCTTAATTTTAAATTTCAACGGCCTCACTGGAACAGGAGTGCTCACTGAGA  
GTAACAAGAAGTTCTTGCCATTTCAACAATTTGGCAGAGACATAGCCGATACTACTGACGCCGTT  
AGGGACCCCCAGACCCTCGAGATTCTCGATATAACGCCCTGCTCCTTCGGTGGAGTTTCCGTGAT  
CACGCCAGGCACCAATACCAGTAACCAGGTGCGCGTGCTGTATCAGGATGTCAACTGTACTGAGG  
TGCCCGTAGCCATCCATGCGGATCAGCTCACACCAACTTGGAGGGTGTAACGACCCGGCTCCAAT  
GTATTCAGACTCGGGCCGGATGCCTTATTGGCGCCGAACACGTGAACAATAGTTACGAATGCGA  
20 TATTCCAATTGGCGCCGGAATCTGTGCTAGCTACCAGACTCAGACGAACTCCCCAGGCAGCCCA  
GCAGCGTTGCCAGCCAGTCAATCATCGCTTATACAATGTCACTTGGAGCCGAAAACTCCGTGGCT  
TACTCAAACAACAGCATCGCCATCCCCACAACTTCAACCATATCCGTGACAACTGAGATTCTGCC  
AGTGTCCATGACTAAGACGTCCGTAGATTGCACTATGTACATATGCGGCGACAGCACAGAATGTT  
CTAATCTGCTGCTGCAATATGGAAGCTTCTGCACTCAACTGAACAGAGCGCTCACAGGCATCGCC  
25 GTGGAGCAGGATAAGAATACCCAGGAGGTGTTGCGCCAAGTTAAGCAGATCTACAAGACCCCCACC  
CATAAAGGATTTTCGGTGGATTCAATTTTAGTCAGATACTCCAGACCCATCTAAGCCATCCAAGA  
GGAGCTTTATCGAGGATCTTTTGTTTAACAAAGTTACTCTGGCCGACGCCGTTTCATCAAGCAG  
TACGGAGATTGCTCGGCGACATCGCTGCTCGTGACCTCATCTGTGCGCAAAAGTTTAAACGGTCT  
GACGGTGCTGCCTCCCCTCCTTACTGATGAAATGATCGCCAGTATACCAGCGCACTCCTCGCTG  
30 GCACCATAACATCCGGTTGGACATTGCGGCGCTGGTGCAGCACTGCAGATACCATTGCGCATGCAA  
ATGGCATATCGTTTCAACGGTATCGGTGTACACAGAATGTCTATATGAGAACCAGAAGCTGAT  
CGAAATCAGTTCAATAGTGCCATCGGAAAAATCCAGGATAGCCTTAGCAGCACAGCCTCAGCCC  
TTGGCAAACTCCAGGATGTGCTGAACCAGAATGCCAGGCTCTCAATACCCTCGTGAAGCAGCTC  
TCATCTAATTTTCGGCGCAATTTCCAGTGTCTCAACGACATCCTCAGCCGCTCGACCCCCCGA  
35 GGCCGAAGTGCAGATTGACAGACTGATTACAGGTGCACTCCAGAGCCTCCAGACTTACGTGACTC  
AGCAGCTGATAAGAGCCGCGGAGATAAGGGCCAGCGCTAACCTGGCTGCCACAAAGATGTCTGAG  
TGCGTGCTGGGCCAGTCCAAGAGAGTAGACTTCTGTGGCAAAGGCTACCATCTGATGAGCTTCCC  
ACAATCCGCACCTCACGGCGTAGTGTTCTCCACGTGACATATGTACCGCTCAGGAGAAGAATT  
TCACTACCGCTCCTGCTATATGCCATGATGGAAGGCTCACTTCCCCCGGGAGGGGGTGTTCGTG  
40 TCCAACGGCACCCATTGGTTTTGTGACTCAGCGGAATTTCTACGAACCCAGATCATAACCACTGA  
CAACACATTTGTGTCCGGAATTGTGACGTGGTCATTGGAATAGTGAACAACACTGTTTATGATC  
CACTGCAGCCAGAATTTGACAGCTTTAAGGAGGAGCTCGACAAGTACTTCAAGAATCATACGTCA  
CCAGATGTGGACCTCGGAGATATTAGCGGTATCAATGCCAGTGTGTCAATATTCAGAAGGAAAT  
AGACCGCTTAATGAGGTGCGCAAAAATCTGAACGAGAGCCTCATCGATCTTCAGGAGCTGGGCA  
45 AATATGAGCAGTACATCAAGTGGCCTTGGTATATTTGGCTTGGCTTCATCGCCGGCCTGATCGCC  
ATAGTAATGGTCACAATTATGCTCTGCTGCATGACCTCTTGCTGCTCCTGTCTGAAAGGCTGCTG  
CTCTTGCGGATCCTGCTGCAAATTTGATGAGGATGACAGTGAACCAGTCTTGAAGGGCGTGAAGC  
TGCACTATACTTAG

PDI-SARS-CoV-2 S protein GSAS+PP wtTM/CT-AA (SEQ ID NO:23)

5 MAKNVAIFGLLFSLLVLPVPSQIFAVNLTTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFL  
 PFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWI FGTTLDSTQSLIIVN  
 NATNVVIKVCFEQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLLMDLEGKQGNF  
 KNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLT  
 10 GDSSSGWTAGAAAYVGYLQPRFTLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTS  
 NFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKCYG  
 VSPTKLNLDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPPDFTGCVIAWNSNNLDSKVG  
 GNYNYLYRLFRKSNLKPFFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV  
 15 VLSFELLHAPATVCGPKKSTNLVKNKCVNFNENGLTGTGVLTESNKKFLPFQQFGRDIADTTDAV  
 RDPQTLLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSN  
 VFQTRAGCLIGAHEVNNSECDIPIGAGICASYQTQTNSPGSASSVASQSI IAYTMSLGAENSV  
 YSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECNNLLQYGSFCTQLNRALTGIA  
 VEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLDAGFIKQ  
 20 YGDCGLDIAARDLICAQKFNGLTVLPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQ  
 MAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSTASALGKLQDVVNQNAQALNTLVKQL  
 SSNFGALSSVLNDILSRLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSE  
 CVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHHDGKAHFPREGVFEV  
 SNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS  
 PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIA  
 IVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLGKVKLHYT

#### IF(PDI)-CoV(opt2).c (SEQ ID NO:24)

TCTCAGATCTTCGCGGTGAATCTTACGACGCGAACACAGTTACCACCCGCAT

#### 25 IF(AVB)-CoV(opt2).r (SEQ ID NO:25)

ACGACACGACTAAGGCCTCTAAGTATAGTGCAGCTTCACGCCCTTCAGGAC

#### PDI-Modified SARS-CoV-2 S protein GSAS+PP H5iTM/CT-DNA (SEQ ID NO:26)

30 ATGGCGAAAAACGTTGCGATTTTCGGCTTATTGTTTTCTCTTCTTGTGTTGGTTCCCTTCTCAGAT  
 CTTCGCGGTGAATCTTACGACGCGAACACAGTTACCACCCGCATATACAAATAGCTTCACTCGGG  
 GTGTTTATTACCCCGACAAAGTGTTTCAGGTCTCCGTGCTCCACTCAACACAGGACCTCTTTCTT  
 CCTTTCTTTTCTAACGTGACATGGTTTCATGCCATTATCATGTATCCGGGCTAAGGTAAGAG  
 GTTCGATAATCCTGTGCTCCCTTTCAATGACGGCGTTTACTTTGCAAGCACAGAGAAGAGTAACA  
 35 TCATCCGAGGTTGGATCTTTGGCACTACCTCGATTCAAAGACGCAGAGCCTCCTCATTGTGAAC  
 AATGCCACTAACGTGGTGATCAAAGTTTGCGAGTTTCAGTTCTGCAATGACCCCTTCTTGGGGGT  
 GTACTATCATAAGAACAACAAGTCTTGATGGAATCTGAATTCCGCGTCTATAGCAGCGCCAACA  
 ACTGCACCTTTGAATACGTGTCCAGCCCTTCTTATGGACCTGGAGGGAAAGCAGGGAAACTTT  
 AAGAATCTGAGAGAGTTTCGTGTTTAAAAATATCGACGGCTATTTTAAGATCTATTCTAAGCACAC  
 GCCTATTAATCTCGTGCGGATCTTCCACAAGGCTTCAGCGCCCTGGAACCACTCGTGACCTCC  
 40 CAATTGGTATCAACATCACTAGATTTTCAGACTCTGCTTGCCTCCACCGATCCTATCTGACACCC  
 GGAGACTCCTCTAGCGGCTGGACTGCCGGCGCTGCCGCTTATTACGTTGGTTATCTTCAGCCACG  
 CACGTTCCCTGCTGAAGTATAACGAGAATGGTACTATTACCGATGCCGTGGATTGTGCCCTTGACC  
 CCCTGTCCGAAACTAAGTGACACTCAAGTCATTCACTGTGGAAAAAGGAATCTACCAGACAAGC  
 AATTTTCGGGTCCAGCCTACTGAGAGCATTGTGCGCTTTCCTAACATCACAAATCTTTGCCCTT  
 45 CGGAGAGGTTTTCAATGCTACACGGTTTGCCCTCCGTGTATGCCTGGAACCGCAAGAGAATTTCCA

ATTGCGTGCCGATTACTCCGTGCTCTACAATAGTGCAAGCTTTAGCACCTTTAAGTGCTATGGC  
GTATCCCCTACTAAGCTTAACGACTTGTGTTTCACAAACGTGTATGCCGACTCCTTTGTGATACG  
GGGCGACGAAGTTAGACAGATAGCACCAGGACAGACGGGAAAGATAGCTGACTACAACATAAAGC  
5 TTCTTGATGACTTCACTGGCTGCGTTATCGCGTGGAATTCTAACAACCTGGACTCAAAAGTCGGC  
GGCAACTATAACTATCTCTATCGGCTGTTCCGCAAGAGTAACCTTAAGCCCTTTGAGAGAGATAT  
AAGCACTGAAAATCTACCAGGCTGGCAGTACGCCCTGTAATGGCGTGGAAGGCTTTAATTGTTATT  
TTCCACTGCAATCCTATGGTTTTTCAGCCAACCAATGGCGTGCGGTACCAACCATAACCGCGTCGTG  
GTGCTCTCCTTTGAACTGCTCCACGCTCCCGCGACTGTCTGCGGCCCAAGAAGTCCACGAACCT  
10 TGTGAAGAATAAGTGCGTTAATTTTAAATTTCAACGGCCTCACTGGAACAGGAGTGCTCACTGAGA  
GTAACAAGAAGTTCCTGCCATTTCAACAATTTGGCAGAGACATAGCCGATACTACTGACGCCGTT  
AGGGACCCCCAGACCCTCGAGATTCTCGATATAACGCCCTGCTCCTTCGGTGGAGTTTCCGTGAT  
CACGCCAGGCACCAATACCAGTAACCAGGTCGCCGTGCTGTATCAGGATGTCAACTGTACTGAGG  
TGCCCGTAGCCATCCATGCGGATCAGCTCACACCAACTTGGAGGGTGACAGCACCGGCTCCAAT  
GTATTCCAGACTCGGGCCGGATGCCTTATTGGCGCCGAACACGTGAACAATAGTTACGAATGCGA  
15 TATTCCAATTGGCGCCGGAATCTGTGCTAGCTACCAGACTCAGACGAACTCCCCAGGCAGCGCCA  
GCAGCGTTGCCAGCCAGTCAATCATCGCTTATACAATGTCACTTGGAGCCGAAAACTCCGTGGCT  
TACTCAAAACAACAGCATCGCCATCCCCACAACTTCACCATATCCGTGACAACTGAGATTCTGCC  
AGTGTCCATGACTAAGACGTCCGTAGATTGCACTATGTACATATGCGGCGACAGCACAGAATGTT  
CTAATCTGCTGCTGCAATATGGAAGCTTCTGCACTCACTGAACAGAGCGCTCACAGGCATCGCC  
20 GTGGAGCAGGATAAGAATACCCAGGAGGTGTTGCCCAAGTTAAGCAGATCTACAAGACCCCCACC  
CATAAAGGATTTCCGGTGGATTCAATTTTAGTCAGATACTCCAGACCCATCTAAGCCATCCAAGA  
GGAGCTTTATCGAGGATCTTTTGTTTAACAAGTTACTCTGGCCGACGCCGGTTTCATCAAGCAG  
TACGGAGATTGCCTCGGCGACATCGCTGCTCGTGACCTCATCTGTGCGCAAAAGTTTAACGGTCT  
GACGGTGCTGCCTCCCCTCCTTACTGATGAAATGATCGCCAGTATACCAGCGCACTCCTCGCTG  
25 GCACCATAACATCCGGTTGGACATTCGGCGCTGGTGCAGCACTGCAGATACCATTGCCATGCAA  
ATGGCATATCGTTTCAACGGTATCGGTGTACACAGAATGTCTTATATGAGAACCAGAAGCTGAT  
CGCAAATCAGTTCAATAGTGCCATCGGAAAAATCCAGGATAGCCTTAGCAGCACAGCCTCAGCCC  
TTGGCAAACCTCCAGGATGTCGTGAACCAGAATGCCAGGCTCTCAATACCCTCGTGAAGCAGCTC  
TCATCTAATTTCCGGCGCAATTTCCAGTGTCTCAACGACATCCTCAGCCGCTCGACCCCCCGA  
30 GGCCGAAGTGACAGATTGACAGACTGATTACAGGTGACTCCAGAGCCTCCAGACTTACGTGACTC  
AGCAGCTGATAAGAGCCGCCGAGATAAGGGCCAGCGCTAACCTGGCTGCCACAAAGATGTCTGAG  
TGCGTGCTGGGCCAGTCCAAGAGAGTAGACTTCTGTGGCAAAGGCTACCATCTGATGAGCTTCCC  
ACAATCCGCACCTCACGGCGTAGTGTTCTCCACGTGACATATGTACCGCTCAGGAGAAGAATT  
TCACTACCGCTCCTGCTATATGCCATGATGGAAGGCTCACTTCCCCCGGGAGGGGGTGTTCGTG  
35 TCCAACGGCACCCATTGGTTTGTGACTCAGCGGAATTTCTACGAACCCAGATCATAACCACTGA  
CAACACATTTGTGTCCGAAATTGTGACGTGGTCATTGGAATAGTGAACAACACTGTTTATGATC  
CACTGCAGCCAGAACTTGACAGCTTTAAGGAGGAGCTCGACAAGTACTTCAAGAATCATACGTCA  
CCAGATGTGGACCTCGGAGATATTAGCGGTATCAATGCCAGTGTGTCAATATTGAGAAGGAAAT  
AGACCGCCTTAATGAGGTGCGCAAAAATCTGAACGAGAGCCTCATCGATCTTCAGGAGCTGGGCA  
40 AATATGAGCAGTACATCAAGTGGCCTTGGTATCAAATACTGTCAATTTATTCAACAGTGGCGAGT  
TCCCTAGCACTGGCAATCATGATGGCTGGTCTATCTTTATGGATGTGCTCCAATGGATCGTTACA  
ATGCAGAATTTGCATTTAA

PDI-Modified SARS-CoV-2 S protein GSAS+PP H5iTM/CT-AA (SEQ ID NO:27)

MAKNVAIFGLLFSLLVLVPSQIFAVNLTTTRTQLPPAYTNSFTRGVVYPDKVFRSSVLHSTQDLFL  
45 PFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWI FGTTLD SKTQSLLI VN  
NATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPF LMDLEGKQGNF  
KNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTL LALHRSYLT P  
GDSSSGWTAGAAAYVGYLQPRFTLLKYNENGTITDAVDCALDPLSETKCT LKSFTVEKGIYQTS  
50 NFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNKRISNCVADYSVLVNSASFSTFKCYG

5 VSPTKLNLDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVG  
 GNINYLYRLFRKSNLKPFFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV  
 VLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAV  
 RDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSN  
 VFQTRAGCLIGAHEVNNSEYCDIPIGAGICASYQTQTNSPGSASSVASQSI IAYTMSLGAENSV  
 YSNNSIAIPTNFTISVTTEILPVSMKTSDVCTMYICGDSTECNNLLQYGSFCTQLNRALTGIA  
 VEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLDAGFIKQ  
 YGDCGLDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSWTFGAGAAALQIPFAMQ  
 10 MAYRFNGIGVTVQNVLYENQKLIANQFNSAIGKIQDSLSTASALGKLQDVVNQNAQALNTLVKQL  
 SSNFGAISSVLNDILSRDPPAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSE  
 CVLGQSKRVDFCGKGYHLMSPQSAPHGVVFLHVTVVPAQEKNFTTAPAICHGDKAHFPREGV  
 SNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS  
 PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYQILSIYSTVAS  
 15 SLALAIMMAGLSLWMCSSNGSLQCRICI

# IF(Avb)-H5I.r (SEQ ID NO:28)

ACGACACGACTAAGGCCTTTAAATGCAAATTCTGCATTGTAACGATCC

# PDI-Modified SARS-CoV-2 S protein GSAS+PP wtTM/H5iCT-DNA (SEQ ID NO:29)

20 ATGGCGAAAAACGTTGCGATTTCGCGCTTATTGTTTTCTCTTCTTGTGTTGGTTCCTTCTCAGAT  
 CTTTCGCGGTGAATCTTACGACGCGAACACAGTTACCACCCGCATATACAAATAGCTTCACTCGGG  
 GTGTTTTATTACCCCGACAAAGTGTTTCAGGTCCTCCGTGCTCCACTCAACACAGGACCTCTTTCTT  
 CCTTTCTTTTCTAACGTGACATGGTTTTCATGCCATTTCATGTATCCGGCACTAACGGTACTAAGAG  
 GTTCGATAATCCTGTGCTCCCTTTCAATGACGGCGTTTACTTTGCAAGCACAGAGAAGAGTAACA  
 25 TCATCCGAGGTTGGATCTTTGGCACTACCCTCGATTCAAAGACGCAGAGCCTCCTCATTGTGAAC  
 AATGCCACTAACGTGGTGATCAAAGTTTGCGAGTTTCAGTTCTGCAATGACCCTTTCTTGGGGGT  
 GTACTATCATAAGAACAACAAGTCTTGATGGAATCTGAATTCCGCGCTCTATAGCAGCGCCAACA  
 ACTGCACCTTTGAATACGTGTCCCAGCCCTTCCTTATGGACCTGGAGGGAAAAGCAGGGAAACTTT  
 AAGAACTCTGAGAGAGTTTCGTGTTTAAAAATATCGACGGCTATTTTAAGATCTATTCTAAGCACAC  
 30 GCCTATTAATCTCGTGCGGATCTTCCACAAGGCTTCAGCGCCCTGGAACCACTCGTGGACCTCC  
 CAATTGGTATCAACATCACTAGATTTAGACTCTGCTTGCCCTCCACCGATCCTATCTGACACCC  
 GGAGACTCCTCTAGCGGCTGGACTGCCGCGCTGCCGCTTATTACGTTGGTTATCTTCAGCCACG  
 CACGTTCTGCTGAAGTATAACGAGAATGGTACTATTACCGATGCCGTGGATTGTGCCCTTGACC  
 CCCTGTCCGAAACTAAGTGACACTCAAGTCATTCACTGTGGAAAAAGGAATCTACCAGACAAGC  
 35 AATTTTCGGGTCCAGCCTACTGAGAGCATTGTGCGCTTTCCTAACATCACAAATCTTTGCCCTT  
 CGGAGAGGTTTTCAATGCTACACGGTTTGCCCTCCGTGTATGCCTGGAACCGCAAGAGAATTTCCA  
 ATTGCGTGGCCGATTACTCCGTGCTCTACAATAGTGCAAGCTTTAGCACCTTTAAGTGCTATGGC  
 GTATCCCTACTAAGCTTAACGACTTGTGTTTACAAACGTGTATGCCGACTCCTTTGTGATACG  
 GGGCGACGAAGTTAGACAGATAGCACCAGGACAGACGGGAAAGATAGCTGACTACAACTATAAGC  
 40 TTCTTGATGACTTCACTGGCTGCGTTATCGCGTGGAATTCTAACACCTGGACTCAAAAGTCGGC  
 GGCAACTATAACTATCTCTATCGGCTGTTCCGCAAGAGTAACCTTAAGCCCTTTGAGAGAGATAT  
 AAGCACTGAAATCTACCAGGCTGGCAGTACGCCCTGTAATGGCGTGGAAGGCTTTAATTGTTATT  
 TTCCACTGCAATCCTATGGTTTTTCAGCCAACCAATGGCGTGGGCTACCAACCATACCGCGTCGTG  
 GTGCTCTCCTTTGAAGTGTCCACGCTCCCGCGACTGTCTGCGGCCCAAGAAGTCCACGAACCT  
 45 TGTGAAGAATAAGTGCGTTAATTTTAAATTTCAACGGCCTCACTGGAACAGGAGTGCTCACTGAGA  
 GTAACAAGAAGTTCCTGCCATTTCAACAATTTGGCAGAGACATAGCCGATACTACTGACGCCGTT  
 AGGGACCCCCAGACCCTCGAGATTCTCGATATAACGCCCTGCTCCTTCGGTGGAGTTTCCGTGAT

CACGCCAGGCACCAATACCAGTAACCAGGTGCGCGTGCTGTATCAGGATGTCAACTGTACTGAGG  
 TGCCCGTAGCCATCCATGCGGATCAGCTCACACCAACTTGGAGGGTGTACAGCACCGGCTCCAAT  
 GTATTCCAGACTCGGGCCGGATGCCTTATTGGCGCCGAACACGTGAACAATAGTTACGAATGCGA  
 TATTCCAATTGGCGCCGGAATCTGTGCTAGCTACCAGACTCAGACGAACTCCCCAGGCAGCGCCA  
 5 GCAGCGTTGCCAGCCAGTCAATCATCGCTTATACAATGTCACTTGGAGCCGAAAACTCCGTGGCT  
 TACTCAAACAACAGCATCGCCATCCCCACAACTTCACCATATCCGTGACAACTGAGATTCTGCC  
 AGTGTCCATGACTAAGACGTCCGTAGATTGCACTATGTACATATGCGGCGACAGCACAGAATGTT  
 CTAATCTGCTGCTGCAATATGGAAGCTTCTGCACTCAACTGAACAGAGCGCTCACAGGCATCGCC  
 GTGGAGCAGGATAAGAATACCCAGGAGGTGTTTCGCCCAAGTTAAGCAGATCTACAAGACCCACC  
 10 CATAAAGGATTTTCGGTGGATTCAATTTTAGTCAGATACTCCAGACCCATCTAAGCCATCCAAGA  
 GGAGCTTTATCGAGGATCTTTTGTTTAACAAAGTTACTCTGGCCGACGCCGGTTTCATCAAGCAG  
 TACGGAGATTGCCTCGGCGACATCGCTGCTCGTGACCTCATCTGTGCGCAAAAGTTTAACGGTCT  
 GACGGTGTGCTGCCCTCCCTCCTTACTGATGAAATGATCGCCAGTATACCAGCGCACTCCTCGCTG  
 GCACCATAACATCCGGTTGGACATTCGGCGCTGGTGCAGCACTGCAGATACCATTGCGCATGCAA  
 15 ATGGCATAATCGTTTCAACGGTATCGGTGTACACAGAATGTCTTATATGAGAACCAGAAGCTGAT  
 CGCAAATCAGTTCAATAGTGCCATCGGAAAAATCCAGGATAGCCTTAGCAGCACAGCCTCAGCCC  
 TTGGCAAACCTCCAGGATGTCGTGAACCAGAATGCCAGGCTCTCAATACCCTCGTGAAGCAGCTC  
 TCATCTAATTTTCGGCGCAATTTCCAGTGTCTCAACGACATCCTCAGCCGCTCGACCCCCCGA  
 GGCCGAAGTGCAGATTGACAGACTGATTACAGGTGCACTCCAGAGCCTCCAGACTTACGTGACTC  
 20 AGCAGCTGATAAGAGCCGCCGAGATAAGGGCCAGCGCTAACCTGGCTGCCACAAAGATGTCTGAG  
 TGGCTGCTGGGCCAGTCCAAGAGAGTAGACTTCTGTGGCAAAGGCTACCATCTGATGAGCTTCCC  
 ACAATCCGCACCTCACGGCGTAGTGTTCTCCACGTGACATATGTACCGGCTCAGGAGAAGAATT  
 TCACTACCGCTCCTGCTATATGCCATGATGGAAGGCTCACTTCCCCCGGGAGGGGGTGTTCGTG  
 TCCAACGGCACCCATTGGTTTGTGACTCAGCGGAATTTCTACGAACCCAGATCATAACCACTGA  
 25 CAACACATTTGTGTCCGGAATTTGTGACGTGGTCATTGGAATAGTGAACAACACTGTTTATGATC  
 CACTGCAGCCAGAACTTGACAGCTTTAAGGAGGAGCTCGACAAGTACTTCAAGAATCATACGTCA  
 CCAGATGTGGACCTCGGAGATATTAGCGGTATCAATGCCAGTGTGTCAATATTCAGAAGGAAAT  
 AGACCGCCTTAATGAGGTGCGCAAAAATCTGAACGAGAGCCTCATCGATCTTCAGGAGCTGGGCA  
 AATATGAGCAGTACATCAAGTGGCCTTGGTATATTTGGCTTGGCTTCATCGCCGGCCTGATCGCC  
 30 ATAGTAATGGTCACAATTATGCTCTCTTTATGGATGTGCTCCAATGGATCGTTACAATGCAGAAT  
 TTGCATTTAA

#### PDI-Modified SARS-CoV-2 S protein GSAS+PP wtTM/H5iCT-AA (SEQ ID NO:30)

MAKNVAIFGLLFSLLVLVPSQIFAVNLTTTRTQLPPAYTNSFTRGVVYPDKVFRSSVLHSTQDLFL  
 PFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIIRGWI FGTTLDSTQSLLIIVN  
 35 NATNVVIVKCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMLDLEGKQGNF  
 KNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLT  
 GDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTS  
 NFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNKRISNCVADYSVLYNSASFSTFKCYG  
 VSPTKLNLDLCTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPPDFTGCVIAWNSNNLDSKVG  
 40 GNYNYLYRLFRKSNLKPFFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV  
 VLSFELLHAPATVCGPKKSTNLVKNKCVNFENGLTGTGVLTESNKKFLPFQQFGRDIADTTDAV  
 RDPQTLLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSN  
 VFQTRAGCLIGAHEVNNSYECDIPIGAGICASYQTQTNPSGSASSVASQSI IAYTMSLGAENSV  
 45 YSNNISIAIPTNFTISVTTEILPVSMKTSDVCTMYICGDSTECNLLLQYGSFCTQLNRALTGIA  
 VEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLDAGFIKQ  
 YGDCLGDIAARDLICAQKFNGLTVLPLLTDEMIAQYTSALLAGTITSGWTFGAGAAALQIPFAMQ  
 MAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSTASALGKLQDVVNQNAQALNTLVKQL  
 SSNFGAISSVLNDILSRLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSE  
 50 CVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDKGAHFPREGV  
 SNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVIVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS

PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIA  
IVMVTIMLSLWMCSNGLQCRICI

Cloning vector 8501 from left to right T-DNA (SEQ ID NO:31)

5 TGGCAGGATATATTGTGGTGTAAACAAATTGACGCTTAGACAACCTTAATAACACATTGCGGACGT  
TTTTAATGTACTGAATTAACGCCGAATCCCGGGCTGGTATATTTATATGTTGTCAAATAACTCAA  
AAACCATAAAAAGTTTAAAGTTAGCAAGTGTGTACATTTTTACTTGAACAAAAATATTCACCTACTA  
CTGTTATAAATCATTATTAAACATTAGAGTAAAGAAATATGGATGATAAGAACAAAGAGTAGTGAT  
10 ATTTTGACAACAATTTTGTGCAACATTTGAGAAAATTTTGTGTTCTCTCTTTTCATTGGTCAA  
AAACAATAGAGAGAGAAAAAGGAAGAGGGGAGAATAAAAAACATAATGTGAGTATGAGAGAGAAAGT  
TGTACAAAAGTTGTACCAAAATAGTTGTACAAATATCATTGAGGAATTTGACAAAAGCTACACAA  
ATAAGGGTTAATTGCTGTAAATAAATAAGGATGACGCATTAGAGAGATGTACCATTAGAGAATTT  
TTGGCAAGTCATTAAAAAGAAAGAATAAATTATTTTTAAATTTAAAGTTGAGTCATTTGATTAA  
ACATGTGATTATTTAATGAATTGATGAAAGAGTTGGATTAAAGTTGTATTAGTAATTAGAATTTG  
15 GTGTCAAATTTAATTTGACATTTGATCTTTTCTATATATTGCCCCATAGAGTCAGTTAACTCAT  
TTTTATATTTCATAGATCAAATAAGAGAAATAACGGTATATTAATCCCTCCAAAAAATAACG  
GTATATTTACTAAAAAATCTAAGCCACGTAGGAGGATAACAGGATCCCGTAGGAGGATAACATC  
CAATCCAACCAATCACAACAATCCTGATGAGATAACCCACTTTAAGCCCACGCATCTGTGGCACA  
TCTACATTATCTAAATCACACATTCTTCCACACATCTGAGCCACACAAAAACCAATCCACATCTT  
20 TATCACCCATTTCTATAAAAAATCACACTTTGTGAGTCTACACTTTGATTCCCTTCAAACACATAC  
AAAGAGAAGAGACTAATTAATTAATTAATCATCTTGAGAGAAAATGGAACGAGCTATACAAGGAA  
ACGACGCTAGGGAACAAGCTAACAGTGAACGTTGGGATGGAGGATCAGGAGGTACCACTTCTCCC  
TTCAAACCTCTGACGAAAGTCCGAGTTGGACTGAGTGGCGGCTACATAACGATGAGACGAATTC  
GAATCAAGATAAATCCCTTTGGTTTCAAGGAAAGCTGGGGTTTCGGGAAAGTTGTATTTAAGAGAT  
25 ATCTCAGATACGACAGGACGGAAGCTTCACTGCACAGAGTCCTTGGATCTTGGACGGGAGATTGCG  
GTTAACTATGCAGCATCTCGATTTTTCGGTTTCGACCAGATCGGATGTACCTATAGTATTCGGTT  
TCGAGGAGTTAGTATCACCGTTTCTGGAGGGTCGCGAACTCTTCAGCATCTCTGTGAGATGGCAA  
TTCGGTCTAAGCAAGAACTGCTACAGCTTGCCCCAATCGAAGTGGAAGTAATGTATCAAGAGGA  
TGCCCTGAAGTACTCAAACCTTCGAAAAAGAAAGCGAGTAAGTTAAATGCTTCTTCGTCTCCT  
30 ATTTATAATATGGTTTGTATTGTTAATTTTGTCTTGTAGAAAGAGCTTAATTAATCGTTGTTGT  
TATGAAATACTATTTGTATGAGATGAACTGGTGAATGTAATTCATTTACATAAGTGGAGTCAGA  
ATCAGAATGTTTCTCCATAACTAAGTACATGAAGACCTGCCGCGTACAATTTGTCTTATATTT  
GAACAACTAAAAATTGAACATCTTTTGCCACAACCTTTATAAGTGGTTAATATAGCTCAAATATATG  
GTCAAGTTCAATAGATTAATAATGGAATATCAGTTATCGAAATTCATTAACAATCAACTTAACG  
35 TTATTAAGTACTAATTTTATATCATCCCCTTTGATAAATGATAGTACACCAATTAGGAAGGAGCA  
TGCTCGCCTAGGAGATTGTCTGTTTCCCGCCTTCAGTTTGCAAGCTGCTCTAGCCGTGTAGCCAAT  
ACGCAAACCGCCTCTCCCCGCGCGTTGGGAATTACTAGCGCGTGTGCGACAAGCTTGCATGCCGGT  
CAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAAGACC  
AAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTGCCCCA  
40 GCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTG  
CGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCAC  
CCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGT  
GATAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAAGA  
CCAAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTGCC  
45 CAGCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCAT  
TGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCC  
ACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGAT  
GTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCTTCTCTCT  
ATATAAGGAAGTTCATTTTATTGGAGAGGCACACAATTTGCTTTAGTGATTAACTTTCTTTTA

5 CAACAAATTAAAGGTCTATTATCTCCCAACAACATAAGAAAACAATGGCGAAAAACGTTGCGATT  
TTCGGCTTATTGTTTTCTCTTCTTGTGTTGGTTCCCTTCTCAGATCTTCGCGACGTCACCTCCTCAG  
CCTAACGACACCCCCATCTGTCTATCCACTGGCCCCGATCTGCTGCCAACTAACTCCATG  
GTGACCCCTGGGATGCCTGGTCAAGGGCTATTTCCCTGAGCCAGTGACAGTGACCTGGAACCTCTGG  
10 ATCCCTGTCCAGCGGTGTGCACACCTTCCCAGCTGTCCTGCAGTCTGACCTCTACACTCTGAGCA  
GCTCAGTGACTGTCCCTCCAGCACCTGGCCCAGCGAGACCGTCACCTGCAACGTTGCCCACCCG  
GCCAGCAGCACCAAGGTGGACAAGAAAATTGTGCCAGGGATTGTGGTTGTAAGCCTTGCCATATG  
TACAGTCCCAGAAGTATCATCTGTCTTCATCTTCCCCCAAAGCCCAAGGATGTGCTCACCATTA  
15 CTCTGACTCCTAAGGTCACGTGTGTTGTGGTAGACATCAGCAAGGATGATCCCCGAGGTCCAGTTC  
AGCTGGTTTGTAGATGATGTGGAGGTGCACACAGCTCAGACGCAACCCGGGAGGAGCAGTTCAA  
CAGCACTTTCCGCTCAGTCAGTGAACCTCCCATCATGCACCAGGACTGGCTCAATGGCAAGGAGA  
CGTCCAGATTTTGGCGATCTATTCAACTGTGCCAGTTCATTGGTACTGGTAGTCTCCCTGGGGG  
CAATCAGTTTCTGGATGTGCTCTAATGGGTCTCTACAGTGTAGAATATGTATTTAAAGGCCTTAG  
20 TCGTGTCTGTTTTCAAATAATATAATCCTTTTAGGGTTTTAGTTAGTTAAATTTTCTGTTGCTC  
CTGTTTAGCAGGTCTGTGCCTTCAGCAAGCACACAAAAACAGAGTGTATTTAAGTTGTTTGT  
TAGTGATTCAAAAAAAAAAATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTT  
GCCGGTCTTGCGATGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGTAATAATTAACAT  
GTAATGCATGACGTTATTTATGAGATGGGTTTTATGATTAGAGTCCCGCAATTATACATTTAAT  
25 ACGCGATAGAAAACAAAATATAGCGCGCAAACCTAGGATAAATTATCGCGCGCGGTGTCTATGAT  
TTACTAGATCTCTAGAGTCTCAAGCTTGGCGCGCCACGTGACTAGTGGCACTGGCCGTCTGTTTT  
ACAACGTCGTGACTGGGAAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTT  
TCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTG  
AATGGCGAATGCTAGAGCAGCTTGAGCTTGGATCAGATTGTCGTTTCCCGCCTTCAGTTTAAACT  
30 ATCAGTGTTTGACAGGATATATTGGCGGGTAAACCTAAGAGAAAAGAGCGTTTA

Construct 8586 from 2X35S prom to NOS term (SEQ ID NO:32)

30 GTCAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAAGA  
CCAAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTGCC  
CAGCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCAT  
TGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAGATGGACCCCC  
ACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGAT  
35 GTGATAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAA  
GACCAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTG  
CCCAGCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATC  
ATTGCGATAAAGGAAAGGCCATCGTTGAAGATGCCCTCTGCCGACAGTGGTCCCAAGATGGACCC  
40 CCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTG  
ATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCTTCCT  
CTATATAAGGAAGTTCATTTTCAATTTGGAGAGGCACACAATTTGCTTTAGTGATTAAACTTTCTTT  
TACAACAAATTAAAGGTCTATTATCTCCCAACAACATAAGAAAACAATGGCGAAAAACGTTGCGA  
45 TTTTCGGCTTATTGTTTTCTCTTCTTGTGTTGGTTCCCTTCTCAGATCTTCGCGGTGAATCTTACG  
ACGCGAACACAGTTACCACCCGCATATACAAATAGCTTCACTCGGGGTGTTTATTACCCCGACAA  
AGTGTTCAAGTTCCTCCGTGCTCCACTCAACACAGGACCTCTTCTTCCCTTTCTTTTCTAACGTGA  
CATGGTTTCATGCCATTTCATGTATCCGGCACTAACGGTACTAAGAGGTTTCGATAATCCTGTGCTC  
CCTTTCAATGACGGCGTTTACTTTGCAAGCACAGAGAAGAGTAACATCATCCGAGGTTGGATCTT  
TGGCACTACCTCGATTCAAAGACGCAGAGCCTCCTCATTGTGAACAATGCCACTAACGTGGTGA  
TCAAAGTTTGGGAGTTTCAGTTCTGCAATGACCCCTTCTTGGGGGTGTACTATCATAAGAACAAC  
AAGTCTTGATGGAATCTGAATTCGCGCTCTATAGCAGCGCCAACAACTGCACCTTTGAATACGT  
GTCCAGCCCTTCTTATGGACCTGGAGGGAAAGCAGGGAACTTTAAGAATCTGAGAGAGTTTCG  
TGTTTAAAAATATCGACGGCTATTTTAAGATCTATTCTAAGCACACGCCTATTAATCTCGTGCGC

5 GATCTTCCACAAGGCTTCAGCGCCCTGGAACCACTCGTGGACCTCCCAATTGGTATCAACATCAC  
TAGATTTTCAGACTCTGCTTGCCCTCCACCGATCCTATCTGACACCCGGAGACTCCTCTAGCGGCT  
GGACTGCCGGCGCTGCCGCTTATTACGTTGGTTATCTTCAGCCACGCAGTTCCTGCTGAAGTAT  
AACGAGAATGGTACTATTACCGATGCCGTGGATTGTGCCCTTGACCCCTGTCCGAAACTAAGTG  
10 CACACTCAAGTCATTCACTGTGGA AAAAGGAATCTACCAGACAAGCAATTTTCGGGTCCAGCCTA  
CTGAGAGCATTTGTGCGCTTTCCTAACATCACAATCTTTGCCCTTCGGAGAGGTTTTCAATGCT  
ACACGGTTTGCCTCCGTGTATGCCTGGAACCGCAAGAGAATTTCCAATTGCGTGGCCGATTACTC  
CGTGCTCTACAATAGTGCAAGCTTTAGCACCTTTAAGTGCTATGGCGTATCCCTACTAAGCTTA  
ACGACTTGTGTTTCACAAACGTGTATGCCGACTCCTTTGTGATACGGGGCGACGAAGTTAGACAG  
15 ATAGCACCCAGGACAGACGGGAAAGATAGCTGACTACAACATAAGCTTCCCTGATGACTTCACTGG  
CTGCGTTATCGCGTGGAAATTCTAACAACCTGGACTCAAAAGTCGGCGGCAACTATAACTATCTCT  
ATCGGCTGTTCGCAAGAGTAACCTTAAGCCCTTTGAGAGAGATATAAGCACTGAAATCTACCAG  
GCTGGCAGTACGCCCTGTAATGGCGTGGAAAGGCTTTAATTGTTATTTTCCACTGCAATCCTATGG  
TTTTTCAGCCAACCAATGGCGTGGGCTACCAACCATAACCGCTCGTGGTGTCTCCTTTGAACTGC  
20 TCCACGCTCCCGCGACTGTCTGCGGCCCAAGAAGTCCACGAACCTTGTGAAGAATAAGTGCGTT  
AATTTTAATTTCAACGGCCTCACTGGAACAGGAGTGCTCACTGAGAGTAACAAGAAGTTTCTGCC  
ATTTCAACAATTTGGCAGAGACATAGCCGATACTACTGACGCCGTTAGGGACCCCCAGACCCTCG  
AGATTCTCGATATAACGCCCTGCTCCTTCGGTGGAGTTTCCGTGATCACGCCAGGCACCAATACC  
AGTAACCAGGTCGCCGTGCTGTATCAGGATGTCAACTGTACTGAGGTGCCCGTAGCCATCCATGC  
25 GGATCAGCTCACACCAACTTGGAGGGTGTACAGCACCGGCTCCAATGTATTTCCAGACTCGGGCCG  
GATGCCCTTATTGGCGCCGAACACGTGAACAATAGTTACGAATGCGATATTTCCAATTGGCGCCGGA  
ATCTGTGCTAGCTACCAGACTCAGACGAACCTCCCAGGCAGCGCCAGCAGCGTTGCCAGCCAGTC  
AATCATCGCTTATACAATGTCACTTGGAGCCGAAAACCTCCGTGGCTTACTCAAACAACAGCATCG  
CCATCCCCACAACTTCACCATATCCGTGACAACCTGAGATTCTGCCAGTGTCCATGACTAAGACG  
30 TCCGTAGATTGCACTATGTACATATGCGGCGACAGCACAGAATGTTCTAATCTGCTGCTGCAATA  
TGGAAGCTTCTGCACTCAACTGAACAGAGCGCTCACAGGCATCGCCGTGGAGCAGGATAAGAATA  
CCCAGGAGGTGTTTCGCCCAAGTTAAGCAGATCTACAAGACCCCAACCATAAAGGATTTTCGGTGG  
TTCAATTTTAGTCAGATACTCCCAGACCCATCTAAGCCATCCAAGAGGAGCTTTATCGAGGATCT  
TTTGTTTAACAAAGTTACTCTGGCCGACGCCGTTTCATCAAGCAGTACGGAGATTGCCTCGGCG  
35 ACATCGCTGCTCGTGACCTCATCTGTGCGCAAAAGTTTAACGGTCTGACGGTGTCTGCCCTCCCTC  
CTTACTGATGAAATGATCGCCAGTATACCAGCGCACTCCTCGCTGGCACCATAACATCCGGTTG  
GACATTTCGGCGCTGGTGCAGCACTGCAGATACCATTTCGCCATGCAAAATCGTATCGTTTCAACG  
GTATCGGTGTGCACACAGAATGTCTATATGAGAACCAGAAGCTGATCGCAAAATCAGTTCAATAGT  
GCCATCGGAAAAATCCAGGATAGCCTTAGCAGCACAGCCTCAGCCCTTGGCAAACTCCAGGATGT  
40 CGTGAACCAGAATGCCAGGCTCTCAATACCCTCGTGAAGCAGCTCTCATCTAATTTTCGGCGCAA  
TTTCCAGTGTCTCAACGACATCCTCAGCCGCTCGACCCCCCGAGGCCGAAGTGCAGATTGAC  
AGACTGATTACAGGTCGACTCCAGAGCCTCCAGACTTACGTGACTCAGCAGCTGATAAGAGCCGC  
CGAGATAAGGGCCAGCGCTAACCTGGCTGCCACAAAGATGTCTGAGTGCCTGCTGGGCCAGTCCA  
AGAGAGTAGACTTCTGTGGCAAAGGCTACCATCTGATGAGCTTCCCACAATCCGCACCTCACGGC  
45 GTAGTGTTTCTCCACGTGACATATGTACCGGCTCAGGAGAAGAATTTCACTACCGCTCCTGCTAT  
ATGCCATGATGGAAAGGCTCACTTCCCCGGGAGGGGGTGTTCGTGTCCAACGGCACCCATTGGT  
TTGTGACTCAGCGGAATTTCTACGAACCCAGATCATAACCACTGACAACACATTTGTGTCCGGA  
AATTGTGAGGTGGTCATTGGAATAGTGAACAACACTGTTTTATGATCCACTGCAGCCAGAACTTGA  
CAGCTTTAAGGAGGAGCTCGACAAGTACTTCAAGAATCATACGTCACCAGATGTGGACCTCGGAG  
50 ATATTAGCGGTATCAATGCCAGTGTGTCAATATTCAGAAGGAAATAGACCGCCTTAATGAGGTC  
GCCAAAAATCTGAACGAGAGCCTCATCGATCTTCAGGAGCTGGGCAAAATATGAGCAGTACATCAA  
GTGGCCTTGGTATATTTGGCTTGGCTTCATCGCCGGCCTGATCGCCATAGTAATGGTCACAATTA  
TGCTCTGCTGCATGACCTCTTGCTGCTCCTGTCTGAAAGGCTGCTGCTCTTGCGGATCCTGCTGC  
AAATTTGATGAGGATGACAGTGAACCAGTCTGAAGGGCGTGAAGCTGCACTATACTTAGAGGCC  
TTAGTCGTGCTGTTTTTCAAATAATATAATCCTTTTAGGGTTTTAGTTAGTTTAAATTTTCTGTT  
GCTCCTGTTTAGCAGGTCGTGCCTTCAGCAAGCACAAAAACAGAGTGTTTATTTTAAAGTTGTT  
TGTTTAGTGATTCAAAAAAAAATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATCC

TGTTGCCGGTCTTGCGATGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGTAATAATTA  
ACATGTAATGCATGACGTTATTTATGAGATGGGTTTTTATGATTAGAGTCCCGCAATTATACATT  
TAATACGCGATAGAAAACAAAATATAGCGCGCAAACCTAGGATAAAATTATCGCGCGCGGTGTCATC  
TATGTTACTAGAT

5 Cloning vector 8500 from left to right T-DNA (SEQ ID NO:33)

10 TGGCAGGATATATTGTGGTGTAAACAAATTGACGCTTAGACAACCTTAATAACACATTGCGGACGT  
TTTAAATGTACTGAATTAACGCCGAATCCCGGGCTGGTATATTTATATGTTGTCAAATAACTCAA  
AAACCATAAAAAGTTTAAAGTTAGCAAGTGTGTACATTTTTACTTGAACAAAAATATTCACCTACTA  
CTGTTATAAATCATTATTAAACATTAGAGTAAAGAAATATGGATGATAAAGAACAGAGTAGTGAT  
15 ATTTTGACAACAATTTTGTGCAACATTTGAGAAAATTTGTTGTTCTCTCTTTTCATTGGTCAA  
AAACAATAGAGAGAGAAAAAGGAAGAGGGAGAATAAAAAACATAATGTGAGTATGAGAGAGAAAGT  
TGACAAAAAGTTGTACCAAAATAGTTGTACAAATATCATTGAGGAATTTGACAAAAGCTACACAA  
ATAAGGGTTAATTGCTGTAAATAAATAAGGATGACGCATTAGAGAGATGTACCATTAGAGAATTT  
TTGGCAAGTCATTAAAAAGAAAGAATAAATTATTTTAAATTTAAAGTTGAGTCATTTGATTAA  
15 ACATGTGATTATTTAATGAATTGATGAAAGAGTTGGATTAAAGTTGTATTAGTAATTAGAATTTG  
GTGTCAAATTTAATTTGACATTTGATCTTTTCTATATATTGCCCCATAGAGTCAGTTAACTCAT  
TTTTATATTTCATAGATCAAATAAGAGAAATAACGGTATATTAATCCCTCCAAAAAAGAACG  
GTATATTTACTAAAAAATCTAAGCCACGTAGGAGGATAACAGGATCCCGTAGGAGGATAACATC  
CAATCCAACCAATCACAACAATCCTGATGAGATAACCCACTTTAAGCCCACGCATCTGTGGCACA  
20 TCTACATTATCTAAATCACACATTCTTCCACACATCTGAGCCACACAAAAACCAATCCACATCTT  
TATCACCCATCTCTATAAAAAATCACACTTTGTGAGTCTACACTTTGATTCCCTTCAAACACATAC  
AAAGAGAAGAGACTAATTAATTAATTAATCATCTTGAGAGAAAATGGAACGAGCTATACAAGGAA  
ACGACGCTAGGGAACAAGCTAACAGTGAACGTGGGATGGAGGATCAGGAGGTACCCTTCTCCC  
TTCAAACCTTCTGACGAAAGTCCGAGTTGGACTGAGTGGCGGCTACATAACGATGAGACGAATTC  
25 GAATCAAGATAATCCCTTGGTTTCAAGGAAAGCTGGGGTTTCGGGAAAGTTGTATTTAAGAGAT  
ATCTCAGATACGACAGGACGGAAGCTTCACTGCACAGAGTCCTTGGATCTTGGACGGGAGATTG  
GTTAACTATGCAGCATCTCGATTTTTCGGTTTCGACCAGATCGGATGTACCTATAGTATTCGGTT  
TCGAGGAGTTAGTATCACCGTTTCTGGAGGGTTCGCGAACTCTTCAGCATCTCTGTGAGATGGCAA  
TTCGGTCTAAGCAAGAAGTGTACAGCTTGCCCCAATCGAAGTGGAAGTAATGTATCAAGAGGA  
30 TGCCCTGAAGGTACTCAAACCTTCGAAAAAGAAAGCGAGTAAGTTAAATGCTTCTTTCGTCTCCT  
ATTTATAATATGGTTTGTATTGTTAATTTTGTCTTGTAGAAAGAGCTTAATTAATCGTTGTTGT  
TATGAAATACTATTTGTATGAGATGAAGTGGTGAATGTAATTCATTTACATAAGTGGAGTCAGA  
ATCAGAATGTTTCTCCATAACTAAGTACATGAAGACCTGCCGCTACAATTTGCTTATATTT  
GAACAACTAAAATTGAACATCTTTTGCCACAACCTTTATAAGTGGTTAATATAGCTCAAATATATG  
35 GTCAAGTTCAATAGATTAATAATGGAAATATCAGTTATCGAAATTCATTAACAATCAACTTAACG  
TTATTAATACTAATTTTATATCATCCCTTTGATAAATGATAGTACACCAATTAGGAAGGAGCA  
TGCTCGCCTAGGAGATTGTCGTTTCCCGCCTTCAGTTTGCAAGCTGCTCTAGCCGTGTAGCCAAT  
ACGCAAACCGCCTCTCCCCGCGCGTTGGGAATTACTAGCGCGTGTGACAAAGCTTGCATGCCGGT  
CAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAAGACC  
40 AAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAACCTCCTCGGATTCCATTGCCCA  
GCTATCTGTCACCTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTG  
CGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAGATGGACCCCCAC  
CCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGCTTCAAAGCAAGTGGATTGATGT  
GATAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAAGA  
45 CCAAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAACCTCCTCGGATTCCATTGCC  
CAGCTATCTGTCACCTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCAT  
TGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAGATGGACCCCC  
ACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGCTTCAAAGCAAGTGGATTGAT  
GTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCTCT

5 ATATAAGGAAGTTCATTTTCATTTGGAGAGGCACTTTTCTAATCAATCATCAAACAGAACGCAGAA  
AATTTCTAAAAACAAAAAAGGCATACAAATGGCGAAAAACGTTGCGATTTTCGGCTTATTGT  
TTTCTCTTCTGTGTTGGTTCCCTTCTCAGATCTTCGCGACGTCACCTCCTCAGCCAAAACGACACC  
CCCATCTGTCTATCCACTGGCCCCCTGGATCTGCTGCCCAAACCTAATCCATGGTGACCCTGGGAT  
10 GCCTGGTCAAGGGCTATTTCCCTGAGCCAGTGACAGTGACCTGGAACCTCTGGATCCCTGTCCAGC  
GGTGTGCACACCTTCCCAGCTGTCTGACGTCTGACCTCTACACTCTGAGCAGCTCAGTGACTGT  
CCCCCTCCAGCACCTGGCCCAGCGAGACCGTCACCTGCAACGTTGCCACCCGGCCAGCAGCACCA  
AGGTGGACAAGAAAATTGTGCCCAGGGATTGTGGTTGTAAGCCTTGCAATGTACAGTCCCAGAA  
15 GTATCATCTGTCTTCATCTTCCCCCAAGCCCAAGGATGTGCTCACCATTACTCTGACTCCCTAA  
GGTCACGTGTGTTGTGGTAGACATCAGCAAGGATGATCCCGAGGTCCAGTTCAGCTGGTTTGTAG  
ATGATGTGGAGGTGCACACAGCTCAGACGCAACCCCGGGAGGAGCAGTTCAACAGCACTTTCCGC  
TCAGTCAGTGAACTTCCCATCATGCACCAGGACTGGCTCAATGGCAAGGAGACGTCCAGATTTTG  
GCGATCTATTCAACTGTCGCCAGTTCATTGGTACTGGTAGTCTCCCTGGGGGCAATCAGTTTCTG  
20 GATGTGCTCTAATGGGTCTCTACAGTGTAGAATATGTATTTAAAGGCCTTAGTCGTGTCGTTTTT  
CAAATAATATAATCCTTTTAGGGTTTTAGTTAGTTTAAATTTTCTGTTGCTCCTGTTTAGCAGGT  
CGTGCTTCAGCAAGCACACAAAACAGAGTGTTTTATTTAAGTTGTTTGTGTTAGTGATTCAAAA  
AAAAATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTTGCCGGTCTTGCGA  
TGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGTAATAATTAACATGTAATGCATGACG  
25 TTATTTATGAGATGGGTTTTATGATTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAAA  
CAAAATATAGCGCGCAAACTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCTCT  
AGAGTCTCAAGCTTGGCGCGCCACGTGACTAGTGGCACTGGCCCTCGTTTTACAACGTCGTGAC  
TGGGAAAACCTGGCGTTACCCAACTTAATCGCCTTGACGACATCCCCCTTTCCGCCAGCTGGCG  
TAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGCT  
30 AGAGCAGCTTGAGCTTGATCAGATTGTCGTTTCCCGCCTTCAGTTTAAACTATCAGTGTTTGAC  
AGGATATATTGGCGGGTAAACCTAAGAGAAAAGAGCGTTTA

Construct 8589 from 2X35S prom to NOS term (SEQ ID NO:34)

30 GTCAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAAGA  
CCAAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTGCC  
CAGCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCAT  
TGCGATAAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCC  
ACCCACGAGGAGCATCGTGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGAT  
35 GTGATAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAA  
GACCAAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTG  
CCCAGCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATC  
ATTGCGATAAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCC  
CCACCCACGAGGAGCATCGTGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTG  
40 ATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCTTCTCT  
CTATATAAGGAAGTTCATTTTCATTTGGAGAGGCACTTTTCTAATCAATCATCAAACAGAACGCAG  
AAAATTTCTAAAAACAAAAAAGGCATACAAATGGCGAAAAACGTTGCGATTTTCGGCTTATT  
GTTTTCTCTTCTGTGTTGGTTCCCTTCTCAGATCTTCGCGGTGAATCTTACGACGCGAACACAGT  
TACCACCCGCATATACAAATAGCTTCACTCGGGGTGTTTTATTACCCCGACAAAGTGTTTCAGGTCC  
45 TCCGTGCTCCACTCAACACAGGACCTCTTTCTTCTTTCTTTTCTAACGTGACATGGTTTTCATGC  
CATTCATGTATCCGGCACTAACGGTACTAAGAGGTTTCGATAATCCTGTGCTCCCTTTCAATGACG  
GCGTTTACTTTGCAAGCACAGAGAAGAGTAACATCATCCGAGGTTGGATCTTTGGCACTACCCCTC  
GATTCAAAGACGCAGAGCCTCCTCATTGTGAACAATGCCACTAACGTGGTGATCAAAGTTTGCAG  
GTTTCAGTTCTGCAATGACCCTTTCTTGGGGGTGTACTATCATAAGAACAACAGTCTTGGATGG  
AATCTGAATTCCGCGTCTATAGCAGCGCCAACAACGACCTTTGAATACGTGTCCAGCCCTTC  
CTTATGGACCTGGAGGGAAAGCAGGGAACTTTAAGAATCTGAGAGAGTTCGTGTTAAAAATAT

CGACGGCTATTTTAAGATCTATTCTAAGCACACGCCTATTAATCTCGTGCGGATCTTCCACAAG  
GCTTCAGCGCCCTGGAACCACTCGTGGACCTCCCAATTGGTATCAACATCACTAGATTTTCAGACT  
CTGCTTGCCCTCCACCGATCCTATCTGACACCCGGAGACTCCTCTAGCGGCTGGACTGCCGGCGC  
TGCCGCTTATTACGTTGGTTATCTTCAGCCACGCACGTTCCCTGCTGAAGTATAACGAGAATGGTA  
5 CTATTACCGATGCCGTGGATTGTGCCCTTGACCCCCCTGTCCGAAACTAAGTGCACACTCAAGTCA  
TTCCTGTGGAAAAAGGAATCTACCAGACAAGCAATTTTCGGGTCCAGCCTACTGAGAGCATTGT  
GCGCTTTCCTAACATCACAAATCTTTGCCCCCTTCGGAGAGGTTTTCAATGCTACACGGTTTGCCT  
CCGTGTATGCCTGGAACCGCAAGAGAATTTCCAATTGCGTGGCCGATTACTCCGTGCTCTACAAT  
AGTGCAAGCTTTAGCACCTTTAAGTGCTATGGCGTATCCCCTACTAAGCTTAACGACTTGTGTTT  
10 CACAAACGTGTATGCCGACTCCTTTGTGATACGGGGCGACGAAGTTAGACAGATAGCACCAGGAC  
AGACGGGAAAGATAGCTGACTACAACATAAGCTTCCCTGATGACTTCACTGGCTGCGTTATCGCG  
TGGAATTCTAACAACCTGGACTCAAAAGTCGGCGGCAACTATAACTATCTCTATCGGCTGTTCCG  
CAAGAGTAACCTTAAGCCCTTTGAGAGAGATATAAGCACTGAAATCTACCAGGCTGGCAGTACGC  
CCTGTAATGGCGTGGAAGGCTTTAATTGTTATTTTCCACTGCAATCCTATGGTTTTTCAGCCAACC  
15 AATGGCGTGGGCTACCAACCATAACCGCGTCTGGTGTCTCCTTTGAACTGCTCCACGCTCCCGC  
GACTGTCTGCGGCCCCAAGAAGTCCACGAACCTTGTGAAGAATAAGTGCCTTAATTTTAATTTCA  
ACGGCCTCACTGGAACAGGAGTGCTCACTGAGAGTAACAAGAAGTTCCTGCCATTTCAACAATTT  
GGCAGAGACATAGCCGATACTACTGACGCCGTTAGGGACCCCCAGACCCTCGAGATTCTCGATAT  
AACGCCCTGCTCCTTCGGTGGAGTTTCCGTGATCACGCCAGGCACCAATACCAGTAACCAGGTGCG  
20 CCGTGCTGTATCAGGATGTCAACTGTACTGAGGTGCCGCTAGCCATCCATGCGGATCAGCTCACA  
CCAACTTGGAGGGTGTACAGCACCGGCTCCAATGTATTCCAGACTCGGGCCGGATGCCTTATTGG  
CGCCGAACACGTGAACAATAGTTACGAATGCGATATTCCAATTGGCGCCGAATCTGTGCTAGCT  
ACCAGACTCAGACGAACCTCCCAGGCAGCGCCAGCAGCGTTGCCAGCCAGTCAATCATCGCTTAT  
ACAATGTCACTTGGAGCCGAAACTCCGTGGCTTACTCAAACAACAGCATCGCCATCCCCACAAA  
25 CTTACCATATCCGTGACAACTGAGATTCTGCCAGTGTCCATGACTAAGACGTCCGTAGATTGCA  
CTATGTACATATGCGGCGACAGCACAGAATGTTCTAATCTGCTGCTGCAATATGGAAGCTTCTGC  
ACTCAACTGAACAGAGCGCTCACAGGCATCGCCGTGGAGCAGGATAAGAATACCCAGGAGGTGTT  
CGCCCAAGTTAAGCAGATCTACAAGACCCACCCATAAAGGATTTCCGTGGATTCAATTTTAGTC  
AGATACTCCAGACCCATCTAAGCCATCCAAGAGGAGCTTTATCGAGGATCTTTTGTTTAACAAA  
30 GTTACTCTGGCCGACGCCGTTTTCATCAAGCAGTACGGAGATTGCCTCGGCGACATCGCTGCTCG  
TGACCTCATCTGTGCGCAAAAGTTTAACGGTCTGACGGTGTGCTGCCCTCCCTCCTTACTGATGAAA  
TGATCGCCAGTATACAGCGCACTCCTCGTGGCACCATAACATCCGTTGGACATTCGGCGCT  
GGTGCAGCACTGCAGATACCATTTCGCCATGCAAATGGCATATCGTTTCAACGGTATCGGTGTCAC  
ACAGAATGTCTATATGAGAACCAGAAGCTGATCGCAAATCAGTTCAATAGTGCCATCGGAAAAA  
35 TCCAGGATAGCCTTAGCAGCACAGCCTCAGCCCTTGGCAAACCTCCAGGATGTCTGTAACCAGAAT  
GCCCAGGCTCTCAATACCCTCGTGAAGCAGCTCTCATCTAATTTTCGGCGCAATTTCCAGTGTCTC  
CAACGACATCCTCAGCCGCTCGACCCCCCGAGGCCGAAGTGCAGATTGACAGACTGATTACAG  
GTCGACTCCAGAGCCTCCAGACTTACGTGACTCAGCAGCTGATAAGAGCCGCCGAGATAAGGGCC  
AGCGCTAACCTGGCTGCCACAAAGATGTCTGAGTGCCTGCTGGGCCAGTCCAAGAGAGTAGACTT  
40 CTGTGGCAAAGGCTACCATCTGATGAGCTTCCCACAATCCGCACCTCACGGCGTAGTGTCTCTCC  
ACGTGACATATGTACCGGCTCAGGAGAAGAATTTCACTACCGCTCCTGCTATATGCCATGATGGA  
AAGGCTCACTTCCCCCGGGAGGGGGTGTTCGTGTCCAACGGCACCCATTGGTTTGTGACTCAGCG  
GAATTTCTACGAACCCAGATACATAACCACTGACAACACATTTGTGTCCGGAATTTGTGACGTGG  
TCATTGGAATAGTGAACAACACTGTTTATGATCCACTGCAGCCAGAACTTGACAGCTTTAAGGAG  
45 GAGCTCGACAAGTACTTCAAGAATCATACGTACCAGATGTGGACCTCGGAGATATTAGCGGTAT  
CAATGCCAGTGTGTCAATATTGAGAAGGAAATAGACCGCCTTAATGAGGTGCGCAAAAATCTGA  
ACGAGAGCCTCATCGATCTTCAGGAGCTGGGCAATATGAGCAGTACATCAAGTGGCCTTGGTAT  
ATTTGGCTTGGCTTCATCGCCGGCTGATCGCCATAGTAATGGTCACAATTATGCTCTGCTGCAT  
GACCTCTTGCTGCTCCTGTCTGAAAGGCTGCTGCTCTTGCGGATCCTGCTGCAAATTTGATGAGG  
50 ATGACAGTGAACAGTCTGAAAGGGCTGAAGCTGCACTATACTTAGAGGCCCTAGTGTGTCGT  
TTTTCAAATAATATAATCCTTTTAGGGTTTTAGTTAGTTTAAATTTTCTGTTGCTCCTGTTTAGC  
AGGTGCTGCCTTCAGCAAGCACACAAAAACAGAGTGTATTTTAAGTTGTTTGTGTTAGTGATTC

AAAAAAAAAATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTTGCCGGTCTT  
GCGATGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGTAATAATTAACATGTAATGCAT  
GACGTTATTTATGAGATGGGTTTTTATGATTAGAGTCCCAGCAATTATACATTTAATACGCGATAG  
AAAACAAAATATAGCGCGCAAAGTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGAT

5

Cloning vector 8716 from left to right T-DNA (SEQ ID NO:35)

TGGCAGGATATATTGTGGTGTAAACAAATTGACGCTTAGACAACCTTAATAACACATTGCGGACGT  
TTTTAATGTACTGAATTAACGCCGAATCCCGGGCTGGTATATTTATATGTTGTCAAATAACTCAA  
AAACCATAAAAAGTTTAAAGTTAGCAAGTGTGTACATTTTTACTTGAACAAAAATATTCACCTACTA  
CTGTTATAAATCATTATTAAACATTAGAGTAAAGAAATATGGATGATAAGAACAAGAGTAGTGAT  
ATTTTGACAACAATTTTGTGCAACATTTGAGAAAATTTTGTGTTCTCTCTTTTCATTGGTCAA  
AAACAAATAGAGAGAGAAAAAGGAAGAGGGAGAATAAAAAACATAATGTGAGTATGAGAGAGAAAGT  
TGTACAAAAGTTGTACCAAAATAGTTGTACAAATATCATTGAGGAATTTGACAAAAGCTACACAA  
ATAAGGGTTAATTGCTGTAAATAAATAAGGATGACGCATTAGAGAGATGTACCATTAGAGAATTT  
TTGGCAAGTCATTA AAAAGAAAGAATAAATTATTTTTAAAATTTAAAAGTTGAGTCATTTGATTAA  
ACATGTGATTATTTAATGAATTGATGAAAGAGTTGGATTAAAGTTGTATTAGTAATTAGAATTTG  
GTGTCAAATTTAATTTGACATTTGATCTTTTCCCTATATATTGCCCCATAGAGTCAGTTAACTCAT  
TTTTATATTTTCATAGATCAAATAAGAGAAATAACGGTATATTAATCCCTCCAAAAA AAAAAACG  
GTATATTTACTAAAAAATCTAAGCCACGTAGGAGGATAACAGGATCCCCGTAGGAGGATAACATC  
CAATCCAACCAATCACAACAATCCTGATGAGATAACCCACTTTAAGCCACGCATCTGTGGCACA  
TCTACATTATCTAAATCACACATTCTTCCACACATCTGAGCCACACAAAAACCAATCCACATCTT  
TATCACCCATCTCTATAAAAAATCACACTTTGTGAGTCTACACTTTGATTCCCTTCAAACACATAC  
AAAGAGAAGAGACTAATTAATTAATTAATCATCTTGAGAGAAAATGGAACGAGCTATACAAGGAA  
ACGACGCTAGGGAACAAGCTAACAGTGAACGTTGGGATGGAGGATCAGGAGGTACCACCTTCTCCC  
TTCAAACCTCTGACGAAAAGTCCGAGTTGGACTGAGTGGCGGCTACATAACGATGAGACGAATTC  
GAATCAAGATAATCCCTTGGTTTCAAGGAAAGCTGGGGTTTCGGGAAAGTTGTATTTAAGAGAT  
ATCTCAGATACGACAGGACGGAAGCTTCACTGCACAGAGTCCTTGGATCTTGGACGGGAGATTTCG  
GTTAACTATGCAGCATCTCGATTTTTCGGTTTCGACCAGATCGGATGTACCTATAGTATTCGGTT  
TCGAGGAGTTAGTATCACCGTTTCTGGAGGGTCGCGAACTCTTCAGCATCTCTGTGAGATGGCAA  
TTCGTTCTAAGCAAGAACTGCTACAGCTTGCCCCAATCGAAGTGGAAAGTAATGTATCAAGAGGA  
TGCCCTGAAGGTACTCAAACCTTCGAAAAAGAAAGCGAGTAAGTTAAAATGCTTCTTCGTCTCCT  
ATTTATAATATGGTTTGTATTGTTAATTTTGTCTTGTAGAAGAGCTTAATTAATCGTTGTTGT  
TATGAAATACTATTTGTATGAGATGAACTGGTGTAAATGTAATTCATTTACATAAGTGGAGTCAGA  
ATCAGAATGTTTCCCTCCATAACTAAGTACATGAAGACCTGCCGCGTACAATTGCTTTATATTT  
GAACAACTAAAATTGAACATCTTTTGCCACAACCTTTATAAGTGGTTAATATAGCTCAAATATATG  
GTCAAGTTCAATAGATTAATAATGGAAATATCAGTTATCGAAATTCATTAACAATCAACTTAACG  
TTATTAATACTACTAATTTTATATCATCCCCTTTGATAAATGATAGTACACCAATTAGGAAGGAGCA  
TGCTCGCCTAGGAGATTGTCGTTTCCCGCCTTCAGTTTGCAAGCTGCTCTAGCCGTGTAGCCAAT  
ACGCAAACCGCCTCTCCCCGCGCGTTGGGAATTACTAGCGCGTGTGACAAAGCTTGCATGCCGGT  
CAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAAGACC  
AAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAACCTCCTCGGATTCCATTGCCCA  
GCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTG  
CGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAGATGGACCCCCAC  
CCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGT  
GATAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAAGA  
CCAAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAACCTCCTCGGATTCCATTGCC  
CAGCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCAT  
TGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAGATGGACCCCC  
ACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGAT

10

15

20

25

30

35

40

45

5 GTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCCTTCCTCT  
ATATAAGGAAGTTCATTTTCATTTGGAGAGGCACTCCATTTGAATCTATCAAACCAAAACACATTG  
AGCAAAATGGCGAAAAACGTTGCGATTTTCGGCTTATTGTTTTCTCTTCTGTGTTGGTTCCTTC  
TCAGATCTTCGCGACGTCACTCCTCAGCCAAAACGACACCCCCATCTGTCTATCCACTGGCCCCCT  
GGATCTGCTGCCCCAACTAACTCCATGGTGACCCCTGGGATGCCTGGTCAAGGGCTATTTCCCTGA  
GCCAGTGACAGTGACCTGGAACCTCTGGATCCCTGTCCAGCGGTGTGCACACCTTCCCAGCTGTCC  
TGCAGTCTGACCTCTACACTCTGAGCAGCTCAGTGACTGTCCCCCCTCAGCACCTGGCCCAGCGAG  
ACCGTCACTGCAACGTTGCCCCACCCGGCCAGCAGCACCAAGGTGGACAAGAAAATTGTGCCCAG  
GGATTGTGGTTGTAAGCCTTGCATATGTACAGTCCCAGAAGTATCATCTGTCTTCATCTTCCCCC  
10 CAAAGCCCCAAGGATGTGCTCACCATTACTCTGACTCCTAAGGTACCGTGTGTTGTGGTAGACATC  
AGCAAGGATGATCCCGAGGTCCAGTTCAGCTGGTTTGTAGATGATGTGGAGGTGCACACAGCTCA  
GACGCAACCCCGGGAGGAGCAGTTCAACAGCACTTTCGGCTCAGTCAGTGAACCTTCCCATCATGC  
ACCAGGACTGGCTCAATGGCAAGGAGACGTCCAGATTTTGGCGATCTATCAACTGTGCGCAGTT  
CATTGGTACTGGTAGTCTCCCTGGGGGCAATCAGTTTCTGGATGTGCTCTAATGGGTCTCTACAG  
15 TGTAAGATATGTATTTAAAGGCCCTTAGTTCGTGTCGTTTTTCAAATAATATAATCCTTTTAGGGTT  
TTAGTTAGTTTAAATTTTCTGTTGCTCCTGTTTAGCAGGTCTGCTTCAGCAAGCACACAAAAA  
CAGAGTGTTTATTTTAAAGTTGTTTGTGTTAGTGATTCAAAAAAAAAATCGTTCAAACATTTGGCAA  
TAAAGTTTCTTAAGATTGAATCCTGTTGCCGGTCTTGCGATGATTATCATATAATTTCTGTTGAA  
TTACGTTAAGCATGTAATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTTATGA  
20 TTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGCAACTAGGAT  
AAATTTATCGCGCGCGGTGTCTATCTATGTTACTAGATCTCTAGAGTCTCAAGCTTGGCGCGCCCCAC  
GTGACTAGTGCGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAAC  
TTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGGCCCGACCGAT  
CGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGCTAGAGCAGCTTGAGCTTGGATCAGAT  
25 TGTCGTTTCCCGCCTTCAGTTTAACTATCAGTGTTTGACAGGATATATTGGCGGGTAAACCTAA  
GAGAAAAGACGTTTA

Construct 8591 from 2X35S prom to NOS term (SEQ ID NO:36)

30 GTCAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAAGA  
CCAAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAACCTCCTCGGATTCCATTGCC  
CAGCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCAT  
TGCGATAAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCC  
ACCCACGAGGAGCATCGTGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGAT  
GTGATAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAA  
35 GACCAAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAACCTCCTCGGATTCCATTG  
CCCAGCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATC  
ATTGCGATAAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCC  
CCACCCACGAGGAGCATCGTGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTG  
ATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCCTTCCT  
40 CTATATAAGGAAGTTCATTTTCATTTGGAGAGGCACTCCATTTGAATCTATCAAACCAAAACACAT  
TGAGCAAAATGGCGAAAAACGTTGCGATTTTCGGCTTATTGTTTTCTCTTCTGTGTTGGTTCCT  
TCTCAGATCTTCGCGGTGAATCTTACGACGCGAACACAGTTACCACCCGCATATACAAATAGCTT  
CACTCGGGGTGTTTATTACCCGACAAAGTGTTCAGGTCCTCCGTGCTCCACTCAACACAGGACC  
TCTTTCTTCCTTTCTTTTCTAACGTGACATGGTTTTCATGCCATTTCATGTATCCGGCACTAACGGT  
45 ACTAAGAGGTTGCGATAATCCTGTGCTCCCTTTCATGACGGCGTTTACTTTGCAAGCACAGAGAA  
GAGTAACATCATCCGAGGTTGGATCTTTGGCACTACCCCTCGATTCAAAGACGCAGAGCCTCCTCA  
TTGTGAACAATGCCACTAACGTGGTGATCAAAGTTTGGAGTTTCAGTTCTGCAATGACCCCTTTC  
TTGGGGGTGTACTATCATAAGAACAACAAGTCTTGGATGGAATCTGAATTCCCGCTCTATAGCAG  
CGCCAACAACCTGCACCTTTGAATACGTGTCCAGCCCTTCCCTTATGGACCTGGAGGGGAAAGCAGG  
GAAACTTTAAGAATCTGAGAGAGTTCGTGTTTAAAAATATCGACGGCTATTTTAAGATCTATTCT

5 AAGCACACGCCTATTAATCTCGTGCGGATCTTCCACAAGGCTTCAGCGCCCTGGAACCACTCGT  
GGACCTCCCAATTGGTATCAACATCACTAGATTTAGACTCTGCTTGCCCTCCACCGATCCTATC  
TGACACCCGGAGACTCCTCTAGCGGCTGGACTGCCGGCGCTGCCGCTTATTACGTTGGTTATCTT  
CAGCCACGCACGTTCTGCTGAAGTATAACGAGAATGGTACTATTACCGATGCCGTGGATTGTGC  
10 CCTTGACCCCTGTCCGAAACTAAGTGCACACTCAAGTCATTCACTGTGGAAAAAGGAATCTACC  
AGACAAGCAATTTTCGGGTCCAGCCTACTGAGAGCATTGTGCGCTTTCCTAACATCACAAATCTT  
TGCCCTTCGGAGAGGTTTTCAATGCTACACGGTTTGCTCCGTGTATGCCTGGAACCGCAAGAG  
AATTTCCAATTGCGTGGCCGATTACTCCGTGCTCTACAATAGTGCAAGCTTTAGCACCTTTAAGT  
15 GCTATGGCGTATCCCTACTAAGCTTAACGACTTGTGTTTACAAACGTGTATGCCGACTCCTTT  
GTGATACGGGGCGACGAAGTTAGACAGATAGCACCAGGACAGACGGGAAAGATAGCTGACTACAA  
CTATAAGCTTCTGATGACTTCACTGGCTGCGTTATCGCGTGGAATTCTAACAACTGGACTCAA  
AAGTCGGCGGCAACTATAACTATCTCTATCGGCTGTTCCGCAAGAGTAACCTTAAGCCCTTTGAG  
AGAGATATAAGCACTGAAATCTACCAGGCTGGCAGTACGCCCTGTAATGGCGTGGAAGGCTTTAA  
20 TTGTTATTTTCCACTGCAATCCTATGGTTTTTCAGCCAACCAATGGCGTGGGCTACCAACCATAACC  
GCGTCGTGGTGCTCTCCTTTGAACTGCTCCACGCTCCCGCGACTGTCTGCGGCCCAAGAAGTCC  
ACGAACCTTGTGAAGAATAAGTGCCTTAATTTTAAATTTCAACGGCCTCACTGGAACAGGAGTGCT  
CACTGAGAGTAACAAGAAGTTCTGCCATTTCAACAATTTGGCAGAGACATAGCCGATACTACTG  
ACGCCGTTAGGGACCCCCAGACCCTCGAGATTCTCGATATAACGCCCTGCTCCTTCGGTGGAGTT  
25 TCCGTGATCACGCCAGGCACCAATACCAGTAACCAGGTCGCCGTGCTGTATCAGGATGTCAACTG  
TACTGAGGTGCCCGTAGCCATCCATGCGGATCAGCTCACACCAACTTGGAGGGTGTACAGACCCG  
GCTCCAATGTATTCCAGACTCGGGCCGGATGCCTTATTGGCGCCGAACAGTGAACAATAGTTAC  
GAATGCGATATTCCAATTGGCGCCGGAATCTGTGCTAGCTACCAGACTCAGACGAACTCCCCAGG  
CAGCGCCAGCAGCGTTGCCAGCCAGTCAATCATCGCTTATACAATGTCACTTGGAGCCGAAAAC  
30 CCGTGGCTTACTCAAACAACAGCATCGCCATCCCCACAACTTCACCATATCCGTGACAACTGAG  
ATTCTGCCAGTGTCATGACTAAGACGTCCGTAGATTGCACTATGTACATATGCGGCGACAGCAC  
AGAATGTTCTAATCTGCTGCTGCAATATGGAAGCTTCTGCACTCAACTGAACAGAGCGCTCACAG  
GCATCGCCGTGGAGCAGGATAAGAATACCCAGGAGGTGTTGCGCCAAAGTTAAGCAGATCTACAAG  
ACCCACCCATAAAGGATTTCCGGTGGATTCAATTTTAGTCAGATACTCCCAGACCCATCTAAGCC  
ATCCAAGAGGAGCTTTATCGAGGATCTTTTGTTTAACAAAGTTACTCTGGCCGACGCCGGTTTCA  
35 TCAAGCAGTACGGAGATTGCCTCGGCGACATCGCTGCTCGTGACCTCATCTGTGCGCAAAAGTTT  
AACGGTCTGACGGTGCTGCCTCCCCCTCCTTACTGATGAAATGATCGCCAGTATACCAGCGCACT  
CCTCGTGGCACCATAACATCCGTTGGACATTGCGCGCTGGTGCAGACTGCAGATACCATTCG  
CCATGCAAAATGGCATATCGTTTCAACGGTATCGGTGTCACACAGAATGTCTATATGAGAACCAG  
AAGCTGATCGCAAAATCAGTTCAATAGTGCCATCGGAAAAATCCAGGATAGCCTTAGCAGCACAGC  
40 CTCAGCCCTTGGCAAACTCCAGGATGTCTGTAACCAGAATGCCAGGCTCTCAATACCCTCGTGA  
AGCAGCTCTCATCTAATTTCCGGCGCAATTTCCAGTGTCTCAACGACATCCTCAGCCGCTCGAC  
CCCCCGAGGCCGAAGTGCAGATTGACAGACTGATTACAGGTGCACTCCAGAGCCTCCAGACTTA  
CGTGACTCAGCAGCTGATAAGAGCCGCCGAGATAAGGGCCAGCGCTAACCTGGCTGCCACAAAGA  
TGTCTGAGTGCGTGCTGGGCCAGTCCAAGAGAGTAGACTTCTGTGGCAAAAGGCTACCATCTGATG  
45 AGCTTCCCACAATCCGCACCTCACGGCGTAGTGTTCCTCCACGTGACATATGTACCGGCTCAGGA  
GAAGAATTTCACTACCGCTCCTGCTATATGCCATGATGGAAGGCTCACTTCCCCCGGGAGGGGG  
TGTTCTGTGTTCAACGGCACCCATTGGTTTTGTGACTCAGCGGAATTTCTACGAACCCAGATCATA  
ACCACTGACAACACATTTGTGTCCGGAATTTGTGACGTGGTCATTGGAATAGTGAACAACACTGT  
TTATGATCCACTGCAGCCAGAATTGACAGCTTTAAGGAGGAGCTCGACAAGTACTTCAAGAATC  
50 ATACGTCAACCAGATGTGGACCTCGGAGATATTAGCGGTATCAATGCCAGTGTTGTCAATATTCAG  
AAGGAAATAGACCGCCTTAATGAGGTGCGCAAAAATCTGAACGAGAGCCTCATCGATCTTCAGGA  
GCTGGGCAAAATATGAGCAGTACATCAAGTGGCCTTGGTATATTTGGCTTGGCTTCATCGCCGGCC  
TGATCGCCATAGTAATGGTCACAATTATGCTCTGCTGCATGACCTCTTGCTGCTCCTGTCTGAAA  
GGCTGCTGCTCTTGCGGATCCTGCTGCAATTTGATGAGGATGACAGTGAACAGTCTCTGAAGGG  
CGTGAAGCTGCACTATACTTAGAGGCCTTAGTCGTGCTGTTTTTCAATAATATAATCCTTTTAG  
GGTTTTAGTTAGTTTAAATTTTCTGTTGCTCCTGTTTAGCAGGTGCTGCCTTCAGCAAGCACACA  
AAAACAGAGTGTTTATTTTAAAGTTGTTTGTGTTAGTGATTCAAAAAAAAAATCGTTCAACATTTG

GCAATAAAGTTTCTTAAGATTGAATCCTGTTGCCGGTCTTGCGATGATTATCATATAATTTCTGT  
TGAATTACGTTAAGCATGTAATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTT  
ATGATTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGCAAAC  
GGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGAT

- 5 C-Terminal Region of Modified SARS-CoV-2 S protein with H5i Hemagglutinin CT, Variation 2 (SEQ ID NO:37)

WYIWLGFIAGLIAIVMVTIMAGLSLWMCSNGSLQCRICI

- C-Terminal Region of Modified SARS-CoV-2 S protein with H5i Hemagglutinin CT, Variation 3 (SEQ ID NO:38)

- 10 WYIWLGFIAGLIAIVMVTIMLCCMCSNGSLQCRICI

- C-Terminal Region of Modified SARS-CoV-2 S protein with H5i Hemagglutinin CT, Variation 4 (SEQ ID NO:39)

WYIWLGFIAGLIAIVMVTIMLCCSNGSLQCRICI

- 3' UTR AvB (Arracacha virus B) (SEQ ID NO:40)

- 15 TAGTCGTGTCGTTTTCAAATAATATAATCCTTTTAGGGTTTGTAGTTAGTTTAAATTTCTGTTG  
CTCCTGTTTAGCAGGTGTCGCTTCAGCAAGCACAAAAACAGAGTGTTATTTTAAGTTGTTT  
GTTTAGTGATTCAAAAAAAAAA

- 3' UTR trBNYVV (Beet necrotic yellow vein virus) (SEQ ID NO: 41)

- 20 CCTATCTTGATGAAGGTTGTTGTGGTTTTCTCATTACTGTTTTATTATTGTTTGAGTTGCTTATG  
TCGGTCTTGATTATGTGGTGCATAATTATTGAACAAATTGTTTGTGGGTTGTAATGTACTGAC  
TGGGTGTGAATTGTACCAGTCGTTAAAGGGTTTACTATCAGTATATTGATAT

- 3' UTR SBMV (Southern bean mosaic virus) (SEQ ID NO: 42)

- 25 TGAGGAGTTGTATAATAATACCTGCACCCTTCTCTTTGGCAGGGAGGGTGTTTCGTTTTACAAAT  
GCCACGCGCTTGAGGGAGAATGCACGTTAATCATCCCTCCGCTAGTGATGGAGCGTAATCCAAAA  
GT

- 3' UTR TuRSV (Turnip ringspot virus) (SEQ ID NO: 43)

- 30 TGATTTATAATAGCCATAGATTAAAGTTTAAATGTATTACGTTTGTATTTTATTCTCTTTTTTAAG  
TTTCCTATGTTGTTTTAAATTAAATATCTGTATAATTAGTAGATGTAAATCTGCTTTGTGCGTTT  
GACAGTCTGTGGAACGCACTGGTTCATGAGATAGGACCACCTAGGAGGTAGGACTCTGGGTTTT  
AATTATCTCATTTCTTAGTTATACCGTATTATATATATGATTTAGTAGTAATTGTTTTCTCTTGA  
TATGTATTATTACTTTTTTTATT

- 3' UTR CPMV (Cowpea Mosaic virus) (SEQ ID NO: 44)

- 35 ATTTTCTTTAGTTTGAATTTACTGTTATTCGGTGTGCATTTCTATGTTTGGTGAGCGGTTTTCTG  
TGCTCAGAGTGTTTATTTTATGTAATTTAATTTCTTTGTGAGCTCCTGTTTAGCAGGTGCTCC  
CTTCAGCAAGGACACAAAAAGATTTTAATTTTATT

## 5

10

## 15

20

## 25

30

## 35

40

45

5 YGDCCLGDIAARDLICAQKFNGLTVLPLLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQ  
MAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSTPSALGKLQDVVNQNAQALNTLVKQL  
SSNFGAISSVLNDILSRDPPEAEVQIDRLITGRLQSLQTYVTQQILRAAEIRASANLAATKMSE  
CVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDKGAHFPREGV FV  
SNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS  
PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIA  
IVMVTIMLSLWMCSNGSLQCRICI

(PDI) Modified SARS-CoV-2 S protein GSAS+6P with H5i Hemagglutinin CT AA (SEQ ID NO:49)

10 MAKNVAIFGLLFSLLVLPVPSQIFAVNLTTRTQLPPAYTNSFTRGVVYPDKVFRSSVLHSTQDLFL  
PFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWI FGTTLDSTQSLLI V N  
NATNVVIVKCEFFQFCNDPFLGVVYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMLDLEGGKQGNF  
KNLREFVFNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLT  
15 GDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSTVEKGIYQTS  
NFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYG  
VSPTKLNLDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKL PDDFTGCVIAWNSNNLDSKVG  
GNYNLYRLFRKSNLKPFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV  
VLSFELLHAPATVCGPKKSTNLVKNKCVNFENGLTGTGVLTESNKKFLPFQQFGRDIADTTDAV  
20 RDPQTEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSN  
VFQTRAGCLIGAHEVNNSYECDIPIGAGICASYQTQTNSPGSASSVASQSI IAYTMSLGAEN SVA  
YSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTEC SNLLLQYGSFCTQLNRALTGIA  
VEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSPIEDLLFNKVTADAGFIKQ  
YGDCCLGDIAARDLICAQKFNGLTVLPLLLTDEMIAQYTSALLAGTITSGWTFGAGPALQIPFPMQ  
25 MAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSTPSALGKLQDVVNQNAQALNTLVKQL  
SSNFGAISSVLNDILSRDPPEAEVQIDRLITGRLQSLQTYVTQQILRAAEIRASANLAATKMSE  
CVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDKGAHFPREGV FV  
SNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS  
PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIA  
IVMVTIMLSLWMCSNGSLQCRICI

30 (PDI) Modified SARS-CoV-2 S protein GSAS+PP+923 with H5i Hemagglutinin CT AA (SEQ ID NO:50)

MAKNVAIFGLLFSLLVLPVPSQIFAVNLTTRTQLPPAYTNSFTRGVVYPDKVFRSSVLHSTQDLFL  
PFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWI FGTTLDSTQSLLI V N  
NATNVVIVKCEFFQFCNDPFLGVVYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMLDLEGGKQGNF  
35 KNLREFVFNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLT  
GDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSTVEKGIYQTS  
NFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYG  
VSPTKLNLDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKL PDDFTGCVIAWNSNNLDSKVG  
GNYNLYRLFRKSNLKPFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV  
40 VLSFELLHAPATVCGPKKSTNLVKNKCVNFENGLTGTGVLTESNKKFLPFQQFGRDIADTTDAV  
RDPQTEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSN  
VFQTRAGCLIGAHEVNNSYECDIPIGAGICASYQTQTNSPGSASSVASQSI IAYTMSLGAEN SVA  
YSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTEC SNLLLQYGSFCTQLNRALTGIA  
VEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTADAGFIKQ  
45 YGDCCLGDIAARDLICAQKFNGLTVLPLLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQ  
MAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDFSSTASALGKLQDVVNQNAQALNTLVKQL  
SSNFGAISSVLNDILSRDPPEAEVQIDRLITGRLQSLQTYVTQQILRAAEIRASANLAATKMSE  
CVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDKGAHFPREGV FV  
SNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS

PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIA  
IVMVTIMLSLWMCNGLQCRICI

(PDI) Modified SARS-CoV-2 S protein GSAS+4P+923 with H5i Hemagglutinin CT AA (SEQ ID NO:51)

5 MAKNVAIFGLLFSLLVLPVPSQIFAVNLTTTRTQLPPAYTNSFTRGVVYPDKVFRSSVLHSTQDLFL  
PFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWI FGTTLDSTQSLIIVN  
NATNVVIVKCEFFQFCNDPFLGVVYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMLDLEGKQGNF  
KNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLT  
10 GDSSSGWTAGAAAYVGYLQPRFTLLKYNENGTITDAVDCALDPLSETKCTLKSTVEKGIYQTS  
NFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYG  
VSPTKLNLDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPPDFTGCVIAWNSNNLDSKVG  
GNYNLYRLFRKSNLKPFFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV  
VLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAV  
RDPQTLLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSN  
15 VFQTRAGCLIGAHEVNNSECDIPIGAGICASYQTQTNSPGSASSVASQSI IAYTMSLGAENSV  
YSNNSIAIPTNFTISVTTEILPVSMKTSTVDCTMYICGDSTECNLLLQYGSFCTQLNRALTGIA  
VEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSPIEDLLFNKVTLDAGFIKQ  
YGDCLGDIAARDLICAQKFNGLTVLPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQ  
MAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSFSSTPSALGKLQDVVNQNAQALNTLVKQL  
20 SSNFGAISSVLNDILSRLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSE  
CVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHGDKAHFPREGV  
SNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS  
PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIA  
IVMVTIMLSLWMCNGLQCRICI

25 (PDI) Modified SARS-CoV-2 S protein GSAS+6P+923 with H5i Hemagglutinin CT AA (SEQ ID NO:52)

MAKNVAIFGLLFSLLVLPVPSQIFAVNLTTTRTQLPPAYTNSFTRGVVYPDKVFRSSVLHSTQDLFL  
PFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWI FGTTLDSTQSLIIVN  
NATNVVIVKCEFFQFCNDPFLGVVYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMLDLEGKQGNF  
30 KNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLT  
GDSSSGWTAGAAAYVGYLQPRFTLLKYNENGTITDAVDCALDPLSETKCTLKSTVEKGIYQTS  
NFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYG  
VSPTKLNLDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPPDFTGCVIAWNSNNLDSKVG  
GNYNLYRLFRKSNLKPFFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV  
35 VLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAV  
RDPQTLLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSN  
VFQTRAGCLIGAHEVNNSECDIPIGAGICASYQTQTNSPGSASSVASQSI IAYTMSLGAENSV  
YSNNSIAIPTNFTISVTTEILPVSMKTSTVDCTMYICGDSTECNLLLQYGSFCTQLNRALTGIA  
VEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSPIEDLLFNKVTLDAGFIKQ  
40 YGDCLGDIAARDLICAQKFNGLTVLPLLTDEMIAQYTSALLAGTITSGWTFGAGPALQIPFPMQ  
MAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSFSSTPSALGKLQDVVNQNAQALNTLVKQL  
SSNFGAISSVLNDILSRLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSE  
CVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHGDKAHFPREGV  
SNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS  
45 PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIA  
IVMVTIMLSLWMCNGLQCRICI

(PDI) Modified SARS-CoV-2 S protein GSAS+PP with H1 Cal Hemagglutinin CT AA (SEQ ID NO:53)

MAKNVAIFGLLFSLLVLPVPSQIFAVNLTTRTQLPPAYTNSFTRGVVYPDKVFRSSVLHSTQDLFL  
 PFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWI FGTTLDSKTQSLIIVN  
 NATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMLDLEGKQGNF  
 KNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLTP  
 5 GDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTS  
 NFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKCYG  
 VSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPPDFTGCVIAWNSNNLDSKVG  
 GNYNYLYRLFRKSNLKPFFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV  
 10 VLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAV  
 RDPQTLLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSN  
 VFQTRAGCLIGAHEVNNSECDIPIGAGICASYQTQTNSPGSASSVASQSI IAYTMSLGAENSV  
 YSNNSIAIPTNFTISVTTEILPVSMKTSTVDCTMYICGDSTECNNLLQYGSFCTQLNRALTGIA  
 VEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLDAGFIKQ  
 YGDCLGDI AARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQ  
 15 MAYRFNGIGVTQNVLYENQKLIANQFN SAIGKIQDSLSTASALGKLQDVVNQNAQALNTLVKQL  
 SSNFGAISSVLNDILSRDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSE  
 CVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHGDKAHFPREGV FV  
 SNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS  
 PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIA  
 20 IVMVTIMLSFWMCSNGSLQCRICI

**(PDI) Modified SARS-CoV-2 S protein GSAS+PP with H3 Minn Hemagglutinin CT AA (SEQ ID NO:54)**

MAKNVAIFGLLFSLLVLPVPSQIFAVNLTTRTQLPPAYTNSFTRGVVYPDKVFRSSVLHSTQDLFL  
 PFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWI FGTTLDSKTQSLIIVN  
 25 NATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMLDLEGKQGNF  
 KNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLTP  
 GDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTS  
 NFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKCYG  
 30 VSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPPDFTGCVIAWNSNNLDSKVG  
 GNYNYLYRLFRKSNLKPFFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV  
 VLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAV  
 RDPQTLLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSN  
 VFQTRAGCLIGAHEVNNSECDIPIGAGICASYQTQTNSPGSASSVASQSI IAYTMSLGAENSV  
 YSNNSIAIPTNFTISVTTEILPVSMKTSTVDCTMYICGDSTECNNLLQYGSFCTQLNRALTGIA  
 35 VEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLDAGFIKQ  
 YGDCLGDI AARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQ  
 MAYRFNGIGVTQNVLYENQKLIANQFN SAIGKIQDSLSTASALGKLQDVVNQNAQALNTLVKQL  
 SSNFGAISSVLNDILSRDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSE  
 CVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHGDKAHFPREGV FV  
 40 SNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS  
 PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIA  
 IVMVTIMLMWACQKGNIRCNICI

**(PDI) Modified SARS-CoV-2 S protein GSAS+PP with H6 HK Hemagglutinin CT AA (SEQ ID NO:55)**

MAKNVAIFGLLFSLLVLPVPSQIFAVNLTTRTQLPPAYTNSFTRGVVYPDKVFRSSVLHSTQDLFL  
 PFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWI FGTTLDSKTQSLIIVN  
 NATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMLDLEGKQGNF  
 KNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLTP  
 45 GDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTS

5 NFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKCYG  
 VSPTKLNLDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVG  
 GNYNYLYRLFRKSNLKPFFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV  
 VLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAV  
 10 RDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSN  
 VFQTRAGCLIGAHEVNNSECDIPIGAGICASYQTQTNSPGSASSVASQSI IAYTMSLGAENSV  
 YSNNSIAIPTNFTISVTTEILPVSMKTSDVCTMYICGDSTECNNLLQYGSFCTQLNRALTGIA  
 VEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLDAGFIKQ  
 YGDCGLDIAARDLICAQKFNGLTVLPLLTDEMIAQYTSALLAGTITSGWTFGAGAAALQIPFAMQ  
 15 MAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSTASALGKLQDVVNQNAQALNTLVKQL  
 SSNFGAISSVLNDILSRLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSE  
 CVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNTTAPAICHGKAHFPREGV FV  
 SNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS  
 PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIA  
 IVMVTIMLGLWMCNNGSMQCRICI

**(PDI) Modified SARS-CoV-2 S protein GSAS+PP with H7 Guangdong Hemagglutinin CT AA (SEQ ID NO:56)**

20 MAKNVAIFGLLFSLLVLPVPSQIFAVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFL  
 PFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWI FGTTLDLSDKTQSLIIVN  
 NATNVVIVKVEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMLDLEGGKQGNF  
 KNLREFVFKNIDGYFKIYSKHTPINLVRDL PQGFSALEPLVDLPIGINITRFQTLALHRSYLT  
 25 GDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTEKGIYQTS  
 NFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKCYG  
 VSPTKLNLDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVG  
 GNYNYLYRLFRKSNLKPFFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV  
 VLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAV  
 30 RDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSN  
 VFQTRAGCLIGAHEVNNSECDIPIGAGICASYQTQTNSPGSASSVASQSI IAYTMSLGAENSV  
 YSNNSIAIPTNFTISVTTEILPVSMKTSDVCTMYICGDSTECNNLLQYGSFCTQLNRALTGIA  
 VEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLDAGFIKQ  
 YGDCGLDIAARDLICAQKFNGLTVLPLLTDEMIAQYTSALLAGTITSGWTFGAGAAALQIPFAMQ  
 35 MAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSTASALGKLQDVVNQNAQALNTLVKQL  
 SSNFGAISSVLNDILSRLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSE  
 CVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNTTAPAICHGKAHFPREGV FV  
 SNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS  
 PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIA  
 IVMVTIMLVFICVKNNGMRCTICI

**(PDI) Modified SARS-CoV-2 S protein GSAS+PP with H9 HK Hemagglutinin CT AA (SEQ ID NO:57)**

40 MAKNVAIFGLLFSLLVLPVPSQIFAVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFL  
 PFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWI FGTTLDLSDKTQSLIIVN  
 NATNVVIVKVEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMLDLEGGKQGNF  
 KNLREFVFKNIDGYFKIYSKHTPINLVRDL PQGFSALEPLVDLPIGINITRFQTLALHRSYLT  
 45 GDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTEKGIYQTS  
 NFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKCYG  
 VSPTKLNLDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVG  
 GNYNYLYRLFRKSNLKPFFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV  
 VLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAV  
 RDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSN

VFQTRAGCLIGAHEVNNSYECDIPIGAGICASYQTQTNSPGSASSVASQSI IAYTMSLGAENSV  
 YSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTEC SNLLQYGSFCTQLNRALTGIA  
 VEQDKNTQEVEFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLENKVT  
 5 YGDCLDIAARDLICAQKFNGLT VLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQ  
 MAYRFNGIGVTQNVLYENQKLIANQFN SAIGKIQDSLSSSTASALGKLQDVVNQNAQALNTLVKQL  
 SSNFGAISSVLNDILSRDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSE  
 CVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHGKAHFPREGV  
 SNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVGIVNNTVYDPLQPELDSFKEELDKYFKNHTS  
 10 PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGF  
 IAGLIAIVMVTIMLLFWAMSNNGSCRCNICI

**(PDI) Modified SARS-CoV-2 S protein GSAS+PP with B/Wash Hemagglutinin CT AA (SEQ ID NO:58)**

MAKNVAIFGLLFSLLVLVPSQIFAVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFL  
 PFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSL  
 15 NATNVVIKVFCEQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPF  
 LMDLEGKQGNFNKLNREFVFKNIDGYFKIYSKH  
 TPINLVRDLPQGFSALEPLVDLPIGINITRFQTL  
 LALHRSYLTFGDSSSGWTAGAAAYYVG  
 YLQPRTFLLKYNENGTITDAVDCALDPLSETKCT  
 LKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCP  
 FGEVFNATRFASVYAWNRKRISNCVADYSVLYNSA  
 20 SFSTFKCYGVSPTKLNLDLCFTNVYADSFVIRGDEVRQ  
 IAPGQTGKIADYNYKL PDDFTGCVIAWNSNNLDSKVG  
 GNYNYLYRLFRKSNLKPFERDISTEIIYQAGSTPCNG  
 VEGFNCFPLQSYGFQPTNGVGYQPYRVV  
 VLSFELLHAPATVCGPKKSTNLVKNKCVNFNENGLTGT  
 GVLTESNKKFLPFQQFGRDIADTTDAV  
 RDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQD  
 VNCTEVPVAIHADQLTPTWRVYSTG  
 25 SNVFQTRAGCLIGAHEVNNSYECDIPIGAGICASYQTQTNSPGSASSVASQSI IAYTMSLGAENSV  
 YSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTEC SNLLQYGSFCTQLNRALTGIA  
 VEQDKNTQEVEFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLENKVT  
 YGDCLDIAARDLICAQKFNGLT VLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQ  
 MAYRFNGIGVTQNVLYENQKLIANQFN SAIGKIQDSLSSSTASALGKLQDVVNQNAQALNTLVKQL  
 SSNFGAISSVLNDILSRDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSE  
 CVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHGKAHFPREGV  
 30 SNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVGIVNNTVYDPLQPELDSFKEELDKYFKNHTS  
 PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGF  
 IAGLIAIVMVTIMLVVYMVSRDNVSCSICL

**(PDI) Modified SARS-CoV-2 S protein GSAS+PP with H5i Hemagglutinin CT (alternative 1) AA (SEQ ID NO:59)**

MAKNVAIFGLLFSLLVLVPSQIFAVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFL  
 PFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSL  
 35 NATNVVIKVFCEQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPF  
 LMDLEGKQGNFNKLNREFVFKNIDGYFKIYSKH  
 TPINLVRDLPQGFSALEPLVDLPIGINITRFQTL  
 LALHRSYLTFGDSSSGWTAGAAAYYVG  
 YLQPRTFLLKYNENGTITDAVDCALDPLSETKCT  
 LKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCP  
 FGEVFNATRFASVYAWNRKRISNCVADYSVLYNSA  
 40 SFSTFKCYGVSPTKLNLDLCFTNVYADSFVIRGDEVRQ  
 IAPGQTGKIADYNYKL PDDFTGCVIAWNSNNLDSKVG  
 GNYNYLYRLFRKSNLKPFERDISTEIIYQAGSTPCNG  
 VEGFNCFPLQSYGFQPTNGVGYQPYRVV  
 VLSFELLHAPATVCGPKKSTNLVKNKCVNFNENGLTGT  
 GVLTESNKKFLPFQQFGRDIADTTDAV  
 RDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQD  
 VNCTEVPVAIHADQLTPTWRVYSTG  
 45 SNVFQTRAGCLIGAHEVNNSYECDIPIGAGICASYQTQTNSPGSASSVASQSI IAYTMSLGAENSV  
 YSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTEC SNLLQYGSFCTQLNRALTGIA  
 VEQDKNTQEVEFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLENKVT  
 YGDCLDIAARDLICAQKFNGLT VLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQ  
 MAYRFNGIGVTQNVLYENQKLIANQFN SAIGKIQDSLSSSTASALGKLQDVVNQNAQALNTLVKQL

5

SSNFGAISSVLNDILSRDLPPEAEVQIDRLITGRLQSLQTYVTQQILRAAEIRASANLAATKMSE  
 CVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHGKAHFPREGVFFV  
 SNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS  
 PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIA  
 IVMVTIMAGLSLWMCSNGSLQCRICI

(PDI) Modified SARS-CoV-2 S protein GSAS+PP with H5i Hemagglutinin CT (alternative 2)  
 AA (SEQ ID NO:60)

10

15

20

25

MAKNVAIFGLLFSLVLVPSQIFAVNLTTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFL  
 PFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDLSDTQSLILVN  
 NATNVVIVKCFEQFCNDPFLGVYHKNKSWMESEFRVYSSANNCTFEYVSQPFMLDLEGKQGNF  
 KNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLT  
 GDSSSGWTAGAAAYVGYLQPRFTLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTS  
 NFRVQPTESIVRFPNITNLCPFGVEFVNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYG  
 VSPTKLNLDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLDDFTGCVIAWNSNNLDSKVG  
 GNYNYLYRLFRKSNLKPFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV  
 VLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAV  
 RDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSN  
 VFQTRAGCLIGAHEVNNSYECDIPIGAGICASYQTQTNPSGSASSVASQSI IAYTMSLGAENSV  
 YSNNISIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECNNLLQYGSFCTQLNRALTGIA  
 VEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLENKVTADAGFIKQ  
 YGDCGLDIAARDLICAQKFNGLTVLPLLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQ  
 MAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSSTASALGKLQDVVNQNAQALNTLVKQL  
 SSNFGAISSVLNDILSRDLPPEAEVQIDRLITGRLQSLQTYVTQQILRAAEIRASANLAATKMSE  
 CVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHGKAHFPREGVFFV  
 SNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS  
 PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIA  
 IVMVTIMAGLSLWMCSNGSLQCRICI

(PDI) Modified SARS-CoV-2 S protein GSAS+PP with H5i Hemagglutinin CT (alternative 3)  
 AA (SEQ ID NO:61)

30

35

40

45

MAKNVAIFGLLFSLVLVPSQIFAVNLTTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFL  
 PFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDLSDTQSLILVN  
 NATNVVIVKCFEQFCNDPFLGVYHKNKSWMESEFRVYSSANNCTFEYVSQPFMLDLEGKQGNF  
 KNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLT  
 GDSSSGWTAGAAAYVGYLQPRFTLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTS  
 NFRVQPTESIVRFPNITNLCPFGVEFVNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYG  
 VSPTKLNLDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLDDFTGCVIAWNSNNLDSKVG  
 GNYNYLYRLFRKSNLKPFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV  
 VLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAV  
 RDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSN  
 VFQTRAGCLIGAHEVNNSYECDIPIGAGICASYQTQTNPSGSASSVASQSI IAYTMSLGAENSV  
 YSNNISIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECNNLLQYGSFCTQLNRALTGIA  
 VEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLENKVTADAGFIKQ  
 YGDCGLDIAARDLICAQKFNGLTVLPLLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQ  
 MAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSSTASALGKLQDVVNQNAQALNTLVKQL  
 SSNFGAISSVLNDILSRDLPPEAEVQIDRLITGRLQSLQTYVTQQILRAAEIRASANLAATKMSE  
 CVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHGKAHFPREGVFFV  
 SNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS  
 PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIA  
 IVMVTIMLCCMCSNGSLQCRICI

(PDI) Modified SARS-CoV-2 S protein GSAS+PP with H5i Hemagglutinin CT (alternative 4)  
AA (SEQ ID NO:62)

MAKNVAIFGLLFSLVLVPSQIFAVNLTTTRTQLPPAYTNSFTRGVVYPDKVFRSSVLHSTQDLFL  
 5 PFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLILVN  
 NATNVVIVKVEFQFCNDPFLGVVYHKNNKSWMESEFRVYSSANNCTFEYVSQPFMLDLEGKQGNF  
 KNLREFVFKNIDGYFKIYSKHTPINLVRDLPGQFSALEPLVDLPIGINITRFQTLALHRSYLTP  
 GDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTS  
 NFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYG  
 10 VSP TKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLDDFTGCVIAWNSNNLDSKVG  
 GNYNYLYRLFRKSNLKPFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV  
 VLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQGFGRDIADTTDAV  
 RDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSN  
 VFQTRAGCLIGAHEVNNSYECDIPIGAGICASYQTQTNSPGSASSVASQSI IAYTMSLGAENSV  
 15 YSNNISIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECNNLLQYGSFCTQLNRALTGIA  
 VEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLENKVTLADAGFIKQ  
 YGDCLDGDI AARDLICAQKFNGLTVLPLLLTDEMIAQYTSALLAGTITSGWTFGAGAAALQIPFAMQ  
 MAYRFNGIGVTQNVLYENQKLIANQFN SAIGKIQDSLSSSTASALGKLQDVVNQNAQALNTLVKQL  
 SSNFGAISSVLNDILSRDLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSE  
 20 CVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHGDKAHFPREGVVF  
 SNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVIVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTS  
 PDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIA  
 IVMVTIMLCCSNGSLQCRICI

N-terminal region of the native SARS-CoV-2 S protein (native signal peptide sequence  
underlined) (SEQ ID NO:63)

25 MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNS

Modified TMCT with intervening peptide sequence (X)<sub>n</sub> (SEQ ID NO: 64)

WYIWLGFIAGLIAIVMTIM(X)<sub>n</sub>CSNGSXXCXICI

PDI-S Protein GSAS+4P-DNA (SEQ ID NO: 65)

atggcgaaaaacggttgcgattttcggccttattgttttctcttcttctgtgttggttccttctcagat  
 30 cttcgcggtgaatccttacgacgcgaacacagttaccacccgcatatacaaatagcttcactcggg  
 gtgtttattaccccgacaaagtgttcaggtcctccgtgctccactcaacacaggacctctttctt  
 ctttcttttctaactgtgacatggtttcatgccattcatgtatccggcactaacggtactaagag  
 gttcgataatcctgtgctccctttcaatgacggcggtttactttgcaagcacagagaagagtaaca  
 35 tcatccgaggttggtatctttggcactaccctcgattcaaagacgcagagcctcctcattgtgaac  
 aatgccactaacgtggtgatcaaagtgttcgaggtttcagttctgcaatgaccctttcttgggggt  
 gtactatcataagaacaacaagtcttgatggaatctgaattccgctctatagcagcgccaaca  
 actgcacctttgaatacgtgtccagcccttccttatggacctggagggaagcagggaaacttt  
 aagaatctgagagagttcgtgtttaaaaatatcgacggctatttttaagatctattctaagcacac  
 40 gcctattaatctcgtgcgcatcttccacaaggcttcagcgccctggaaccactcgtggacctcc  
 caattggtatcaacatcactagatttcagactctgcttgccctccaccgatcctatctgacaccc  
 ggagactcctctagcggctggactgccggcgctgccgcttattacgttggttatcttcagccacg  
 cacgttcctgctgaagtataacgagaatggtactattaccgatgccgtggattgtgcccttgacc  
 45 ccctgtccgaaactaagtgcacactcaagtcattcactgtggaaaaaggaatctaccagacaagc  
 aattttcgggtccagcctactgagagcattgtgcgctttcctaaccacaaaatctttgcccctt  
 cggagaggttttcaatgctacacgggtttgcctccgtgtatgcctggaaccgcaagagaatttcca

attgctggccgattactccgtgctctacaatagtgaagcttttagcacctttaagtgtatggc  
 gtatcccctactaagcttaacgacttgtgtttcacaaacgtgtatgccgactcctttgtgatacg  
 gggcgacgaagttagacagatagcaccaggacagacgggaaagatagctgactacaactataagc  
 5 ttcctgatgacttcactggctgcttattcgctggaattctaacaacctggactcaaaagtccgc  
 ggcaactataactatctctatcggtgttccgcaagagtaaccttaagccctttgagagagatat  
 aagcactgaaatctaccaggtggcagtagccctgtaatggcgtggaaggtttaattgttatt  
 ttccactgcaatcctatggttttcagccaaccaatggcgtgggtaccaaccataccgcgtcgtg  
 gtgctctcctttgaactgctccacgctcccgactgtctgcggcccaagaagtccacgaacct  
 10 tgtgaagaataagtgcgttaatttttaatttcaacggcctcactggaacaggagtgtcactgaga  
 gtaacaagaagtccctgccatttcaacaatttggcagagacatagccgatactactgacgcgtt  
 agggacccccagaccctcgagattctcgatataacgcctgtccttcggtggagtttccgtgat  
 cagccaggcaccaataaccagtaaccaggtgcgcgtgtgtatcaggatgtcaactgtactgagg  
 tgcctgtagccatccatgcggtacagctcacaccaacttggaggtgtacagcaccggctccaat  
 15 gtattccagactcgggcccggatgccttattggcgccgaacacgtgaacaatagttacgaatgcga  
 tattccaattggcgccggaatctgtgctagctaccagactcagacgaactcccaggcagcgcca  
 gcagcgttgccagccagtcaatcatcgcttatacaatgtcacttggagccgaaaactccgtggct  
 tactcaaacaacagcatcgccatccccacaaacttcaccatatccgtgacaactgagattctgcc  
 agtgtccatgactaagacgtccgtagattgcactatgtacatatgcggcgacagcacagaatgtt  
 ctaatctgtctgctgcaatatggaagcttctgcactcaactgaacagagcgtcacaggcatcgcc  
 20 gtggagcaggataagaatacccaggaggtgttcgccaagttaagcagatctacaagaccccacc  
 cataaaggatttcggtggattcaatttttagtcagatactccagaccatctaagccatccaaga  
 ggagcccatcgaggatcttttgtttaacaaagtactctggcgacgcgggttcatcaagcag  
 tacggagattgcctcgcgacatcgctgctgctgacctcatctgtgcgcaaaagttaaaggctc  
 25 gacggtgctgcctcccctccttactgatgaaatgatcgccagctataaccagcgactcctcgctg  
 gcaccataacatccggttggacattcggcgctggtgcagcactgcagataccattcgccatgcaa  
 atggcatatcgtttcaacgggtatcggtgtcacacagaatgtcctatatgagaaccagaagctgat  
 cgcaaatcagttcaatagtgccatcggaataatccaggatagccttagcagcacaccctcagccc  
 ttggcaaaactccaggatgtcgtgaaccagaatgccaggctctcaataccctcgtgaagcagctc  
 30 tcatctaatttcggcgcaatttccagtgctcctcaacgacatcctcagccgcctcgacccccccga  
 ggccgaagtgcagattgacagactgattacaggtcgactccagagcctccagacttacgtgactc  
 agcagctgataagagccgcccagagataagggccagcgctaacctggctgccacaaagatgtctgag  
 tgcgtgctgggcccagtcgaagagtagacttctgtggcaaggctaccatctgatgagcttccc  
 acaatccgcacctcacggcgtagtgttccctccacgtgacatatgtaccggctcaggagaagaatt  
 35 tcaactaccgctcctgctatatgccatgatggaaggctcacttccccgggaggggggtgttcgtg  
 tccaacggcaccattggtttgtgactcagcggaatttctacgaaccccagatcataaccactga  
 caacacatttgtgtccggaaattgtgacgtggtcattggaatagtgaacaacactgtttatgatc  
 cactgcagccagaacttgacagctttaaggaggagctcgacaagtacttcaagaatcatacgtca  
 ccagatgtggacctcgagatattagcggtatcaatgccagtggtgtcaatattcagaaggaaat  
 40 agaccgccttaatgaggtcgccaaaaatctgaacgagagcctcatcgatcttcaggagctgggca  
 aatatgagcagtagcatcaagtggccttggatattttggcttggcttcatcgccggcctgatcgcc  
 atagtaatggtcacaattatgctctctttatggatgtgctccaatggatcggttacaatgcagaat  
 ttgcatttaa

#### PDI-S Protein GSAS+6P-DNA (SEQ ID NO: 66)

atggcgaaaaacgttgcgattttcggttattgttttctcttcttgtgttggttccttctcagat  
 45 cttcgcggtgaatcttacgacgcgaacacagttaccacccgcataatacaaatagcttactcggg  
 gtgtttattaccccgacaaagtgttcaggctcctccgtgctccactcaacacaggacctcttctt  
 cctttcttttctaactgtacatggtttcatgccattcatgtatccggcactaacggtactaagag  
 gttcgataatcctgtgctccctttcaatgacggcgtttactttgcaagcacagagaagagtaaca  
 50 tcatccgaggttggatctttggcactaccctcgattcaaagacgcagagcctcctcattgtgaac  
 aatgccactaacgtggtgatcaaagtgttcgagtttcagttctgcaatgacccttcttgggggt

gtactatcataagaacaacaagtcttggatggaatctgaattccgcgtctatagcagcgccaaca  
actgcacctttgaatacgtgtcccagcccttccttatggacctggagggaaagcagggaacttt  
aagaatctgagagagttcgtgtttaaaaatatcgacggctattttaagatctattctaagcacac  
gcctattaatctcgtgcgcatcttccacaaggcttcagcgccctggaaccactcgtggacctcc  
5 caattgggtatcaacatcactagatttcagactctgcttgccctccaccgatcctatctgacaccc  
ggagactcctctagcggtggaactgcccgcgtgcccgttattacgttggttatcttcagccaag  
cacgttcctgctgaagtataacgagaatggtactattaccgatgccgtggattgtgcccttgacc  
cctgtccgaaactaagtgcacactcaagtcattcactgtggaaaaaggaatctaccagacaagc  
aattttcgggtccagcctactgagagcattgtgcgctttcctaacatcacaaatctttgcccctt  
10 cggagaggttttcaatgctacacggtttgctccgtgtatgcctggaaacgcaagagaatttcca  
attgctggccgattactccgtgctctacaatagtgaagcttttagcacctttaagtgtatggc  
gtatcccctactaagcttaacgacttgtgtttcacaacgtgtatgccgactcctttgtgatacg  
ggcgacgaagttagacagatagcaccaggacagacgggaaagatagctgactacaactataagc  
ttcctgatgacttcactggctgctgttatcgctggaattctaacaacctggactcaaaagtgcgc  
15 ggcaactataactatctctatcggtgttccgcaagagtaaccttaagccctttgagagagatat  
aagcactgaaatctaccaggtggcagtagcgcctgtaatggcgtggaaggtttaattgttatt  
ttccactgcaatcctatggttttcagccaaccaatggcgtgggctaccaaccataccgcgtcgtg  
gtgctctcctttgaactgctccacgctcccgcgactgtctgcggccccaagaagtccacgaacct  
tgtgaagaataagtgcgttaatttttaattttcaacggcctcactggaacaggagtgtcactgaga  
20 gtaacaagaagtccctgccatttcaacaatttggcagagacatagccgatactactgacgcggtt  
agggaccccagaccctcgagattctcgatataacgcctgctccttcggtggagtttccgtgat  
cacgccaggcacaataaccagtaaccagggtgcgcgtgctgtatcaggatgtcaactgtactgagg  
tgcccgtagccatccatgcccagctcacaccaacttgagggtgtacagcacccggtccaat  
gtattccagactcgggcccggatgccttattggcgccgaacacgtgaacaatagttacgaatgcga  
25 tattccaattggcgccggaatctgtgctagctaccagactcagacgaactcccaggcagcgcca  
gcagcgttgccagccagtcacatcgcttatacaatgtcacttgagcggaaaactccgtggct  
tactcaacaacagcatcgccatcccacaaacttcacatctcgtgacaactgagattctgcc  
agtgtccatgactaagacgtccgtagattgcactatgtacatatgcggcgacagcacagaatggt  
ctaactctgctgctgcaatatggaagcttctgcactcaactgaacagagcgctcacaggcatcgcc  
30 gtggagcaggataagaatacccaggagggtgttcgcccaagttaagcagatctacaagaccccacc  
cataaaggatttcggtggattcaatttttagtcagatactcccagaccatctaagccatccaaga  
ggagcccatcgaggatcttttgtttaacaaagttaactctggccgacgcgggtttcactcagcag  
tcaggagattgcctcggcgacatcgctgctcgtagcctcatctgtgcgcaaaagtttaacggtct  
gacggtgctgcctcccctccttactgatgaaatgatcgcccagtataaccagcgactcctcgctg  
35 gcaccataacatccggttgacattcggcgctggtcccgcactgcagataccattcccctgcaa  
atggcatatcgtttcaacgggtatcggtgtcacacagaatgtcctatatgagaaccagaagctgat  
cgaaatcagttcaatagtccatcggaataatccaggatagccttagcagcacacctcagccc  
ttggcaaaactccaggatgtcgtgaaccagaatgccagggtctcaataccctcgtgaagcagctc  
tcacttaatttcggcgcaatttccagtgctcctcaacgacatcctcagccgctcgacccccccga  
40 ggccgaagtgcagattgacagactgattacaggctgactccagagcctccagacttacgtgactc  
agcagctgataagagccgagagataagggccagcgctaacctggctgccacaaagatgtctgag  
tgctgctgggccagtccaagagagtagacttctgtggcaaaaggctaccatctgatgagcttccc  
acaatccgcacctcacggcgtagtgttctccacgtgacatatgtaccggctcaggagaagaatt  
tactaccgctcctgctatatgcatgatggaaaggctcacttccccgggaggggtgttcgtg  
45 tccaacggcacccattggtttgtgactcagcggaatttctacgaaccccagatcataaccactga  
caacacatttgtgtccggaaattgtgacgtggtcattggaatagtgaacaacactgtttatgatc  
cactgcagccagaacttgacagctttaaggaggagctcgacaagtaacttaagaatcatacgtca  
ccagatgtggacctcgagatatattagcggtatcaatgccagtggtgtcaatattcagaaggaaat  
agaccgccttaatgaggtcgcaaaaaatctgaacgagagcctcatcgatcttcaggagctgggca  
50 aatatgagcagtagcatcaagtggccttggtatatttggcttggttcacgcggcctgatcgcc  
atagtaattggtcacaattatgctctctttatggatgtgctccaatggatcggttacaatgcagaat  
ttgcatttaa

## PDI-S Protein GSAS+2P+L923F-DNA (SEQ ID NO: 67)

atggcgaaaaacgttgcgattttcggccttattgttttctcttcttgtgttggttccttctcagat  
cttcgcggtgaatcttacgacgcgaacacagttaccacccgcataatacaaatagcttcactcggg  
gtgtttattaccccgacaaagtgttcagggtcctccgtgctccactcaacacaggacctctttctt  
5 cctttcttttctaactgacatggtttcatgccattcatgtatccggcactaacggtactaagag  
gttcgataatcctgtgctccctttcaatgacggcggtttactttgcaagcacagagaagagtaaca  
tcatccgaggttgatctttggcactaccctcgattcaaagacgcagagcctcctcattgtgaac  
aatgccactaacgtggtgatcaaagtttgcgagtttcagttctgcaatgaccctttcttgggggt  
gtactatcataagaacaacaagtcttgatggaatctgaattccgcgtctatagcagcgccaaca  
10 actgcacctttgaatacgtgtcccagcccttccttatggacctggaggaaaagcagggaacttt  
aagaatctgagagagttcgtgtttaaaaatatcgacggctatttttaagatctattctaagcacac  
gcctattaatctcgtgcgcgatcttccacaaggcttcagcgccctggaaccactcgtggacctcc  
caattggtatcaacatcactagatttcagactctgcttgccctccaccgatcctatctgacaccc  
ggagactcctctagcggctggactgcccgcgctgcccgttattacgttggttatcttcagccacg  
15 cacgttcctgctgaagtataacgagaatggtactattaccgatgccgtggattgtgcccctgacc  
cctgtccgaaactaagtgcacactcaagtcattcactgtggaaaaaggaatctaccagacaagc  
aattttcgggtccagcctactgagagcattgtgcgctttcctaactcacaaatctttgcccctt  
cggagaggttttcaatgctacacgggtttgcctccgtgtatgcctggaaccgcgaagagaatttcca  
attgctgtggcgattactccgtgctctacaatagtgaagctttagcacctttaagtgtctatggc  
20 gtatccctactaagcttaacgacttggtgtttcacaacacgtgtatgccgactcctttgtgatacg  
ggcgacgaagttagacagatagcaccaggacagacgggaaagatagctgactacaactataagc  
ttcctgatgacttcactggctgcggttatcgcggtggaattctaacaacctggactcaaaagtccgc  
ggcaactataactatctctatcggtgttccgcaagagtaaccttaagccctttgagagagatat  
aagcactgaaatctaccaggctggcagtagcgcctgtaatggcggtggaaggctttaattgttatt  
25 ttccactgcaatcctatggtttttcagccaaccaatggcggtgggctaccaaccataaccgcgtcgtg  
gtgctctcctttgaactgctccacgctcccgcgactgtctgcggccccaagaagtccacgaacct  
tgtgaagaataagtgcgttaatttttaatttcaacggcctcactggaacaggagtgctcactgaga  
gtaacaagaagtccctgccatttcaacaatttggcagagacatagccgatactactgacgcggtt  
agggacccccagaccctcgagattctcgatataacgcctgctccttcggtggagtttccgtgat  
30 cacgccaggcaccaataaccagtaaccagggtcgccgtgctgtatcaggatgtcaactgtactgagg  
tgccgttagccatccatgcggtacagctcacaccaacttgagggtgtacagcaccggctccaat  
gtattccagactcgggcccggatgccttatttggcgccgaacacgtgaacaatagttacgaatgcga  
tattccaatttggcgccggaatctgtgctagctaccagactcagacgaactcccaggcagcgcca  
gcagcgttgccagccagtcaatcatcgcttatacaatgtcacttgagccgaaaaactccgtgggt  
35 tactcaaacaacagcatcgccatccccacaaacttcaccatatccgtgacaactgagattctgcc  
agtgtccatgactaagacgtccgtagattgcactatgtacatatgcccgcagacagacagaatggt  
ctaactctgctgctgcaatatggaagcttctgcaactgaacagagcgctcacaggcatcgcc  
gtggagcaggataagaataaccaggaggtgttcgcccaggttaagcagatctacaagacccccacc  
cataaaggatttccgtggattcaatttttagtcagatactcccagaccatctaagccatccaaga  
40 ggagctttatcgaggatcttttgtttaacaaagttaactctggccgacgcgggtttcatcaagcag  
tacggagattgcctcggcgacatcgctgctcgtgacctcatctgtgcgcaaaagtttaacggctc  
gacggtgctgcctcccctccttactgatgaaatgatcgccaggtataaccagcactcctcgctg  
gcaccataacatccggttgacattcggcgctggtgcagcactgcagataaccattcgccatgcaa  
atggcatatcggttcaacgggtatcggtgtcacacagaatgtcctatatgagaaccagaagctgat  
45 cgcaaatcagttcaatagtgccatcggaataatccaggatagcttcagcagcacagcctcagccc  
ttggcaaaactccaggatgtcgtgaaccagaatgccaggctctcaataccctcgtgaagcagctc  
tcatctaatttccggcgcaatttccagtgctcctcaacgacatcctcagccgctcgacccccccga  
ggccgaagtgcagattgacagactgattacaggctcgaactccagagcctccagacttacgtgactc  
agcagctgataagagccgcgagataagggccagcgctaacctggctgccacaaagatgtctgag  
50 tgcgtgctgggcccagtcgaagagagtagacttctgtggcaaggctaccatctgatgagcttccc  
acaatccgcacctcacggcgtagtggttctccacgtgacatatgtaccggctcaggagaagaatt

5 t c a c t a c c g t c c t g c t a t a t g c c a t g a t g g a a g g c t c a c t t c c c c g g g a g g g g t g t t c g t g  
t c c a a c g g c a c c c a t t g g t t t g t g a c t c a g c g g a a t t t c t a c g a a c c c c a g a t c a t a a c c a c t g a  
c a a c a c a t t t g t g t c c g g a a a t t g t g a c g t g g t c a t t g g a a t a g t g a a c a a c a c t g t t t a t g a t c  
c a c t g c a g c c a g a a c t t g a c a g c t t t a a g g a g g a g c t c g a c a a g t a c t t c a a g a a t c a t a c g t c a  
c c a g a t g t g g a c c t c g g a g a t a t t a g c g g t a t c a a t g c c a g t g t t g t c a a t a t t c a g a a g g a a a t  
a g a c c g c c t t a a t g a g g t c g c a a a a a t c t g a a c g a g a g c c t c a t c g a t c t t c a g g a g c t g g g c a  
a a t a t g a g c a g t a c a t c a a g t g g c c t t g g t a t a t t t g g c t t g g c t t c a t c g c c g g c c t g a t c g c c  
a t a g t a a t g g t c a c a a t t a t g c t c t c t t t a t g g a t g t g c t c c a a t g g a t c g t t a c a a t g c a g a a t  
t t g c a t t t a a

10 PDI-S Protein GSAS+4P+L923F-DNA (SEQ ID NO: 68)

15 a t g g c g a a a a a c g t t g c g a t t t t c g g c t t a t t g t t t t c t c t t c t t g t g t t g g t t c c t t c t c a g a t  
c t t c g c g g t g a a t c t t a c g a c g c g a a c a c a g t t a c c a c c c g c a t a t a c a a a t a g c t t c a c t c g g g  
g t g t t t a t t a c c c c g a c a a a g t g t t c a g g t c c t c c g t g c t c c a c t c a a c a c a g g a c c t c t t t c t t  
c t t t c t t t t c t a a c g t g a c a t g g t t t c a t g c c a t t c a t g t a t c c g g c a c t a a c g g t a c t a a g a g  
g t t c g a t a a t c c t g t g c t c c c t t t c a a t g a c g g c g t t t a c t t t g c a a g c a c a g a g a a g a g t a a c a  
t c a t c c g a g g t t g g a t c t t t g g c a c t a c c c t c g a t t c a a a g a c g c a g a g c c t c c t c a t t g t g a a c  
a a t g c c a c t a a c g t g g t g a t c a a a g t t t g c g a g t t t c a g t t c t g c a a t g a c c c t t t c t t g g g g g t  
g t a c t a t c a t a a g a a c a a a g t c t t g g a t g g a a t c t g a a t t c c g c g t c t a t a g c a g c g c c a a c a  
a c t g c a c c t t t g a a t a c g t g t c c c a g c c t t c c t t a t g g a c c t g g a g g g a a a g c a g g g a a a c t t t  
20 a a g a a t c t g a g a g a g t t c g t g t t t a a a a a t a t c g a c g g c t a t t t t a a g a t c t a t t c t a a g c a c a c  
g c c t a t t a a t c t c g t g c g c g a t c t t c c a c a a g g c t t c a g c g c c c t g g a a c c a c t c g t g g a c c t c c  
c a a t t g g t a t c a a c a t c a c t a g a t t t c a g a c t c t g c t t g c c c t c c a c c g a t c c t a t c t g a c a c c c  
g g a g a c t c c t c t a g c g g c t g g a c t g c c g g c g t g c c g c t t a t t a c g t t g g t a t c t t c a g c c a c g  
c a c g t t c c t g c t g a a g t a t a a c g a g a a t g g t a c t a t t a c c g a t g c c g t g g a t t g t g c c c t t g a c c  
25 c c c t g t c c g a a a c t a a g t g c a c a c t c a a g t c a t t c a c t g t g g a a a a g g a a t c t a c c a g a c a a g c  
a a t t t t c g g g t c c a g c c t a c t g a g a g c a t t g t g c g c t t t c c t a a c a t c a c a a a t c t t t g c c c c t t  
c g g a g a g g t t t t c a a t g c t a c a c g g t t t g c c t c c g t g t a t g c c t g g a a c c g c a a g a g a a t t t c c a  
a t t g c g t g g c c g a t t a c t c c g t g c t c t a c a a t a g t g c a a g c t t t a g c a c c t t t a a g t g c t a t g g c  
30 g t a t c c c t a c t a a g c t t a a c g a c t t g t g t t t c a c a a c g t g t a t g c c g a c t c c t t t g t g a t a c g  
g g g c g a g a g t t a g a c a g a t a g c a c c a g g a c a g c g g g a a g a t a g t g a c t a c a a c t g a t a a g c  
t t c c t g a t g a c t t c a c t g g c t g c g t t a t c g c g t g g a a t t c t a a c a a c c t g g a c t c a a a a g t c g g c  
g g c a a c t a t a a c t a t c t c t a t c g g c t g t t c c g c a a g a g t a a c c t t a a g c c c t t t g a g a g a g a t a t  
a a g c a c t g a a a t c t a c c a g g c t g g c a g t a c g c c c t g t a a t g g c g t g g a a g g c t t t a a t t g t t a t t  
t t c c a c t g c a a t c c t a t g g t t t t c a g c c a a c c a a t g g c g t g g g c t a c c a a c c a t a c c g c g t c g t g  
35 g t g c t c t c c t t t g a a c t g c t c c a c g c t c c c g c g a c t g t c t g c g g c c c c a a g a a g t c c a c g a a c c t  
t g t g a a g a a t a a g t g c g t t a a t t t t a a t t t c a a c g g c c t c a c t g g a a c a g g a g t g c t c a c t g a g a  
g t a a c a a g a a g t t c c t g c c a t t t t c a a c a a t t t g g c a g a g a c a t a g c c g a t a c t a c t g a c g c c g t t  
a g g g a c c c c c a g a c c c t c g a g a t t c t c g a t a t a a c g c c c t g c t c c t t c g g t g g a g t t t c c g t g a t  
c a c g c c a g g c a c c a a t a c c a g t a a c c a g g t c g c c g t g c t g t a t c a g g a t g t c a a c t g t a c t g a g g  
40 t g c c c g t a g c c a t c c a t g c g g a t c a g c t c a c a c c a a c t t g g a g g t g t a c a g c a c c g g c t c c a a t  
g t a t t c c a g a c t c g g g c g g a t g c c t t a t t g g c c g a a c a c g t g a a c a a t a g t t a c g a a t g c g a  
t a t t c c a a t t g g c g c g g a a t c t g t g t a g t a c c a g a c t c a g a c g a a c t c c c c a g g c a g c g c c a  
g c a g c g t t g c c a g c c a g t c a a t c a t c g c t t a t a c a a t g t c a c t t g g a g c c g a a a a c t c c g t g g c t  
t a c t c a a a c a a c a g c a t c g c c a t c c c c a c a a a c t t c a c c a t a t c c g t g a c a a c t g a g a t t c t g c c  
45 a g t g t c c a t g a c t a a g a c g t c c g t a g a t t g c a c t a t g t a c a t a t g c g g c g a c a g c a c a g a a t g t t  
c t a a t c t g c t g c t g c a a t a t g g a a g c t t c t g c a c t c a a c t g a a c a g a g c g c t c a c a g g c a t c g c c  
g t g g a g c a g g a t a a g a a t a c c a g g a g g t g t t c g c c c a a g t t a a g c a g a t c t a c a a g a c c c c a c c  
c a t a a a g g a t t t c g g t g g a t t c a a t t t t a g t c a g a t a c t c c c a g a c c c a t c t a a g c c a t c c a a g a  
g g a g c c c a t c g a g g a t c t t t t g t t t a a c a a a g t t a c t c t g g c c g a c g c c g g t t t c a t c a a g c a g  
50 t a c g g a g a t t g c c t c g g c g a c a t c g t g c t c g t g a c c t c a t c t g t g c g c a a a a g t t t a a c g g t c t

5 gacggtgctgcctccccctccttactgatgaaatgatcgcccagtataaccagcgcaactcctcgctg  
 gcaccataacatccggttggacattcggcgctgggtgcagcactgcagataccattcgccatgcaa  
 atggcataatcgtttcaacgggtatcggtgtcacacagaatgtcctatatgagaaccagaagctgat  
 cgcaaatcagttcaatagtgccatcggaataatccaggatagcttcagcagcacaccctcagccc  
 ttggcaaaactccaggatgtcgatgaaccagaatgccaggctctcaataccctcgatgaagcagctc  
 tcatctaatttcggcgcaatttcagtggtcctcaacgacatcctcagccgctcgacccccccga  
 ggccgaagtgcagattgacagactgattacaggctgactccagagcctccagacttacgtgactc  
 agcagctgataagagccgagagataagggccagcgctaacctggctgccacaaagatgtctgag  
 10 tgcgtgctgggcccagtcgaagagagtagacttctgtggcaaggctaccatctgatgagcttccc  
 acaatccgcacctcacggcgtagtggttcctccacgtgacatatgtaccggctcaggagaagaatt  
 tcaactaccgctcctgctatatgccatgatggaaaggctcacttccccgggaggggggtgttcgtg  
 tccaacggcaccatttggttggactcagcggaatttctacgaacccagatcataaccactga  
 caacacatttggtgcccgaattgtgacgtggtcattggaatagtgaacaacactgtttatgatc  
 cactgcagccagaacttgacagctttaaggaggagctcgacaagtacttcaagaatcatacgtca  
 15 ccagatgtggacctcggagatattagcgggtatcaatgccagtggttgtaaatcagaaggaaat  
 agaccgccttaatgaggtcgccaaaaatctgaacgagagcctcatcgatcttcaggagctgggca  
 aatatgagcagtacatcaagtggccttggtatatttggttggttcatcgccggcctgatcgcc  
 atagtaatgggtcacaattatgctctctttatggatgtgctccaatggatcggtacaatgcagaat  
 ttgcatttaa

20 **PDI-S Protein GSAS+6P+L923F-DNA (SEQ ID NO: 69)**

atggcgaaaaacggttgcatgttttcggcttattgttttctcttcttgtgttggttccttctcagat  
 cttcgcggtgaatcttacgacgcgaacacagttaccacccgcataatacaaatagcttactcggg  
 25 gtgtttattaccccgacaaagtgttcaggtcctccgtgctccactcaacacaggacctctttctt  
 cctttcttttctaactgacatgggttcatgccattcatgtatccggcactaacggtactaagag  
 gttcgataatcctgtgctccctttcaatgacggcggtttactttgcaagcacagagaagagtaaca  
 tcatccgaggttggatctttggcactaccctcgattcaaagacgcagagcctcctcattgtgaac  
 aatgccactaacgtggtgatcaaagtgttcgaggtttcagttctgcaatgaccctttcttgggggt  
 gtactatcataagaacaacaagtcttggatggaatctgaattccgctctatagcagcgccaaca  
 30 actgcacctttgaatacgtgtcccagcccttcttattggacctggagggaagcagggaaacttt  
 aagaatctgagagagttcggtgtttaaaaatatcgacggctatttttaagatctattctaagcacac  
 gcctattaatctcggtgcgcatcttccacaaggcttcagcgccctggaaccactcggtggacctcc  
 caattggtatcaacatcactagatttcagactctgcttgcctccaccgatcctatctgacaccc  
 ggagactcctctagcggtggtgactgcccggcgctgccgcttattacgttggttatcttcagccacg  
 35 cacgttcctgtgaagtataacgagaatggtactattaccgatgccgtggattgtgcccttgacc  
 ccctgtccgaaactaagtgcacactcaagtcatcactgtggaaaaagggaatctaccagacaagc  
 aattttcggtccagcctactgagagcattgtgctgttccctaacatcacaaatctttgcccctt  
 cggagaggttttcaatgctacacggtttgcctccgtgtatgcctggaaccgcaagagaatttcca  
 attgctggtggcgattactccgtgctctacaatagtgaagcttttagcacctttaagtgtatggc  
 40 gtatcccctactaagcttaacgacttgtgtttcacaaacgtgtatgccgactcctttgtgatacg  
 gggcgacgaagttagacagatagcaccaggacagacgggaaagatagctgactacaactataagc  
 ttcctgatgacttcactggtgctgttccgctggaattctaacaacctggactcaaaagtcggc  
 ggcaactataactatctctatcggtgttccgcaagagtaaccttaagccctttgagagagatat  
 aagcactgaaatctaccaggtggcagtagccctgtaatggcgtggaaggctttaattgttatt  
 45 ttccactgcaatcctatgggttttcagccaaccaatggcgtgggctaccaaccataaccgcgtcgtg  
 gtgctctcctttgaactgctccacgctcccgcgactgtctgcggccccaagaagtccacgaacct  
 tgtgaagaataagtgcgttaatttttaatttcaacggcctcactggaacaggagtgctcactgaga  
 gtaacaagaagtctcctgccatttcaacaatttggcagagacatagccgatactactgacgcggtt  
 agggacccccagaccctcgagattctcgatataacgccctgctccttcggtggagtttccgtgat  
 50 cagccaggccaccaataaccagtaaccaggtcgccgtgctgtatcaggatgtcaactgtactgagg  
 tgcccgtagccatccatgcggatcagctcacaccaacttgagggtgtacagcaccggctccaat

gtattccagactcgggcccggatgccttattggcgccgaacacgtgaacaatagttacgaatgcga  
 tattccaattggcgccggaatctgtgctagctaccagactcagacgaactcccaggcagcgcca  
 gcagcgttgccagccagtcacatcgcttatacaatgtcacttgaggccgaaaactccgtggct  
 tactcaacaacagcatcgccatcccacaaacttcacatcatcgtgacaactgagattctgcc  
 5 agtgtccatgactaagacgtccgtagattgcactatgtacatatgcggcgacagcacagaatggt  
 ctaatctgctgctgcaatatggaagcttctgcactcaactgaacagagcgctcacaggcatcgcc  
 gtggagcaggataagaatacccaggagggtgttcgcccaagttaagcagatctacaagacccccacc  
 cataaaggatttcggtggattcaatttttagtcagatactcccagacccatctaagccatccaaga  
 ggagccccatcgaggatcttttgttaacaaagttaactctggccgacgcgggtttcatcaagcag  
 10 tacggagattgcctcggcgacatcgctgctcgtgacctcatctgtgcgcaaaagttaacggctc  
 gacgggtgctgcctcccctccttactgatgaaatgatcgcccagtataccagcgactcctcgctg  
 gcaccataacatccggttgacattcggcgctggtcccgactgcagataccattccccatgcaa  
 atggcatatcgtttcaacgggtatcggtgtcacacagaatgtcctatatgagaaccagaagctgat  
 cgaaatcagttcaatagtgcctatcggaataatccaggatagcttcagcagcacacctcagccc  
 15 ttggcaaacccaggtatgctgtaaccagaatgccaggctctcaataccctcgtgaagcagctc  
 tcatctaatttcggcgcaatttcagtgctcctcaacgacatcctcagccgcctcgacccccccga  
 ggccgaagtgcagattgacagactgattacaggctgactccagagcctccagacttacgtgactc  
 agcagctgataagagccgcgagataagggccagcgctaacctggctgccacaaagatgtctgag  
 tgctgctgggccagtcgaagagagtagacttctgtggcaaaaggctaccatctgatgagcttccc  
 20 acaatccgcacctcacggcgtagtggtcctccacgtgacatatgtaccggctcaggagaagaatt  
 tcaataccgctcctgctatatgcatgatggaaggctcacttccccgggagggggtgttcgtg  
 tccaacggcacccatttggttgtgactcagcggaatttctacgaaccccagatcataaccactga  
 caacacatttgtgtccggaaattgtgacgtggtcattggaatagtgaacaacactgtttatgatc  
 cactgcagccagaacttgacagctttaaggaggagctcgacaagtacttcaagaatcatacgtca  
 25 ccagatgtggacctcgagatattagcgggtatcaatgccagtggtgtcaatattcagaaggaaat  
 agaccgccttaatgaggtcgccaaaaatctgaacgagagcctcatcgatcttcaggagctgggca  
 aatatgagcagtagcatcaagtggccttggtatatttggcttggttcacgcggcctgatcgcc  
 atagtaattggtcacaattatgctctctttatggatgtgctccaatggatcggtacaatgcagaat  
 ttgcatttaa

30 **PDI-Modified S protein with H5i Hemagglutinin CT (V1) DNA (SEQ ID NO: 70)**

atggcgaaaaacggttgcatgttttcggcttattgttttctcttcttgtgttggttccttctcagat  
 cttcgcggtgaatcttacgacggaacacagttaccaccgcgcatatacaaatagcttcactcggg  
 gtgtttattaccccgacaaagtgttcagggtcctccgtgctccactcaacacaggacctctttctt  
 35 cttttcttttctaacgtgacatgggttcatgccattcatgtatccggcactaacggtactaagag  
 gttcgataatcctgtgctccctttcaatgacggcggtttactttgcaagcacagagaagagtaaca  
 tcatccgaggttggtatctttggcactaccctcgattcaaagacgcagagcctcctcattgtgaac  
 aatgccactaacgtgggtgatcaaagtttgcgagtttcagttctgcaatgacctttcttgggggt  
 gtactatcataagaacaacaagtcttggatggaatctgaattccgctctatagcagcgccaaca  
 40 actgcacctttgaatacgtgtcccagcccttcttattggacctggagggaagcagggaacttt  
 aagaatctgagagagttcgtgttttaaaatatacgacggctattttaagatctattctaagcacac  
 gcctattaatctcgtgcgcgatcttcacaaggcttcagcgccctggaaccactcgtggacctcc  
 caattggtatcaacatcactagatttcagactctgcttgccctccaccgatcctatgtgacccc  
 ggagactcctctagcggctggactgcggcgctgcccgttattacgttggttatcttcagccacg  
 45 cacgttcctgctgaagtataacgagaatggtactattaccgatgccgtggattgtgcccttgacc  
 cctgtccgaaactaagtgcacactcaagtcattcactgtggaaaaaggaatctaccagacaagc  
 aattttcgggtccagcctactgagagcattgtgcgctttcctaacatcacaaatctttgcccctt  
 cggagaggttttcaatgctacacggtttgcctccgtgtatgcctggaaccgcaagagaatttcca  
 attgctgtggccgattactccgtgctctacaatagtgaagcttttagcactttaagtgtatggc  
 50 gtatcccctactaagcttaacgacttgtgtttcacaaacgtgtatgccgactcctttgtgatacg  
 ggcgcgacgaagttagacagatagcaccaggacagacgggaaagatagctgactacaactataagc

5 ttcctgatgacttcactggctgcggttatcgcggtgaattctaacaacctggactcaaaagtgcggc  
 ggcaactataactatctctatcggtgttccgcaagagtaaccttaagccctttgagagagatat  
 aagcactgaaatctaccaggctggcagtagcgccctgtaatggcggtggaaggctttaattgttatt  
 10 ttcactgcaatcctatgggttttcagccaaccaatggcggtgggtaccacaccataaccgctgctg  
 gtgctctcctttgaactgctccacgctcccgcgactgtctgcggccccaagaagtccacgaacct  
 tgtgaagaataagtgcgttaatttttaatttcaacggcctcactggaacaggagtgtcactgaga  
 gtaacaagaagtccctgccatttcaacaatttggcagagacatagccgatactactgacgcccgtt  
 agggacccccagaccctcgagattctcgatataacgcccgtgctccttcggtggagtttccggtgat  
 15 cagccagggaccaataaccagtaaccagggtcgccgtgctgtatcaggatgtcaactgtactgagg  
 tgcccgtagccatccatgcggatcagctcacaccaacttggagggtgtacagcaccggctccaat  
 gtattccagactcgggcccggatgccttatttggcgccgaacacgtgaacaatagttaacgaatgcga  
 tattccaatttggcgccggaatctgtgctagctaccagactcagacgaactcccaggcagcgcca  
 gcagcgttgccagccagtcaatcatcgcttatacaatgtcacttggagccgaaaactccgtggct  
 tactcaaaacaacagcatcgccatccccacaaacttcaccatatccgtgacaactgagattctgcc  
 20 agtgtccatgactaagacgtccgtagattgcactatgtacatatgcgcgacagcacagaatgtt  
 ctaatctgctgctgcaatatggaagcttctgcaactgaacagagcgctcacaggcatcgcc  
 gtggagcaggataagaatacccaggagggtgttcgcccaggttaagcagatctacaagacccccacc  
 cataaaggatttccgtggattcaatttttagtcagatactcccagaccatctaagccatccaaga  
 ggagctttatcgaggatcttttgtttaacaaagttaactctggccgacgcccgtttcatcaagcag  
 25 tacggagattgcctcggcgacatcgctgctcgtagacctcatctgtgcgcaaaagttaacggctc  
 gacggtgctgcctccccctccttactgatgaaatgatcgccagtataccagcgactcctcgctg  
 gcaccataaacatccggttggacattcggcgctgggtgcagcactgcagataccattcgccatgcaa  
 atggcatatcgtttcaacgggtatcggtgtcacacagaatgtcctatatgagaaccagaagctgat  
 cgcaaatcagttcaatagtgcctatcggaataatccaggatagccttagcagcacagcctcagccc  
 30 ttggcaaaactccaggatgtcgtgaaccagaatgccaggctctcaataccctcgtgaagcagctc  
 tcatctaatttccggcgcaatttccagtgctcctcaacgacatcctcagccgctcgacccccccga  
 ggccgaagtgcagattgacagactgattacagggtcgactccagagcctccagacttacgtgactc  
 agcagctgataagagccgcccagagataagggccagcgctaacctggctgccacaaagatgtctgag  
 tgctgctgggcccagtcgaagagagtagacttctgtggcaaaaggctaccatctgatgagcttccc  
 35 acaatccgcacctcacggcgtagtggttccctccacgtgacatatgtaccggctcaggagaagaatt  
 tcaactaccgctcctgctatatgcatgatggaagggtcacttcccccgaggggggtgttcgtg  
 tccaacggcacccattgggttgtgactcagcggaatttctacgaacccagatcataaccactga  
 caacacatttgtgtccggaaattgtgacgtggtcatttggaaatagtgaacaacactgtttatgatc  
 cactgcagccagaacttgacagctttaaggaggagctcgacaagtacttcaagaatcatacgtca  
 40 ccagatgtggacctcgagatattagcgggtatcaatgccagtggtgtcaatatcagaaggaaat  
 agaccgccttaatgaggtcgccaaaaatctgaacgagagcctcatcgatcttcaggagctgggca  
 aatatgagcagtagcatcaagtggccttggtatatttggcttggcttcatcgccggcctgatcgcc  
 atagtaatggtcacaattatgatggccggcctctctttatggatgtgctccaatggatcggttaca  
 atgcagaatttgcatttaa

40 **PDI-Modified S protein with H5i Hemagglutinin CT (V2) DNA (SEQ ID NO: 71)**

45 atggcgaaaaacggttgcgattttcggttattgttttctcttcttgtgttggttccttctcagat  
 cttcgcggtgaatcttacgacgcgaacacagttaccaccgcgcatatacaaatagcttcactcggg  
 gtgtttattacccccgacaaagtgttcagggtcctccgtgctccactcaacacaggacctctttctt  
 cctttcttttctaactgacatgggttcatgccattcatgtatccggcactaacggtactaagag  
 50 gttcgataatcctgtgctccctttcaatgacggcgtttactttgcaagcacagagaagagtaaca  
 tcatccgaggttggatctttggcactaccctcgattcaaagacgcagagcctcctcattgtgaac  
 aatgccactaacgtgggtgatcaaagtttgcgagtttcagttctgcaatgaccctttcttgggggt  
 gtactatcataagaacaacaagtcttggatggaatctgaattccgctctatagcagcgccaaca  
 actgcacctttgaatacgtgtccagcccttcttattggacctggagggaaagcagggaacttt  
 aagaatctgagagaggttcgtgttttaaaaatcgcagcgctatttttaagatctattctaagcacac

gcctattaatctcgtgcgcgatcttccacaaggcttcagcgccttgaaccactcgtggacctcc  
 caattggtatcaacatcactagatttcagactctgcttgcctccaccgatcctatctgacaccc  
 ggagactcctctagcggctggactgccggcgctgccgcttattacgttggttatcttcagccacg  
 caggttcctgctgaagtataacgagaatggtactattaccgatgccgtggattgtgcccttgacc  
 5 cctgtccgaaactaagtgcacactcaagtcattcactgtggaaaaaggaatctaccagacaagc  
 aattttcgggtccagcctactgagagcattgtgcttccctaacatcacaaatctttgcccctt  
 cggagaggttttcaatgctacacgggttgcctccgtgtatgcctggaaccgcaagagaatttcca  
 attgctggtggcgattactccgtgctctacaatagtgaagcttttagcacctttaagtgtatggc  
 gtatcccctactaagcttaacgacttgtgtttcacaaacgtgtatgccgactcctttgtgatacg  
 10 gggcgacgaagttagacagatagcaccaggacagacgggaaagatagctgactacaactataagc  
 ttcttgatgacttcactggctgcttattcgctggaattctaacaacctggactcaaaagtcggc  
 ggcaactataactatctctatcggtgttccgcaagagtaaccttaagccctttgagagagatat  
 aagcactgaaatctaccaggtggcagtagcctgtaatggcgtggaaggtttaattgttatt  
 ttccactgcaatcctatgggttttcagccaaccaatggcgtgggctaccaaccataaccgctcgtg  
 15 gtgctctcctttgaactgctccacgctcccgcgactgtctgcggccccaagaagtcacgaacct  
 tgtgaagaataagtgcgttaatttttaatttcaacggcctcactggaacaggagtgctcactgaga  
 gtaacaagaagttcctgccatttcaacaatttggcagagacatagccgatactactgacgccgtt  
 agggacccccagaccctcgagattctcgatataacgcctgctccttcggtggagtttccgtgat  
 cagccagggcaccaataaccagtaaccaggtcgccgtgctgtatcaggatgtcaactgtactgagg  
 20 tgcccgtagccatccatgcggtacagctcacaccaacttgagggtgtacagcaccggctccaat  
 gtattccagactcgggcgggatgccttattggcgccgaacacgtgaacaatagttacgaatgca  
 tattccaattggcgccggaatctgtgctagctaccagactcagacgaactcccaggcagcgcca  
 gcagcgttgccagccagtcattcatcgcttatacaatgtcacttggagccgaaaaactccgtggct  
 tactcaaaacaacagcatcgccatccccacaaacttcaccatatccgtgacaactgagattctgcc  
 25 agtgtccatgactaagacgtccgtagattgcactatgtacatatgcggcgacagcacagaatggt  
 ctaatctgctgctgcaatatggaagcttctgcaactgaacagagcgctcacaggcatcgcc  
 gtggagcaggataagaatacccaggaggtgttcgcccaagttaagcagatctacaagacccccacc  
 cataaaggatttccgtggattcaatttttagtcagatactcccagaccatctaagccatccaaga  
 ggagctttatcgaggatcttttgtttaacaaagttactctggccgacgcccgtttcatcaagcag  
 30 tacggagattgcctcggcgacatcgctgctcgtgacctcatctgtgcgcaaaagttaaagggtct  
 gacggtgctgcctcccctccttactgatgaaatgatcgcccagtataccagcgcactcctcgctg  
 gcaccataacatccggttgacattcggcgctgggtgcagcactgcagataccattcgccatgcaa  
 atggcatatcggttcaacggatcggtgtcacacagaatgtcctatatgagaaccagaagctgat  
 cgaaatcagttcaatagtgccatcggaaaaaatccaggatagccttagcagcacagcctcagccc  
 35 ttggcaaaactccaggatgtcgtgaaccagaatgccagggtctcaataccctcgtgaagcagctc  
 tcatctaatttcggcgcaatttccagtgctcctcaacgacatcctcagccgcctcgaccccccgga  
 ggccgaagtgcagattgacagactgattacaggtcgactccagagcctccagacttacgtgactc  
 agcagctgataagagccgcccagataagggccagcgtaacctggctgccacaaagatgtctgag  
 tgcgtgctgggcccagtcgaagagtagacttctgtggcaaggctaccatctgatgagcttccc  
 40 acaatccgcacctcacggcgtagtgcttccacgtgacatatgtaccggctcaggagaagaatt  
 tcaactaccgctcctgctatatgcatgatggaagggtcacttcccccgagggggtgttcgtg  
 tccaacggcacccattgggttgtgactcagcggaatttctacgaacccagatcataaccactga  
 caacacatttgtgtccggaattgtgacgtggtcattggaatagtgaaacaactgtttatgatc  
 cactgcagccagaacttgacagctttaaggaggagctcgacaagtacttcaagaatcatacgtca  
 45 ccagatgtggacctcgagatattagcggtatcaatgccagtggtgtcaatattcagaaggaaat  
 agaccgccttaatgaggtcgccaaaaatctgaacgagagcctcatcgatcttcaggagctgggca  
 aatatgagcagtagacatcaagtggccttggtatatttggcttggcttcatcgccggcctgatcgcc  
 atagtaatggtcacaattatggccggcctctctttatggatgtgctccaatggatcggttacaatg  
 cagaatttgcatttaa

50 PDI-Modified S protein with H5i Hemagglutinin CT (V3) DNA (SEQ ID NO: 72)

atggcgaaaaacgttgcgatttttcggcttattgttttctcttcttgtgttggttccttctcagat  
cttcgcggtgaatcttacgcgcgaacacagttaccacccgcataatacaaatagcttactcggg  
gtgtttattaccccgacaaagtgttcaggtcctccgtgctccactcaacacaggacctcttctt  
cctttcttttctaactgacatggtttcatgccattcatgtatccggcactaacggtactaagag  
5 gttcgataatcctgtgctccctttcaatgacggcggtttactttgcaagcacagagaagagtaaca  
tcacccgaggttggatctttggcactaccctcgattcaaagacgcagagcctcctcattgtgaac  
aatgccactaacgtggtgatcaaagtttgcgagtttcagttctgcaatgacccttcttgggggt  
gtactatcataagaacaacaagtcttggatggaatctgaattccgcgtctatagcagcgccaaca  
actgcacctttgaatacgtgtcccagcccttcttctatggacctggagggaagcagggaaacttt  
10 aagaatctgagagagttcgtgttataaaatatcgacggctattttaagatctattctaagcacac  
gcctattaatctcgtgcgcgatcttccacaaggcttcagcgccctggaaccactcgtggacctcc  
caattggtatcaacatcactagatttcagactctgcttgcctccaccgatcctatctgacaccc  
ggagactcctctagcggtggtgactgccggcgctgccgttattacgttgggtatcttcagccacg  
cacgttcctgctgaagtataacgagaatggtactattaccgatgccgtggattgtgcccttgacc  
15 cctgtccgaaactaagtgcacactcaagtcattcactgtggaaaaaggaatctaccagacaagc  
aattttcgggtccagcctactgagagcattgtgcgctttcctaacatcacaaatcttggccctt  
cggagaggttttcaatgctacacggtttgcctccgtgtatgcctggaaccgcgaagagaatttcca  
attgcgtggccgattactccgtgctctacaatagtgaagcttttagcacctttaagtgtatggc  
gtatcccctactaagcttaacgacttgtgtttcacaaacgtgtatgccgactcctttgtgatacg  
20 gggcgacgaagttagacagatagcaccaggacagacgggaaagatagctgactacaactataagc  
ttcctgatgacttcactggctgcttctcgcgtggaattctaacaacctggactcaaaagtcggc  
ggcaactataactatctctatcggtgttccgcaagagtaaccttaagccctttgagagagatat  
aagcactgaaatctaccaggtggcagtagccctgtaatggcgtggaaggctttaattgttatt  
ttccactgcaatcctatggttttcagccaaccaatggcgtgggctaccaaccataaccgcgtcgtg  
25 gtgctctcctttgaactgctccacgctcccgcgactgtctgcggccccaagaagtccacgaacct  
tgtgaagaataagtgcgttaatttttaatttcaacggcctcactggaacaggagtgtcactgaga  
gtaacaagaagtccctgccatttcaacaatttggcagagacatagccgatactactgacgcggt  
agggacccccagaccctcgagatttctcgatataacgccctgctccttcggtggagtttccgtgat  
cacgccaggcaccaataaccagtaaccaggtcgccgtgctgtatcaggatgtcaactgtactgagg  
30 tgcccgtagccatccatgcggtacagctcacaccaacttgagggtgtacagcacccggtccaat  
gtattccagactcgggcccggatgccttatttggcgccgaacacgtgaacaatagttacgaatgcga  
tattccaattggcgccggaatctgtgctagctaccagactcagacgaactccccaggcgccga  
cgacggttgccagccagtcattcatcgcttatacaaatgtcacttggagccgaaaactccgtggct  
tactcaaaacaacagcatcgccatccccacaaacttcaccatatccgtgacaactgagattctgcc  
35 agtgtccatgactaagacgtccgtagattgcactatgtacatatgcggcgacagcacagaatgtt  
ctaactctgctgctgcaatatggaagcttctgcactcaactgaacagagcgctcacaggcatcgcc  
gtggagcaggataagaataccagaggtgttccgccaagttaagcagatctacaagacccccacc  
cataaaggatttcggtggattcaatttttagtcagatactcccagaccatctaagccatccaaga  
ggagctttatcgaggatcttttgtttaacaaagttactctggccgacgcgggtttcatcaagcag  
40 tacggagattgcctcggcgacatcgctgctcgtgacctcatctgtgcgcaaaagtttaacggtct  
gacggtgctgcctcccctccttactgatgaaatgatcgcccagatataccagcgcaactcctcgtg  
gcaccataacatccggttgacattcggcgctggtgcagcactgcagataccattcgccatgcaa  
atggcatatcgtttcaacggtatcggtgtcacacagaatgtcctatagagacacagaagctgcc  
cgaaatcagttcaatagtccatcggaataatccaggatagccttagcagaccagcctcagccc  
45 ttggcaaaactccaggatgtcgtgaaccagaatgccaggctctcaataccctcgtgaagcagctc  
tcactaatttcggcgcaatttccagtgctcctcaacgacatcctcagccgctcgaccccccgga  
ggccgaagtgcagattgacagactgattacaggtcgactccagagcctccagacttacgtgactc  
agcagctgataagagccgcccagataagggccagcgctaacctggctgccacaaagatgtctgag  
tgcgtgctgggcccagccaagagagtagacttctgtggcaaaaggctaccatctgatgagcttccc  
50 acaatccgcacctcacggcgtagtgcttccacgtgacatatgtaccggtcaggagaagaatt  
tactaccgctcctgctatatgccatgatggaaggctcacttccccggggagggggtgttcgtg  
tccaacggcacccattggtttgtgactcagcggaaatttctacgaacccagatcataaccactga

5 caacacatttgtgtccggaattgtgacgtgggtcattggaatagtgaacaacactgtttatgata  
 cactgcagccagaacttgacagctttaaggaggagctcgacaagtacttcaagaatcatacgtca  
 ccagatgtggacctcgagatattagcggatcaatgccagtggtgtcaatattcagaaggaaat  
 agaccgccttaatgaggtcgccaaaaatctgaacgagagcctcatcgatcttcaggagctgggca  
 aatatgagcagtacatcaagtggccttggtatatttggcttggcttcacgcggcctgatcgcc  
 atagtaatggtcacaattatgctctgctgcatgtgctccaatggatcggtacaatgcagaatttg  
 catttaa

**PDI-Modified S protein with H5i Hemagglutinin CT (V4) DNA (SEQ ID NO: 73)**

10 atggcgaaaaacgttgcgattttcggttattgttttctcttcttgtgttggttccttctcagat  
 ctctcggtgaatcttacgacgcgaacacagttaccaccgcgcatatacaaatagcttactcggg  
 gtgtttattaccccgacaaagtgttcaggtcctcggtgctccactcaacacaggacctcttctt  
 ccttcttttctaactgacatgggttcatgccattcatgtatccggcactaacggtactaagag  
 gtctgataatcctgtgctcccttcaatgacggcggttactttgcaagcacagagaagagtaaca  
 15 tcatccgaggttggatctttggcactaccctcgattcaaagacgcagagcctcctcattgtgaac  
 aatgccactaacgtgggtgatcaaagtttgcgagtttcagttctgcaatgacccttcttgggggt  
 gtactatcataagaacaacaagtcttggatggaatctgaattccgctctatagcagcgccaaca  
 actgcacctttgaatacgtgtcccagcccttcttctatggacctggaggaaagcagggaacttt  
 aagaatctgagagagttcgtgtttaaaaatatcgacggctattttaagatctattctaagcacac  
 20 ccctattaatctcgtgcgcgatcttccacaaggcttcagcgccctggaaccactcgtggacctcc  
 caattggtatcaacatcactagatttcagactctgcttgcctccaccgatcctatctgacaccc  
 ggagactcctctagcggtggaactgcccgcgtgcccgttattacgttgggtatcttcagccacg  
 cacgttctgctgaagtataacgagaatggtactattaccgatgccgtggattgtgcccttgacc  
 cctgtccgaaactaagtgcacactcaagtcattcactgtgaaaaaggaatctaccagacaagc  
 25 aatttccgggtccagcctactgagagcattgtgcgcttctcctaactcacaaatcttggccctt  
 cggagaggttttcaatgctacacggtttgcctccgtgtatgcctggaaccgcgaagagaatttcca  
 attgctggtggcgttactccgtgctctacaatagtgaagcttttagcacctttaagtgtatggc  
 gtatcccctactaagcttaacgacttgtgtttcacaaacgtgtatgccgactccttctgtgatacg  
 gggcgacgaagttagacagatagcaccaggacagacgggaaagatagctgactacaactataagc  
 30 ttctgtgacttactggtgctgttccgctggaattctaacaacctggactcaaaagtgcgc  
 ggcaactataactatctctatcggctgttccgcaagagtaaccttaagcccttggagagagatat  
 aagcactgaaatctaccaggtggcagtagcgcctgtaatggcgtggaaggtttaattgttatt  
 ttccactgcaatcctatggttttcagccaaccaatggcgtgggctaccaaccataaccgcgtcgtg  
 gtgctctccttgaactgctccacgctcccgcgactgtctgcggccccaagaagtccacgaacct  
 35 tgtgaagaataagtgcgttaatttttaatttcaacggcctcactggaacaggagtgtcactgaga  
 gtaacaagaagtctcctgccatttcaacaatttggcagagacatagccgatactactgacgcggtt  
 agggacccccagaccctcgagattctcgatataacgcctgctcctcgttgaggttccgtgat  
 cagccaggcaccaataaccagtaaccaggtgcgcgtgctgtatcaggatgtcaactgtactgagg  
 tgcccgtagccatccatgaggatcagctcacaccaacttggagggtgtacagcacccggtccaat  
 40 gtattccagactcgggccggatgccttattggcgccgaacacgtgaacaatagttacgaatgcga  
 tattccaattggcgccggaatctgtgctagctaccagactcagacgaactcccaggcagcgcca  
 gcagcgttggcagccagtcacatcgcttatacaatgtcacttggagcgaaaaactccgtggct  
 tactcaacaacagcatcgccatccccacaaacttcaccatatccgtgacaactgagattctgccc  
 agtgtccatgactaagacgtccgtagattgcactatgtacatatgcggcgacagcacagaatggt  
 45 ctaatctgctgctgcaatatggaagcttctgcactcaactgaacagagcgctcacaggcatcgcc  
 gtggagcaggataagaataaccaggaggtgttcgcccagtttaagcagatctacaagacccccacc  
 cataaaggatttccgtggattcaatttttagtcagatactcccagacccatctaagccatccaaga  
 ggagctttatcgaggatcttttgtttaacaaagttactctggccgacgcgggtttcatcaagcag  
 tacggagattgcctcggcgacatcgctgctcgtgacctcatctgtgcgcaaaagtttaacggctc  
 50 gacgggtgctgcctcccctccttactgatgaaatgatcgccagctataaccagcgccactcctcgctg  
 gcaccataacatccggttggacattcggcgctgggtgcagcactgcagataccattcgccatgcaa

atggcatatcgtttcaacgggtatcggtgtcacacagaatgtcctatatgagaaccagaagctgat  
cgcaaatcagttcaatagtgccatcggaaaaatccaggatagccttagcagcacagcctcagccc  
ttggcaaacctccaggatgtcgtgaaccagaatgccaggctctcaataccctcgtgaagcagctc  
tcatctaatttcggcgcaatttccagtgctcctcaacgacatcctcagccgcctcgaccccccgga  
5 ggccgaagtgcagattgacagactgattacaggtcgactccagagcctccagacttacgtgactc  
agcagctgataagagccgcgagataagggccagcgctaacctggctgccacaaagatgtctgag  
tgctgtgtgggccagtcgaagagagtagacttctgtggcaaaggctaccatctgatgagcttccc  
acaatccgcacctcacggcgtagtggtcctccacgtgacatatgtaccggctcaggagaagaatt  
tactaccgctcctgctatatgccatgatggaaaggctcacttccccgggagggggtgttcgtg  
10 tccaacggcacccattgggttgtagctcagcggaatttctacgaacccagatcataaccactga  
caacacatttgtgtccggaaattgtgacgtggctcattggaatagtgaacaacactgtttatgatc  
cactgcagccagaacttgacagctttaaggaggagctcgacaagtacttcaagaatcatacgtca  
ccagatgtggacctcgagatattagcgggtatcaatgccagtggtgtcaatattcagaaggaaat  
agaccgccttaatgaggtcgccaaaaatctgaacgagagcctcatcgatcttcaggagctgggca  
15 aatatgagcagtagcatcaagtggccttggtatatttggcttggcttcacgcggcctgatcgcc  
atagtaatggtcacaattatgctctgctgctccaatggatcgttacaatgcagaatttgcattta  
a

**PDI-S-protein+H1 Cal DNA (SEQ ID NO: 74)**

atggcgaaaaacggttgcgattttcggcttattgttttctcttcttgtgttggttccttctcagat  
20 ctctcgcggtgaatcttacgacgcgaacacagttaccacccgcataatacaaatagcttcactcggg  
gtgtttattaccccgacaaagtgttcagggtcctccgtgctccactcaacacaggacctcttctt  
cctttcttttctaacgtgacatgggttcatgccattcatgtatccggcactaacggtactaagag  
gttcgataaatcctgtgctccctttcaatgacggcggttactttgcaagcacagagaagagtaaca  
tcatccgaggttggatctttggcactaccctcgattcaaagacgcagagcctcctcattgtgaac  
25 aatgccactaacgtgggtgatcaaagtttgcgagtttcagttctgcaatgaccctttcttgggggt  
gtactatcataagaacaacaagtcttggatggaatctgaattccgcgtctatagcagcgccaaca  
actgcacctttgaatacgtgtcccagcccttcttatggacctggagggaaagcagggaacttt  
aagaatctgagagagttcgtgtttaaaaatatcgacggctatttttaagatctattctaagcacac  
gcctattaatctcgtgcgcatcttccacaaggcttcagcgccctggaaccactcgtggacctcc  
30 caattgggtatcaacatcactagatttcagactctgcttgcctccaccgatcctatctgacaccc  
ggagactcctctagcggctggactgccggcgctgccgcttattacgttggttatcttcagccacg  
cacgttcctgctgaagtataacgagaatggtactattaccgatgcggtggattgtgcccttgacc  
ccctgtccgaaactaagtgcacactcaagtcattcactgtggaaaaaggaatctaccagacaagc  
aattttcgggtccagcctactgagagcattgtgctgttccctaacatcacaatatctttgccctt  
35 cggagaggttttcaatgctacacgggttgcctccgtgtatgcctggaaccgcaagagaatttcca  
attgctgtggccgattactccgtgctctacaatagtgaagcttttagcacctttaagtgtatggc  
gtatccctactaagcttaacgacttgtgtttcacaaacgtgtatgccgactcctttgtgatacg  
ggcgacgaagttagacagatagcaccaggacagacgggaaagatagctgactacaactataagc  
ttcctgatgacttcactggctgctgttatcgctggaattctaacaacctggactcaaaagtccgc  
40 ggcaactataactatctctatcggctgttccgcaagagtaaccttaagccctttgagagagatat  
aagcactgaaatctaccaggtggcagtagccctgtaatggcggtgaaggctttaattgttatt  
ttccactgcaatcctatgggtttcagccaaccaatggcggtgggtaccaaccataaccgcgtcgtg  
gtgctctcctttgaactgtccacgctcccgactgtctgcggcccaagaagtccacgaacct  
tgtgaagaataagtgcgttaatttttaatttcaacggcctcactggaacaggagtgctcactgaga  
45 gtaacaagaagtctcctgccatttcaacaatttggcagagacatagccgatactactgacgccgtt  
agggacccccagaccctcgagattctcgatataacgcctgtccttcggtggagtttccgtgat  
cacgccaggccaataaccagtaaccaggtcgccgtgctgtatcaggatgtcaactgtactgagg  
tgcccgtagccatccatgcggtacagctcacaccaacttggagggtgtacagcaccggctccaat  
gtattccagactcgggcccggatgccttatttggcgccgaacacgtgaacaatagttacgaatgcga  
50 tattccaattggcgccggaatctgtgctagtaccagactcagacgaactccccaggcagcgcca

gcagcgttgccagccagtcgaatcatcgcttatacaatgtcacttggagccgaaaaactccgtggct  
tactcaaacaacagcatcgccatccccacaaacttcaccatatccgtgacaactgagattctgcc  
agtgtccatgactaagacgtccgtagattgcactatgtacatatgcggcgacagcacagaatggt  
ctaactctgctgctgcaatatggaagcttctgcactcaactgaacagagcgctcacaggcatcgcc  
5 gtggagcaggataagaataccaggaggtgttcgcccagttaagcagatctacaagacccccacc  
cataaaggatttccgtggattcaatttttagtcagatactcccagacccatctaagccatccaaga  
ggagctttatcgaggatcttttgtttaacaaagttaactctggccgacgcccgtttcatcaagcag  
tacggagattgcctcggcgacatcgctgctcgtagacctcatctgtgcgcaaaagtttaacggctc  
gacggtgctgcctccccctccttactgatgaaatgatcgcccagtataccagcgcactcctcgctg  
10 gcaccataacatccggttggacattcggcgctgggtgcagcactgcagataccattcgccatgcaa  
atggcatatcgtttcaacgggtatcggtgtcacacagaatgtcctatatgagaaccagaagctgat  
cgcaaatcagttcaatagtgccatcggaataatccaggatagccttagcagcacagcctcagccc  
ttggcaaaactccaggatgtcgtgaaccagaatgccaggctctcaataccctcgtgaagcagctc  
tcatctaatttccggcgcaatttccagtgctcctcaacgacatcctcagccgcctcgacccccccga  
15 ggccgaagtgcagattgacagactgattacaggtcgactccagagcctccagacttacgtgactc  
agcagctgataagagccgagagataagggccagcgctaacctggctgccacaaagatgtctgag  
tgcgtgctgggcccagtcgaagagagtagacttctgtggcaaaaggctaccatctgatgagcttccc  
acaatccgcacctcacggcgtagtggtcctccacgtgacatatgtaccggctcaggagaagaatt  
tactaccgctcctgctatatgccatgatggaaaggctcacttcccccgaggggggtgttcgtg  
20 tccaacggcacccattgggttgtgactcagcgggaatttctacgaacccagatcataaccactga  
caacacatttgtgtccggaattgtgacgtgggtcattggaatagtgaacaacactgtttatgac  
cactgcagccagaacttgacagctttaaggaggagctcgacaagtacttcaagaatcatacgtca  
ccagatgtggacctcgagatattagcggatcaatgccagtggtgtcaatattcagaaggaaat  
agaccgccttaatgaggtcgccaaaaatctgaacgagagcctcatcgatcttcaggagctgggca  
25 aatatgagcagtagacatcaagtggccttgggtatatttggcttggcttcatcgccggcctgatcgcc  
atagtaatggtcacaattatgctcagcttctggatgtgctctaattgggtctctacagtgtagaat  
atgtatttaa

PDI-S-protein+H3 Minn DNA (SEQ ID NO: 75)

atggcgaaaaacggttgcgattttcggttattgttttctcttcttctgtgttggttccttctcagat  
30 cttcgcggtgaatcttacgacgcgaacacagttaccaccgcataatacaaatagcttcactcggg  
gtgtttattacccccgacaaagtgttcagggtcctccgtgctccactcaacacaggacctctttctt  
cctttcttttctaactgacatggtttcatgccattcatgtatccggcactaacggtactaagag  
gttcgataaatcctgtgctccctttcaatgacggcggtttactttgcaagcacagagaagagtaaca  
tcatccgaggttggatctttggcactaccctcgattcaaagacgcagagcctcctcattgtgaac  
35 aatgccactaacgtggtgatcaaagtgttcgaggtttcagttctgcaatgacctttcttgggggt  
gtactatcataagaacaacaagtcttggatggaatctgaattccgcgtctatagcagcgccaaca  
actgcacctttgaatacgtgtcccagcccttcccttatggacctggagggaagcagggaaacttt  
aagaatctgagagagttcgtgtttaaaaatatcgacggctattttaagatctattctaagcacac  
gcctattaatctcgtgcgcatcttccacaaggcttcagcgccctggaaccactcgtggacctcc  
40 caattggtatcaacatcactagatttcagactctgcttgcctccaccgatcctatctgacaccc  
ggagactcctctagcggctggactgccggcgctgccgcttattacgttgggtatcttcagccacg  
cacgttcctgctgaagtataacgagaatgggtactattaccgatgccgtggattgtgaccttgacc  
cctgtccgaaaactaagtgcacactcaagtcattcactgtggaaaaaggaatctaccagacaagc  
aattttcgggtccagcctactgagagcattgtgctgttcttaacatcacaaatctttgcccctt  
45 cggagaggttttcaatgctacacgggttgcctccgtgtatgcctggaaccgcaagagaatttcca  
attgctgtggccgattactccgtgctctacaatagtgaagcttttagcacctttaagtgtatggc  
gtatccccactaagcttaacgacttgtgtttcacaaacgtgtatgccgactcctttgtgatac  
ggcgacgaagttagacagatagcaccaggacagacgggaaagatagctgactacaactataagc  
ttcctgatgacttcaactggtgcttgcgtggaattctaacaacctggactcaaaagtgcggc  
50 ggcaactataactatctctatcggtgttccgcaagagtaaccttaagccctttgagagagatat

aagcactgaaatctaccaggetggcagtagccctgtaatggcgtggaaggctttaattgttatt  
 ttccactgcaatcctatgggttttcagccaaccaatggcgtgggctaccaaccataaccgcgtcgtg  
 gtgctctcctttgaactgctccacgctcccgcgactgtctgcggcccaagaagtccacgaacct  
 5 tgtgaagaataagtgcgttaatttttaatttcaacggcctcactggaacaggagtgtcactgaga  
 gtaacaagaagtccctgccatttcaacaatttggcagagacatagccgatactactgacgcggtt  
 agggacccccagaccctcgagattctcgatataacgcctgtccttcggtggagtttccggtgat  
 cagccaggcaccataaccagtaaccagggtcgccgtgctgtatcaggatgtcaactgtactgagg  
 tgcccgtagccatccatgaggatcagctcacaccaacttggagggtgtacagcaccgggtccaat  
 10 gtattccagactcgggcccggatgccttattggcgccgaacacgtgaacaatagttacgaatgcga  
 tattccaattggcgccggaatctgtgctagctaccagactcagacgaaactccccaggcagcgcca  
 gcagcgttgccagccagtcaatcatcgcttatacaatgtcacttggagccgaaaactccgtgggt  
 tactcaaacacagcatcgccatccccacaaacttcaccatatccgtgacaactgagattctgcc  
 agtgtccatgactaagacgtccgtagattgactatgtacatatgcggcgacagcacagaatgtt  
 15 ctaatctgctgctgcaatatggaagcttctgactcaactgaacagagcgtcacaggcatcgcc  
 gtggagcaggataagaatacccaggaggtgttcgccaagttaagcagatctacaagacccccacc  
 cataaaggatttcggtggattcaatttttagtcagatactcccagaccatctaagccatccaaga  
 ggagctttatcgaggatcttttgtttaacaaagttaactctggccgacgcccgtttcatcaagcag  
 tacggagattgcctcgccgacatcgctgctcgtgacctcatctgtgcgcaaaagttaaagggtct  
 20 gacggtgctgcctccccctccttactgatgaaatgatcgccagtataccagcgcactcctcgctg  
 gcaccataacatccggttgacattcgccgctggtgcagcactgcagataccattcgccatgcaa  
 atggcataatcggttcaacgggtatcggtgtcacacagaatgtcctatatgagaaccagaagctgat  
 cgcaaatcagttcaatagtgccatcggaataatccaggatagccttagcagcacagcctcagccc  
 ttggcaaaactccaggatgtcgtgaaccagaatgccagggtctcaataccctcgtgaagcagctc  
 25 tcatctaatttcggcgcaatttccagtgctcctcaacgacatcctcagccgcctcgacccccccga  
 ggccgaagtgcagattgacagactgattacagggtcgactccagagcctccagacttacgtgactc  
 agcagctgataagagccgcccagagataagggccagcgctaacctggctgccacaaagatgtctgag  
 tgcgtgctgggcccagtcgaagagagtagacttctgtggcaaggctaccatctgatgagcttccc  
 acaatccgcacctcacggcgtagtgctcctccacgtgacatatgtaccggctcaggagaagaatt  
 30 tcaactaccgctcctgctatatgccatgatggaaaggctcacttcccccgaggagggtgttcgtg  
 tccaacggcaccatttggttgtgactcagcgggaatttctacgaacccagatcataaccactga  
 caacacatttgtgtccggaaattgtgacgtggtcatttggaatagtgacaacactgtttatgatc  
 cactgcagccagaacttgacagctttaaaggaggagctcgacaagtaacttcaagaatcatacgtca  
 ccagatgtggacctcgagatttagcggtatcaatgccagtggttgtaaatcagaaggaaat  
 35 agaccgccttaatgaggtcgccaaaaatctgaacgagagcctcatcgatcttcaggagctgggca  
 aatatgagcagtagatcaagtgcccttggtatatttggcttggcttcatcgccggcctgatcgcc  
 atagtaatggtcacaattatgctcatgtgggcctgtcagaagggaacatcagatgcaacatctg  
 catctaa

# PDI-S-protein+H6 HK DNA (SEQ ID NO: 76)

atggcgaaaaacgttgcgattttcggttattgttttctcttcttgtgttggttccttctcagat  
 40 cttcgcggtgaatcttacgacgcgaacacagttaccacccgcatatacaaatagcttcactcggg  
 gtgtttattacccccgacaaagtgttcaggtcctccgtgctccactcaacacaggacctcttctt  
 ctttcttttctaacgtgacatgggttcatgccattcatgtatccggcactaaccggtactaagag  
 gttcgataatcctgtgctcccttcaatgacggcggttactttgcaagcacagagaagagtaaca  
 45 tcatccgaggttggtatcttggcactaccctcgattcaaagacgcagagcctcctcattgtgaac  
 aatgccactaacgtggtgatcaaagtgttcgaggttccagttctgcaatgacccttcttgggggt  
 gtactatcataagaacaacaagtcttggtggaatctgaattccgctctatagcagcgccaaca  
 actgcacctttgaatacgtgtccagcccttcttatggacctggagggaagcagggaaacttt  
 aagaatctgagagagttcgtgtttaaaaatatcgacggctatttttaagatctattctaagcacac  
 50 gcctattaatctcgtgcgcgatcttccacaaggcttcagcgccctggaaccactcgtggacctcc  
 caattggtatcaacatcactagatttccagactctgcttgcctccaccgatcctatctgacaccc

5 ggagactcctctagcgggtggaactgccggcgctgccgcttattacgttggttatcttcagccacg  
cacgttcctgctgaagtataacgagaatggtactattaccgatgccgtggattgtgcccttgacc  
cctgtccgaaactaagtgcacactcaagtcattcactgtggaaaaaggaatctaccagacaagc  
aattttcggtccagcctactgagagcattgtgcgctttcctaacatcacaatctttgcccctt  
cggagaggttttcaatgctacacggtttgcctccgtgtatgcctggaaccgcaagagaatttcca  
attgcgtggccgattactccgtgctctacaatagtgaagcttttagcacctttaagtgtatggc  
gtatcccctactaagcttaacgacttgtgtttcacaaacgtgtatgccgactcctttgtgatacg  
ggcgacgaagttagacagatagcaccaggacagacgggaaagatagctgactacaactataagc  
ttcctgatgacttcactggctgcgttatcgcggtgaattctaacaacctggactcaaaagtccgc  
10 ggcaactataactatctctatcggctgttccgcaagagtaaccttaagccctttgagagagatat  
aagcactgaaatctaccaggctggcagtagccctgtaatggcgtggaaggctttaattgttatt  
ttccactgcaatcctatggttttcagccaaccaatggcgtgggtaccacaccataaccgcgtcgtg  
gtgctctcctttgaactgctccacgctcccgcgactgtctgcggccccaagaagtccacgaacct  
tgtgaagaataagtgcgttaatttttaatttcaacggcctcactggaacaggagtgtcactgaga  
15 gtaacaagaagtccctgccattttcaacaatttggcagagacatagccgatactactgacgcggtt  
agggacccccagaccctcgagattctcgatataacgccctgctccttcggtggagtttccgtgat  
cacgccaggccaccaataaccagtaaccaggctgcgcgtgctgtatcaggatgtcaactgtactgagg  
tgcccgtagccatccatgcggatcagctcacaccaacttgagggtgtacagcacccggtccaat  
gtattccagactcgggccggatgccttattggcgccgaacacgtgaacaatagttacgaatgcga  
20 tattccaattggcgccggaatctgtgctagctaccagactcagacgaactcccaggcagcgcca  
gcagcgttgccagccagtcacatcgcttatacaatgtcacttgagccgaaaaactccgtggct  
tactcaaacacacagcatcgccatccccacaaacttcaccatatccgtgacaactgagattctgcc  
agtgtccatgactaagacgtccgtagattgcactatgtacatatgcggcgacagcacagaatggt  
ctaactctgctgctgcaatatggaagcttctgcactcaactgaacagagcgctcacaggcatcgcc  
25 gtggagcaggataagaataccaggagggtgttcgcccaagttaagcagatctacaagacccccacc  
cataaaggatttcggtggattcaatttttagtcagatactcccagacccatctaagccatccaaga  
ggagctttatcgaggatcttttgtttaacaaagttactctggccgacgcgggtttcatcaagcag  
tacggagattgcctcggcgacatcgctgctcgtgacctcatctgtgcgcaaaagtttaacggtct  
gacggtgctgcctcccctccttactgatgaaatgatcgcccagtataaccagcgactcctcgctg  
30 gcaccataacatccggttgacattcggcgctggtgcagcactgcagataccattcgccatgcaa  
atggcataatcggttcaacgggtatcggtgtcacacagaatgtcctatatgagaaccagaagctgat  
cgaaatcagttcaatagtgccatcggaataatccaggatagccttagcagcacagcctcagccc  
ttggcaaaactccaggatgctggaacagaatgccaggctctcaataccctcggtgaagcagctc  
tcatctaatttcggcgcaatttccagtgctcctcaacgacatcctcagccgcctcgacccccccga  
35 ggccgaagtgcagattgacagactgattacaggctcactccagagcctccagacttacgtgactc  
agcagctgataagagccgagataagggccagcgtaacctggctgccacaaagatgtctgag  
tgctgtgctggccagtcgaagagtagacttctgtggcaaaggctaccatctgatgagcttccc  
acaatccgcacctcacggcgtagtgctcctccacgtgacatatgtaccggtcaggagaagaatt  
tactaccgctcctgctatatgcatgatggaaggctcacttcccccgagggggtgttctgtg  
40 tccaacggcacccattggtttgtgactcagcggaatttctacgaacccagatcataaccactga  
caacacatttgtgtccggaaattgtgacgtggtcattggaatagtgaacaacactgtttatgatc  
cactgcagccagaacttgacagctttaaggaggagctcgacaagtaacttcaagaatcatacgtca  
ccagatgtggaactcgagatattagcggtatcaatgccagtggttgaatattcagaaggaaat  
agaccgccttaatgaggtcgccaaaaatctgaacgagagcctcatgatcttcaggagctgggca  
45 aatatgagcagtagcatcaagtggccttggtatatttggcttggttcatcgccggcctgatcgcc  
atagtaatgggtcacaattatgctcggctcttggatgtgttcaaatgggttcaatgcagtgaggat  
atgtatataa

PDI-S-protein+H7 Guangdong DNA (SEQ ID NO: 77)

50 atggcgaaaaacgttgcgattttcggcttattgttttctcttcttgtgttggttccttctcagat  
cttcgcggtgaatcttacgacgcgaacacagttaccacccgcataatacaaatagcttcactcggg

gtgtttattaccccgacaaagtgttcagggtcctccgtgctccactcaacacaggacctctttctt  
cctttcttttctaactgacatggtttcatgccattcatgtatccggcactaacggtactaagag  
gttcgataatcctgtgctccctttcaatgacggcgtttactttgcaagcacagagaagagtaaca  
tcatccgaggttgatctttggcactaccctcgattcaaagacgcagagcctcctcattgtgaac  
5 aatgccactaacgtggatcaaagtttgaggtttcagttctgcaatgaccctttcttgggggt  
gtactatcataagaacaacaagtcttggatggaatctgaattccgctctatagcagcgccaaca  
actgcacctttgaatacgtgtcccagcccttcttatggacctggagggaagcagggaacttt  
aagaatctgagagagttcgtgtttaaaaatatcgacggctattttaagatctattctaagcacac  
gcctattaatctcgtgcgcatcttcacaaggcttcagcgccctggaacctcgtggacctcc  
10 caattggtatcaacatcactagatttcagactctgcttgccctccaccgatcctatctgacaccc  
ggagactcctctagcggctggactgccggcgctgccgcttattacgttggttatcttcagccacg  
cacgttcctgctgaagtataacgagaatggtactattaccgatgccgtggattgtgcccttgacc  
cctgtccgaaactaagtgcacactcaagtcattcactgtgaaaaaggaatctaccagacaagc  
aattttcggtccagcctactgagagcattgtgctgttccctaacatcacaaatctttgcccctt  
15 cggagaggttttcaatgctacacgggttgccctccgtgtatgccctggaaccgcaagagaatttcca  
attgctgtggccgattactccgtgctctacaatagtgaagcttttagcactttaagtgtatggc  
gtatccctactaagcttaacgacttgtgtttcacaaacgtgtatgccgactcctttgtgatacg  
ggcgacgaagttagacagatagcaccaggacagacgggaaagatagctgactacaactataagc  
ttcctgatgacttcactggctgctgttatcgctggaattctaacaacctggactcaaaagtccggc  
20 ggcaactataactatctctatcggctgttcgcgaagagtaaccttaagccctttgagagagatat  
aagcactgaaatctaccaggctggcagtagccctgtaattggcgtggaaggctttaattgttatt  
ttccactgcaatcctatgggttttcagccaaccaatggcgtgggtaccaaccatacccgctcgtg  
gtgctctcctttgaactgctccacgctcccgactgtctgcggcccaagaagtccacgaacct  
tgtgaagaataagtgcgttaatttttaatttcaacggcctcactggaacaggagtgtcactgaga  
25 gtaacaagaagttcctgccatttcaacaatttggcagagacatagccgatactactgacgcggtt  
agggacccccagaccctcgagattctcgatataacgcctgtccttcggtggagtctccgtgat  
cacgccaggcaccaataaccagtaaccaggctgcgcgtgtgtatcaggatgtcaactgtactgagg  
tgcccgtagccatccatgccgatcagctcacaccaacttgagggtgtacagcacccggctccaat  
gtattccagactcgggcccggatgccttattggcgccgaacacgtgaacaatagttacgaatgcga  
30 tattccaattggcgccggaatctgtgctagctaccagactcagacgaactcccaggcagcgcca  
gcagcgttgccagccagtcaatcatcgcttatacaatgtcacttgagccgaaaactccgtggct  
tactcaacaacagcatgccatccccacaaattcaccatatccgtgacaactgagattctgtcc  
agtgtccatgactaagacgctccgtagattgcactatgtacatatgcggcgacagcagaatggtt  
ctaactctgctgctgcaatatggaagcttctgcactcaactgaacagagcgtcacaggcatcgcc  
35 gtggagcaggataagaataaccaggaggtgttcgccaagttaagcagatctacaagacccccacc  
cataaaggattttcggtggattcaatttttagtcagatactccagacccatctaagccatccaaga  
ggagctttatcgaggatcttttgtttaacaaagttaactctggccgacgcgggtttcatcaagcag  
tacggagattgcctcggcgacatcgctgctcgtgacctcatctgtgcgcaaaagtttaacggtct  
gacggtgctgcctccccctccttactgatgaaatgatcgcccagtataaccagcgactcctcgctg  
40 gcaccataacatccggttgacattcggcgctggtgcagcactgcagataccattcgccatgcaa  
atggcataatcggttcaacgggtatcggtgtcacacagaatgtcctatatgagaaccagaagctgat  
cgcaaatcagttcaatagtgcctcggaaaaatccaggatagccttagcagcacagcctcagccc  
ttggcaaaactccaggatgtcgtgaaccagaatgccaggctctcaatccctcgtgaagcagctc  
tcatctaatttcggcgcaatttccagtgctcctcaacgacatcctcagccgctcgaccccccgga  
45 ggccgaagtgcagattgacagactgattacaggctcactccagagcctccagacttacgtgactc  
agcagctgataagagccgagagataagggccagcgctaacctggctgccacaaagatgtctgag  
tgctgtgctgggacagccaagagagtagacttctgtggcaaaaggctaccatctgatgagcttccc  
acaatccgcacctcacggcgtagtgttctccacgtgacatatgtaccggctcaggagaagaatt  
tactaccgctcctgctatatgcatgatggaaggctcacttccccgggaggggggtgttcgtg  
50 tccaacggcaccattgggttgtgactcagcggaatttctacgaacccagatcataaccactga  
caacacatttgtgtccggaaattgtgacgtggtcattggaatagtgaacaacactgtttatgatc  
cactgcagccagaacttgacagctttaaggaggagctcgacaagtacttcaagaatcatacgtca

5 ccagatgtggacctcggagatattagcgggtatcaatgccagtggtgtcaatattcagaaggaaat  
agaccgccttaatgaggtcgccaaaaatctgaacgagagcctcatcgatcttcaggagctgggca  
aatatgagcagtagcatcaagtggccttggtatatttggttggttcacgcggcctgatcgcc  
atagtaatgggtcacaattatgctcgtcttcacatgtgtgaagaatggaacatgcggtgcactat  
ttgtatataa

PDI-S-protein+H9 HK DNA (SEQ ID NO: 78)

10 atggcgaaaaacggttgcgattttcggcctattgttttctcttcttggttggttccttctcagat  
cttcgcggtgaatcttacgacgcgaacacagttaccaccgcacatatacaaaatagcttcactcggg  
gtgtttattaccccgacaaagtgttcagggtcctccgtgctccactcaacacaggacctctttctt  
cctttcttttctaacgtgacatggtttcatgccattcatgtatccggcactaacggtactaagag  
gttcgataatcctgtgctccctttcaatgacggcggtttactttgcaagcacagagaagagtaaca  
tcacccgaggttggtatctttggcactaccctcgattcaaagacgcagagcctcctcattgtgaac  
aatgccactaacgtggtgatcaaagtttgcgagtttcagttctgcaatgaccctttcttgggggt  
gtactatcataagaacaacaagtcttggtggaatctgaattccgcgtctatagcagcgccaaca  
15 actgcacctttgaatacgtgtcccagcccttccttatggacctggagggaaagcaggggaaacttt  
aagaatctgagagagttcgtgtttaaaaatatcgacggctattttaagatctattctaagcacac  
gcctattaatctcgtgcgcgatcttccacaaggcttcagcgccctggaaccactcgtggacctcc  
caattggtatcaacatcactagatttcagactctgcttgccctccaccgatcctatctgacaccc  
ggagactcctctagcggctggactgcggcgctgcgcttattacgttggttatcttcagccacg  
20 cacgttcctgctgaagtataacgagaatggtactattaccgatgcggtggattgtgcccttgacc  
ccctgtccgaaaactaagtgcacactcaagtcattcactgtggaaaaaggaatctaccagacaagc  
aattttcgggtccagcctactgagagcattgtgcgctttcctaacatcacaaatctttgcccctt  
cggagaggttttcaatgctacacggtttgcctccgtgtatgcctggaaccgcaagagaatttcca  
attgctgtggccgattactccgtgctctacaatagtgaagcttttagcacctttaagtgtatggc  
25 gtatccctactaagcttaacgacttgtgtttcacaaacgtgtatgccgactcctttgtgatacg  
ggcgacgaagttagacagatagcaccaggacagacgggaaagatagctgactacaactataagc  
ttcctgatgacttcactggctgcttattcgcggtggaattctaacaacctggactcaaaagtgcgc  
ggcaactataactatctctatcggctgttcgcaagagtaaccttaagccctttgagagagatat  
aagcactgaaatctaccaggctggcagtagcgcctgtaatggcggtggaaggctttaattgttatt  
30 ttccactgcaatcctatggttttcagccaaacaaatggcggtgggtaccacacataccgcgtcgtg  
gtgctctcctttgaactgctccacgctcccgcgactgtctgcggcccaagaagtccacgaacct  
tgtgaagaataagtgcgttaatttttaatttcaacggcctcactggaacaggagtgctcactgaga  
gtaacaagaagtccctgccatttcaacaatttggcagagacatagccgatactactgacgcggtt  
agggacccccagaccctcgagattctcgatataacgcctgctccttcggtggagtttccgtgat  
35 cacgccaggccaataaccagtaaccaggctgcgcgtgctgtatcaggatgtcaactgtactgagg  
tgcccgtagccatccatgcggatcagctcacaccaacttgagggtgtacagcaccggctccaat  
gtattccagactcgggcccggatgccttattggcgccgaacacgtgaacaatagttacgaatgcga  
tattccaattggcgccggaatctgtgctagctaccagactcagacgaactcccaggcagcgcca  
40 gcagcgttgccagccagtcacatcgcgttatacaatgtcacttgagccgaaaactccgtggct  
tactcaaaacaacagcatcgccatccccacaaacttcaccatatccgtgacaactgagattctgcc  
agtgtccatgactaagacgtccgtagattgcactatgtacatatgcggcgacacacagaatggt  
ctaactctgctgctgcaatatggaagcttctgcactcaactgaacagagcgtccacaggcatcgcc  
gtggagcaggataagaataaccaggaggtgttcgccaagttaagcagatctacaagacccccacc  
cataaaggatttcggtggattcaatttttagtcagatactccagacccatctaagccatccaaga  
45 ggagctttatcgaggatcttttgtttaacaaagttaactctggccgacgcgggtttcatcaagcag  
tacggagattgcctcggcgacatcgctgctcgtgacctcatctgtgcgcaaaagttaacggtct  
gacggtgctgcctcccctccttactgatgaaatgatcgccagtataccagcgactcctcgtctg  
gcaccataacatccggttgacattcggcgctggtgcagcactgcagataccattcgccatgcaa  
atggcatatcgtttcaacgggtatcggtgtcacacagaatgtcctatatgagaaccagaagctgat  
50 cgcaaatcagttcaatagtgcacatcggaaaaaatccaggatagccttagcagcacagcctcagccc

5           ttggcaaaactccaggatgtcgtgaaccagaatgccaggctctcaataccctcgtgaagcagctc  
          tcactctaatttcggcgcaatttcagtgctctcaacgacatcctcagccgctcgaccccccgga  
          ggccgaagtgcagattgacagactgattacaggtcgactccagagcctccagacttacgtgactc  
          agcagctgataagagccgagagataagggccagcgctaacctggctgccacaaagatgtctgag  
10           tgcgtgctggggccagtccaagagagtagacttctgtggcaaaggctaccatctgatgagcttccc  
          acaatccgcacctcacggcgtagtggtcctccacgtgacatatgtaccggctcaggagaagaatt  
          tcactaccgctcctgctatatgccatgatggaaaggctcacttccccgggagggggtgttcgtg  
          tccaacggcaccattgggttgtagctcagcggaatttctacgaacccagatcataaccactga  
          caacacatttggtgctccggaattgtgacgtggcattggaatagtgaacaacactgtttatgac  
15           cactgcagccagaacttgacagctttaaggaggagctcgacaagtacttcaagaatcatacgtca  
          ccagatgtggacctcgagatattagcggatcaatgccagtgttgtaaatattcagaaggaaat  
          agaccgccttaatgaggtcgcaaaaaatctgaacgagagcctcatcgatcttcaggagctgggca  
          aatatgagcagtagacatcaagtggccttggtatatttggcttggcttcacgcggcctgatcgcc  
          atagtaatggtcacaattatgctcctgttctgggccatgtccaatggatcttcagatgcaacat  
          ttgtatataa

**PDI-S-protein+ B/ Wash DNA (SEQ ID NO: 79)**

20           atggcgaaaaacgttgcgattttcggcttattgttttctcttcttctgtgttggttccttctcagat  
          cttcgcggtgaatcttacgacgcgaacacagttaccaccgcataatacaaatagcttcactcggg  
          gtgtttattaccccgacaaagtgttcaggtcctccgtgctccactcaacacagagacctcttctctt  
          cctttcttttctaacgtgacatggtttcatgccattcatgtatccggcactaacggtactaagag  
          gttcgataatcctgtgctccctttcaatgacggcggtttactttgcaagcacagagaagagtaaca  
          tcattccgaggttgatctttggcactaccctcgattcaaagacgcagagcctcctcattgtgaac  
25           aatgccactaacgtggtgatcaaagtttgcgagtttcagttctgcaatgacccttcttgggggt  
          gtactatcataagaacaacaagtcttggtggaatctgaattccgcgtctatagcagcgccaaca  
          actgcacctttgaatacgtgtccagcccttcttatggacctggagggaaagcagggaacttt  
          aagaatctgagagagttcgtgtttaaaaatatcgacggctatttttaagatctattctaagcacac  
          gcctattaatctcgtgcgcatcttccacaaggcttcagcgccctggaaccactcgtggacctcc  
          caattggtatcaacatcactagatttcagactctgcttgccctccaccgatcctatctgacaccc  
30           ggagactcctctagcggctggactgcccgcgtgcccgttattacgttggtattcttcagccacg  
          cacgttctcgtgaagtataacgagaatggtactattaccgatgcggtgattgtgcccttgacc  
          ccctgtccgaaaactaagtgcacactcaagtcattcactgtggaaaaaggaatctaccagacaagc  
          aattttcgggtccagcctactgagagcattgtgcttctcctaacatcacaaatctttgcccctt  
          cggagaggttttcaatgctacacgggttgctcctcggtgatgcctggaaccgcaagagaatttcca  
35           attgctggtggcgttactccgtgctctacaatagtgaagcttttagcacctttaagtgtatggc  
          gtatcccctactaagcttaacgacttgtgtttcacaaacgtgtatgccgactcctttgtgatacg  
          ggcgacgaagttagacagatagcaccaggacagacgggaaagatagctgactacaactataagc  
          ttcctgatgacttcactggctgcttattcgctggaattctaacaacctggactcaaaagtgcggc  
          ggcaactataactatctctatcggtgttcgcaagagtaaccttaagccctttgagagagatat  
40           aagcactgaaatctaccaggtggcagtagccctgtaatggcgtggaaggctttaattgttatt  
          ttccactgcaatcctatggttttcagccaaccaatggcgtgggtaccaaccataaccgcgtcgtg  
          gtgctctcctttgaactgtccacgctcccgcgactgtctgcggcccaagaagtcacgaacct  
          tgtgaagaataagtgcgttaatttttaatttcaacggcctcactggaacagagtgtcactgaga  
          gtaacaagaagttcctgccatttcaacaatttggcagagacatagccgatactactgacgcggtt  
45           agggacccccagaccctcgagattctcgatataacgcctgctccttcggtggagtttccgtgat  
          cacgccaggcaccaataaccagtaaccaggtcgccgtgctgtatcaggatgtcaactgtactgagg  
          tgcccgtagccatccatgcggatcagctcacaccaacttgagggtgtacagcaccggctccaat  
          gtattccagactcgggcccggatgccttattggcgccgaacacgtgaacaatagttacgaatgcga  
          tattccaattggcgccggaatctgtgctagctaccagactcagacgaactcccaggcagcgcca  
50           gcagcgttgccagccagtcaatcatcgcttatacaatgtcacttggagccgaaaactccgtggct  
          tactcaaacaacagcatcgccatccccacaacttcaccatatccgtgacaactgagattctgccc

5 agtgtccatgactaagacgtccgtagattgcactatgtacatatgcggcgacagcacagaatggt  
 ctaatctgctgctgcaatatggaagcttctgcactcaactgaacagagcgctcacaggcatcgcc  
 gtggagcaggataagaatacccaggaggtgttcgcccagttaagcagatctacaagaccccacc  
 cataaaggatttccggtggattcaatttttagtcagatactcccagacccatctaagccatccaaga  
 10 ggagctttatcgaggatcttttgtttaacaaagttactctggccgacgcccgtttcatcaagcag  
 tacggagattgcctcggcgacatcgctgctcgtgacctcatctgtgcgcaaaagtttaacggtct  
 gacggtgctgcctccccctccttactgatgaaatgatcgcccagtataccagcgcaactcctcgctg  
 gcaccataacatccggttggacattcggcgctggtgcagcactgcagataccattcgccatgcaa  
 atggcatatcgtttcaacgggtatcggtgtcacacagaatgtcctatatgagaaccagaagctgat  
 15 cgcaaatcagttcaatagtgccatcggaataatccaggatagccttagcagcacagcctcagccc  
 ttggcaaaactccaggatgtcgtgaaccagaatgccaggctctcaataccctcgtgaagcagctc  
 tcatctaatttccggcgcaatttccagtgctcctcaacgacatcctcagccgctcgaccccccgga  
 ggccgaagtgcagattgacagactgattacaggctcgaactccagagcctccagacttacgtgactc  
 agcagctgataagagccgcccagagataagggccagcgctaacctggctgccacaaagatgtctgag  
 20 tgcgtgctgggcccagtcgaagagagtagacttctgtggcaaaggctaccatctgatgagcttccc  
 acaatccgcacctcacggcgtagtggtcctccacgtgacatatgtaccggtcaggagaagaatt  
 tcaactaccgctcctgctatatgccatgatggaaggctcacttccccgggaggggggtgttcgtg  
 tccaacggcacccattggtttgtgactcagcggaatttctacgaacccagatcataaccactga  
 caacacatttgtgtccggaaattgtgacgtggtcattggaatagtgaacaacactgtttatgatc  
 25 cactgcagccagaacttgacagctttaaggaggagctcgacaagtaacttcaagaatcatacgtca  
 ccagatgtggacctcgagatattagcgggtatcaatgccagtggtgtcaatattcagaaggaaat  
 agaccgcttaatgaggtcgccaaaaatctgaacgagagcctcatcgatcttcaggagctgggca  
 aatatgagcagtagcatcaagtggccttggtatatttggcttggcttcatcgccggcctgatcgcc  
 atagtaatgggtcacaattatgctcgttggtttatatggtctccagagacaatgtttcttgctccat  
 ttgtctataa

IF-H1HawaiiCT.r (SEQ ID NO: 80)

acgacacgactaaggccttttaatacatattctacactgtagagaccc

IF-H3MinnesotaCT.r (SEQ ID NO: 81)

acgacacgactaaggccttttagatgcagatggtgcatctgatgttgcccttctg

30 IF-HongKongCT.r (SEQ ID NO: 82)

acgacacgactaaggcctttatatacatatcctgcactgcattgaaccattt

IF-GuangdongCT.r (SEQ ID NO: 83)

acgacacgactaaggcctttatatacaaatagtgcaccgcatgtttcca

IF-H9HKCT.r (SEQ ID NO: 84)

35 acgacacgactaaggcctttatatacaaatggtgcatctgcaagatccat

IF-BWashCT.r (SEQ ID NO: 85)

acgacacgactaaggcctttatagacaaatggagcaagaaacattgtctc

IF(nbHEL40)-PDI.c (SEQ ID NO: 86)

ccaaaacacattgagcaaaatggcgaaaaacggttgcgattttcggttat

IF(AvB+wtCT).r (SEQ ID NO: 87)

ACGACACGACTAAGGCCTTTAGGTATAATGGAGTTTCACCCCTTCAGAA

5

PDI-SARS-COV-1 wtTMCT-DNA (SEQ ID NO: 88)

atggcgaaaaacggttgcgattttcggttatattgttttctcttcttgtgttggttccttctcagat  
cttcgctgTCCGATCTGGATCGGTGCACTACATTTGACGATGTGCAGGCACCTAATTATACTCAGC  
ACACCTCTTCCATGCGGGGCGTGTACTACCCGACGAGATTTTTAGAAAGTGACACACTGTACCTG  
10 ACCCAAGACCTTTTTCTCCCATTTTATAGCAATGTCACGGGATTCCACACTATCAATCACACATT  
CGGAAACCTGTTATCCCTTTTAAAGACGGTATCTACTTTGCTGCTACTGAGAAATCGAATGTTG  
TTCGCGGTGGGTGTTTCGGCTCAACCATGAACAACAAGAGTCAGTCAGTAATAATTATAAAACAAC  
TCAACCAATGTCGTCATCAGGGCTTGCAACTTCGAGCTCTGCGATAACCCCTTCTTTGCGGTGTC  
CAAACCGATGGGCACTCAGACCCATACCATGATTTTTGACAATGCCTTTAATTGTACTTTTGAAT  
15 ACATCAGCGATGCCCTTCTCGTGGATGTGTCCGAGAAGAGCGGTAACCTCAAGCATTTCGCGGAG  
TTCGTTTTTCAAAAACAAAGACGGGTTTTGTATGTCTACAAAGGCTATCAGCCCATTGATGTGGT  
CAGGGATCTGCCAGTGGTTTTCAACACACTGAAGCCCATCTTCAAGTTGCCACTCGGCATCAACA  
TCACTAATTTCCGCGCCATCCTAACTGCTTTTTCGCCAGCGCAGGATATTTGGGGGACATCCGCC  
GCAGCTTACTTCTGTTGGATATCTGAAGCCACCCTTTTATGCTAAAGTACGACGAGAATGGCAC  
20 CATCACCAGTGGCGTGGACTGCTCAGAGAATCCTCTCGCAGAGCTCAAGTGTTCAGTGAAGTCAT  
TTGAGATCGACAAAGGGATCTACCAACATCTAAGTTTTCGCGTGTGCTTCCGGTGACGTAGTC  
CGTTTCCCAAATATCACTAAGTTGTGCCCTTTTGGTGAAGTATTCAACGCAACCAAATTTCCCGAG  
TGTGTATGCGTGGGAGCGCAAAAAGATCAGTAAGTGTGTGGCTGACTATAGTGTCTGTACAATA  
GCACCTTCTTACGACCTTCAAGTGTATGGAGTGAGCGCTACAAAACGATCTTTGTTTC  
25 TCAACGTGTACGCCGATTCAATTTGTCTGTTAAAGGTGACGATGTGAGGCAGATCGCTCCAGGCCA  
GACAGGTGTGATTGCTGACTATAATTACAACTGCCAGACGACTTCATGGGGTGCGTGTAGCTT  
GGAATACAGAAACATTTGACGCCACCTCCACGGGAAATTACAATTACAGTATCGTTACCTTCGC  
CATGGAAAGTTGAGACCTTTCGAGCGTGATATAAGTAACGTGCCCTTTAGTCCAGATGGAAAACC  
CTGCACACCCCTGCTCTCAATTGCTATTGGCTCTCAATGACTACGGCTTTTACACAACACTG  
30 GCATCGGATACCGCCTTACCGGGTCTGTGGTGTCTAGTTTTGAGTTGCTTAACGCACCCGCCACC  
GTGTGTGGTCTTAACTTTCTACTGACCTGATTAATAAACCAATGCGTCAACTTCAATTTTAACGG  
GCTGACCGGCACCGGTGTCTGACCCCTAGCTCTAAGAGATTCCAGCCTTTTCAGCAGTTTCGGGA  
GGGATGTGAGCGACTTTACCGACTCTGTGAGGGATCCAAAGACCAGCGAGATACTGGATATCTCG  
CCCTGCGAGTTTCGGTGGCGTGTCCGTTATTACACCTGGCACCAACGCCTCCTCAGAGGTGGCGGT  
35 GCTCTATCAAGATGTCAACTGCACTGATGTGTCAACTGCCATCCATGCCGATCAGCTGACCCCG  
CCTGGCGCATCTACAGTACCGGAACAACGTTTTTCAGACCCAGCCGGCTGTCTAATCGGCGCA  
GAGCAGTTGACACATCCTACGAATGTGACATACCTATCGGGGCAGGCATTTGCGCTAGCTACCA  
TACCGTGTCACTGTTGGCTTCCACGTCAAAAAGTCAATCGTTGCCTACACGATGAGTCTGGGGG  
CTGACTCATCTATCGCCTACAGCAACAATACCATTGCAATTCCCAAACTTCAGTATCTCCATC  
40 ACAACAGAGGTGATGCCCGTTTCTATGGCTAAAACATCAGTCGATTGCAATATGTATATATGCGG  
CGATAGTACTGAGTGCGCAATCTCTTGTACAGTACGGCTCCTTTTGTACCCAGCTGAACCGAG  
CACTGTCTGGAATCGCCGAGAACAGGATCGCAATACCCGGGAAGTCTTCGCCCAGGTGAAGCAG  
ATGTACAAAACGCCCCTCTCAAGTATTTTCGGCGGATTCAACTTTTCTCAGATTTTGCCTGACCC  
GCTCAAGCCAACAAAACGATCTTTTATCGAAGACCTTCTGTTTAAACAAGGTCACACTGGCGGATG  
45 CTGGGTTCATGAAAACAGTACGGTGAATGCCTGGGGGACATCAATGCCAGAGATCTGATCTGCGCC  
CAGAAATTCAATGGCTTAACAGTCCTCCACCTCTCTTGACCGACGATATGATCGCTGCGTACAC  
CGCTGCTCTGGTATCGGGCACCGCGACTGCTGGCTGGACCTTTGGTGCCGGAGCCGCACTCCAGA  
TCCCATTCGCCATGCAGATGGCCTACCGCTTCAACGGAATCGGGGTACCCAGAACGTGCTGTAT

5 GAGAACCAGAAACAGATCGCCAATCAGTTCAATAAGGCAATTAGTCAGATTTCAGGAGAGTCTTAC  
CACTACCAGCACCGCCCTGGGCAAGCTGCAAGATGTTGTGAACCAGAAATGCGCAGGCATTAAACA  
CTCTGGTTAAACAGCTGAGCTCAAATTTTGGTGCAATCTCTTCAGTTCTGAACGATATCCTGAGT  
CGGCTGGATCCGCCAGAGGCTGAAGTGCAAATTGATCGTTTGATCACCGGGAGGCTACAATCTCT  
10 GCAGACGTACGTGACCCAGCAGCTCATCCGGGCAGCCGAAATTCGCGCATCAGCCAACCTCGCTG  
CAACTAAGATGTCTGAGTGCGTGCTGGGCCAGAGTAAGAGGGTGGACTTTTGTGGTAAGGGATAC  
CACCTCATGTCTTTCCGCAAGCGGCTCCCCACGGCGTGGTTTTCTTACACGTTACCTATGTGCC  
ATCCCAAGAACGCAATTTACCACCGCTCCAGCTATCTGTCATGAGGGCAAAGCATATTTCCCCA  
GGGAAGGAGTATTTGTGTTTAAATGGCACGTCCTGGTTTATAACCCAACGTAACCTTTTTCTCCCCA  
15 CAGATTATCACAACCGACAACACATTCTGTCTGGGAATTGTGACGTCGTGATCGGGATCATTAA  
CAATACCGTTTACGATCCCTTGACAGCCGAGCTTGACTCCTTTAAAGAGGAACTAGACAAATACT  
TTAAGAATCACACCTCACCGGACGTAGATTTGGGAGACATCTCTGGAATTAATGCCTCTGTGGTG  
AATATCCAGAAGGAGATCGACCGCTGAATGAAGTCGCCAAGAACCTCAACGAGTCCCTGATAGA  
TCTGCAAGAACTGGGCAAATATGAACAGTACATCAAATGGCCGTGGTACGTGTGGTTGGGCTTTA  
20 TCGCTGGACTTATTGCAATCGTGATGGTGACGATTCTGCTCTGCTGTATGACTTCCTGCTGCTCT  
TGTCTGAAGGGCGCCTGTAGCTGTGGTTCTGCTGCAAGTTCGACGAGACGACTCCGAACCAGT  
TCTGAAGGGGGTGAAACTCCATTATACCTAA

**PDI-SARS-COV-1 H5iTMCT-DNA (SEQ ID NO: 89)**

20 atggcgaaaaacggttgcgatcttctcggttattgttttctcttcttctgtgttggttccttctcagat  
cttcgcgTCCGATCTGGATCGGTGCACTACATTTGACGATGTGCAGGCACCTAATTATACTCAGC  
ACACCTCTTCCATGCGGGGCGTGTACTACCCGACGAGATTTTTAGAAGTGACACACTGTACCTG  
ACCCAAGACCTTTTTCTCCCATTTTATAGCAATGTACGGGATTCCACACTATCAATCACACATT  
25 CGGAAACCTGTTATCCCTTTAAAGACGGTATCTACTTTGCTGCTACTGAGAAATCGAATGTTG  
TTCGCGGTTGGGTGTTTCGGCTCAACCATGAACAACAAGAGTCAGTCAGTAATAATTATAAAACAAC  
TCAACCAATGTCTGTCATCAGGGCTTGCAACTTCGAGCTCTGCGATAACCCCTTCTTTGCGGTGTC  
CAAACCGATGGGCACTCAGACCCATACCATGATTTTGACAATGCCTTTAATTGTACTTTTCAAT  
ACATCAGCGATGCCTTCTCGCTGGATGTGTCCGAGAAGAGCGGTAACCTCAAGCATTTCGGGGAG  
30 TTCGTTTTCAAAAACAAGACGGGTTTTGTATGCTACAAAGGCTATCAGCCCATTTGATGTTGGT  
CAGGGATCTGCCCAGTGGTTTCAACACACTGAAGCCCATCTTCAAGTTGCCATCGGCATCAACA  
TCACTAATTTCCGCGCCATCCTAACTGCTTTTTCGCCAGCGCAGGATATTTGGGGGACATCCGCC  
GCAGCTTACTTCTGTGGGATATCTGAAGCCACCACTTTTATGCTAAAGTACGACGAGAATGGCAC  
CATCACCGATGCCGTGGACTGCTCACAGAATCCTCTCGCAGAGCTCAAGTGTTCAGTGAAGTCAT  
35 TTGAGATCGACAAAGGGATCTACCAACATCTAACTTTTCGCTGCTGCTTCCGGTGACGTAGTC  
CGTTTCCCAAATATCACTAACTTGTGCCCTTTTGGTGAAGTATTCAACGCAACCAAATTCCCCAG  
TGTGTATGCGTGGGAGCGCAAAAAGATCAGTAACGTGTGTGGCTGACTATAGTGTCTGTACAATA  
GCACCTTCTTCAGCACCTTCAAGTGTATGGAGTGAGCGCTACAAAACCTGAACGATCTTTGTTTT  
TCAAACGTGTACGCCGATTCAATTTGTGTTAAAGGTGACGATGTGAGGCAGATCGCTCCAGGCCA  
40 GACAGGTGTGATTGCTGACTATAATTACAACTGCCAGACGACTTCATGGGGTGCGTGCTAGCTT  
GGAATACAAGAAACATTGACGCCACCTCCACGGGAATTACAATTACAAGTATCGTTACCTTCGC  
CATGGAAAGTTGAGACCTTCGAGCGTGATATAAGTAACGTGCCCTTTAGTCCAGATGGAAAACC  
CTGCACACCCCTGCTCTCAATTGCTATTGGCCTCTCAATGACTACGGCTTTTACACAACTACTG  
GCATCGGATACACGCTTACCGGGTCTGGTGCTCAGTTTTGAGTTGCTTAACGCACCCGCCACC  
45 GTGTGTGGTCTTAACTTTCTACTGACCTGATTAATAAACCAATGCGTCAACTTCAATTTTAACGG  
GCTGACCGGCACCGGTGTCTGACCCCTAGCTCTAAGAGATTCCAGCCTTTTCAGCAGTTCGGGA  
GGGATGTGAGCGACTTTACCGACTCTGTGAGGATCCAAAGACCAGCGAGATACTGGATATCTCG  
CCCTGCAGTTTCGGTGGCGTGTCCGTTATTACACCTGGCACCAACGCCTCCTCAGAGGTGGCGGT  
GCTCTATCAAGATGTCAACTGCACTGATGTGTCAACTGCCATCCATGCCGATCAGCTGACCCCG  
50 CCTGGCGCATCTACAGTACCGGAACAACGTTTTTCAGACCCAGGCCGGCTGTCTAATCGGCGCA  
GAGCACGTTGACACATCCTACGAATGTGACATACCTATCGGGGCAGGCATTTGCGCTAGCTACCA

5 TACCGTGTCACTGTTGGCTTCCACGTACACAAAAGTCAATCGTTGCCTACACGATGAGTCTGGGGG  
CTGACTCATCTATCGCCTACAGCAACAATACCATTGCAATTTCCACAAAACCTCAGTATCTCCATC  
ACAACAGAGGTGATGCCCGTTTCTATGGCTAAAACATCAGTCGATTGCAATATGTATATATGCGG  
CGATAGTACTGAGTGCGCCAATCTCTTGTTACAGTACGGCTCCTTTTGTACCCAGCTGAACCGAG  
10 CACTGTCTGGAATCGCCGAGAACAGGATCGCAATACCCGGGAAGTCTTCGCCCAGGTGAAGCAG  
ATGTACAAAAACGCCCCTCTCAAGTATTTTCGGCGGATTCAACTTTTCTCAGATTTTGCCTGACCC  
GCTCAAGCCAACAAAACGATCTTTTATCGAAGACCTTCTGTTTAAACAAGGTACACTGGCGGATG  
CTGGGTTCATGAAACAGTACGGTGAATGCCTGGGGGACATCAATGCCAGAGATCTGATCTGCGCC  
CAGAAATTCATGGCTTAACAGTCCCTCCACCTCTCTTGACCGACGATATGATCGCTGCGTACAC  
15 CGCTGCTCTGGTATCGGGCACCGGACTGCTGGCTGGACCTTTGGTGGCGAGCCGCACTCCAGA  
TCCCATTTCGCCATGCAGATGGCCTACCGCTTCAACGGAATCGGGGTCACCCAGAACGTGCTGTAT  
GAGAACCAGAAAACAGATCGCCAATCAGTTCAATAAGGCAATTAGTCAGATTCAAGGAGAGTCTTAC  
CACTACCAGCACCGCCCTGGGCAAGCTGCAAGATGTTGTGAACCAGAATGCGCAGGCATTAAACA  
CTCTGGTTAAACAGCTGAGCTCAAATTTTGGTGCAATCTCTTCAGTTCTGAACGATATCCTGAGT  
20 CGGCTGGATCCGCCAGAGGCTGAAGTGCAAATTCATCGTTTGATCACCAGGAGGCTACAATCTCT  
GCAGACGTACGTGACCCAGCAGCTCATCCGGGCAGCCGAAATTCGCGCATCAGCCAACCTCGCTG  
CAACTAAGATGTCTGAGTGCGTGCTGGGCCAGAGTAAGAGGGTGGACTTTTGTGGTAAGGGATAC  
CACCTCATGTCCTTTCCGCAAGCGGCTCCCCACGGCGTGGTTTTCTTACACGTTACCTATGTGCC  
ATCCCAAGAACGCAATTTACCACCGCTCCAGCTATCTGTCATGAGGGCAAAGCATATTTCCCCA  
25 GGAAGGAGTATTTGTGTTTAAATGGCACGTCTGGTTTATAACCCAACGTAACCTTTTCTCCCCA  
CAGATTATCACAACCGACAACACATTCTGTGCTCGGGAATTGTGACGTGCTGATCGGGATCATTAA  
CAATACCGTTTACGATCCCTTGACAGCCGAGCTTGACTCCTTTAAAGAGGAACTAGACAAATACT  
TTAAGAATCACACCTCACCGGACGTAGATTGGGAGACATCTCTGGAATTAATGCCTCTGTGGTG  
AATATCCAGAAGGAGATCGACCGCTGAATGAAGTCGCCAAGAACCTCAACGAGTCCCTGATAGA  
30 TCTGCAAGAACTGGGCAAATATGAACAGTACATCAAATGGCCGTGGTACcaaatactgtcaattt  
attcaacagtggcgagttccctagcactggcaatcatgatggctggtctatctttatggatgtgc  
tccaatggatcgttacaatgcagaatttgcattTAA

**PDI-SARS-COV-1 H5iCT-DNA (SEQ ID NO: 90)**

30 atggcgaaaaacggttgcgatatttcggcttattgttttctcttcttgtgttggttccttctcagat  
cttcgctTCCGATCTGGATCGGTGCACTACATTTGACGATGTGCAGGCACCTAATTATACTCAGC  
ACACCTCTTCCATGCGGGGCGTGTACTACCCGACGAGATTTTATAGAAAGTGACACACTGTACCTG  
ACCCAAGACCTTTTCTCCCATTTTATAGCAATGTCACGGGATTCCACACTATCAATCACACATT  
35 CGGAAACCTGTTATCCCTTTAAAGACGGTATCTACTTTGCTGCTACTGAGAAATCGAATGTTG  
TTCGCGGTTGGGTGTTTCGGCTCAACCATGAACAACAAGAGTCAGTCAGTAATAATTATAAACAAC  
TCAACCAATGTCGTCATCAGGGCTTGCAACTTCGAGCTCTGCGATAACCCCTTCTTTGCGGTGTC  
CAAACCGATGGGCACTCAGACCCATACCATGATTTTGAACAATGCCCTTAAATTGTAATTTTGAAT  
ACATCAGCGATGCCTTCTCGCTGGATGTGTCCGAGAAGAGCGGTAACCTCAAGCATTTCGCGGAG  
40 TTCTGTTTCAAAAACAAAGACGGGTTTTTGTATGTCTACAAAGGCTATCAGCCCATTGATGTGGT  
CAGGGATCTGCCCAGTGGTTTCAACACACTGAAGCCCATCTTCAAGTTGCCACTCGGCATCAACA  
TCACTAATTTCCGCGCCATCCTAACTGCTTTTTTCGCCAGCGCAGGATATTTGGGGGACATCCGCC  
GCAGCTTACTTCGTGGGATATCTGAAGCCACCCTTTTATGCTAAAGTACGACGAGAATGGCAC  
CATCACCGATGCCGTGGACTGCTCACAGAATCCTCTCGCAGAGCTCAAGTGTTCAAGTGAAGTCAT  
45 TTGAGATCGACAAAGGATCTACCAACATCTAACTTTTCGCGTCGTGCCTTCCGGTGACGTAGTC  
CGTTTCCCAAATATCACTAACTTGTGCCCTTTTGGTGAAGTATTCAACGCAACCAAATTTCCCGAG  
TGTGTATGCGTGGGAGCGCAAAAAGATCAGTAACCTGTGTGGCTGACTATAGTGTTCTGTACAATA  
GCACCTTCTTTCAGCACCTTCAAGTGTTATGGAGTGAGCGCTACAAAACGTAACGATCTTTGTTTC  
TCAAACGTGTACGCCGATTCAATTTGTGCTTAAAGGTGACGATGTGAGGCAGATCGCTCCAGGCCA  
50 GACAGGTGTGATTGCTGACTATAATTACAACTGCCAGACGACTTCATGGGGTGCGTGCTAGCTT  
GGAATACAAGAAACATTGACGCCACCTCCACGGGAATTACAATTACAAGTATCGTTACCTTCGC

CATGGAAAGTTGAGACCCTTCGAGCGTGATATAAGTAACGTGCCCTTTAGTCCAGATGGAAAACC  
 CTGCACACCCCTGCTCTCAATTGCTATTGGCCTCTCAATGACTACGGCTTTTACACAACACTACTG  
 GCATCGGATACCAGCCTTACCGGGTCGTGGTGCTCAGTTTTGAGTTGCTTAACGCACCCGCCACC  
 GTGTGTGGTCCCTAAACTTTCTACTGACCTGATTAATAAACCAATGCGTCAACTTCAATTTTAACGG  
 5 GCTGACCGGCACCGGTGTCTGACCCCTAGCTCTAAGAGATTCCAGCCTTTTCAGCAGTTTCGGGA  
 GGGATGTGAGCGACTTTACCGACTCTGTCAGGGATCCAAAGACCAGCGAGATACTGGATATCTCG  
 CCCTGCAGTTTCGGTGGCGTGTCCGTTATTACACCTGGCACCACGCCTCCTCAGAGGTGGCGGT  
 GCTCTATCAAGATGTCAACTGCACTGATGTGTCAACTGCCATCCATGCCGATCAGCTGACCCCG  
 CCTGGCGCATCTACAGTACCGGGAACAACGTTTTTCAGACCCAGGCCGGCTGTCTAATCGGCGCA  
 10 GAGCACGTTGACACATCCTACGAATGTGACATACCTATCGGGGCAGGCATTTGCGCTAGCTACCA  
 TACCGTGTCACTGTTGGCTTCCACGTCAAAAAGTCAATCGTTGCCTACACGATGAGTCTGGGGG  
 CTGACTCATCTATCGCCTACAGCAACAATACCATTGCAATTTCCACAAAACCTTCAGTATCTCCATC  
 ACAACAGAGGTGATGCCCGTTTCTATGGCTAAAACATCAGTCGATTGCAATATGTATATATGCGG  
 15 CGATAGTACTGAGTGCGCCAATCTCTTGTTACAGTACGGCTCCTTTTGTACCCAGCTGAACCGAG  
 CACTGTCTGGAATCGCCGCAGAACAGGATCGCAATACCCGGGAAGTCTTCGCCCAGGTGAAGCAG  
 ATGTACAAAACGCCCACTCTCAAGTATTTTCGGCGGATTCAACTTTTCTCAGATTTTGCCTGACCC  
 GCTCAAGCCAACAAAACGATCTTTTATCGAAGACCTTCTGTTTAAACAAGGTCACTGCGCGATG  
 CTGGGTTTCATGAAACAGTACGGTGAATGCCTGGGGGACATCAATGCCAGAGATCTGATCTGCGCC  
 20 CAGAAATTCATGGCTTAACAGTCCTCCCACCTCTCTTGACCGACGATATGATCGCTGCGTACAC  
 CGCTGCTCTGGTATCGGGCACCGCGACTGCTGGCTGGACCTTTGGTGCCGGAGCCGCACTCCAGA  
 TCCCATTCGCCATGCAGATGGCCTACCGCTTCAACGGAATCGGGGTACCCAGAACGTGCTGTAT  
 GAGAACCAGAAAACAGATCGCCAATCAGTTCAATAAGGCAATTAGTCAGATTTCAGGAGAGTCTTAC  
 CACTACCAGCACCGCCCTGGGCAAGCTGCAAGATGTTGTGAACCAGAATGCGCAGGCATTAAACA  
 25 CTCTGGTTAAACAGCTGAGCTCAAATTTTGGTGCAATCTCTTCAGTTCTGAACGATATCCTGAGT  
 CGGCTGGATCCGCCAGAGGCTGAAGTGCAATTGATCGTTTGATCACCGGGAGGCTACAATCTCT  
 GCAGACGTACGTGACCCAGCAGCTCATCCGGGCAGCCGAAATTCGCGCATCAGCCAACCTCGCTG  
 CAACTAAGATGTCTGAGTGCCTGCTGGGCCAGAGTAAGAGGGTGGACTTTTGTGGTAAGGGATAC  
 CACCTCATGTCTTTCCGCAAGCGGCTCCCCACGGCGTGGTTTTCTTACACGTTACCTATGTGCC  
 30 ATCCCAAGAACGCAATTTACACACCGCTCCAGCTATCTGTCTATGAGGGCAAAGCATATTTCCCA  
 GGAAGGAGTATTTGTGTTAATGGCACGTCTGGTTTTATAACCAACGTAACCTTTTCTCCCCA  
 CAGATTATCACAACCGACAACACATTGCTGTCTGGGAATTGTGACGTGCTGATCGGGACATTAA  
 CAATACCGTTTACGATCCCTTGACGCCGAGCTTGACTCCTTTAAAGAGGAACAGACAAATACT  
 TTAAGAATCACACCTCACCGGACGTAGATTTGGGAGACATCTCTGGGAATTAATGCCTCTGTGGTG  
 35 AATATCCAGAAGGAGATCGACCGCTGAATGAAGTCGCCAAGAACCTCAACGAGTCCCTGATAGA  
 TCTGCAAGAAGTGGGCAAATATGAACAGTACATCAATGGCCGTGGTACGTGTGGTTGGGCTTTA  
 TCGCTGGACTTATTGCAATCGTGATGGTGACGATTCTGCTCtctttatggatgtgctccaatgga  
 tcgttacaatgcagaatttgcattTAA

#### PDI-SARS-COV-1 H5iCT(V4)-DNA (SEQ ID NO: 91)

40 atggcgaaaaacgttgcgatttttcggcttattgttttctcttcttctgtgttggttccttctcagat  
 cttcgcgTCCGATCTGGATCGGTGCACTACATTTGACGATGTGCAGGCACCTAATTATACTCAGC  
 ACACCTCTTCCATGCGGGGCGTGTACTACCCGACGAGATTTTTAGAGTGACACACTGTACCTG  
 ACCCAAGACCTTTTTCTCCATTTTATAGCAATGTCACGGGATTCCACACTATCAATCACACATT  
 CGGAAACCTGTATCCCTTTAAAGACGGTATCTACTTTGCTGCTACTGAGAAATCGAATGTTG  
 45 TTCGCGGTTGGGTGTTCCGCTCAACCATGAACAACAAGAGTCAGTCAGTAATAATTATAACAAC  
 TCAACCAATGTCGTATCAGGGCTTGCAACTTCGAGCTCTGCGATAACCCCTTCTTTGCGGTGTC  
 CAAACCGATGGGCACTCAGACCCATACCATGATTTTTGACAATGCCCTTAATTGTACTTTTGAAT  
 ACATCAGCGATGCCTTCTCGCTGGATGTGTCCGAGAAGAGCGGTAACCTCAAGCATTTCGCGGAG  
 TTCGTTTTCAAAAACAAAGACGGTTTTTGTATGTCTACAAAGGCTATCAGCCCATTGATGTGGT  
 50 CAGGGATCTGCCCAGTGGTTTTCAACACACTGAAGCCCATCTTCAAGTTGCCACTCGGCATCAACA

5 TCACTAATTTCCGCGCCATCCTAACTGCTTTTTTCGCCAGCGCAGGATATTTGGGGGACATCCGCC  
GCAGCTTACTTCGTGGGATATCTGAAGCCCACCCTTTTATGCTAAAGTACGACGAGAATGGCAC  
CATCACCGATGCCGTGGACTGCTCACAGAATCCTCTCGCAGAGCTCAAGTGTTTCAGTGAAGTCAT  
10 TTGAGATCGACAAAGGGATCTACCAAACATCTAACCTTTTCGCGTCGTGCCCTCCGGTGACGTAGTC  
CGTTTCCCAAATATCACTAACTTGTGCCCTTTTGGTGAAGTATTC AACGCAACCAAATTTCCCGAG  
TGTGTATGCGTGGGAGCGCAAAAAGATCAGTAACTGTGTGGCTGACTATAGTGTTCTGTACAATA  
GCACCTTCTTCAGCACCTTCAAGTGTTATGGAGTGAGCGCTACAAAACCTGAACGATCTTTGTTTC  
TCAAACGTGTACGCCGATTCAATTTGTCTTAAAGGTGACGATGTGAGGCAGATCGCTCCAGGCCA  
15 GACAGGTGTGATTGCTGACTATAATTACAACTGCCAGACGACTTCATGGGGTGCGTGCTAGCTT  
GGAATACAAGAAACATTGACGCCACCTCCACGGGAAATTACAATTACAAGTATCGTTACCTTCGC  
CATGGAAAGTTGAGACCCCTTCGAGCGTGATATAAGTAACGTGCCCTTTAGTCCAGATGGAAAACC  
CTGCACACCCCTGCTCTCAATTGCTATTGGCCTCTCAATGACTACGGCTTTTACACAACCTACTG  
GCATCGGATACCGCCTTACCGGGTCTGTTGCTCAGTTTTGAGTTGCTTAACGCACCCGCCACC  
20 GTGTGTGGTCTAAACTTTCTACTGACCTGATTA AAAACCAATGCGTCAACTTCAATTTTAACGG  
GCTGACCGGCACCGGTGTCTGACCCCTAGCTCTAAGAGATTCCAGCTTTTTCAGCAGTTCGGGA  
GGGATGTGAGCGACTTTACCGACTCTGT CAGGGATCCAAAGACCAGCGAGATACTGGATATCTCG  
CCCTGCAGTTTCGGTGGCGTGTCCGTTATTACACCTGGCACCAACGCCTCCTCAGAGGTGGCGGT  
GCTCTATCAAGATGTCAACTGCACTGATGTGTCAACTGCCATCCATGCCGATCAGCTGACCCCG  
25 CCTGGCGCATCTACAGTACCGGGAACAACGTTTTTCAGACCCAGGCCGGCTGTCTAATCGGCGCA  
GAGCAGTTGACACATCCTACGAATGTGACATACTATCGGGGCAGGCATTTGCGCTAGCTACCA  
TACCGTGTCACTGTTGGCTTCCACGT CACAAAAGTCAATCGTTGCCCTACACGATGAGTCTGGGG  
CTGACTCATCTATCGCCTACAGCAACAATACCATTGCAATTTCCACAAAACCTTCAGTATCTCCATC  
ACAACAGAGGTGATGCCCCGTTTCTATGGCTAAAACATCAGTCGATTGCAATATGTATATATGCGG  
30 CGATAGTACTGAGTGCGCCAATCTCTTGTTACAGTACGGCTCCTTTTGTACCCAGCTGAACCGAG  
CACTGTCTGGAATCGCCGAGAACAGGATCGCAATACCCGGGAAGTCTTCGCCCAGGTGAAGCAG  
ATGTACAAAACGCCCCTCTCAAGTATTTTCGGCGGATTCAACTTTTCTCAGATTTTGCCTGACCC  
GCTCAAGCCAACAAAACGATCTTTTATCGAAGACCTTCTGTTTAAACAGGTACACTGGCGGATG  
CTGGGTTTCATGAAAACAGTACGGTGAATGCCTGGGGGACATCAATGCCAGAGATCTGATCTGCGCC  
35 CAGAAATTCAATGGCTTAACAGTCCTCCCACCTCTCTTGACCGACGATATGATCGCTGCGTACAC  
CGCTGCTCTGGTATCGGGCACCGCGACTGCTGGCTGGACCTTTGGTGCCGGAGCCGCACTCCAGA  
TCCCATTCGCCATGCAGATGGCCTACCGCTTCAACGGAATCGGGGTACCCAGAACGTGCTGTAT  
GAGAACCAGAAACAGATCGCCAATCAGTTCAATAAGGCAATTAGTCAGATTCAGGAGAGTCTTAC  
CACTACCAGCACCGCCCTGGGCAAGCTGCAAGATGTTGTGAACCAGAATGCGCAGGCATTAAACA  
40 CTCTGGTTAAACAGCTGAGCTCAAATTTTGGTGCAATCTCTTCAGTTCTGAACGATATCCTGAGT  
CGGCTGGATCCGCCAGAGGCTGAAGTGCAATTTGATCGTTTGATCACCGGGAGGCTACAATCTCT  
GCAGACGTACGTGACCCAGCAGCTCATCCGGGCAGCCGAAATTCGCGCATCAGCCAACCTCGCTG  
CAACTAAGATGTCTGAGTGCGTGCTGGGCCAGAGTAAGAGGGTGAGCTTTTGTGGTAAGGGATAC  
CACCTCATGTCTTTCCGCAAGCGGCTCCCCACGGCGTGTTTTCTTACACGTTACCTATGTGCC  
45 ATCCCAAGAACGCAATTTACACACCGCTCCAGCTATCTGT CATGAGGGCAAAAGCATATTTCCCCA  
GGGAAGGAGTATTTGTGTTAATGGCACGTCCTGGTTTATAACCCAACGTAACCTTTTCTCCCCA  
CAGATTATCACAACCGACAACACATTCTGTCTGGGAATTGTGACGTCGTGATCGGGATCATTAA  
CAATACCGTTTACGATCCCTTGCAGCCCGAGCTTGACTCCTTTAAAGAGGAACTAGACAAATACT  
TTAAGAATCACACCTCACCGGACGTAGATTTGGGAGACATCTCTGGAATTAATGCCTCTGTGGTG  
AATATCCAGAAGGAGATCGACCGCTGAATGAAGTCGCCAAGAACCCTAACGAGTCCCTGATAGA  
TCTGCAAGAACTGGGCAAATATGAACAGTACATCAAATGGCCGTGGTACGTGTGGTTGGGCTTTA  
TCGCTGGACTTATTGCAATCGTGATGGTGACGATTCTGCTCt gct gct ccaat ggat cgt taca  
tgcagaatttgcattTAA

PDI-SARS-COV-1 H1cCT-DNA (SEQ ID NO: 92)

atggcgaaaaacgttgcgattttcggcttattgttttctcttcttgtgttggttccttctcagat  
cttcgcgTCCGATCTGGATCGGTGCACTACATTTGACGATGTGCAGGCACCTAATTATACTCAGC  
ACACCTCTTCCATGCGGGGCGTGTACTACCCGACGAGATTTTTAGAAAGTGACACACTGTACCTG  
ACCCAAGACCTTTTTCTCCCATTTTTATAGCAATGTCACGGGATTCCACACTATCAATCACACATT  
5 CGGAAACCTGTATCCCCCTTTAAAGACGGTATCTACTTTGCTGCTACTGAGAAATCGAATGTTG  
TTCGCGGTGGGTGTTTCGGCTCAACCATGAACAACAAGAGTCAGTCAGTAATAATTATAAACAAC  
TCAACCAATGTCGTCATCAGGGCTTGCAACTTCGAGCTCTGCGATAACCCCTTCTTTGCGGTGTC  
CAAACCGATGGGCACTCAGACCCATAACCATGATTTTTGACAATGCCTTTAATTGTACTTTTCAAT  
10 ACATCAGCGATGCCCTTCTCGCTGGATGTGTCCGAGAAGAGCGGTAACCTCAAGCATTTGCGGGAG  
TTCGTTTTCAAAAAACAAGACGGGTTTTTGTATGTCTACAAAGGCTATCAGCCCATTGATGTGGT  
CAGGGATCTGCCCAGTGGTTTTCAACACACTGAAGCCCATCTTCAAGTTGCCACTCGGCATCAACA  
TCACTAATTTCCGCGCCATCCTAACTGCTTTTTCGCCAGCGCAGGATATTTGGGGGACATCCGCC  
GCAGCTTACTTCTGTTGGGATATCTGAAGCCCACCCTTTTATGCTAAAGTACGACGAGAATGGCAC  
15 CATCACCGATGCCGTGGACTGCTCACAAGAACCTCTCGCAGAGCTCAAGTGTTCAGTGAAGTCAT  
TTGAGATCGACAAAGGGATCTACCAACATCTAACTTTTCGCGTCTGTCCTTCCGGTGACGTAGTC  
CGTTTCCCAAATATCACTAACTTGTGCCCTTTTGGTGAAGTATTCAACGCAACCAAATTTCCCGAG  
TGTGTATGCGTGGGAGCGCAAAAAGATCAGTAACTGTGTGGCTGACTATAGTGTCTGTACAATA  
GCACCTTCTTACGACCTTCAAGTGTATGGAGTGAGCGCTACAAAACCTGAACGATCTTTGTTTTT  
20 TCAACGTGTACGCCGATTCAATTTGTCTTAAAGGTGACGATGTGAGGCAGATCGCTCCAGGCCA  
GACAGGTGTGATTGCTGACTATAATTACAACTGCCAGACGACTTCATGGGGTGCGTGTAGCTT  
GGAATACAAGAAACATTGACGCCACCTCCACGGGAAATTACAATTACAGTATCGTTACCTTCGC  
CATGGAAAGTTGAGACCCCTTCGAGCGTGATATAAGTAACGTGCCCTTTAGTCCAGATGGAAAACC  
CTGCACACCCCTGCTCTCAATTGCTATTGGCTCTCAATGACTACGGCTTTTACACAACACTG  
25 GCATCGGATACCGCCTTACCGGGTCTGTTGCTCAGTTTTGAGTTGCTTAACGCACCCGCCACC  
GTGTGTGGTCTTAACTTTCTACTGACCTGATTAATAAACCAATGCGTCAACTTCAATTTTAACGG  
GCTGACCGGCACCGGTGTCTGACCCCTAGCTCTAAGAGATTCCAGCCTTTTCAGCAGTTCGGGA  
GGGATGTGAGCGACTTTACCGACTCTGTGAGGGATCCAAAGACCAGCGAGATACTGGATATCTCG  
CCCTGCGAGTTTCGGTGGCGTGTCCGTTATTACACCTGGCACCAACGCCCTCCTCAGAGGTGGCGGT  
30 GCTCTATCAAGATGTCAACTGCACTGATGTGTCAACTGCCATCCATGCCGATCAGCTGACCCCG  
CCTGGCGCATCTACAGTACCGGGAACAACGTTTTTCAGACCCAGGCCGGCTGTCTAATCGGCGCA  
GAGCAGTTGACACATCCTACGAATGTGACATACCTATCGGGGCAGGCATTTGCGCTAGCTACCA  
TACCGTGTCACTGTTGGCTTCCACGTCAAAAAGTCAATCGTTGCCATACAGTATCTGCGGGG  
CTGACTCATCTATCGCCTACAGCAACAATACCATTGCAATTCCCACAACTTCAGTATCTCCATC  
35 ACAACAGAGGTGATGCCCCGTTTTCTATGGCTAAAACATCAGTCGATTGCAATATGTATATATGCGG  
CGATAGTACTGAGTGCGCCAATCTCTTGTACAGTACGGCTCCTTTTGTACCCAGCTGAACCGAG  
CACTGTCTGGAATCGCCGAGAACAGGATCGCAATACCCGGGAAGTCTTCGCCCAGGTGAAGCAG  
ATGTACAAAACGCCCACTCTCAAGTATTTTCGGCGGATTCAACTTTTCTCAGATTTTGCCTGACCC  
GCTCAAGCCAACAAAACGATCTTTTATCGAAGACCTTCTGTTTAAACAGGTACACTGGCGGATG  
40 CTGGGTTTCATGAAAACAGTACGGTGAATGCCTGGGGGACATCAATGCCAGAGATCTGATCTGCGCC  
CAGAAATTCAATGGCTTAACAGTCTCTCCACCTCTCTTGACCGACGATATGATCGCTGCGTACAC  
CGCTGCTCTGGTATCGGGCACCGCGACTGCTGGCTGGACCTTTGGTGCCGGAGCCGCACTCCAGA  
TCCCATTCGCCATGCAGATGGCCTACCGCTTCAACGGAATCGGGGTACCCAGAACGTGCTGTAT  
GAGAACCAGAAACAGATCGCCAATCAGTTCAATAAGGCAATTAGTCAGATTTCAGGAGAGTCTTAC  
CACTACCAGCACCGCCCTGGGCAAGCTGCAAGATGTTGTGAACCAGAAATGCGCAGGCATTAAACA  
45 CTCTGGTTAAACAGCTGAGCTCAAATTTTGGTGCAATCTCTTCAGTTCTGAACGATATCCTGAGT  
CGGCTGGATCCGCCAGAGGCTGAAGTGCAATTTGATCGTTTGTATCACCGGGAGGCTACAATCTCT  
GCAGACGTACGTGACCCAGCAGCTCATCCGGGCAGCCGAAATTCGCGCATCAGCCAACCTCGCTG  
CAACTAAGATGTCTGAGTGCCTGCTGGGCCAGAGTAAGAGGGTGGAATTTTGTGGTAAGGGATAC  
CACCTCATGTCTTTCCGCAAGCGGCTCCCCACGGCGTGGTTTTTCTTACAGTTACCTATGTGCC  
50 ATCCCAAGAACGCAATTTACACACGCTCCAGCTATCTGTCATGAGGGCAAGCATATTTCCCCA  
GGGAAGGAGTATTTGTGTTTAAATGGCACGTCCTGGTTTTATAACCAACGTAACCTTTTTCTCCCA  
CAGATTATCACAAACGACAACACATTCTGTCTGGGAATTGTGACGTCGTGATCGGGATCATTAA

5 CAATACCGTTTACGATCCCTTGCAGCCCGAGCTTGACTCCTTTAAAGAGGAACTAGACAAATACT  
TTAAGAATCACACCTCACCGGACGTAGATTTGGGAGACATCTCTGGAATTAATGCCTCTGTGGTG  
AATATCCAGAAGGAGATCGACCGCTGAATGAAGTCGCCAAGAACCTCAACGAGTCCCTGATAGA  
TCTGCAAGAACTGGGCAAATATGAACAGTACATCAAATGGCCGTGGTACGTGTGGTTGGGCTTTA  
TCGCTGGACTTATTGCAATCGTGATGGTGACGATTCTGCTCagcttctggatgtgctctaattggg  
tctctacagtgtagaatatgtattTAA

**PDI-SARS-COV-1 wtTMCT-AA (SEQ ID NO: 93)**

10 MAKNVAIFGLLFSLLVLPVPSQIFASDLDRCTTFDDVQAPNYTQHTSSMRGVVYPDEIFRSDTLYL  
TQDLFLPFYSNVTGFHTINHTFGNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQSVIIINN  
STNVVIRACNFELCDNPFFAVSKPMGTQTHMTI FDNAFNCTFEYISDAFSLDVSEKSGNFKHLRE  
FVFNKNDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINITNFRAILTAFSPAQDIWGTS  
AAYFVGYLKPTTFMLKYDENGITDAVDCSQNPLAELKCSVKSEIDKGIYQTSNFRVVPSPGDVV  
15 RFPNITNLCPFGEVFNATKFPVSVAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNLDLFC  
SNVYADSFVVKGDDVRQIAPGQTGVIADYNYKL PDDFMGCVLAWNTRNIDATSTGNYNKYRYLR  
HGKLRPFERDISNVFPSPDGKPCPPALNCYWPLNDYGFYTTTGIGYQPYRVVLSFELLNAPAT  
VCGPKLSTDLIKNCVNFNFNGLTGTGVLTPSSKRFQPFQFGRDVSDFTDVSRDPKTSEILDIS  
PCSFGGVSVITPGTNASSEVAVLYQDVNCTDVSTAIHADQLTPAWRIYSTGNVVFQTQAGCLIGA  
EHVDTSECDIPIGAGICASYHTVSLLASTSQKSIVAYTMSLGADSSIAYSNNTIAIPTNFSISI  
20 TTEVMPVSMAKTSVDCNMYICGDSTECANLLLQYGSFCTQLNRALSGIAAEQDRNTREVFAQVKQ  
MYKTPTLKYFGGFNFSQILPDPLKPTKRSFIEDLLFNKVTLADAGFMKQYGECLGDINARDLICA  
QKFNGLTVLPLLLTDDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLY  
ENQKQIANQFNKAISQIQESLTTTSTALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILS  
RLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGY  
25 HLMSFPQAAPHGVVFLHVTYVPSQERNFTTAPAICHEGKAYFPREGVVFVNGTSWFITQRNFFSP  
QIITTDNTFVSGNCDVVIGIINNTVYDPLQPELDSFKEELDKYFKNHTSPDVLGDISGINASVV  
NIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYVWLGFIAGLIAIVMTILLCCMTSCCS  
CLKGACSCGSCCKFDEDDSEPVLKGVKLHYT

30 **PDI-SARS-COV-1 H5iTMCT-AA (SEQ ID NO: 94)**

MAKNVAIFGLLFSLLVLPVPSQIFASDLDRCTTFDDVQAPNYTQHTSSMRGVVYPDEIFRSDTLYL  
TQDLFLPFYSNVTGFHTINHTFGNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQSVIIINN  
STNVVIRACNFELCDNPFFAVSKPMGTQTHMTI FDNAFNCTFEYISDAFSLDVSEKSGNFKHLRE  
FVFNKNDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINITNFRAILTAFSPAQDIWGTS  
35 AAYFVGYLKPTTFMLKYDENGITDAVDCSQNPLAELKCSVKSEIDKGIYQTSNFRVVPSPGDVV  
RFPNITNLCPFGEVFNATKFPVSVAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNLDLFC  
SNVYADSFVVKGDDVRQIAPGQTGVIADYNYKL PDDFMGCVLAWNTRNIDATSTGNYNKYRYLR  
HGKLRPFERDISNVFPSPDGKPCPPALNCYWPLNDYGFYTTTGIGYQPYRVVLSFELLNAPAT  
VCGPKLSTDLIKNCVNFNFNGLTGTGVLTPSSKRFQPFQFGRDVSDFTDVSRDPKTSEILDIS  
40 PCSFGGVSVITPGTNASSEVAVLYQDVNCTDVSTAIHADQLTPAWRIYSTGNVVFQTQAGCLIGA  
EHVDTSECDIPIGAGICASYHTVSLLASTSQKSIVAYTMSLGADSSIAYSNNTIAIPTNFSISI  
TTEVMPVSMAKTSVDCNMYICGDSTECANLLLQYGSFCTQLNRALSGIAAEQDRNTREVFAQVKQ  
MYKTPTLKYFGGFNFSQILPDPLKPTKRSFIEDLLFNKVTLADAGFMKQYGECLGDINARDLICA  
QKFNGLTVLPLLLTDDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLY  
45 ENQKQIANQFNKAISQIQESLTTTSTALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILS  
RLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGY  
HLMSFPQAAPHGVVFLHVTYVPSQERNFTTAPAICHEGKAYFPREGVVFVNGTSWFITQRNFFSP  
QIITTDNTFVSGNCDVVIGIINNTVYDPLQPELDSFKEELDKYFKNHTSPDVLGDISGINASVV

NIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYQILSIYSTVASSLALAIMMAGLSLWMC  
SNGSLQCRICI

**PDI-SARS-COV-1 H5iCT-AA (SEQ ID NO: 95)**

5 MAKNVAIFGLLFSLLVLVPSQIFASDLDRCTTFDDVQAPNYTQHTSSMRGVVYPDEIFRSDTLYL  
TQDLFLFPYSNVTGFHTINHTFGNPVIFPKDGIYFAATEKSNVVRGWVFGSTMNNKSQSVIIINN  
STNVVIRACNFELCDNPFPAVSKPMGTQTHMTI FDNAFNCTFEYISDAFSLDVSEKSGNFKHLRE  
FVFKNKDGLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINITNFRAILTAFSPAQDIWGTSA  
AAYFVGYLKPTTFMLKYDENGITDAVDCSQNPLAELKCSVKSEIDKGIYQTSNFRVVPSPGDVV  
RFPNITNLCPFGEVFNATKFPVYAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNLDLCF  
10 SNVYADSFVVGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLR  
HGKLRPFERDISNVFPSPDGKPCPPALNCYWPLNDYGFTTTGIGYQPYRVVLSFELLNAPAT  
VCGPKLSTDLIKNCVNFENGLTGTGVLT PSSKRFQPFQFGRDVSDFTD SVRDPKTSEILDIS  
PCSFGGVSVITPGTNASSEVAVLYQDVNCTDVSTAIHADQLTPAWRIYSTGNNVFQTQAGCLIGA  
EHVDSYECDIPIGAGICASYHTVSLLASTSQKSIVAYTMSLGADSSIAYSNTTIAIPTNFSISI  
15 TTEVMPVSMAKTSVDCNMYICGDSTECANLLLQYGSFCTQLNRALSGIAAEQDRNTREVFAQVKQ  
MYKTPTLKYFGGFNFSQILPDPLKPTKRSFIEDLLFNKVTLADAGFMKQYGECLGDINARDLICA  
QKFNGLTVLPLLTDDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLY  
ENQKQIANQFNKAISQIQESLTTTSTALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILS  
RLDPPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGY  
20 HLMSFPQAAPHGVVFLHVTVVPSQERNFTTAPAICHEGKAYFPREGVVFVNGTSWFITQRNFFSP  
QIITTDNTFVSGNCDVVIGIINNTVYDPLQPELDSFKEELDKYFKNHTSPDVLGDISGINASVV  
NIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYVWLGFIAGLIAIVMTILLSLWMCNG  
SLQCRICI

**PDI-SARS-COV-1 H5iCT(V4)-AA (SEQ ID NO: 96)**

25 MAKNVAIFGLLFSLLVLVPSQIFASDLDRCTTFDDVQAPNYTQHTSSMRGVVYPDEIFRSDTLYL  
TQDLFLFPYSNVTGFHTINHTFGNPVIFPKDGIYFAATEKSNVVRGWVFGSTMNNKSQSVIIINN  
STNVVIRACNFELCDNPFPAVSKPMGTQTHMTI FDNAFNCTFEYISDAFSLDVSEKSGNFKHLRE  
FVFKNKDGLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINITNFRAILTAFSPAQDIWGTSA  
AAYFVGYLKPTTFMLKYDENGITDAVDCSQNPLAELKCSVKSEIDKGIYQTSNFRVVPSPGDVV  
30 RFPNITNLCPFGEVFNATKFPVYAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNLDLCF  
SNVYADSFVVGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLR  
HGKLRPFERDISNVFPSPDGKPCPPALNCYWPLNDYGFTTTGIGYQPYRVVLSFELLNAPAT  
VCGPKLSTDLIKNCVNFENGLTGTGVLT PSSKRFQPFQFGRDVSDFTD SVRDPKTSEILDIS  
PCSFGGVSVITPGTNASSEVAVLYQDVNCTDVSTAIHADQLTPAWRIYSTGNNVFQTQAGCLIGA  
35 EHVDSYECDIPIGAGICASYHTVSLLASTSQKSIVAYTMSLGADSSIAYSNTTIAIPTNFSISI  
TTEVMPVSMAKTSVDCNMYICGDSTECANLLLQYGSFCTQLNRALSGIAAEQDRNTREVFAQVKQ  
MYKTPTLKYFGGFNFSQILPDPLKPTKRSFIEDLLFNKVTLADAGFMKQYGECLGDINARDLICA  
QKFNGLTVLPLLTDDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLY  
ENQKQIANQFNKAISQIQESLTTTSTALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILS  
40 RLDPPAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGY  
HLMSFPQAAPHGVVFLHVTVVPSQERNFTTAPAICHEGKAYFPREGVVFVNGTSWFITQRNFFSP  
QIITTDNTFVSGNCDVVIGIINNTVYDPLQPELDSFKEELDKYFKNHTSPDVLGDISGINASVV  
NIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYVWLGFIAGLIAIVMTILLCCSNGSLQ  
CRICI

**PDI-SARS-COV-1 H1cCT-AA (SEQ ID NO: 97)**

MAKNVAIFGLLFSLLVLPVPSQIFASDLDRCTTFDDVQAPNYTQHTSSMRGVVYPDEIFRSDTLYL  
 TQDLFLPFYSNVTGFHTINHTFGNPVIFPKDGIYFAATEKSNVVRGWVFGSTMNNKSQSVIIINN  
 STNVVIRACNFELCDNPFFAVSKPMGTQTHMTIFDNAFNCTFEYISDAFSLDVSEKSGNFKHLRE  
 FVFKNKDGLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINITNFRAILTAFSPAQDIWGTSA  
 5 AAYFVGYLKPTTFMLKYDENGITDAVDCSQNPLAELKCSVKSEIDKGIYQTSNFRVVPSPGDVV  
 RFPNITNLCPFGEVFNATKFPSVYAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNLDLCF  
 SNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLR  
 HGKLRPFERDISNVPFSPDGKPCPPALNCYWPLNDYGFYTTTGIGYQPYRVVLSFELLNAPAT  
 10 VCGPKLSTDLIKNQCVNFNFNGLTGTGVLT PSSKRFQPFQFGRDVSDFTD SVRDPKTSEILDIS  
 PCSFGLSVITPGTNASSEVAVLYQDVNCTDVSTAIHADQLTPAWRIYSTGNNVFQTQAGCLIGA  
 EHVDSYECDIPIGAGICASYHTVSLLASTSQKSIVAYTMSLGADSSIAYSNNITIAIPTNFSISI  
 TTEVMPVSMAKTSVDCNMYICGDSTECANLLQYGSFCTQLNRALSGIAAEQDRNTREVFAQVKQ  
 MYKTPTLKYFGGFNFSQILPDPLKPTKRSFIEDLLFNKVTLADAGFMKQYGECLGDINARDLICA  
 15 QKFNGLTVLPLLTDDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLY  
 ENQKQIANQFNKAISQIQESLTTTSTALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILS  
 RLDPEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGY  
 HLMSFPQAAPHGVVFLHVITYVPSQERNFTTAPAICHEGKAYFPREGVFVFNGTSWFITQRNFFSP  
 QIITTDNTFVSGNCDVVIGIINNTVYDPLQPELDSFKEELDKYFKNHTSPDVLGDISGINASVV  
 20 NIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYVWLGFIAGLIAIVMTILLSFWMCSNG  
 SLQCRICI

**IF(AvB+wtCT-MERS).r (SEQ ID NO: 98)**

ACGACACGACTAAGGCCTTCAGTGAACGTGGACCTTGTGAGGCTCAAGGTCATACTCCTC

**IF(H1cCT-wtTM).r (SEQ ID NO: 99)**

25 ACGACACGACTAAGGCCTTCAAATACATATTCTACACTGTAGAGACCCA

**IF(H5ITMCT).r (SEQ ID NO: 100)**

ACGACACGACTAAGGCCTTCAAATGCAAATTCTGCATTGTAACGATCC

30 **PDI-MERS-wtTMCT-DNA (SEQ ID NO: 101)**

atggcgaaaaaacgttgcgattttcggcttattgttttctcttcttctgtgttggttccttctcagat  
 cttcgcgTATGTCGATGTGGGTCCCGATAGTGTTAAGTCCGCCTGCATCGAAGTGGACATTACAGC  
 AGACCTTCTTCGATAAGACTTGGCCTCGGCCAATTGATGTGTCCAAGGCCGACGGCATTATCTAC  
 CCCCAGGTTCGGACATATTCCAACATAACTATCACCTATCAGGGGCTATTCCCTTATCAGGGCGA  
 35 CCATGGGGACATGTACGTTTACAGCGCTGGTCACGCTACAGGGACGACCCCCAGAAAGCTCTTCG  
 TGGCGAACTATAGTCAGGACGTGAAACAGTTTGCCAACGGTTTTGTAGTGCATCGGGGACGCC  
 GCTAACTCCACTGGTACTGTTATTATCAGCCCTCCACGAGTGCCACAATTCGAAAGATCTATCC  
 GGCCTTCATGCTAGGATCCTCTGTGGGCAATTTAGCGACGGTAAGATGGGTGCGTTCTTCAACC  
 ACACGCTTGTGCTGCTTCCCGATGGGTGCGGTACTTTGCTGAGGGCCTTTTACTGTATCCTAGAG  
 40 CCCCATCCGGCAACCACTGCCCCGCCGGGAACCTCGTATACTTCCTTTGCCACTTATCATACTCC  
 AGCCACGGATTGTAGCGATGGGAACACAATAGGAACGCCAGTTTGAATTCCTTTAAAGAGTACT  
 TCAACTTGCGGAATTGTACCTTCATGTATACATATAACATTACTGAGGACGAAATCTCGAATGG  
 TTCGGAATCACTCAAACAGCCCAGGGAGTGCACCTCTTTAGTTCTCGCTATGTGGACTTATATGG  
 AGGCAATATGTTTCAATTCGCCACCTTACCGTCTACGATACGATCAAGTATTACTCGATCATAC

5 CCCACTCCATTAGGTCCATTTCAGAGCGATCGCAAGGCATGGGCCGCATTCTATGTGTATAAGCTC  
CAGCCCCTGACCTTCCTCTTGGATTTCTCCGTGGACGGCTACATCAGAAGGGCTATCGATTGCGG  
GTTCAACGACCTCAGCCAGCTGCATTGTTCTTATGAGAGCTTTGACGTGGAAGCGGAGTTTACT  
CAGTCTCTTCCTTTGAGGCTAAACCTTCAGGTAGCGTCGTAGAGCAAGCAGAGGGTGTGGAGTGC  
10 GATTTCTCACCCTGCTCAGCGGAACCCACCCAGGTCTACAACCTTTAAGCGGCTCGTGTTTAC  
AACTGTAACCTATAACTTGACTAAGTTGCTGTCACCTCTTTCCGTGAATGATTTTACATGCTCCC  
AAATCAGCCCAGCCGCTATTGCGTCTAATTGCTATTCCTCATTGATCCTGGATTACTTCAGTTAC  
CCCCTCTCTATGAAGAGCGATCTCTCGGTTAGTAGCGCTGGGCCATTTCCTCAGTTTAACTACAA  
15 ACAATCCTTTTCCAATCCAACATGCCTGATCTTAGCTACTGTACCCACAACTGACTACTATTA  
CGAAGCCACTCAAGTACTCATACTAATAAAGTGCAGCCGATTCCCTCAGTGATCGCACCAGAA  
GTGCCGCAGCTTGTAACCGGAACAGTACTCCCCATGCGTCTCTATTGTGCCTTCTACAGTGTG  
GGAAGACGGCGATTATTATAGAAAGCAGCTGTCGCCACTGGAAGGTGGCGGGTGGCTAGTTGCCA  
GTGGGTCCACAGTTGCCATGACCGAGCAACTTCAGATGGGGTTTGGCATAACAGTGCAGTATGGT  
20 ACCGATACGAACAGCGTGTGTCCAAAATTGGAATTTGCTAACGACACCAAGATCGCCTCCAGTT  
GGGAAATTGTGTTGAATATTCCCTGTACGGAGTGTGAGGCCGGGGGGTGTTCAAAATTGCACCG  
CCGTGGGAGTGAGGCAGCAAAGATTTCGTGTACGACGCATACCAGAACTAGTCCGATACTATTCT  
GACGATGGAAACTACTACTGTCTGCGCGCTTGCGTCTCAGTGCCCGTGAGTGTGCATATATGATAA  
GGAGACCAAGACTCACGCTACTCTCTTTGGTTCTGTGCGGTGCGAACACATTTCTCTACAATGT  
25 CCCAGTATAGTTCGCTCCACTCGGTCTATGTTAAAGCGCAGAGACAGTACCTACGGCCCTCTACAG  
ACACCTGTGGGGTGCGTTCTCGGCCCTTGTCATTTAGCCTGTTTGTGGAGGATTGTAAGCTGCC  
CCTTGGTCAAAGCTTATGCGCACTGCCGATACGCCAGCACACTTACACCAGCTTCAGTGGGGT  
CCGTCCCCGGGGAAATGAGATTGGCCTCGATCGCTTTCAACCACCCCATACAGGTGGATCAGCTC  
AACTCGTCATACTTCAAGCTAAGCATCCCTACTAATTTCTCCTTTGGTGTGACTCAGGAGTACAT  
30 TCAGACCACAATTCAAAGGTGACCGTTGACTGCAAGCAGTATGTGTGCAACGGGTTCCAGAAAT  
GTGAACAGCTGCTCCGGGAGTATGGCCAGTTCTGTTCTAAAATCAACCAGGCCCTCCACGGAGCA  
AACCTTAGGCAGGACGATTCTGTGAGAAACCTCTTTGCCAGCGTCAAGAGTTCTCAGAGTTCCCC  
TATTATACCTGGCTTCGGCGGGGATTTCAACCTGACACTACTTGAACCTGTAAGCATATCAACCG  
GAAGTCGCAGTGCCCGTTCCGCCATCGAGGATCTGCTCTTCGACAAAGTAACTATTGCAGATCCC  
GGATACATGCAGGGGTATGACGACTGCATGCAGCAGGGTCCAGCCTCTGCAAGGGATCTGATATG  
35 CGCACAGTATGTGCTGGGTACAAAGTGTTGCCTCCTCTCATGGACGTGAACATGGAAGCGGCCCT  
ATACCTCCTCACTTCTAGGCTCCATAGCGGGCGTGGGATGGACCGCAGGGCTTTCAAGCTTCGCC  
GCAATTCCTTTGCTCAATCTATCTTCTACAGGCTTAATGGCGTTGGAATCAACCCAGCAGGTGTT  
AAGCGAAAACAGAAATTGATTGCCAATAAGTTTAACCAAGCTTTGGGGCCATGCAGACAGGCT  
TTACAACCACAAACGAGGCTTTCCATAAAGTACAGGATGCGGTAACAATAACGCACAAGCCCTG  
40 TCAAAGCTGGCTTCAGAGCTCTCAAATACATTTGGCGCTATATCCGCGTCTATCGGCGATATCAT  
ACAACGGTTGGACCCACCCGAACAGGACGCACAGATTGATCGTTTGATCAACGGGAGGCTTACCA  
CCTTAAACGCTTTTGTGGCCAGCAACTGGTGGCGTCTGAGAGCGCCGCTTGAGCGCTCAGCTG  
GCAAAGGATAAAGTGAATGAATGCGTGAAAGCTCAATCAAAGAGAAGTGGGTTTTGTGGGCAGGG  
TACTCATATTGTTTCTTTGTGGTGAACGCCCAATGGACTCTACTTTATGCATGTTGGATACT  
45 ACCCGAGCAACCACATCGAGGTGCTTTCCGCCTATGGGCTTTGTGACGCAGCAAACCTACTAAC  
TGTATCGCGCCAGTTAATGGCTACTTTATTAACAATAACACACGCATTGTGGATGAATGGAG  
TTACACAGGGTCCAGCTTCTACGCTCCAGAGCCTATCACCTCTCTGAACACAAAGTATGTGGCAC  
CTCAGGTCACATATCAGAACATCTCGACAACCTGCCCGCCCACTCTTGGGCAACTCCACAGGG  
ATCGACTTCCAGGACGAGCTTGACGAATTCTTCAAGAACGTGTCCACCAGTATCCCTAATTTTGG  
50 TTCGCTGACCCAAATTAACACAACCCTGCTCGATCTGACATATGAAATGCTTTCACTACAGCAGG  
TGGTCAAAGCGTTGAACGAGTCGTATATCGACCTGAAAGAGTTAGGGAATTACACATACTATAAC  
AAATGGCCCTGGTATATTTGGTTAGGATTCATTGCCGGGCTGGTGGCCCTTGCTTGTGCGTATT  
TTTCATCTTGTGCTGTACCGGTTGCGGTACGAATTGCATGGGAAAACGAAATGTAATCGGTGCT  
GCGATCGCTATGAGGAGTATGACCTTGAGCCTCACAAAGTCCACGTTCACTGA

PDI-MERS-H5iTMCT-DNA (SEQ ID NO: 102)

atggcgaaaaacgttgcgattttcggcttattgttttctcttcttgtgttggttccttctcagat  
cttcgcgTATGTCGATGTGGGTCCCGATAGTGTTAAGTCCGCCTGCATCGAAGTGACATTACAGC  
AGACCTTCTTCGATAAGACTTGGCCTCGGCCAATTGATGTGTCCAAGGCCGACGGCATTATCTAC  
CCCCAAGGTCGGACATATTCCAACATAACTATCACCTATCAGGGGCTATTCCTTATCAGGGCGA  
5 CCATGGGGACATGTACGTTTACAGCGCTGGTCACGCTACAGGGACGACCCCCAGAAGCTCTTCG  
TGGCGAACTATAGTCAGGACGTGAAACAGTTTGCCAACGGTTTTGTAGTGCATCGGGGACGCC  
GCTAACTCCACTGGTACTGTTATTATCAGCCCTTCCACGAGTGCCACAATTGCAAAGATCTATCC  
GGCCTTCATGCTAGGATCCTCTGTGGGCAATTTTAGCGACGGTAAGATGGGTTCGTTCTTCAACC  
10 ACACGCTTGTGCTGCTTCCCGATGGGTGCGGTACTTTGCTGAGGGCCTTTTACTGTATCCTAGAG  
CCCCGATCCGGCAACCACTGCCCCGCCGGGAACCTCGTATACTTCCCTTGGCACTTATCATACTCC  
AGCCACGGATTGTAGCGATGGGAACACAATAGGAACGCCAGTTTGAATTCCTTTAAAGAGTACT  
TCAACTTGCGGAATTGTACCTTCATGTATACATATAACATTACTGAGGACGAAATTCCTGAATGG  
TTCGGAATCACTCAAACAGCCAGGGAGTGCACCTCTTTAGTTCTCGCTATGTGGACTTATATGG  
AGGCAATATGTTTCAATTGCGCACCTTACCCGTCTACGATACGATCAAGTATTACTCGATCATAC  
15 CCCACTCCATTAGGTCCATTACAGAGCGATCGCAAGGCATGGGCCGCATTCTATGTGTATAAGCTC  
CAGCCCTGACCTTCCCTCTTGGATTTCTCCGTGGACGGCTACATCAGAAGGGCTATCGATTGCGG  
GTTCAACGACCTCAGCCAGCTGCATTGTTCTTATGAGAGCTTTGACGTGGAAGCGGAGTTTACT  
CAGTCTCTTCTTTGAGGCTAAACCTTCAGGTAGCGTCGTAGAGCAAGCAGAGGGTGTGGAGTGC  
GATTTCTCACCACTGCTCAGCGGAACCCACCCAGGTCTACAACCTTAAAGCGGCTCGTGTTCAC  
20 AAATCAGCCCAGCCGCTATTGCGTCTAATTGCTATTTCCTCATTGATCCGATTACTTCAGTTAC  
CCCCCTCTCTATGAAGAGCGATCTCTCGGTTAGTAGCGCTGGGCCTATTTCCAGTTTAACTACAA  
ACAATCCTTTTCCAATCCAACATGCCTGATCTTAGCTACTGTACCCCAACCTGACTACTATTA  
CGAAGCCACTCAAGTACTCATACATTAATAAGTGCAGCCGATTCCCTCAGTGATGATCGCACCGAA  
25 GTGCCGCAGCTTGTAACCGGAACCACTACTCCCATGCGTCTCTATTGTGCCTTCTACAGTGTG  
GGAAGACGGCGATTATTATAGAAAGCAGCTGTGCCACTGGAAGGTGGCGGGTGGCTAGTTGCCA  
GTGGGTCCACAGTTGCCATGACCGAGCAACTTCAGATGGGGTTTGGCATAACAGTGCAGTATGGT  
ACCGATACGAACAGCGTGTGTCCAAAATTGGAATTTGCTAACGACACCAAGATCGCCTCCAGTT  
GGGAAATTGTGTTGAATATTCCCTGTACGGAGTGTGAGGCCGGGGGGTGTTCAAAATTGCACCG  
30 CCGTGGGAGTGAGGCAGCAAAGATTTCGTGTACGACGCATACCAGAATCTAGTCGGATACTATTCT  
GACGATGGAACTACTACTGTCTGCGCGCTTGCCTCTCAGTGCCCGTGAGTGTATATATGATAA  
GGAGACCAAGACTCACGCTACTCTTTGGTTCTGTGCGGTGCGAACACATTTCCCTTACAATGT  
CCCAGTATAGTCGCTCCACTCGGTCTATGTTAAAGCGCAGAGACAGTACCTACGGCCCTCTACAG  
ACACCTGTGGGGTGCCTTCTCGGCCTTGTCAATTCTAGCCTGTTTGTGGAGGATTGTAAGCTGCC  
35 CCTTGGTCAAAGCTTATGCGCACTGCCGATACGCCAGCACACTTACACCAGCTTCAGTGGGGT  
CCGTCCCCGGGAAATGAGATTGGCCTCGATCGCTTCAACCACCCCATACAGGTGGATCAGCTC  
AACTCGTCATACTTCAAGCTAAGCATCCCTACTAATTTCTCCTTTGGTGTGACTCAGGAGTACAT  
TCAGACCACAATTCAAAAGGTGACCGTTGACTGCAAGCAGTATGTGTGCAACGGGTTCAGAAAT  
GTGAACAGCTGCTCCGGGAGTATGGCCAGTTCTGTTCTAAAATCAACCAGGCCCTCCACGGAGCA  
40 AACCTTAGGCAGGACGATTCTGTGAGAAACCTCTTTGCCAGCGTCAAGAGTTCTCAGAGTTCCCC  
TATTATACCTGGCTTCGGCGGGGATTTCAACCTGACACTACTTGAACCTGTAAGCATATCAACCG  
GAAGTCGAGTGCCCGTTCCGCCATCGAGGATCTGCTCTTCGACAAAGTAACTATTGCAGATCCC  
GGATACATGCAGGGGTATGACGACTGCATGCAGCAGGTCAGCCTCTGCAAGGGATCTGATATG  
CGCACAGTATGTGCTGGGTACAAAGTGTTGCCTCCTCTCATGGACGTGAACATGGAAGCGGCCT  
45 ATACCTCCTCACTTCTAGGCTCCATAGCGGGCGTGGGATGGACCGCAGGGCTTTCAAGCTTCGCC  
GCAATTCCTTTGCTCAATCTATCTTCTACAGGCTTAATGGCGTTGGAATCACCCAGCAGGTGTT  
AAGCGAAAACAGAAATTGATTGCCAATAAGTTTAACCAAGCTTTGGGGCCATGCAGACAGGCT  
TTACAACCACAAACGAGGCTTTCCATAAAGTACAGGATGCGGTAAACAATAACGCACAAGCCCTG  
TCAAAGCTGGCTTCAGAGCTCTCAAATACATTTGGCGCTATATCCGCGTCTATCGGCGATATCAT  
50 ACAACGTTGGACCCACCCGAACAGGACGCACAGATTGATCGTTTGATCAACGGGAGGCTTACCA  
CCTTAAACGCTTTTGTGGCCCAGCAACTGGTGCGGTCTGAGAGCGCCGCTTGAGCGCTCAGCTG  
GCAAAGGATAAAGTGAATGAATGCGTGAAAGCTCAATCAAAGAGAAGTGGGTTTTGTGGGCAGGG

5 TACTCATATTGTTTCCTTTGTGGTGAACGCCCCAAATGGACTCTACTTTATGCATGTTGGATACT  
ACCCGAGCAACCACATCGAGGTCGTTTCCGCCTATGGGCTTTGTGACGCAGCAAACCTACTAAC  
TGTATCGCGCCAGTTAATGGCTACTTTATTAAAACAAATAACACACGCATTGTGGATGAATGGAG  
TTACACAGGGTCCAGCTTCTACGCTCCAGAGCCTATCACCTCTCTGAACACAAAAGTATGTGGCAC  
CTCAGGTCACATATCAGAACATCTCGACAAACCTGCCCCCCCCACTCTTGGGCAACTCCACAGGG  
ATCGACTTCAGGACGAGCTTGACGAATTCCTCAAGAACGTGTCCACCAGTATCCCTAATTTTGG  
TTCGTGACCCAAATTAACACAACCCTGCTCGATCTGACATATGAAATGCTTTCCTACAGCAGG  
TGGTCAAAGCGTTGAACGAGTCGTATATCGACCTGAAAGAGTTAGGGAATTACACATACTATAAC  
10 AAATGGCCCTGGTATcaataactgtcaatttattcaacagtggcgagttccctagcactggcaat  
catgatggctggtctatctttatggatgtgctccaatggatcgttacaaatgcagaatttgcattT  
GA

## PDI-MERS-H5iCT-DNA (SEQ ID NO: 103)

15 atggcgaaaaacgttgcgatthttcggttattgttttctcttcttgtgttggttccttctcagat  
cttcgctTATGTCGATGTGGGTCCCAGTAGTGTAAAGTCCGCCTGCATCGAAGTGGACATTACAGC  
AGACCTTCTTCGATAAGACTTGGCCTCGGCCAATTGATGTGTCCAAGGCCGACGGCATTATCTAC  
CCCCAAGGTCGGACATATTCCAACATAACTATCACCTATCAGGGGCTATTCCTTATCAGGGCGA  
CCATGGGGACATGTACGTTTACAGCGCTGGTCACGCTACAGGGACGACCCCCAGAAGCTCTTCG  
TGGCGAACTATAGTCAGGACGTGAAACAGTTTGCCAACGGTTTTGTAGTCCGCATCGGGGCAGCC  
20 GCTAACTCCACTGGTACTGTTATTATCAGCCCTTCCACGAGTGCCACAATTTCGAAAGATCTATCC  
GGCCTTCATGCTAGGATCCTCTGTGGGCAATTTTAGCGACGGTAAGATGGGTTCGGTTCTTCAACC  
ACACGCTTGTGCTGCTTCCCGATGGGTGCGGTACTTTGCTGAGGGCCTTTTACTGTATCCTAGAG  
CCCCGATCCGGCAACCACTGCCCCGCCGGGAACCTCGTATACTTCCTTTGCCACTTATCATACTCC  
AGCCACGGATTGTAGCGATGGGAACACAATAGGAACGCCAGTTTGAATTCCTTTAAAGAGTACT  
TCAACTTGCAGGAATTGTACCTTCATGTATACATATAACATTACTGAGGACGAAATTCTCGAATGG  
25 TTCGGAATCACTCAAACAGCCCAGGGAGTGCACCTCTTTAGTTCTCGCTATGTGGACTTATATGG  
AGGCAATATGTTTCAATTTCGCCACCTTACCCGTCTACGATACGATCAAGTATTACTCGATCATAAC  
CCCACTCCATTAGGTCCATTACAGAGCGATCGCAAGGCATGGGCCGCAATTCTATGTGTATAAGCTC  
CAGCCCCTGACCTTCTCTTGGATTTCTCCGTGGACGGCTACATCAGAAGGGCTATCGATTGCGG  
GTTCAACGACCTCAGCGAGCTGCATTTGTTCTTATGAGAGCTTTGACGTGGAAGCGGAGTTTACT  
30 AGTCTCTTCTTCTTTGAGGCTAAACCTTCAGGTAGCGTCTGATAGCAAGCAGAGGGTGTGGAGTGC  
GATTTCTCACCCTGCTCAGCGGAACCCCCACCCAGGTCTACAACCTTTAAGCGGCTCGTGTTTAC  
AACTGTAACTATAACTTGACTAAGTTGCTGTCACTCTTTTCCGTGAATGATTTTACATGCTCCC  
AAATCAGCCCAGCCGCTATTGCGTCTAATTGCTATTCCTCATTGATCCTGGATTACTTCAGTTAC  
CCCCTCTCTATGAAGAGCGATCTCTCGGTTAGTAGCGCTGGGCCTATTTCCAGTTTAACTACAA  
35 ACAATCCTTTTCCAATCCAACATGCCTGATCTTAGCTACTGTACCCACAACCTGACTACTATTA  
CGAAGCCACTCAAGTACTCATAATTAATAAGTGCAGCCGATTCCTCAGTGATGATCGCACCGAA  
GTGCCGAGCTTGTAAACGCGAACCAGTACTCCCCATGCGTCTCTATTGTGCCTTCTACAGTGTG  
GGAAGACGGCGATTATTATAGAAAGCAGCTGTGCCACTGGAAGGTGGCGGGTGGCTAGTTGCCA  
40 GTGGGTCCACAGTTGCCATGACCGAGCAACTTCAGATGGGGTTTGGCATAACAGTGCAGTATGGT  
ACCGATACGAACAGCGTGTGTCCAAAATTGGAATTTGCTAACGACACCAAGATCGCCTCCCAGTT  
GGGAAATTGTGTTGAATATTCCTGTACGGAGTGTGAGGCCGGGGGTGTTCCAAAATTGCACCG  
CCGTGGGAGTGAGGCAGCAAAGATTCTGTGTACGACGCATACCAGAATCTAGTCGGATACTATTCT  
GACGATGGAAACTACTACTGTCTGCGCGCTTGCCTCTCAGTGCCCGTGAGTGTATATATGATAA  
45 GGAGACCAAGACTCACGCTACTCTCTTTGGTTCTGTGCGGTGCGAACACATTTCTCTACAATGT  
CCCAGTATAGTCGCTCCACTCGGTCTATGTTAAAGCGCAGAGACAGTACCTACGGCCCTCTACAG  
ACACCTGTGGGGTGCGTTCTCGGCCTTGTCAATTCTAGCCTGTTTGTGGAGGATTGTAAGCTGCC  
CCTTGGTCAAAGCTTATGCGCACTGCCGATACGCCAGCACACTTACACCAGCTTCAGTGGGGT  
CCGTCCCCGGGGGAAATGAGATTGGCCTCGATCGCTTTCACCCACCCCATACAGGTGGATCAGCTC  
50 AACTCGTCACTTCAAGCTAAGCATCCCTACTAATTTCTCCTTTGGTGTGACTCAGGAGTACAT  
TCAGACCACAATTCAAAAGGTGACCGTTGACTGCAAGCAGTATGTGTGCAACGGGTTCCAGAAAT

5 GTGAACAGCTGCTCCGGGAGTATGGCCAGTTCTGTTCTAAAATCAACCAGGCCCTCCACGGAGCA  
 AACCTTAGGCAGGACGATTCTGTGAGAAACCTCTTTGCCAGCGTCAAGAGTTCTCAGAGTTCCCC  
 TATTATACCTGGCTTCGGCGGGGATTTCAACCTGACACTACTTGAACCTGTAAGCATATCAACCG  
 GAAGTCGCAGTGCCCGTTCCGCCATCGAGGATCTGCTCTTCGACAAAGTAACTATTGCAGATCCC  
 10 GGATACATGCAGGGGTATGACGACTGCATGCAGCAGGGTCCAGCCTCTGCAAGGGATCTGATATG  
 CGCACAGTATGTGCTGGGTACAAAGTGTTGCCTCCTCTCATGGACGTGAACATGGAAGCGGCCT  
 ATACCTCCTCACTTCTAGGCTCCATAGCGGGCGTGGGATGGACCGCAGGGCTTTCAGCTTCGCC  
 GCAATTCCCTTTGCTCAATCTATCTTCTACAGGCTTAATGGCGTTGGAATCACCCAGCAGGTGTT  
 15 AAGCGAAAACCAGAAATTGATTGCCAATAAGTTTAACCAAGCTTTGGGGGCCATGCAGACAGGCT  
 TTACAACCACAAAACGAGGCTTTCCATAAAGTACAGGATGCGGTAAACAATAACGCACAAGCCCTG  
 TCAAAGCTGGCTTCAGAGCTCTCAAATACATTTGGCGCTATATCCGCGTCTATCGGCGATATCAT  
 ACAACGGTTGGACCCACCCGAACAGGACGCACAGATTGATCGTTTGATCAACGGGAGGCTTACCA  
 CCTTAAACGCTTTTGTGGCCAGCAACTGGTGGGTCTGAGAGCGCCGCTTGAGCGCTCAGCTG  
 20 GCAAAGGATAAAGTGAATGAATGCGTGAAAGCTCAATCAAAGAGAAGTGGGTTTTGTGGGCAGGG  
 TACTCATATTGTTTCTTTGTGGTGAACGCCCAAATGGACTCTACTTTATGCGATGTTGGATACT  
 ACCCGAGCAACCACATCGAGGTGCTTTCCGCCTATGGGCTTTGTGACGCAGCAAACCTACTAAC  
 TGTATCGCGCCAGTTAATGGCTACTTTATTAACAATAACACACGCATTGTGGATGAATGGAG  
 TTACACAGGGTCCAGCTTCTACGCTCCAGAGCCTATCACCTCTCTGAACACAAAGTATGTGGCAC  
 25 CTCAGGTCACATATCAGAACATCTCGACAAACCTGCCCCCCCCACTCTTGGGCAACTCCACAGGG  
 ATCGACTTCCAGGACGAGCTTGACGAATTCTTCAAGAACGTGTCCACCAGTATCCCTAATTTTGG  
 TTCGCTGACCCAAATTAACACAACCCTGCTCGATCTGACATATGAAATGCTTTCACTACAGCAGG  
 TGGTCAAAGCGTTGAACGAGTCGTATATCGACCTGAAAGAGTTAGGGAATTACACATACTATAAC  
 AAATGGCCCTGGTATATTTGGTTAGGATTCATTGCCGGGCTGGTGGCCCTTGCTTGTGCGTATT  
 TTTTCATCTTGtctttatggatgtgctccaatggatcggttacaatgcagaatttgcatTTGA

#### PDI-MERS-H5iCT(V4)-DNA (SEQ ID NO: 104)

atggcgaaaaacgttgcgattttcggcttattgttttctcttcttctgtgttggttccttctcagat  
 cttcgcgTATGTGATGTGGGTCCCGATAGTGTAAAGTCCGCCTGCATCGAAGTGGACATTTCAGC  
 30 AGACTTCTTCGATAAGACTTGGCCTCGGCCAATTGATGTGTCCAAGGCCGACGCGATTATAC  
 CCCCAGGTTCGGACATATTCCAACATAACTATCACCTATCAGGGGCTATTCCCTTATCAGGGCGA  
 CCATGGGGACATGTACGTTTACAGCGCTGGTCACGCTACAGGGACGACCCCCCAGAAGCTCTTCG  
 TGGCGAACTATAGTCAGGACGTGAAACAGTTTGCCAACGGTTTTGTAGTGCATCGGGGACGCC  
 35 GCTAACTCCACTGGTACTGTTATTATCAGCCCTTCCACGAGTGCCACAATTGAAAGATCTATCC  
 GGCCTTCATGCTAGGATCCTCTGTGGGCAATTTTAGCGACGGTAAGATGGGTGCGTTCTTCAACC  
 ACACGCTTGTGCTGCTTCCCGATGGGTGCGGTACTTTGCTGAGGGCCTTTTACTGTATCCTAGAG  
 CCCCATCCGGCAACCACTGCCCCGCCGGGAACCTCGTATACTTCTTGGCACTTATCATACTCC  
 AGCCACGGATTGTAGCGATGGGAACACAATAGGAACGCCAGTTTGAATTCCTTTAAAGAGTACT  
 TCAACTTGCGGAATTGTACCTTCATGTATACATATAACATTACTGAGGACGAAATTCCTGAATGG  
 40 TTCGGAATCACTCAAACAGCCCAGGGAGTGCACCTCTTTAGTTCTCGCTATGTGGACTTATATGG  
 AGGCAATATGTTTCAATTGCGCACCTTACCCGTCTACGATACGATCAAGTATTACTCGATCATA  
 CCCACTCCATTAGGTCCATTGAGAGCGATCGCAAGGCATGGGCCGCTTCTATGTGTATAAGCTC  
 CAGCCCCTGACCTTCTTGGATTTCTCCGTGGACGGCTACATCAGAAGGGCTATCGATTGCGG  
 GTTCAACGACCTCAGCCAGCTGCATTGTTCTTATGAGAGCTTTGACGTGGAAGCGGAGTTTACT  
 45 CAGTCTCTTCTTTGAGGCTAAACCTTCAGGTAGCGTCTGATAGCAAGCAGAGGGTGTGGAGTGC  
 GATTTCTCACCACTGCTCAGCGGAACCCACCCAGGTCTACAACCTTAAAGCGGCTCGTGTTCAC  
 AAACGTAACTATAAATTGACTAAGTTGCTGTCACTCTTTTCCGTGAATGATTTTACATGCTCCC  
 AAATCAGCCCAGCCGCTATTGCGTCTAATTGCTATTCCCTCATTGATCCTGGATTACTTCAGTTAC  
 CCCCCTCTCTATGAAGAGCGATCTCTCGGTTAGTAGCGCTGGGCCATTTTCCCAGTTTAACTACAA  
 50 ACAATCCTTTTCCAATCCAACATGCCTGATCTTAGCTACTGTACCCCAACCTGACTACTATTA  
 CGAAGCCACTCAAGTACTCATACTAATAAGTGCAGCCGATTCCCTCAGTGATGATCGCACCGAA

GTGCCGCAGCTTGTAACGCGAACCAGTACTCCCCATGCGTCTCTATTGTGCCTTCTACAGTGTG  
GGAAGACGGCGATTATTATAGAAAGCAGCTGTCGCCACTGGAAGGTGGCGGGTGGCTAGTTGCCA  
GTGGGTCCACAGTTGCCATGACCGAGCAACTTCAGATGGGGTTTGGCATAACAGTGCAGTATGGT  
ACCGATACGAACAGCGTGTGTCCAAAATTGGAATTTGCTAACGACACCAAGATCGCCTCCCAGTT  
5 GGGAAATTGTGTTGAATATTCCCTGTACGGAGTGTGAGGCCGGGGGGTGTTCAAAAATTGCACCG  
CCGTGGGAGTGAGGCAGCAAAGATTTCGTGTACGACGCATACCAGAACTAGTCGGATACTATTCT  
GACGATGGAAACTACTACTGTCTGCGCGCTTGCGTCTCAGTGCCCGTGAGTGTACATATGATAA  
GGAGACCAAGACTCACGCTACTCTCTTTGGTTCTGTGCGGTGCGAACACATTTCTCTACAATGT  
CCCAGTATAGTCGCTCCACTCGGTCTATGTTAAAGCGCAGAGACAGTACCTACGGCCCTCTACAG  
10 ACACCTGTGGGGTGCCTTCTCGGCCTTGTCAAATTCAGCCTGTTTGTGGAGGATTGTAAGCTGCC  
CCTTGGTCAAAGCTTATGCGCACTGCCCAGTACGCCAGCACACTTACACCAGCTTCAGTGGGGT  
CCGTCCCCGGGAAATGAGATTGGCCTCGATCGCTTTCACCCACCCCATACAGGTGGATCAGCTC  
AACTCGTCATACTTCAAGCTAAGCATCCCTACTAATTTCTCCTTTGGTGTGACTCAGGAGTACAT  
TCAGACCACAATTCAAAGGTGACCGTTGACTGCAAGCAGTATGTGTGCAACGGGTTCCAGAAAT  
15 GTGAACAGCTGCTCCGGGAGTATGGCCAGTTCTGTTCTAAAATCAACCAGGCCCTCCACGGAGCA  
AACCTTAGGCAGGACGATTCTGTGAGAAACCTCTTTGCCAGCGTCAAGAGTTCTCAGAGTTCCCC  
TATTATACCTGGCTTCGGCGGGGATTTCAACCTGACACTACTTGAACCTGTAAGCATATCAACCG  
GAAGTCGAGTGCCCGTTCCGCCATCGAGGATCTGCTCTTCGACAAAGTAACTATTGCAGATCCC  
GGATACATGCAGGGGTATGACGACTGCATGCAGCAGGGTCCAGCCTCTGCAAGGGATCTGATATG  
20 CGCACAGTATGTCGCTGGGTACAAAGTGTTGCCTCCTCTCATGGACGTGAACATGGAAGCGGCCT  
ATACCTCCTCACTTCTAGGCTCCATAGCGGGCGTGGGATGGACCGCAGGGCTTTCAAGCTTCGCC  
GCAATTCCTTTGCTCAATCTATCTTCTACAGGCTTAATGGCGTTGGAATCACCCAGCAGGTGTT  
AAGCGAAAACAGAAATTGATTGCCAATAAGTTTAACCAAGCTTTGGGGGCCATGCAGACAGGCT  
TTACAACCACAAACGAGGCTTTCCATAAAGTACAGGATGCGGTAAACAATAACGCACAAGCCCTG  
25 TCAAAGCTGGCTTCAGAGCTCTCAAATACATTTGGCGCTATATCCGCGTCTATCGGCGATATCAT  
ACAACGTTTGACCCACCCGAACAGGACGCACAGATTGATCGTTTGATCAACGGGAGGCTTACCA  
CCTTAAACGCTTTTGTGGCCAGCAACTGGTGCGGTCTGAGAGCGCCGCTTGAGCGCTCAGCTG  
GCAAAGGATAAAGTGAATGAATGCGTGAAAGCTCAATCAAAGAGAAGTGGGTTTTGTGGGCAGGG  
TACTCATATTGTTTCTTTGTGGTGAACGCCCCAAATGGACTCTACTTTATGCATGTTGGATACT  
30 ACCCGAGCAACCACATCGAGGTGTTTTCCGCCTATGGGCTTTGTGACGCAGCAAACCTACTAAC  
TGATTCGCGCCAGTTAATGGCTACTTTATTAACAATAACACACGCATTGTGGATGAATGGAG  
TTACACAGGGTCCAGCTTCTACGCTCCAGAGCCTATCACCTCTCTGAACACAAAGTATGTGGCAG  
CTCAGGTCACATATCAGAACATCTCGACAAACCTGCCCCCCCCACTCTTGGGCAACTCCACAGGG  
ATCGACTTCCAGGACGAGCTTGACGAATCTTCAAGAACGTGTCCACCAGTATCCCTAATTTTGG  
35 TTCGTGACCCAAATTAACACAACCCTGCTCGATCTGACATATGAAATGCTTTCCTACAGCAGG  
TGGTCAAAGCGTTGAACGAGTCGTATATCGACCTGAAAGAGTTAGGGAATTACACATACTATAAC  
AAATGGCCCTGGTATATTTGGTTAGGATTCATTGCCGGGCTGGTGGCCCTTGCTTGTGCGTATT  
TTTCATCTTGtctgctccaatggatcgttacaatgcagaatttgcattTGA

40 PDI-MERS-H1cCT-DNA (SEQ ID NO: 105)

atggcgaaaaacggttgcgattttgcgcttattgttttctcttcttctgtgttggttccttctcagat  
cttcgcgTATGTCGATGTGGGTCCCGATAGTGTTAAGTCCGCTGCATCGAAGTGGACATTACAGC  
AGACCTTCTTCGATAAGACTTGGCCTCGGCCAATTGATGTGTCCAAGGCCGACGGCATTATCTAC  
45 CCCCAGGTTCGGACATATTCCAACATAACTATCACCTATCAGGGGCTATTCCCTTATCAGGGCGA  
CCATGGGGACATGTACGTTTACAGCGCTGGTCACGCTACAGGGACGACCCCCAGAAGCTCTTCG  
TGGCGAACTATAGTCAGGACGTGAAACAGTTTGCCAACGGTTTTGTAGTGCGCATCGGGGACGCC  
GCTAACTCCACTGGTACTGTTATTATCAGCCCTTCACGAGTGCCACAATTGCAAAGATCTATCC  
GGCCTTCATGCTAGGATCCTCTGTGGGCAATTTTAGCGACGGTAAGATGGGTTCGCTTCTTCAACC  
ACACGCTTGTGCTGCTTCCCGATGGGTGCGGTACTTTGCTGAGGGCCTTTTACTGTATCCTAGAG  
50 CCCCAGTCCGGCAACCACTGCCCCGCGGGAACCTCGTATACTTCCCTTGCCACTTATCATACTCC

AGCCACGGATTGTAGCGATGGGAACTACAATAGGAACGCCAGTTTGAATTCCCTTTAAAGAGTACT  
TCAACTTGC GGAATTGTACCTTCATGTATACATATAACATTACTGAGGACGAAATTCTCGAATGG  
TTCGGAATCACTCAAACAGCCCAGGGAGTGCACCTCTTTAGTTCTCGCTATGTGGACTTATATGG  
AGGCAATATGTTTCAATTGCGCACCTTACCCGTCTACGATACGATCAAGTATTACTCGATCATAAC  
5 CCCACTCCATTAGGTCCATTAGAGCGATCGCAAGGCATGGGCCGCATTCTATGTGTATAAGCTC  
CAGCCCCTGACCTTCCTCTTGGATTTCTCCGTGGACGGCTACATCAGAAGGGCTATCGATTGCGG  
GTTCAACGACCTCAGCCAGCTGCATTGTTCTTATGAGAGCTTTGACGTGGAAGCGGAGTTTACT  
CAGTCTCTTCCCTTTGAGGCTAAACCTTCAGGTAGCGTCGTAGAGCAAGCAGAGGGTGTGGAGTGC  
GATTTCTCACCCTGCTCAGCGGAACCCACCCAGGTCTACAACCTTAAGCGGCTCGTGTTTAC  
10 AAACGTGTAACCTATAAATTGACTAAGTTGCTGTCACTCTTTTCCGTGAATGATTTTACATGCTCCC  
AAATCAGCCCAGCCGCTATTGCGTCTAATTGCTATTCCCTCATTGATCCTGGATTACTTCAGTTAC  
CCCCCTCTCTATGAAGAGCGATCTCTCGGTTAGTAGCGCTGGGCCCTATTTCCAGTTTAACTACAA  
ACAATCCTTTTCCAATCCAACATGCCTGATCTTAGCTACTGTACCCCAACAACCTGACTACTATTA  
15 CGAAGCCACTCAAGTACTCATACTAATAAGTGCAGCCGATTCCCTCAGTGATGATCGCACCGAA  
GTGCCGAGCTTGTAAACGCGAACCAGTACTCCCATGCGTCTCTATTGTGCCTTCTACAGTGTG  
GGAAGACGGCGATTATTATAGAAAGCAGCTGTCGCCACTGGAAGGTGGCGGGTGGCTAGTTGCCA  
GTGGGTCCACAGTTGCCATGACCGAGCAACTTCAGATGGGGTTTGGCATAACAGTGCAGTATGGT  
ACCGATACGAACAGCGTGTGTCCAAAATTGGAATTTGCTAACGACACCAAGATCGCCTCCAGTT  
GGGAAATTGTGTTGAATATTCCCTGTACGGAGTGTGAGGCCGGGGGTGTTCCAAAATTGCACCG  
20 CCGTGGGAGTGAGGCAGCAAAGATTTCGTGTACGACGCATACCAGAATCTAGTCGGATACTATTCT  
GACGATGGAACTACTACTGTCTGCGCGCTTGCGTCTCAGTGCCCGTGAGTGTCTATATATGATAA  
GGAGACCAAGACTCACGCTACTCTTTGGTTCTGTGCGGTGCGAACACATTTCCCTCTACAATGT  
CCCAGTATAGTCGCTCCACTCGGTCTATGTTAAAGCGCAGAGACAGTACCTACGGCCCTCTACAG  
ACACCTGTGGGGTGCCTTCTCGGCCCTGTCAATTCTAGCCTGTTTGTGGAGGATTGTAAGCTGCC  
25 CCTTGGTCAAAGCTTATGCGCACTGCCGATACGCCCAGCACACTTACACCAGCTTCAGTGGGGT  
CCGTCCCCGGGAAATGAGATTGGCCTCGATCGCTTTCAACCACCCCATACAGGTGGATCAGCTC  
AACTCGTCATACTTCAAGCTAAGCATCCCTACTAATTTCTCCTTTGGTGTGACTCAGGAGTACAT  
TCAGACCACAATTCAAAAGGTGACCGTTGACTGCAAGCAGTATGTGTGCAACGGGTTCAGAAAT  
GTGAACAGCTGCTCCGGGAGTATGGCCAGTTCTGTTCTAAAATCAACCAGGCCCTCCACGGAGCA  
30 AACCTTAGGCAGGACGATTCTGTCTAGAAACCTCTTTGCCAGCGTCAAGAGTTCTCAGAGTTCCCC  
TATTATACCTGGCTTCGGCGGGGATTTCAACCTGACACTACTTGAACCTGTAAGCATATCAACCG  
GAAGTCGACGTGCCCGTTCCGCCATCGAGGATCTGCTCTTCGACAAAAGTAACTATGCAGATCCC  
GGATACATGCAGGGGTATGACGACTGCATGCAGCAGGGTCCAGCCTTGCAAGGGATCTGATATG  
CGCACAGTATGTGCTGGGTACAAAGTGTGCTCCTCTCATGGACGTGAACATGGAAGCGGCCCT  
35 ATACCTCCTCACTTCTAGGCTCCATAGCGGGCGTGGGATGGACCGCAGGGCTTTCAAGCTTCGCC  
GCAATTCCTTTGCTCAATCTATCTTCTACAGGCTTAATGGCGTTGGAATCACCCAGCAGGTGTT  
AAGCGAAAACAGAAATTGATTGCCAATAAGTTTAACCAAGCTTTGGGGGCCATGCAGACAGGCT  
TTACAAACCACAAACGAGGCTTTCCATAAAGTACAGGATGCGGTAAACAATAACGCACAAGCCCTG  
TCAAAGCTGGCTTCAGAGCTCTCAAATACATTTGGCGCTATATCCGCGTCTATCGGCGATATCAT  
40 ACAACGGTTGGACCCACCCGAACAGGACGCACAGATTGATCGTTTGATCAACGGGAGGCTTACCA  
CCTTAAACGCTTTTGTGGCCAGCAACTGGTGGGTCTGAGAGCGCCGCTTGAGCGCTCAGCTG  
GCAAAGGATAAAGTGAATGAATGCGTGAAAGCTCAATCAAAGAGAAGTGGGTTTTGTGGGCAGGG  
TACTCATATTGTTTCTTTGTGGTGAACGCCCAATGGACTCTACTTTATGCATGTTGGATATCC  
ACCCGAGCAACCACATCGAGGTGTTTTCCGCCTATGGGCTTTGTGACGCAGCAAAACCTACTAAC  
45 TGTATCGCGCCAGTTAATGGCTACTTTATTAACAAATAACACACGCATTGTGGATGAATGGAG  
TTACACAGGGTCCAGCTTCTACGCTCCAGAGCCTATCACCTCTCTGAACACAAAAGTATGTGGCAC  
CTCAGGTCACATATCAGAACATCTCGACAAACCTGCCCCCCCCACTCTTGGGCAACTCCACAGGG  
ATCGACTTCCAGGACGAGCTTGACGAATCTTCAAGAACGTGTCCACCAGTATCCCTAATTTTGG  
TTCGCTGACCCAAATTAACACAACCCTGCTCGATCTGACATATGAAATGCTTCTACTACAGCAGG  
50 TGGTCAAAGCGTTGAACGAGTCGTATATCGACCTGAAAGAGTTAGGGAATTACACATACTATAAC  
AAATGGCCCTGGTATATTTGGTTAGGATTCATTGCCGGGCTGGTGGCCCTTGCTTGTGCGTATT

TTTCATCTTGagcttctggatgtgctctaattgggtctctacagtgtagaatatgtattTGA

**PDI-MERS-wtTMCT-AA (SEQ ID NO: 106)**

5 MAKNVAIFGLLFSLVLVPSQIFAYVDVGPDSVKSACIEVDIQQTFFDKTWPRPIDVSKADGIIY  
PQGRITYSNITITYQGLFPYQGDHGDYVYSAGHATGTTTPQKLFVANYSQDVKQFANGFVVRIGAA  
ANSTGTVIISPSTSATIRKIYPAFMLGSSVGNFSDGKMGRFFNHTLVLLPDGCGTLLRAFYCILE  
10 PRSGNHCPAGNSYTSFATYHTPATDCSDGNYNRNASLNSFKEYFNLNRCTFMYTYNITEDEILEW  
FGITQTAQGVHLFSSRYVDLYGGNMFFQFATLPVYDTIKYYSIIPHSIRSISQSDRKAWAAFYVYKL  
QPLTFLLDFSVVDGYIRRAIDCGFNDLSQLHCSYESFDVESGVYSVSSFEAKPSGSVVEQAEGVEC  
15 DFSPLLSGTPPQVYNFKRLVFTNCNYNLTCLLSLFSVNDFTCSQISPAAIASNCYSSLILDYFSY  
PLSMKSDLVSSAGPISQFNYKQSFNPTCLILATVPHNLTTITKPLKYSYINKCSRFLSDDRTE  
VPQLVNAVQYSPCVSIVPSTVWEDGDYRQKLSPLEGGGWLVASGSTVAMTEQLQMGFGITVQYG  
TDNSVCPKLEFANDTKIASQLGNCVEYSLYGVSGRGVFQNCCTAVGVRQQRFFVYDAYQNLVGYYS  
20 DDGNYCLRACVSVPVSVIYDKETKTHATLFGSVACEHISSTMSQYSRSTRSMLKRRDSTYGPLQ  
TPVGCVLGLVNSSLFVEDCKLPLGQSLCALPDTPTSTLTPASVGSVPGEMLASIAFNHPIQVDQL  
NSSYFKLSIPTNFSFGVTQEYIQTTIQKVTVDCKQYVCNGFQKCEQLLREYGFQFCSKINQALHGA  
NLRQDDSVRNLFASVKSSQSSPIIPGFGGDFNLTLLEPVSI STGSRARSASIEDLLFDKVTIADP  
GYMQGYDDCMQQGPASARDLICAQYVAGYKVLPLMDVNMEAAYSLLGSIAGVGWTAGLSSFA  
25 AIPFAQSI FYRLNGVGITQQVLSENQKLIANKFNQALGAMQTGFTTTNEAFHKVQDAVNNAQAL  
SKLASELSNTFGAISASIGDIIQRLDPPEQDAQIDRLINGRLTTLNAFVAQQLVRSESAALSAQL  
AKDKVNECVKAQSKRSGFCGQGTHTIVSFVNAPNGLYFMHVGYPSNHIEVVSAYGLCDAANPTN  
CIAPVNGYFIKTNTRIVDEWSTYTGSSFYAPEPITSLNTKYVAPQVTYQNI STNLPPPLGNSTG  
IDFQDELDEFFKNVSTSI PNFGSLTQINTTLLDLTYEMLSLQQVVKALNESYIDLKELGNYTYYN  
30 KWPWYIWLGFIAGLVALALCVFFILCCTGCGTNCMGKLCNRCCDRYEEDLEPHKVHVH

**PDI-MERS-H5iTMCT-AA (SEQ ID NO: 107)**

MAKNVAIFGLLFSLVLVPSQIFAYVDVGPDSVKSACIEVDIQQTFFDKTWPRPIDVSKADGIIY  
PQGRITYSNITITYQGLFPYQGDHGDYVYSAGHATGTTTPQKLFVANYSQDVKQFANGFVVRIGAA  
30 ANSTGTVIISPSTSATIRKIYPAFMLGSSVGNFSDGKMGRFFNHTLVLLPDGCGTLLRAFYCILE  
PRSGNHCPAGNSYTSFATYHTPATDCSDGNYNRNASLNSFKEYFNLNRCTFMYTYNITEDEILEW  
FGITQTAQGVHLFSSRYVDLYGGNMFFQFATLPVYDTIKYYSIIPHSIRSISQSDRKAWAAFYVYKL  
QPLTFLLDFSVVDGYIRRAIDCGFNDLSQLHCSYESFDVESGVYSVSSFEAKPSGSVVEQAEGVEC  
35 DFSPLLSGTPPQVYNFKRLVFTNCNYNLTCLLSLFSVNDFTCSQISPAAIASNCYSSLILDYFSY  
PLSMKSDLVSSAGPISQFNYKQSFNPTCLILATVPHNLTTITKPLKYSYINKCSRFLSDDRTE  
VPQLVNAVQYSPCVSIVPSTVWEDGDYRQKLSPLEGGGWLVASGSTVAMTEQLQMGFGITVQYG  
TDNSVCPKLEFANDTKIASQLGNCVEYSLYGVSGRGVFQNCCTAVGVRQQRFFVYDAYQNLVGYYS  
40 DDGNYCLRACVSVPVSVIYDKETKTHATLFGSVACEHISSTMSQYSRSTRSMLKRRDSTYGPLQ  
TPVGCVLGLVNSSLFVEDCKLPLGQSLCALPDTPTSTLTPASVGSVPGEMLASIAFNHPIQVDQL  
NSSYFKLSIPTNFSFGVTQEYIQTTIQKVTVDCKQYVCNGFQKCEQLLREYGFQFCSKINQALHGA  
NLRQDDSVRNLFASVKSSQSSPIIPGFGGDFNLTLLEPVSI STGSRARSASIEDLLFDKVTIADP  
GYMQGYDDCMQQGPASARDLICAQYVAGYKVLPLMDVNMEAAYSLLGSIAGVGWTAGLSSFA  
45 AIPFAQSI FYRLNGVGITQQVLSENQKLIANKFNQALGAMQTGFTTTNEAFHKVQDAVNNAQAL  
SKLASELSNTFGAISASIGDIIQRLDPPEQDAQIDRLINGRLTTLNAFVAQQLVRSESAALSAQL  
AKDKVNECVKAQSKRSGFCGQGTHTIVSFVNAPNGLYFMHVGYPSNHIEVVSAYGLCDAANPTN  
CIAPVNGYFIKTNTRIVDEWSTYTGSSFYAPEPITSLNTKYVAPQVTYQNI STNLPPPLGNSTG  
IDFQDELDEFFKNVSTSI PNFGSLTQINTTLLDLTYEMLSLQQVVKALNESYIDLKELGNYTYYN  
KWPWYQILSIYSTVASSLALAIMMAGLSLWMCNGLSLQCRICI

## PDI-MERS-H5iCT-AA (SEQ ID NO: 108)

MAKNVAIFGLLFSLVLVPSQIFAYVDVGPDSVKSACIEVDIQQTFFDKTWPRPIDVSKADGIIY  
 PQGRITYSNITITYQGLFPYQGDHGDYVYSAGHATGTTTPQKLFVANYSQDVKQFANGFVVRIGAA  
 ANSTGTVVIISPSTSATIRKIYPAFMLGSSVGNFSDGKMGRFFNHTLVLLPDGCGTLLRAFYCILE  
 5 PRSGNHCPAGNSYTSFATYHTPATDCSDGNYNRNASLNSFKEYFNLNCTFMYTYNITEDEILEW  
 FGITQTAQGVHLFSSRYVDLYGGNMFFQFATLPVYDTIKYYSIIPHSIRSIQSDRKAWAAFYVYKL  
 QPLTFLLDFSVVDGYIRRAIDCGFNDLSQLHCSYESFDVESGVYSVSSFEAKPSGVSVEQAEGVEC  
 DFSPLLSGTPPQVYNFKRLVFTNCNYNLTCLLTLATVPHNLTTITKPLKYSYINKCSRFLSDDRTE  
 10 PLSMKSDLVSSAGPISQFNYKQSFNPTCLILATVPHNLTTITKPLKYSYINKCSRFLSDDRTE  
 VPQLVNAVQYSPCVSIVPSTVWEDGDYRQKLSPLEGGGWLVASGSTVAMTEQLQMGGFITVQYG  
 TDTNSVCPKLEFANDTKIASQLGNCVEYSLYGVSGRGVFQNCCTAVGVRQQRFFVYDAYQNLVGYYS  
 DDGNYYCLRACVSVPVSVIYDKETKTHATLFGSVACEHISSTMSQYSRSTRSMLKRRDSTYGPLQ  
 TPVGCVLGLVNSSLFVEDCKLPLGQSLCALPDTPTSTLTPASVGSVPGEMLASIAFNHPIQVDQL  
 15 NSSYFKLSIPTNFSFGVTQEYIQTTIQKVTVDCKQYVCNGFQKCEQLLREYGQFCSKINQALHGA  
 NLRQDDSVRNLFASVKSSQSSPIIPGFGGDFNLTLLEPVSIISTGSRARSASIEDLLFDKVTIADP  
 GYMQGYDDCMQOGPASARDLICAQYVAGYKVLPLMDVNMEAAYTSSLLGSIAGVGWTAGLSSFA  
 AIPFAQSIIFYRLNGVGITQQVLSENQKLIANKFNQALGAMQTGFTTTNEAFHKVQDAVNNNAQAL  
 SKLASELSNTFGAISASIGDIIQRLDPPEQDAQIDRLINGRLTTLNFAVQAQLVRSESAALSAQL  
 AKDKVNECVKAQSKRSGFCGQGTHTIVSFVVNAPNGLYFMHVGYYPNSHIEVVSAYGLCDAANPTN  
 20 CIAPVNGYFIKTNNTRIVDEWSYTGSSFYAPEPITSLNTKYVAPQVITYQNIISTNLPPPLGNSTG  
 IDFQDELDEFFKNVSTSI PNFGSLTQINTTLLDLTYEMLSLQQVVKALNESYIDLKELGNYTYYN  
 KWPWYIWLGFIAGLVALALCVFFILSLWMCNGLSLQCRICI

## PDI-MERS-H5iCT(V4)-AA (SEQ ID NO: 109)

MAKNVAIFGLLFSLVLVPSQIFAYVDVGPDSVKSACIEVDIQQTFFDKTWPRPIDVSKADGIIY  
 PQGRITYSNITITYQGLFPYQGDHGDYVYSAGHATGTTTPQKLFVANYSQDVKQFANGFVVRIGAA  
 ANSTGTVVIISPSTSATIRKIYPAFMLGSSVGNFSDGKMGRFFNHTLVLLPDGCGTLLRAFYCILE  
 25 PRSGNHCPAGNSYTSFATYHTPATDCSDGNYNRNASLNSFKEYFNLNCTFMYTYNITEDEILEW  
 FGITQTAQGVHLFSSRYVDLYGGNMFFQFATLPVYDTIKYYSIIPHSIRSIQSDRKAWAAFYVYKL  
 30 QPLTFLLDFSVVDGYIRRAIDCGFNDLSQLHCSYESFDVESGVYSVSSFEAKPSGVSVEQAEGVEC  
 DFSPLLSGTPPQVYNFKRLVFTNCNYNLTCLLTLATVPHNLTTITKPLKYSYINKCSRFLSDDRTE  
 PLSMKSDLVSSAGPISQFNYKQSFNPTCLILATVPHNLTTITKPLKYSYINKCSRFLSDDRTE  
 VPQLVNAVQYSPCVSIVPSTVWEDGDYRQKLSPLEGGGWLVASGSTVAMTEQLQMGGFITVQYG  
 35 TDTNSVCPKLEFANDTKIASQLGNCVEYSLYGVSGRGVFQNCCTAVGVRQQRFFVYDAYQNLVGYYS  
 DDGNYYCLRACVSVPVSVIYDKETKTHATLFGSVACEHISSTMSQYSRSTRSMLKRRDSTYGPLQ  
 TPVGCVLGLVNSSLFVEDCKLPLGQSLCALPDTPTSTLTPASVGSVPGEMLASIAFNHPIQVDQL  
 NSSYFKLSIPTNFSFGVTQEYIQTTIQKVTVDCKQYVCNGFQKCEQLLREYGQFCSKINQALHGA  
 NLRQDDSVRNLFASVKSSQSSPIIPGFGGDFNLTLLEPVSIISTGSRARSASIEDLLFDKVTIADP  
 GYMQGYDDCMQOGPASARDLICAQYVAGYKVLPLMDVNMEAAYTSSLLGSIAGVGWTAGLSSFA  
 40 AIPFAQSIIFYRLNGVGITQQVLSENQKLIANKFNQALGAMQTGFTTTNEAFHKVQDAVNNNAQAL  
 SKLASELSNTFGAISASIGDIIQRLDPPEQDAQIDRLINGRLTTLNFAVQAQLVRSESAALSAQL  
 AKDKVNECVKAQSKRSGFCGQGTHTIVSFVVNAPNGLYFMHVGYYPNSHIEVVSAYGLCDAANPTN  
 CIAPVNGYFIKTNNTRIVDEWSYTGSSFYAPEPITSLNTKYVAPQVITYQNIISTNLPPPLGNSTG  
 45 IDFQDELDEFFKNVSTSI PNFGSLTQINTTLLDLTYEMLSLQQVVKALNESYIDLKELGNYTYYN  
 KWPWYIWLGFIAGLVALALCVFFILCCSNGSLQCRICI

## PDI-MERS-H1cCT-AA (SEQ ID NO: 110)

5 MAKNVAIFGLLFSLVLVPSQIFAYVDVGPDSVKSACIEVDIQQTFFDKTWPRPIDVSKADGIIY  
 PQGRYSNITITYQGLFPYQGDHGDYVYSAGHATGTTTPQKLFVANYSQDVKQFANGFVVRIGAA  
 ANSTGTVIISPSTSATIRKIYPAFMLGSSVGNFSDGKMGRFFNHTLVLLPDGCGTLLRAFYCILE  
 PRSGNHCPAGNSYTSFATYHTPATDCSDGNYNRNASLNSFKEYFNLRNCTFMYTYNITEDEILEW  
 10 FGITQTAQGVHLFSSRYVDLYGGNMFFQFATLPVYDTIKYYSIIPHSIRSISQSDRKAWAAFYVYKL  
 QPLTFLLDFSVDGYIRRAIDCGFNDLSQLHCSYESFDVESGVYSSVSFEAKPSGSVVEQAEGVEC  
 DFSPLLSGTPPQVYNFKRLVFTNCNYNLTCLLSLFSVNDFTCQSIPAAIASNCYSSLILDYFSY  
 PLSMKSDLSVSSAGPISQFNYKQSFNPTCLILATVPHNLTTITKPLKYSYINKCSRFLSDDRTE  
 15 VPQLVNANQYSPCVSIVPSTVWEDGDYIRKQLSPLEGGGWLVASGSTVAMTEQLQMGFGITVQYG  
 TDTNSVCPKLEFANDTKIASQLGNCVEYSLYGVSGRGVFQNCTAVGVRRQRFVYDAYQNLVGYY  
 DDGNYCLRACVSVPVSVIYDKETKTHATLFGSVACEHISSTMSQYSRSTRSMLKRRDSTYGPLQ  
 TPVGCVLGLVNSSLFVEDCKLPLGQSLCALPDTPTSTLTPASVGSVPGEMRLASIAFNHP IQVDQL  
 NSSYFKLSIPTNFSFGVTQEYIQTTIQKVTVDCKQYVCNGFQKCEQLLREYGQFCSKINQALHGA  
 20 NLRQDDSVRNLFASVKSSQSSPIIPGFGGDFNLTLLEPVSISTGSRARSASIEDLLFDKVTIADP  
 GYMQGYDDCMQQGPASARDLICAQYVAGYKVLPLMDVNMEAAYSLLGS IAGVGWTAGLSSFA  
 AIPFAQSI FYRLNGVGITQQVLSENQKLIANKFNQALGAMQTGFTTTNEAFHKVQDAVNNNAQAL  
 SKLASELSNTFGAISASIGDIIQRLDPPEQDAQIDRLINGRLTTLNLFVAQQLVRSESAALSAQL  
 AKDKVNECVKAQSKRSGFCGQGTHIVSFVVNAPNGLYFMHVGYPSNHIEVVSAYGLCDAANPTN  
 25 CIAPVNGYFIKTNNTTRIVDEWYSYTGSSFYAPEPITSLNTKYVAPQVTYQNI STNLPPPLLGNSTG  
 IDFQDELDEFKFNVSTSI PNFGSLTQINTTLLDLTYEMLS LQQVVKALNESYIDLKELGNITYYN  
 KWPWYIWLGF IAGLVALALCVFFILSFWMCSNGSLQCRICI

#### Cloning vector 7147 from left to right T-DNA (SEQ ID NO: 111)

25 tggcaggatatattgtggtgtaaaacaaattgacgcttagacaacttaataacacattgcgagcgt  
 ttttaattgtactgaattaacgccgaatcccggtggtatatttataatgttgtaaaataactcaa  
 aaaccataaaagtttaagtttagcaagtggtgtacatctttacttgaacaaaaatattcacactacta  
 ctgttataaaatcattattaaacattagagtaaaagaatatggatgataagaacaagagtagtgat  
 attttgacaacaattttgttgcaacatttgagaaaaattttgttggtctctcttttcattggtcaa  
 30 aaacaatagagagagaaaaaggaagaggagagaataaaaaacataatgtgagtagatgagagagaagt  
 tgtacaaaagttgtacaaaatagttgtacaaaatcatttgaggaatttgacaaaagctacacaa  
 ataaggggttaattgtctgtaataaataaggatgacgcattagagagatgtaccattagagaat  
 ttggcaagtcattaaaaagaaagaataaattatcttttaaaattaaaagttgagtcatttgattaa  
 acatgtgattattttaattgaattgatgaaagagttggattaaagttgtatttagtaattagaatttg  
 35 gtgtcaaatttaatttgacatttgatcttttcttatatattgccccatagagtcagtttaactcat  
 ttttatatttcatagatcaaataagagaaataacgggtatattaatccctccaaaaaataaacg  
 gtatatttactaaaaaatctaagccacgtaggaggataaacaggatccccgtaggaggataacatc  
 caatccaaccaatcacacaatcctgatgagataacccactttaagccacgcatctgtggcaca  
 tctacattatctaaatcacacattcttcacacatctgagccacacaaaaaccaatccacatctt  
 40 tatcaccattctataaaaaatcacactttgtgagtcacactttgattcccttcaaacacatac  
 aaagagaagagactaattaattaattaatcatcttgagagaaaatggaacgagctatacaaggaa  
 acgagctaggggaacaagctaacagtgacagttgggatggaggatcaggaggtaccacttctccc  
 ttcaaaacttctgacgaaagtcgagttggactgagtggtggcggtacataaacgatgagacgaattc  
 gaatcaagataatcccccttggtttcaaggaaagctggggtttcgggaaagttgtatttaagagat  
 atctcagatacgacaggacggaagcttcactgcacagagtccttggtatcttgacgggagattcg  
 45 gttactatgcagcatctcgatcttttcggtttcgaccagatcggtatgtacctatagttatcggtt  
 tcgaggagttagtatcaccggtttctggagggtcgcgaaactcttcagcatctctgtgagatggcaa  
 ttcggtctaagcaagaactgctacagcttgcccaatcgaagtggaaagtaattgtatcaagagga  
 tgccctgaaggtactcaaaccttcgaaaaagaaagcgagtaagttaaaatgcttcttctgctcct  
 50 atttataatatggtttgttattgttaattttgttcttgtagaagagcttaattaatcggtgtgtg  
 tatgaaatactatttgtatgagatgaactggtgtaattgtaattcatttacataagtgagtcaga

atcagaatgtttcctccataactaactagacatgaagacctgccgcgtacaattgtcttatattt  
 gaacaactaaaattgaacatcttttgccacaactttataagtggttaatatagctcaaatatatg  
 gtcaagttcaatagattaataatggaaatatcagttatcgaaattcattaacaatcaacttaacg  
 ttattaactactaattttatatcatccccctttgataaatgatagtacccaattaggaaggagca  
 5 tgctcgctagggagattgtcggtttccgccttcagtttgcaagctgctctagccgtgtagccaat  
 acgcaaaccgcctctccccgcgcgttggaattactagcgcgtgctcgacaagcttgcatgccggt  
 caacatgggtggagcacgacacacttgtctactccaaaaatatcaaagatacagtctcagaagacc  
 aaagggaattgagacttttcaacaaagggttaatatccggaacctcctcggttccattgcccc  
 gctatctgtcactttattgtgaagatagtggaaaaggaaggtggctcctacaaatgccatcattg  
 10 cgataaaggaaaaggccatcggtgaagatgcctctgccgacagtggtcccaagatggacccccac  
 ccacgaggagcatcggtgaaaaagaagacgttccaaccacgtcttcaaagcaagtggattgatgt  
 gataacatgggtggagcacgacacacttgtctactccaaaaatatcaaagatacagtctcagaaga  
 ccaaagggaattgagacttttcaacaaagggttaatatccggaacctcctcggttccattgccc  
 cagctatctgtcactttattgtgaagatagtggaaaaggaaggtggctcctacaaatgccatcat  
 15 tgcgataaaggaaaaggccatcggtgaagatgcctctgccgacagtggtcccaagatggaccccc  
 acccacgaggagcatcggtgaaaaagaagacgttccaaccacgtcttcaaagcaagtggattgat  
 gtgatatctccactgacgtaagggtgacgcacaatcccactatccttcgcaagaccttccctct  
 atataaggaagttcatttcatgttgagaggcactccatttgaatctatcaaacaaaaacacattg  
 agacgtcacgtactcctcagccaaaacgacacccccatctgtctatccactggccccctggatctg  
 20 ctgccccaaactaactccatgggtgaccttggtgatgcctgggtcaagggttatctccctgagccagt  
 acagtgcactggaactctggatccctgtccagcgggtgtgcacaccttccagctgtcctgcagtc  
 tgacctctacactctgagcagctcagtgactgtccctccagcacctggcccagcgagaccgtca  
 cctgcaacgttgcccacccggccagcagcaccaaggtggacaagaaaattgtgcccagggtattgt  
 ggttgaagccttgcatatgtacagtcaccagaagtatcatctgtcttcatcttccccccaaagcc  
 25 caaggatgtgctcaccattactctgactcctaagggtcacgtgtgttggtgtagacatcagcaagg  
 atgatcccgagggtccagttcagctggtttgtagatgatgtggaggtgcacacagctcagacgcaa  
 ccccgaggagcagttcaacagcactttccgctcagtcagtgaaactcccatcatgcaccagga  
 ctggctcaatggcaaggagacgtccagattttggcgatctattcaactgtcgccagttcattgggt  
 actggtagtctccctgggggcaatcagtttctggatgtgctctaattgggtctctacagtgtagaa  
 30 tatgtatttaaaggccttagtcgtgtcggttttcaataatataatccttttagggttttagtta  
 gtttaaattttctgttgcctctgttttagcaggtcggtgccttcagcaagcacacaaaaacagagtg  
 tttattttaagttggtttgtagtattcaaaaaaaatcggtcaaacatttggcaataaaagtt  
 tcttaagattgaatcctgttgccggtcttgcgatgattatcatataatttctgttgattacgtt  
 aagcatgtaataattaacatgtaatgcatgacgttatttatgagatgggtttttatgattagagt  
 35 cccgcaattatacattttaatacgcgatagaaaacaaaatatagcgcgcaaacaggataaattat  
 cgcgcgcggtgtcatctatgttactagatctctagagtctcaagcttggcgcgccacgtgacta  
 gtggcactggccgtcggtttacaacgtcggtgactgggaaaaccctggcggttaccacacttaacg  
 ccttgacgacatcccccttctgccagctggcgtaatagcgaagaggcccgaccgatcgccctt  
 40 cccaacagtttgcgagcctgaatggcgaatgctagagcagcttgagcttgatcagattgtcggt  
 tcccgcttcagtttaactatcagtggttgacaggatatattggcgggtaaacctaaagagaaaa  
 gagcggtta

# Native SARS-CoV-1 S protein wtTM/CT AA P59594 (SEQ ID NO: 112)

MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTLQDLFLPFYSN  
 VTGFHTINHTFGNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNKSQSVIIINNSTNVVIRACNF  
 45 ELCDNPFFAVSKPMGTQHTMIFDNAFNCTFEYISDAFSLDVSEKSGNFKHLREFVFKNKDGLFY  
 VYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINITNFRAILTAFSPAQDIWGTSAAYFVGYLKPT  
 TFMLKYDENGITITDAVDCSQNPLAELKCSVKSFEIDKGIYQTSNFRVVPsgdVVRFPNITNLCPF  
 GEVFNATKFPSVYAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNLDLCSNVYADSFVVK  
 GDDVRQIAPGQTGVIADYNYKL PDDFMGCVLAWNTRNIDATSTGNYNKYRYLRHGKLRPFERDI  
 50 SNVPFSPDGKPTFPALNCYWPLNDYGFYTTTGIGYQPYRVVLSFELLNAPATVCGPKLSTDLI

5 KNQCVNFNFNGLTGTGVLTTPSSKRFQPFQFGRDVSDFDTSVRDPKTSEILDISPCSFGGVSVIT  
 PGTNASSEVAVLYQDVNCTDVSTAIHADQLTPAWRIYSTGNNVFQTQAGCLIGAEHVDTSYECDI  
 PIGAGICASYHTVSLRSTSQKSIVAYTMSLGADSSIAYSNNTIAIPTNFSISITTEVMPVSMK  
 TSVDCNMYICGDSSTECANLLLQYGSFCTQLNRALSGIAAEQDRNTREVFAQVKQMYKTPTLKYFG  
 10 GFNFSQILPDPLKPTKRSFIEDLLFNKVTLADAGFMKQYGECLGDINARDLICAQKFNGLTVLPP  
 LLTDDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKQIANQFN  
 KAISQIQESLTTTSTALGKLQDVVNQNAQALNTLVKQLSSNFGAISVVLNDILSRLDKVEAEVQI  
 DRLITGRQLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSPQAAPH  
 GVVFLHVTVYVPSQERNFTTAPAICHEGKAYFPREGVVFVNGTSWFITQRNFFSPQIITTDNTFVS  
 15 GNCDEVVIGIINNTVYDPLQPELDSFKEELDKYFKNHTSPDVLGDISGINASVNIQKEIDRLNE  
 VAKNLNESLIDLQELGKYEQYIKWPWYVWLGFIAGLIAIVMVTILLCCMTSCCSCLKGACSCGSC  
 CKFDEDDSEPVLGKVKLHYT

Native MERS S protein wtTM/CT AA AFY13307 (SEQ ID NO: 113)

15 MIHSVFLLMFLLTPTESYVDVGPDSVKSACIEVDIQQTFFDKTWPRPIDVSKADGIIYPQGRITYS  
 NITITYQGLFPYQGDHGDYVYSAGHATGTTQKLFVANYSQDVKQFANGFVVRIGAAANSTGTV  
 IISPSTSATIRKIYPAFMLGSSVGNFSDGKMGRFFNHTLVLLPDGCGTLLRAFYCILEPRSGNHC  
 PAGNSYTSFATYHTPATDCSDGNYNRNASLNSFKEYFNLRNCTFMYTYNITEDEILEWFGITQTA  
 20 QGVHLFSSRYVDLYGGNMFQFATLPVYDTIKYYSIIPHSIRSIQSDRKAWAAFVYVKLQPLTFL  
 DFSVDGYIRRAIDCGFNDLSQLHCSYESFDVESGVSVSSFEAKPSGSVVEQAEGVECDFSPLLS  
 GTPPQVYNFKRLVFTNCNYNLTKLLSLFSVNDFTCQSIPAAIASNCYSSILIDYFSYPLSMKSD  
 LSVSSAGPISQFNYKQSFNSPTCLILATVPHNLTITKPLKYSYINKCSRFLSDDRTEVPQLVNA  
 NQYSPCVSIVPSTVWEDGDYRQKQLSPLEGGGWLVASGSTVAMTEQLQMGFGITVQYGTDTNSVC  
 25 PKLEFANDTKIASQLGNCVEYSLYGVSGRGVFQNCITAGVVRQRFVYDAYQNLVGYYSDDGNYYC  
 LRACVSVVSVIYDKETKTHATLFGSVACEHISSTMSQYSRSTRSMLKRRDSTYGPLQTPVGCVL  
 GLVNSSLFVEDCKLPLGQSLCALPDTPSTLTPRSVRSVPGEMRLASIAFNHPIQVDQLNSSYFKL  
 SIPTNFSFGVTQEQYIQTITIQKVTVDCKQYVCNGFQKCEQLLREYGFQCSKINQALHGANLRQDDS  
 VRNLFASVKSSQSSPIIPGFGGDFNLTLLEPVSISTGSRARSATIEDLLFDKVTIADPGYMQGYD  
 DCMQQGPASARDLICAQYVAGYKVLPLMDVNMEAAAYTSLLGSIAGVGWTAGLSSFAAIPFAQS  
 30 IFYRLNGVGITQQVLSENQKLIANKFNQALGAMQTGFTTTNEAFHKVQDAVNNNNAQALSKLASEL  
 SNTFGAISASIGDIQRLDVLEQDAQIDRLINGRLTTLNAFVAQQQLVRSESAALSAQLAKDKVNE  
 CVKAQSKRSGFCGQGTHTIVSFVNAPNGLYFMHVGYYPSNHIEVVSAYGLCDAANPTNCIAPVNG  
 YFIKTNNTRIVDEWSYTGSSFYAPEPITSLNTRYKQVQVQYQNIISTNLPPPLGNSTGIDFQDEL  
 DEFFKNVSTSIPNFGSLTQINTTLLDLTYEMLSLQQVVKALNESYIDLKELGNYTYYNKWPWYIW  
 LGFIAGLVALALCVFFILCCTGCGTNCMGKLCNRCCDRYEEYDLEPHKVHVH

35 Native SARS-CoV-1 S protein wtTM/CT AA P59594 without signal peptide (SEQ ID NO: 114)

SDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLFLPFYSNVTGFHTINHTFGN  
 PVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQSVIIINNSTNVVIRACNFELCDNPFPAVSKP  
 MGTQTHMTIFDNAFNCTFEYISDAFSLDVSEKSGNFKHLREFVFNKDGFLYVYKGYQPIDVVRD  
 40 LPSGFNTLKPIFKLPLGINITNFRAILTAQSPAQDIWGTSAAYFVGYLKPTTFMLKYDENGITIT  
 DAVDCSQNPLAELKCSVKSFIEDKGIYQTSNFRVVPSPGDVVRFPNITNLCPFGEVFNATKFPVSY  
 AWERRKISNCVADYSVLYNSTFFSTFKCYGVSATKLNLDLCSNVYADSFVVKGDVVRQIAPGQTG  
 VIADYNKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVFPSPDGKPC  
 PPALNCYWPNDYGFYTTTGIGYQPYRVVLSFELLNAPATVCGPKLSTDLIKNQCVNFNFNGLT  
 45 GTGVLTTPSSKRFQPFQFGRDVSDFDTSVRDPKTSEILDISPCSFGGVSVITPGTNASSEVAVLY  
 QDVNCTDVSTAIHADQLTPAWRIYSTGNNVFQTQAGCLIGAEHVDTSYECDIPIGAGICASYHTV  
 SLLRSTSQKSIVAYTMSLGADSSIAYSNNTIAIPTNFSISITTEVMPVSMKTSVDCNMYICGDS  
 TECANLLLQYGSFCTQLNRALSGIAAEQDRNTREVFAQVKQMYKTPTLKYFGGFNFSQILPDPLK  
 PTKRSFIEDLLFNKVTLADAGFMKQYGECLGDINARDLICAQKFNGLTVLPPLLTDDMIAAYTAA

5 LVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKQIANQFNKAISQIQESLTTT  
 STALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRDKVEAEVQIDRLITGRLQSLQT  
 YVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQAAPHGVVFLHVITYVPSQ  
 ERNFTTAPAICHEGKAYFPREGVFVFNGTSWFITQRNFFSPQIITDNTFVSGNCDVVIGIINNT  
 VYDPLQPELDSFKEELDKYFKNHTSPDVLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQ  
 ELGKYEQYIKWPWYVWLGFIAGLIAIVMVTILLCCMTSCCSCLKGACSCGSCCKFDEDDSEPVLK  
 GVKLHYT

Native MERS S protein wtTM/CT AA AFY13307 without signal peptide (SEQ ID NO: 115)

10 YVDVGPDSVKSACIEVDIQQTFFDKTWPRPIDVSKADGIIYPQGRITYSNITITYQGLFPYQGDHG  
 DMYVYSAGHATGTTPOKL FVANYSDVKQFANGFVVRIGAAANSTGTVIISPSTSATIRKIYPAF  
 MLGSSVGNFSDGKMGRFFNHTLVLLPDGCGTLLRAFYCILEPRSGNHCPAGNSYTSFATYHTPAT  
 DCSDGNYNRNASLNSFKEYFNLRNCTFMYTYNITEDEILEWFGITQTAQGVHLFSSRYVDLYGGN  
 MFQFATLPVYDTIKYYSIIPHSIRSIQSDRKAWAAFYVYKLQPLTFLLDFSVLDGYIRRAIDCGFN  
 15 DLSQLHCSYESFDVESGVYSVSSFEAKPSGSVVEQAEGVECDFSPLLSGTPPQVYNFKRLVFTNC  
 NYNLTKLLSLFSVNDFTCSQISPAAIASNCYSSLILDYFSYPLSMKSDLVSSAGPISQFNYKQS  
 FSNPTCLILATVPHNLTTITKPLKYSYINKCSRFLSDDRTEVPQLVNANQYSPCVSIVPSTVWED  
 GDYRKLQSLPLEGGGWLVASGSTVAMTEQLQMGFGITVQYGTDTNSVCPKLEFANDTKIASQLGN  
 CVEYSLYGVSGRGVFNCTAVGVVRQRFVYDAYQNLVGYYSDDGNYYCLRACVSVFVSVIYDKET  
 KTHATLFGSVACEHISSTMSQYSRSTRSMLKRRDSTYGPLQTPVGCVLGLVNSSLFVEDCKLPLG  
 20 QSLCALPDTPSTLTPRSVRSVPGEMRLASIAFNHPIQVDQLNSSYFKLSIPTNFSFGVTQEYIQT  
 TIQKVTVDCKQYVCNGFQKCEQLLREYQGFCSKINQALHGANLRQDDSVRNLFASVKSSQSSPII  
 PGFGGDFNLTLLEPVSISTGSRARSASIEDLLFDKVTIADPGYMQGYDDCMQGPASARDLICAQ  
 YVAGYKVLPPMLDVNMEAAYTSSLLGSIAGVGWTAGLSSFAAIPFAQSI FYRLNGVGITQQVLSE  
 NQKLIANKFNQALGAMQTGFTTTNEAFHKVQDAVNNNAQALSKLASELSNTFGAISASIGDIIQR  
 25 LDVLEQDAQIDRLINGRLTTLNAFVAQQLVRSESAALSAQLAKDKVNECVKAQSKRSGFCGQGTH  
 IVSFVNAPNGLYFMHVGYYPSNHIEVVSAYGLCDAANPTNCIAPVNGYFIKTNNTRIVDEWSYT  
 GSSFYAPEPITSLNTKYVAPQVTYQNI STNLPPPLLGNSTGIDFQDELDEFFKNVSTSI PNFGSL  
 TQINTTLLDLTYEMLSLQQVVKALNESYIDLKELGNYTYYNKWPWYIWLGFIAGLVALALCVFFI  
 LCCTGCGTNCMGKCLKCNRCDDRYEEDLEPHKVHVH

30 TMCT region of modified PDI-SARS-COV-1 wtTMCT-AA (SEQ ID NO: 116)

WYVWLGFIAGLIAIVMVTILLCCMTSCCSCLKGACSCGSCCKFDEDDSEPVLKGVKLHYT

TMCT region of modified PDI-SARS-COV-1 H5iTMCT-AA (SEQ ID NO: 117)

WYQILSIYSTVASSLALAIMMAGLSLWMCSNGSLQCRICI

TMCT region of modified PDI-SARS-COV-1 H5iCT-AA (SEQ ID NO: 118)

35 WYVWLGFIAGLIAIVMVTILLSLWMCSNGSLQCRICI

TMCT region of modified PDI-SARS-COV-1 H5iCT(V4)-AA (SEQ ID NO: 119)

WYVWLGFIAGLIAIVMVTILLCCSNGSLQCRICI

TMCT region of modified PDI-SARS-COV-1 H1cCT-AA (SEQ ID NO: 120)

WYVWLGFIAGLIAIVMVTILLSFWMCSNGSLQCRICI

- TMCT region of modified PDI-MERS-wtTMCT-AA (SEQ ID NO: 121)  
WYIWLGFIAGLVALALCVFFILCCTGCGTNCMGKLCNRCCDRYEEYDLEPHKVHVH
- TMCT region of modified PDI-MERS-H5iTMCT-AA (SEQ ID NO: 122)  
WYQILSIYSTVASSLALAIMMAGLSLWMCSNGSLQCRICI
- 5 TMCT region of modified PDI-MERS-H5iCT-AA (SEQ ID NO: 123)  
WYIWLGFIAGLVALALCVFFILSLWMCSNGSLQCRICI
- TMCT region of modified PDI-MERS-H5iCT(V4)-AA (SEQ ID NO: 124)  
WYIWLGFIAGLVALALCVFFILCCSNGSLQCRICI
- 10 TMCT region of modified PDI-MERS-H1cCT-AA (SEQ ID NO: 125)  
WYIWLGFIAGLVALALCVFFILSFWMCSNGSLQCRICI
- TMCT region of modified PDI-S-protein+H1 Cal (SEQ ID NO: 126)  
WYIWLGFIAGLIAIVMVTIMLSFWMCSNGSLQCRICI
- TMCT region of modified PDI-S-protein+H3 Minn (SEQ ID NO: 127)  
WYIWLGFIAGLIAIVMVTIMLMWACQKGNIRCNICI
- 15 TMCT region of modified PDI-S-protein+H6 HK (SEQ ID NO: 128)  
WYIWLGFIAGLIAIVMVTIMLGLWMCSNGSMQCRICI
- TMCT region of modified PDI-S-protein+H7 Guangdong (SEQ ID NO: 129)  
WYIWLGFIAGLIAIVMVTIMLVFICVKNGNMRCTICI
- TMCT region of modified PDI-S-protein+H9 HK (SEQ ID NO: 130)
- 20 TMCT region of modified PDI-S-protein+ B/ Wash (SEQ ID NO: 131)  
WYIWLGFIAGLIAIVMVTIMLVVYMVSRDNVSCSICL
- Consensus Sequence of TM Domain of Coronavirus S-protein (SEQ ID NO: 132)  
WYXWLGFIAGLXAXXX{X}VXXXL ({X} may be absent)
- 25 Consensus Sequence of TM Domain of Coronavirus S-protein (SEQ ID NO: 133)  
WY[I/V]WLGFIAGL[V/I]A[L/I][A/V][L/M]{X}V[F/T][F/I]XL (wherein {X} may be C or absent)
- TM/CT Region of Modified SARS-CoV-1 S protein with intervening peptide sequence X<sub>n</sub> (SEQ ID NO: 134)

WYVWLGFIAGLIAIVMTIL - (X)<sub>n</sub> - CSNGSXXCXICI

TM/CT Region of Modified MERS S protein with intervening peptide sequence X<sub>n</sub> (SEQ ID NO: 135)

WYIWLGFIAGLVALALCVFFIL - (X)<sub>n</sub> - CSNGSXXCXICI

5

IF(AvB+wtCT-OC43).r (SEQ ID NO: 136)

ACGACACGACTAAGGCCTTCAGTCGTCATGCGAGGTCTTAATGACAAGC

PDI-OC43-wtTMCT-DNA (SEQ ID NO: 137)

10 atggcgaaaaacgttgcgattttcggcttattgttttctcttcttgtgttggttccttctcagat  
cttcgcgGTGATCGGCGATCTGAATTGTACCCTGGATCCCCGCCTGAAAGGGAGCTTTAACAACC  
GAGATACAGGACCCCCGTCTATATCCATAGATACAGTGGATGTTACGAACGGGCTCGGCACCTAC  
TATGTGCTAGACCGAGTTTATTTGAACACCACCTTATTCCTCAATGGATACTACCCAACCTCAGG  
15 TAGTACTTACAGAAACATGGCGCTGAAGGGTACGGATCTGCTGAGCACCTATGGTTTAAACCTC  
CCTTCTCTCGGACTTTATTAATGGCATCTTCGCTAAGGTGAAAAACACGAAGGTTTTCAAAGAT  
GGAGTGATGATTTCAGAGTTCCCTGCGATCACCATTGGAAGTACCTTCGTGAATACTTCCTATAG  
CGTGGTGGTTCAACCACGGACAATCAACTCCACCCAGGACGGCGTCAACAAGCTCCAGGGATTGC  
TGGAGGTGTCAGTCTGTCAATATAACATGTGTGAGTACCCACACACTATCTGTCAACCTAATCTA  
GGCAACCACTTTAAGGAACTGTGGCACTACGATACGGGGGTGGTAAGTTGCTTATATAAGAGAAA  
20 TTTCACTATGATGTTAATGCAACGTACCTGTACTTTCACTTCTATCAAGAAGGAGGAACCTTCT  
ACGCATATTTACAGATACCGGCTTTGTGACGAAATCTTATTCAACGTTTACCTCGGAATGGCA  
TTAAGCCATTATTACGTGATGCCTCTCACTTGCATCAGACGCCCTAAGGATGGTTTTTCTCTGGA  
GTACTGGGTCACTCCCCTGACACCACGGCAGTACCTGCTTGCTTTTAACCAGGACGGTATCATTT  
TTAATGCCGTGATTGTATGAGCGATTTTATGAGCGAGATAAAGTGCAAGACCCAATCTATTGCT  
25 CCGCCCACGGGGGTGTACGAACTGAATGGTTACACCGTCCAGCCCGTTGCCGATGTATATAGACG  
GAAACCAGACCTGCCAATTGCAACATCGAAGCTTGGTTAAACGATAAGTCAGTGCCCTCCCCC  
TCAATTGGGAGAGGAAGACTTTCTCCAACGTGAATTTCAACATGTCAAGCCTGATGTCTTTCATT  
CAAGCCGATTTCGTTCACTTGTAATAATATAGATGCAGCAAAGATCTATGGTATGTGCTTCAGTTC  
CATCACAATAGATAAGTTTGAATACCAAACCGTCGCAAGGTGGACCTTCAGCTCGGCAACCTGG  
30 GCTATCTGCAGTCCAGCAATTATAGAATAGACACCACCGCCACATCATGTCAGCTGTACTATAAC  
CTCCCAGCAGCGAACGTCACTGTTAGTAGGTTCAATCCTTCTACCTGGAATAAAAGGTTTGGATT  
CATCGAAGATAGTGTGTTTCGTACCTCAGCCAACAGGAGTGTTCACCAATCAGAGCGTGGTCTACG  
CCCAACATTGCTTCAAGGCACCCAAAAATTTCTGCCCATGTAGCAGTTGCTCCTGCCCGGGTAAG  
AACAATGGGATCGGCACCTGCCAGCAGGCACCAATTCACCTACATGCGATAATCTGTGTACACT  
35 GGATCCTATTACACTTAAGGCCCTGATACCTACAAATGCCCCAGAGCAAGAGCCTGGTCCGTA  
TCGGAGAACACTGTTCCGGACTTGCAAGTAAAAAGCGACTATTGTGGAAATAACTCTTGCACTTGT  
CAGCCACAAGCCTTCCCTCGGTTGGTCCGCTGACTCTTGTTTACAAGGGGATAAGTGTAACATCTT  
CGCAAATTTTCATCTTACACGATGTGAATAACGGCTTAACATGCAGCACAGATCTCCAGAAGGCAA  
ACACAGAGATCGAATTAGGAGTCTGCGTTAATTACGATCTCTACGGGATCTCTGGCCAGGGCATC  
40 TTCGTGGAGGTTAATGCTACCTACTACAATAGTTGGCAAAATCTGCTCTACGATAGCAATGGCAA  
CCTCTATGGATTACAGAGACTATATTACTAACAGGACGTTTCATGATTCACTCGTGCTATTCCGGGC  
GGGTGTCAGCAGCTTATCAGCAAATCTTCAGAGCCAGCTCTGCTATTCCGAAACATAAAATGT  
AATTACGTGTTCAATAATTCACTGACTCGGCAGCTGCAGCCGATTAAATTACAGCTTCGACAGCTA  
45 CCTTGGTTGCGTTGTTAACGCCTACAACCTCCACTGCCATATCAGTTTCAGACCTGCGACCTTACTG  
TGGGCTCTGGCTATTGTGTGCGATTATTCAAAGAACGGGGGAGCGGGTCCGCAATAACAACCTGGC  
TATAGGTTACCAATTTTGAGCCTTTCACCGTGAATAGTGTCAACGATAGCCTGGAGCCTGTGCG

AGGTCTTTATGAGATACAAATCCCCTCCGAGTTCACAATTGGCAACATGGAAGAGTTCATCCAGA  
CGAGTTCCTCCAAAGGTGACGATCGATTGCGCGGCTTTCGTCTGCGGCGACTACGCCGCATGCAAG  
TTACAACCTCGTTGAGTATGGAAGTTTTTTCGATAATATAAACGCAATTCTGACTGAAGTGAACGA  
ACTGCTGGACACCACTCAGTTGCAGGTGGCAAATTCGCTCATGAACGGCGTGACACTGTCAACCA  
5 AACTGAAGGACGGTGTCAATTTCAATGTGGATGACATTAACCTTCAGCCCCGTACTGGGCTGTTTG  
GGTAGTGAGTGTTCTAAGGCTAGCAGCCGCTCCGCCATTGAGGACTTGTGTTTTGATAAAGTTAA  
GCTGAGTGACGTTGGATTTGTTGAGGCGTATAATAACTGTACCGGTGGTGAGAGATAAGGGATC  
TGATCTGTGTCCAGAGTTATAAGGGGATTAAGGTTCTCCCCCGCTACTCTCGGAGAATCAGATA  
10 TCAGGATACACCCCTGGCCGCTACCTCAGCCTCGCTGTTTCCCCCTTGGACCGCTGCCGCCGGTGT  
CCCATTTTATTTGAATGTGCAGTATCGGATCAACGGTCTGGGAGTGACAATGGACGTGCTGTCTC  
AGAACCAGAAACTGATCGCCAATGCATTCAACAATGCTCTGCACGCCATCCAGCAAGGGTTTTGAC  
GCTACAAATTTCTGCCCTCGTAAAAATCCAGGCCGTGGTGAATGCTAACGCCGAAGCCCTTAATAA  
TCTGCTCCAGCAGCTTTCTAACCGCTTTGGAGCTATTTCTGCCCTACTGCAGGAAATTCTATCCA  
15 GACTGGATCCCCCTGAGGCAGAAGCCCAAATCGACCGTCTCATAAACGGCAGACTCACTGCTCTT  
AACGCCCTACGTTAGTCAACAATTGAGCGATTTCGACCTTGGTGAAATTCAGCGCAGCTCAGGCTAT  
GGAGAAGGTGAACGAGTGCGTGAAGTCACAGAGCTCCAGAATCAATTTCTGTGGCAATGGGAACC  
ATATCATCTCCTTGGTTTCAAGATGCTCCCTACGGCCTGTATTTCACTCCACTTCAACTACGTGCC  
ACGAAGTACGTTACAGCCAAAGTGTCCCCCGACTGTGCATCGCTGGTAACAGGGGCATTGCACC  
20 AAAATCCGGCTACTTCGTCAATGTCAACAACACATGGATGTATACTGGGAGTGGTTATTATTACC  
CTGAACCTATAACAGAGAACAATGTAGTAGTCATGTCCACATGCGCCGTCAATTATACTAAGGCC  
CCCTATGTTATGCTCAACACTTCAATTCCCAATCTCCCGGATTTCAAAGAAGAGCTGGATCAGTG  
GTTTAAGAATCAGACATCCGTGGCCCCCTGACTTAAGCTTGGATTATATCAATGTGACTTTTTTATAG  
ACTTACAGGTTCGAGATGAACCGACTCCAGGAAGCTATAAAAGTACTGAACCACTCCTATATCAAT  
25 CTGAAAGATATCGGTACATACGAATATTACGTAAAATGGCCTTGGTATGTGTGGCTACTAATTTG  
CCTTGCGGGCGTGGCTATGCTGGTCTGCTGTTCTTCATTTGCTGCTGTACTGGGTGCGGTACTT  
CCTGCTTCAAAAAGTGCGGTGGTTGCTGCGACGACTACACTGGGTACCAGGAGCTTGTCAATTAAG  
ACCTCGCATGACGACTGA

## PDI-OC43-H5iTMCT-DNA (SEQ ID NO: 138)

atggcgaaaaaacgttgcgatTTTTcggttattgttttctcttcttgtgttggttccttctcagat  
cttcgCGTGATCGGCGATCTGAATTGTACCCTGGATCCCCGCCTGAAAGGGAGCTTTAAACAACC  
GAGATACAGGACCCCCGTCTATATCCATAGATACAGTGGATGTTACGAACGGGCTCGGCACCTAC  
TATGTGCTAGACCGAGTTTATTTGAACACCACCTTATTCCTCAATGGATACTACCCAACCTCAGG  
30 TAGTACTTACAGAAACATGGCGCTGAAGGGTACGGATCTGCTGAGCACCTATGGTTTTAAACCTC  
35 CCTTCTCTCGGACTTTATTAATGGCATCTTCGCTAAGGTGAAAAACGAAGGTTTTCAAGAT  
GGAGTGATGTATTACAGATTCCCTGCGATCACCATTGGAAGTACCTTCGTGAATACTTCTATAG  
CGTGGTGGTTCAACCACGGACAATCAACTCCACCCAGGACGGCGTCAACAAGCTCCAGGGATTGC  
TGGAGGTGTCAGTCTGTCAATATAACATGTGTGAGTACCCACACACTATCTGTCAACCTAATCTA  
40 GGCAACCACTTTAAGGAACGTGTGGCACTACGATACGGGGGTGGTAAGTTGCTTATATAAGAGAAA  
TTTCACCTATGATGTTAATGCAACGTACCTGTACTTTCACTTCTATCAAGAAGGAGGAACCTTCT  
ACGCATATTTACAGATACCGGCTTTGTGACGAAATCTTATTCAACGTTTACCTCGGAATGGCA  
TTAAGCCATTATTACGTGATGCCTCTCACTTGCATCAGACGCCCTAAGGATGGTTTTTCTCTGGA  
GTACTGGGTCACTCCCCTGACACCACGGCAGTACCTGCTTGTCTTTAACCAGGACGGTATCATTT  
TTAATGCCGTGCGATTGTATGAGCGATTTTATGAGCGAGATAAAGTGCAAGACCCAATCTATTGCT  
45 CCGCCACGGGGGTGTACGAACCTGAATGGTTACACCGTCCAGCCCGTTGCCGATGTATATAGACG  
GAAACCAGACCTGCCCAATTGCAACATCGAAGCTTGGTTAAACGATAAGTCAGTGCCCTCCCCC  
TCAATTGGGAGAGGAAGACTTTCTCCAACCTGTAATTTCAACATGTCAAGCCTGATGTCTTTCATT  
CAAGCCGATTTCGTTCACTTGTAAATAATATAGATGCAGCAAAGATCTATGGTATGTGCTTCAGTTC  
CATCACAATAGATAAGTTTGAATACCAAACCGTCGCAAGGTGGACCTTCAGCTCGGCAACCTGG  
50 GCTATCTGCAGTCCAGCAATTATAGAATAGACACCACCGCCACATCATGTGAGCTGTACTATAAC

CTCCCAGCAGCGAACGTCAGTGTTAGTAGGTTCAATCCTTCTACCTGGAATAAAAGGTTTGGATT  
CATCGAAGATAGTGTGTTTCGTACCTCAGCCAACAGGAGTGTTCACCAATCACAGCGTGGTCTACG  
CCCAACATTGCTTCAAGGCACCCAAAAATTTCTGCCCATGTAGCAGTTGCTCCTGCCCGGGTAAG  
AACAATGGGATCGGCACCTGCCCAGCAGGCACCAATTCACCTACATGCCGATAATCTGTGTACACT  
5 GGATCCTATTACACTTAAGGCCCCCTGATACCTACAAATGCCCCAGAGCAAGAGCCTGGTTCGGTA  
TCGGAGAACACTGTTCCGGACTTGCAGTAAAAAGCGACTATTGTGGAAATAACTCTTGCACTTGT  
CAGCCACAAGCCTTCCCTCGGTTGGTCCGCTGACTCTTGTTTACAAGGGGATAAGTGTAAACATCTT  
CGCAAATTTTCATCTTACACGATGTGAATAACGGCTTAACATGCAGCACAGATCTCCAGAAGGCAA  
10 ACACAGAGATCGAATTAGGAGTCTGCGTTAATTACGATCTCTACGGGATCTCTGGCCAGGGCATC  
TTCGTGGAGGTTAATGCTACCTACTACAATAGTTGGCAAAATCTGCTCTACGATAGCAATGGCAA  
CCTCTATGGATTACAGAGACTATATTACTAACAGGACGTTTCATGATTCACCTCGTGCTATTCCGGGC  
GGGTGTCAGCAGCTTATCACGCAAATTCCTTCAGAGCCAGCTCTGCTATTCCGAAACATAAAATGT  
AATTACGTGTTCAATAATTCAGTACTCGGCAGCTGCAGCCGATTAATTACAGCTTCGACAGCTA  
15 CCTTGGTTGCGTTGTTAACGCCTACAACCTCCACTGCCATATCAGTTCAGACCTGCGACCTTACTG  
TGGGCTCTGGCTATTGTGTGCGATTATTCAAAGAACGGGGGAGCGGGTCCGCAATAACAACCTGGC  
TATAGGTTACCAATTTTGAGCCTTTACCCGTGAATAGTGTCAACGATAGCCTGGAGCCTGTCTGG  
AGGTCTTTATGAGATACAAATCCCCTCCGAGTTCACAATTGGAACATGGAAGAGTTTCATCCAGA  
CGAGTTCCTCCAAAGGTGACGATCGATTGCGCGGCTTTCGTCCTGCGGCGACTACGCCGCATGCAAG  
TTACAACCTCGTTGAGTATGGAAGTTTTTTCGATAATATAAACGCAATTCCTGACTGAAGTGAACGA  
20 ACTGCTGGACACCACTCAGTTGCAGGTGGCAAATTCGCTCATGAACGGCGTGACACTGTCAACCA  
AACTGAAGGACGGTGTCAATTTCAATGTGGATGACATTAACTTCAGCCCCGTACTGGGCTGTTTG  
GGTAGTGAGTGTTCTAAGGCTAGCAGCCGCTCCGCCATTGAGGACTTGTGTTTTGATAAAGTTAA  
GCTGAGTGACGTTGGATTTGTTGAGGCGTATAATAACTGTACCGGTGGTGCAGAGATAAGGGATC  
TGATCTGTGTCCAGAGTTATAAAGGGGATTAAGGTTCTCCCCCGCTACTCTCGGAGAATCAGATA  
25 TCAGGATACACCCTGGCCGCTACCTCAGCCTCGCTGTTTCCCCCTTGACCGCTGCCGCCGGTGT  
CCCATTTTATTTGAATGTGCGATATCGGATCAACGGTCTGGGAGTGACAATGGACGTGCTGTCTC  
AGAACCAGAACTGATCGCCAATGCATTCAACAATGCTCTGCACGCCATCCAGCAAGGGTTTGAC  
GCTACAAATTTCTGCCCTCGTAAAAATCCAGGCCGTGGTGAATGCTAACGCCGAAGCCCTTAATAA  
TCTGCTCCAGCAGCTTTCTAACCGCTTTGGAGCTATTTCTGCCCTCACTGCAGGAAATCTATCCA  
30 GACTGGATCCCCCTGAGGCAGAAGCCCCAATCGACCGTCTCATAAACGGCAGACTCACTGCTCTT  
AACGCCTACGTTAGTCAACAATTGAGCGATTGACCTTGGTGAAATTCAGCGCAGCTCAGGCTAT  
GGAGAAGGTGAACGAGTGCCTGAAGTCACAGAGCTCCAGAATCAATTTCTGTGGCAATGGGAACC  
ATATCATCTCCTTGGTTCAGAATGCTCCCTACGGCCTGTATTTCACTCACTTCAACTACGTGCC  
ACGAAGTACGTTACAGCCAAAGTGTCCCCCGACTGTGCATCGCTGGTAACAGGGGCAATTGCACC  
35 AAAATCCGGCTACTTCGTCAATGTCAACAACACATGGATGTATACTGGGAGTGGTTATTATTACC  
CTGAACCTATAACAGAGAACAATGTAGTAGTCATGTCCACATGCGCCGTCAATTATACTAAGGCC  
CCCTATGTTATGCTCAACACTTCAATTCCCAATCTCCCGATTTCAAAGAAGAGCTGGATCAGTG  
GTTTAAGAATCAGACATCCGTGGCCCCCTGACTTAAGCTTGGATTATATCAATGTGACTTTTTTAG  
ACTTACAGGTGCGAGATGAACCGACTCCAGGAAGCTATAAAAGTACTGAACCACTCCTATATCAAT  
40 CTGAAAGATATCGGTACATACGAATATTACGTAAAATGGCCTTGGTATcaaatactgtcaattta  
ttcaacagtggcgagttccctagcactggcaatcatgatggctggtctatctttatggatgtgct  
ccaatggatcggttacaatgcagaatttgcattTGA

**PDI-OC43-H5iCT-DNA (SEQ ID NO: 139)**

atggcgaaaaacgttgcgattttcggccttattgttttctcttcttctgtgttggttccttctcagat  
45 cttcgcgGTGATCGGCGATCTGAATTGTACCCTGGATCCCCGCCTGAAAGGGAGCTTTAACAACC  
GAGATACAGGACCCCCGTCTATATCCATAGATACAGTGGATGTTACGAACGGGCTCGGCACCTAC  
TATGTGCTAGACCGAGTTTATTTGAACACCACCTTATTCCTCAATGGATACTACCCAACCTCAGG  
TAGTACTTACAGAAACATGGCGCTGAAGGGTACGGATCTGCTGAGCACCTATGGTTTTAAACCTC  
50 CCTTCCTCTCGGACTTTATTAATGGCATCTTCGCTAAGGTGAAAAACAGGAAGGTTTTCAAAGAT

5 GGAGTGATGTATTTCAGAGTTCCCTGCGATCACCATTGGAAGTACCTTCGTGAATACTTCCTATAG  
CGTGGTGGTTCAACCACGGACAATCAACTCCACCCAGGACGGCGTCAACAAGCTCCAGGGATTGC  
TGGAGGTGTCAGTCTGTCAATATAACATGTGTGAGTACCCACACACTATCTGTCACCCTAATCTA  
GGCAACCACTTTAAGGAACTGTGGCACTACGATACGGGGTGGTAAGTTGCTTATATAAGAGAAA  
10 TTTACCTATGATGTTAATGCAACGTACCTGTACTTTCACTTCTATCAAGAAGGAGGAACTTTCT  
ACGCATATTTACAGATACCGGCTTTGTGACGAAATTCCTTATTCACGTTTACCTCGGAATGGCA  
TTAAGCCATTATTACGTGATGCCTCTCACTTGCATCAGACGCCCTAAGGATGGTTTTTCTCTGGA  
TACTGGGTCACTCCCCTGACACCACGGCAGTACCTGCTTGCTTTTAACCAGGACGGTATCATTT  
15 TTAATGCCGTGCGATTGTATGAGCGATTTTATGAGCGAGATAAAGTGCAAGACCCAATCTATTGCT  
CCGCCCACGGGGGTGTACGAACTGAATGGTTACACCGTCCAGCCCGTTGCCGATGTATATAGACG  
GAAACCAGACCTGCCCAATTGCAACATCGAAGCTTGGTTAAACGATAAAGTCAGTGCCCTCCCCC  
TCAATTGGGAGAGGAAGACTTTCTCCAACGTAAATTTCAACATGTCAAGCCTGATGTCTTTCATT  
CAAGCCGATTCTGTTCACTTGTAAATAATATAGATGCAGCAAAGATCTATGGTATGTGCTTCAGTTC  
CATCACAATAGATAAGTTTGAATACCAAACCGTCGCAAGGTGGACCTTCAGCTCGGCAACCTGG  
20 GCTATCTGCAGTCCAGCAATTATAGAATAGACACCACCGCCACATCATGTGAGCTGTACTATAAC  
CTCCCAGCAGCGAACGTGAGTGTAGTAGGTTCAATCCTTCTACCTGGAATAAAAGGTTTGGATT  
CATCGAAGATAGTGTGTTTCGTACCTCAGCCAACAGGAGTGTTCACCAATCACAGCGTGGTCTACG  
CCCAACATTTGCTTCAAGGCACCCAAAATTTCTGCCCATGTAGCAGTTGCTCCTGCCCGGGTAAG  
AACAAATGGGATCGGCACCTGCCAGCAGGCACCAATTCACCTACATGCGATAATCTGTGTACACT  
25 GGATCCTATTACACTTAAGGCCCTGATACCTACAAATGCCCCAGAGCAAGAGCCTGGTCGGTA  
TCGGAGAACACTGTTCCGGACTTGCGAGTAAAAAGCGACTATTGTGGAAATAACTCTTGCACTTGT  
CAGCCACAAGCCTTCTCGGTTGGTCCGCTGACTCTTGTTTACAAGGGGATAAGTGTAAACATCTT  
CGCAAATTTTCATCTTACACGATGTGAATAACGGCTTAACATGCAGCACAGATCTCCAGAAGGCAA  
ACACAGAGATCGAATTAGGAGTCTGCGTTAATTACGATCTCTACGGGATCTCTGGCCAGGGCATC  
30 TTCGTGGAGGTTAATGCTACCTACTACAATAGTTGGCAAAATCTGCTCTACGATAGCAATGGCAA  
CCTCTATGGATTACAGAGACTATATTACTAACAGGACGTTTCATGATTCACTCGTGCTATTCCGGGC  
GGGTGTCAGCAGCTTATCACGCAAATTCCTCAGAGCCAGCTCTGCTATTCCGAAACATAAAATGT  
AATTACGTGTTCAATAATTCAGTACTCGGCAGCTGCAGCCGATTAATTACAGCTTCGACAGCTA  
CCTTGGTTGCGTTGTTAACGCCTACAACCTCCACTGCCATATCAGTTCAGACCTGCGACCTTACTG  
35 TGGGCTCTGGCTATTGTGTGCGATTATTCAAAGAACGGGGGAGCGGGTCCGCAATAACAACCTGGC  
TAGGTTACCAATTTTGAGCCTTTACCCTGAATAGTGTCAACGATAGCCTGGAGCCTGTGCGG  
AGGTCCTTATGAGATACAAATCCCTCCGAGTTCACAATTGGCAACATGGAAGAGTTTCATCCAGA  
CGAGTTCCCCAAAGGTGACGATCGATTGCGCGGCTTTTCGTCTGCGCGGACTACGCCGCATGCAAG  
TTACAACCTCGTTGAGTATGGAAGTTTTTGGGATAATATAAACGCAATTCTGACTGAAGTGAACGA  
40 ACTGCTGGACACCACTCAGTTGCAGGTGGCAATTCGCTCATGAACGGCGTGACACTGTCAACCA  
AACTGAAGGACGGTGTCAATTTCAATGTGGATGACATTAACCTCAGCCCCGTAAGGCTGTTTGG  
GGTAGTGAGTGTCTAAGGCTAGCAGCCGCTCCGCCATTGAGGACTTGTGTTTGATAAAGTTAA  
GCTGAGTGACGTTGGATTTGTTGAGGCGTATAATACTGTACCGGTGGTGAGAGATAAAGGGATC  
TGATCTGTGTCCAGAGTTATAAGGGGATTAAGGTTCTCCCCCGCTACTCTCGGAGAATCAGATA  
45 TCAGGATACACCTGGCCGCTACCTCAGCCTCGCTGTTTCCCCCTTGACCGCTGCCGCCGGTGT  
CCCATTTTATTTGAATGTGCAGTATCGGATCAACGGTCTGGGAGTGACAATGGACGTGCTGTCTC  
AGAACCAGAACTGATCGCCAATGCATTCAACAATGCTCTGCACGCCATCCAGCAAGGGTTTGAC  
GCTACAAATTCGCCCCTCGTAAAAATCCAGGCCGTGGTGAATGCTAACGCCGAAGCCCTTAATAA  
TCTGCTCCAGCAGCTTTCTAACCCTTTGGAGCTATTTCTGCCTCACTGCAGGAAATTCATATCCA  
50 GACTGGATCCCCCTGAGGCAGAAGCCCAATCGACCGTCTCATAAACGGCAGACTCACTGCTCTT  
AACGCCTACGTTAGTCAACAATTGAGCGATTGACCTTGGTGAAATTCAGCGCAGCTCAGGCTAT  
GGAGAAGGTGAACGAGTGCCTGAAGTCACAGAGCTCCAGAATCAATTTCTGTGGCAATGGGAACC  
ATATCATCTCCTTGGTTTCAAGATGCTCCCTACGGCCTGTATTTTCATCCACTTCAACTACGTGCCC  
ACGAAGTACGTTACAGCCAAAGTGTCCCCGGACTGTGCATCGCTGGTAACAGGGGCATTGCACC  
AAAATCCGGCTACTTCGTCAATGTCAACAACACATGGATGTATACTGGGAGTGGTTATTATTACC  
CTGAACCTATAACAGAGAACAATGTAGTAGTCATGTCCACATGCGCCGTCAATTATACTAAGGCC  
CCCTATGTTATGCTCAACACTTCAATTCCTCAATCTCCCGATTTCAAAGAAGAGCTGGATCAGTG

5 GTTTAAGAATCAGACATCCGTGGCCCCCTGACTTAAGCTTGGATTATATCAATGTGACTTTTTTTAG  
ACTTACAGGTGAGATGAACCGACTCCAGGAAGCTATAAAAGTACTGAACCACTCCTATATCAAT  
CTGAAAGATATCGGTACATACGAATATTACGTAAAATGGCCTTGGTATGTGTGGCTACTAATTTG  
CCTTGCGGGCGTGGCTATGCTGGTCCTGCTGTTCTTCATTtctttatggatgtgctccaatggat  
cgttacaatgcagaatttgcatTTGA

## PDI-OC43-H5iCT(V4)-DNA (SEQ ID NO: 140)

10 atggcgaaaaacggttgcgattttctggccttattgttttctcttcttctgtgttggttccttctcagat  
cttcggtGTGATCGGCGATCTGAATTGTACCCTGGATCCCCGCCTGAAAGGGAGCTTTAAACAACC  
GAGATACAGGACCCCCGTCTATATCCATAGATACAGTGGATGTTACGAACGGGCTCGGCACCTAC  
TATGTGCTAGACCGAGTTTATTTGAACACCACCTTATTCCTCAATGGATACTACCCAACCTCAGG  
TAGTACTTACAGAAACATGGCGCTGAAGGGTACGGATCTGCTGAGCACCTATGGTTTAAACCTC  
CCTTCTCTCGGACTTTATTAATGGCATCTTCGCTAAGGTGAAAAACACGAAGGTTTTCAAAGAT  
15 GGAGTGATGTATTACAGATTCCCTGCGATCACCATTGGAAGTACCTTCGTGAATACTTCTATAG  
CGTGGTGGTTCAACCACGGACAATCAACTCCACCCAGGACGGCGTCAACAAGCTCCAGGGATTGC  
TGGAGGTGTCAGTCTGTCAATATAACATGTGTGAGTACCCACACACTATCTGTCAACCTAATCTA  
GGCAACCACTTTAAGGAACGTGTGGCACTACGATACGGGGGTGGTAAGTTGCTTATATAAGAGAAA  
TTTCACCTATGATGTTAATGCAACGTACCTGTACTTTCACTTCTATCAAGAAGGAGGAACCTTCT  
ACGCATATTTTACAGATACCGGCTTTGTGACGAAATTCTTATTCACCGTTTACCTCGGAATGGCA  
20 TTAAGCCATTATTACGTGATGCCTCTCACTTGCATCAGACGCCCTAAGGATGGTTTTTCTCTGGA  
GTACTGGGTCACTCCCCTGACACCACGGCAGTACCTGCTTGCTTTTAAACCAGGACGGTATCATTT  
TTAATGCCGTGATTGTATGAGCGATTTTATGAGCGAGATAAAGTGCAAGACCCAATCTATTGCT  
CCGCCACGGGGGTGTACGAACCTGAATGGTTACACCGTCCAGCCCGTTGCCGATGTATATAGACG  
GAAACCAGACCTGCCCAATTGCAACATCGAAGCTTGGTTAAACGATAAGTCAGTGCCCTCCCCC  
25 TCAATTGGGAGAGGAAGACTTTCTCCAACCTGTAATTTCAACATGTCAAGCCTGATGTCTTTCATT  
CAAGCCGATTTCGTTCACTTGTAAATAATATAGATGCAGCAAAGATCTATGGTATGTGCTTCAGTTC  
CATCACAATAGATAAGTTTGAATACCAAACCGTCGCAAGGTGGACCTTCAGCTCGGCAACCTGG  
GCTATCTGCAGTCCAGCAATTATAGAATAGACACCACCGCCACATCATGTGAGTGTACTATAAC  
CTCCCAGCAGCGAACGTGAGTGTAGTGTCAATCCTTCTACCTGGAATAAAGGTTTGGATT  
30 CATCGAAGATAGTGTGTTTCGTACCTCAGCCAACAGGAGTGTTCACCAATCAGACGCTGGTCTACG  
CCCAACATTGCTTCAAGGCACCCAAAAATTTCTGCCCATGTAGCAGTTGCTCCTGCCCGGGTAAG  
AACATGGGATCGGCACCTGCCAGCAGGCACCAATTCACCTACATGCGATAATCTGTGTACACT  
GGATCCTATTACACTTAAGGCCCTGTACCTACAAATGCCCCAGAGCAAGAGCCTGGTCGGTA  
TCGGAGAACACTGTTCCGGACTTGCAGTAAAAGCGACTATTGTGGAAATAACTCTTGCACTTGT  
35 CAGCCACAAGCCTTCTCGGTTGGTCCGCTGACTCTTGTTTACAAGGGGATAAGTGTAACATCTT  
CGCAATTTTCATCTTACACGATGTGAATAACGGCTTAACATGCAGCACAGATCTCCAGAAGGCAA  
ACACAGAGATCGAATTAGGAGTCTGCGTTAATTACGATCTCTACGGGATCTCTGGCCAGGGCATC  
TTCGTGGAGGTTAATGCTACCTACTACAATAGTTGGCAAAATCTGCTCTACGATAGCAATGGCAA  
40 CCTCTATGGATTACAGAGACTATATTACTAACAGGACGTTTCATGATTCACTCGTGCTATTCCGGGC  
GGGTGTCAGCAGCTTATCAGCAAATTCCTCAGAGCCAGCTCTGCTATTCCGAAACATAAAATGT  
AATTACGTGTTCAATAATTCAGTACTCGGCAGCTGCAGCCGATTAATTACAGCTTCGACAGCTA  
CCTTGGTTGCGTTGTAAACGCCTACAACCTCCACTGCCATATCAGTTCAGACCTGCGACCTTACTG  
TGGGCTCTGGCTATTGTGTCGATTATTCAAAGAACGGGGGGAGCGGGTCCGCAATAACAACCTGGC  
TATAGGTTACCAATTTTGAAGCCTTTACCGTGAATAGTGTCAACGATAGCCTGGAGCCTGTCCG  
45 AGGTCTTTATGAGATACAAATCCCCTCCGAGTTCACAATTGGCAACATGGAAGAGTTCATCCAGA  
CGAGTTCCCCAAAGGTGACGATCGATTGCGCGGCTTTTCGTCTGCGGCGACTACGCCGATGCAAG  
TTACAACTCGTTGAGTATGGAAGTTTTTTCGATAATATAAACGCAATTCTGACTGAAGTGAACGA  
ACTGCTGGACACCACTCAGTTGCAGGTGGCAAAATTCGCTCATGAACGGCGTGACACTGTCAACCA  
AAGTGAAGGACGGTGTCAATTTCAATGTGGATGACATTAACCTCAGCCCCGTAAGTGGCTGTTTG  
50 GGTAGTGAGTGTCTAAGGCTAGCAGCCGCTCCGCCATTGAGGACTTGTGTTTGATAAAGTTAA

5 GCTGAGTGACGTTGGATTTGTTGAGGCGTATAATAACTGTACCGGTGGTGCAGAGATAAGGGATC  
TGATCTGTGTCCAGAGTTATAAGGGGATTAAGGTTCTCCCCCGCTACTCTCGGAGAATCAGATA  
TCAGGATACACCCCTGGCCGCTACCTCAGCCTCGCTGTTTTCCCCCTTGGACCGCTGCCGCCGGTGT  
10 CCCATTTTATTTGAATGTGCAGTATCGGATCAACGGTCTGGGAGTGACAATGGACGTGCTGTCTC  
AGAACCAGAAACTGATCGCCAATGCATTCAACAATGCTCTGCACGCCATCCAGCAAGGGTTTTGAC  
GCTACAAATTTCTGCCCTCGTAAAAATCCAGGCCGTGGTGAATGCTAACGCCGAAGCCCTTAATAA  
TCTGCTCCAGCAGCTTTCTAACCGCTTTGGAGCTATTTCTGCCCTACTGCAGGAAATTTCTATCCA  
GACTGGATCCCCCTGAGGCAGAAGCCCAAATCGACCGTCTCATAAACGGCAGACTCACTGCTCTT  
15 AACGCTACGTTAGTCAACAATTGAGCGATTTCGACCTTGGTGAAATTCAGCGCAGCTCAGGCTAT  
GGAGAAGGTGAACGAGTGCCTGAAGTCACAGAGCTCCAGAATCAATTTCTGTGGCAATGGGAACC  
ATATCATCTCCTTGGTTCAGAATGCTCCCTACGGCCTGTATTTTCATCCACTTCAACTACGTGCC  
ACGAAGTACGTTACAGCCAAAGTGTCCCCGGACTGTGCATCGCTGGTAACAGGGGCATTGCACC  
AAAATCCGGCTACTTCGTCAATGTCAACAACACATGGATGTATACTGGGAGTGGTTATTATTACC  
CTGAACCTATAACAGAGAACAATGTAGTAGTCATGTCCACATGCGCCGTCAATTATACTAAGGCC  
20 CCTATGTTATGCTCAACACTTCAATTCCCAATCTCCCGGATTTCAAAGAAGAGCTGGATCAGTG  
GTTTAAGAATCAGACATCCGTGGCCCCCTGACTTAAGCTTGGATTATATCAATGTGACTTTTTTTAG  
ACTTACAGGTGCAGATGAACCGACTCCAGGAAGCTATAAAAGTACTGAACCACTCCTATATCAAT  
CTGAAAGATATCGGTACATACGAATATTACGTAAAATGGCCTTGGTATGTGTGGCTACTAATTTG  
CCTTGCGGGCGTGGCTATGCTGGTCTGCTGTTCTTCATTTgctgctccaatggatcggttacaat  
25 cgagaatttgcattTGA

**PDI-OC43-H1cCT-DNA (SEQ ID NO: 141)**

25 atggcgaaaaacggttgcgattttctggccttattgttttctcttcttctgtgttggttccttctcagat  
cttcgcgGTGATCGGCGATCTGAATTGTACCCTGGATCCCCGCCTGAAAGGGAGCTTTAACAACC  
GAGATACAGGACCCCGTCTATATCCATAGATACAGTGGATGTTACGAACGGGCTCGGCACCTAC  
TATGTGCTAGACCGAGTTTATTTGAACACCACCTTATTCCTCAATGGATACTACCCAACCTCAGG  
TAGTACTTACAGAAACATGGCGCTGAAGGGTACGGATCTGCTGAGCACCCCTATGGTTTTAAACCTC  
CCTTCCTCTCGGACTTTATTAATGGCATCTTCGCTAAGGTGAAAAACAGGAAGGTTTTCAAAGAT  
30 GGAGTGATGTATTCAGAGTTCCCTGCGATCACCATTGGAAGTACCTTCGTGAATACTTCCCTATAG  
CGTGGTGGTTCAACCACGGACAATCAACTCCACCCAGGACGGCGTCAACAAGCTCCAGGGATTGC  
TGGAGGTGTCAGTCTGTCAATATAACATGTGTGAGTACCCACACACTATCTGTACCCCTAATCTA  
GGCAACCACTTTAAGGAACTGTGGCACTACGATACGGGGGTGGTAAGTTGCTTATATAAGAGAAA  
TTTCACCTATGATGTTAATGCAACGTACCTGTACTTTCACTTCTATCAAGAAGGAGGAACCTTCT  
35 ACGCATATTTACAGATAACCGGCTTTGTGACGAAATTCCTTATTCAACGTTTACCTCGGAATGGCA  
TTAAGCCATTATTACGTGATGCCTCTCACTTGCATCAGACGCCCTAAGGATGGTTTTTCTCTGGA  
GTACTGGGTCACTCCCTGACACCACGGCAGTACCTGCTTGCTTTAACCAGGACGGTATCATTT  
TTAATGCCGTGCGATTGTATGAGCGATTTTATGAGCGAGATAAAGTGCAAGACCCAATCTATTGCT  
CCGCCCACGGGGGTGTACGAACCTGAATGGTTACACCGTCCAGCCCGTTGCCGATGTATATAGACG  
40 GAAACCAGACCTGCCCAATTGCAACATCGAAGCTTGGTTAAACGATAAGTCAGTGCCCTCCCCC  
TCAATTGGGAGAGGAAGACTTTCTCCAATGTAATTTCAACATGTCAAGCCTGATGTCTTTCATT  
CAAGCCGATTTCGTTCACTTGTAATAATATAGATGCAGCAAAGATCTATGGTATGTGCTTCAGTTC  
CATCACAATAGATAAGTTTGAATAACAAACCGTCGCAAGGTGGACCTTCAGCTCGGCAACCTGG  
GCTATCTGCAGTCCAGCAATTATAGAATAGACACCACCGCCACATCATGTGCTGCTGTACTATAAC  
CTCCCAGCAGCGAACGTGAGTGTAGTAGGTTCAATCCTTCTACCTGGAATAAAAGGTTTGGATT  
45 CATCGAAGATAGTGTGTTTCGTACCTCAGCCAACAGGAGTGTTCACCAATCACAGCGTGGTCTACG  
CCCAACATTGCTTCAAGGCACCCAAAAATTTCTGCCCATGTAGCAGTTGCTCCTGCCCGGGTAAG  
AACAATGGGATCGGCACCTGCCCAGCAGGCACCAATTCACCTACATGCCGATAATCTGTGTACACT  
GGATCCTATTACACTTAAGGCCCTGATACCTACAAATGCCCCAGAGCAAGAGCCTGGTTCGGTA  
TCGGAGAACACTGTTCCGGACTTGCAGTAAAAAGCGACTATTTGTGGAAATAACTCTTGCACTTGT  
50 CAGCCACAAGCCTTCCTCGGTTGGTCCGCTGACTCTTGTTTACAAGGGGATAAGTGTAAACATCTT

CGCAAATTTTCATCTTACACGATGTGAATAACGGCTTAACATGCAGCACAGATCTCCAGAAGGC  
 ACACAGAGATCGAATTAGGAGTCTGCGTTAATTACGATCTCTACGGGATCTCTGGCCAGGGCATC  
 TTCGTGGAGGTTAATGCTACCTACTACAATAGTTGGCAAAATCTGCTCTACGATAGCAATGGCAA  
 CCTCTATGGATTACAGAGACTATATTACTAACAGGACGTTTCATGATTCACTCGTGCTATTCCGGGC  
 5 GGGTGTGACGAGCTTATCACGCAAATTCCTTCAGAGCCAGCTCTGCTATTCCGAAACATAAAATGT  
 AATTACGTGTTCAATAATTCAGTACTCGGCAGCTGCAGCCGATTAATTACAGCTTCGACAGCTA  
 CCTTGGTTGCGTTGTTAACGCCTACAACCTCCACTGCCATATCAGTTCAGACCTGCGACCTTACTG  
 TGGGCTCTGGCTATTGTGTCGATTATTCAAAGAACGGGGGAGCGGGTCCGCAATAACAACCTGGC  
 10 TATAGGTTACCAATTTTGAGCCTTTACCGTGAATAGTGTCAACGATAGCCTGGAGCCTGTCGG  
 AGGTCTTTATGAGATACAAATCCCCTCCGAGTTCACAATTGGCAACATGGAAGAGTTTCATCCAGA  
 CGAGTTCCCCAAAGGTGACGATCGATTGCGCGGCTTTTCGTCTGCGGCGACTACGCCGCATGCAAG  
 TTACAACCTCGTTGAGTATGGAAGTTTTTTCGATAATATAAACGCAATTCTGACTGAAGTGAACGA  
 ACTGCTGGACACCACTCAGTTGCAGGTGGCAAATTCGCTCATGAACGGCGTGACACTGTCAACCA  
 AACTGAAGGACGGTGTCAATTTCAATGTGGATGACATTAACCTCAGCCCCGTAAGTGGCTGTTT  
 15 GGTAGTGAGTGTTCTAAGGCTAGCAGCCGCTCCGCCATTGAGGACTTGTGTTTGATAAAGTTAA  
 GCTGAGTGACGTTGGATTTGTTGAGGCGTATAATAACTGTACCGGTGGTGACAGATTAAGGGATC  
 TGATCTGTGTCCAGAGTTATAAGGGGATTAAGGTTCTCCCCCGCTACTCTCGGAGAATCAGATA  
 TCAGGATACACCCTGGCCGCTACCTCAGCCTCGCTGTTTCCCCCTTGACCGCTGCCGCCGGTGT  
 CCCATTTTATTTGAATGTGCAAGTATCGGATCAACGGTCTGGGAGTGACAATGGACGTGCTGTCTC  
 20 AGAACCAGAACTGATCGCCAAATGCATTCAACAATGCTCTGCACGCCATCCAGCAAGGGTTTGAC  
 GCTACAAATTTCTGCCCTCGTAAAAATCCAGGCCGTGGTGAATGCTAACGCCGAAGCCCTTAATAA  
 TCTGCTCCAGCAGCTTTCTAACCGCTTTGGAGCTATTTCTGCCTCACTGCAGGAAATTTCTATCCA  
 GACTGGATCCCCCTGAGGCAGAAGCCCAAATCGACCGTCTCATAAACGGCAGACTCACTGCTCTT  
 AACGCCTACGTTAGTCAACAATTGAGCGATTTCGACCTTGGTGAAATTCAGCGCAGCTCAGGCTAT  
 25 GGAGAAGGTGAACGAGTGCGTGAAGTCACAGAGCTCCAGAATCAATTTCTGTGGCAATGGGAACC  
 ATATCATCTCCTTGGTTTCAAGATGCTCCCTACGGCCTGTATTTCACTCCACTTCAACTACGTGCC  
 ACGAAGTACGTTACAGCCAAAGTGTCCCCCGACTGTGCATCGCTGGTAACAGGGGCATTGCACC  
 AAAATCCGGCTACTTCGTCAATGTCAACAACACATGGATGTATACTGGGAGTGGTTATTATTACC  
 CTGAACCTATAACAGAGAACAATGTAGTAGTCATGTCCACATGCGCCGTCAATTATACTAAGGCC  
 30 CCCTATGTTATGCTCAACACTTCAATTTCCCAATCTCCCGGATTTCAAAGAAGAGCTGGATCAGTG  
 GTTTAAGAATCAGACATCCGTGGCCCCCTGACTTAAGCTTGGATTATATCAATGTGACTTTTTTAG  
 ACTTACAGGTTCGAGATGAACCGACTCCAGGAAGCTATAAAAAGTACTGAACCACTCCTATATCAAT  
 CTGAAAAGATATCGGTACATACGAATATTACGTAAAATGGCCTTGGTATGTGTGGCTACTAATTTG  
 35 CCTTGCGGGCGTGGCTATGCTGGTCCTGCTGTTCTTCATTagcttctggatgtgctcctaatgggt  
 ctctacagtgtagaatatgtattTGA

# PDI-OC43-wtTMCT-AA (SEQ ID NO: 142)

MAKNVAFGLLFSLVLVPSQIFAVIGDLNCTLDPRLKGSFNNRDTGPPSISIDTVDVTNGLGTY  
 YVLDREVLYNTTLFLNGYYPTSGSTYRNMALKGTDLLSTLWFKPPFLSDFINGIFAKVKNTKVFKD  
 40 GVMYSEFFAITIGSTFVNTSYSVVVQPTINSTQDGVNKLQGLLEVSVQYNMCEYPHTICHPNL  
 GNHFKEWLHYDTGVVSCLYKRNFYDYNATYLYFHFYQEGGTFYAYFTDTGFVTKFLENVYLGMA  
 LSHYYVMPLTCIRRPKDGFSLEYWVTPLTPRQYLLAFNQDGIIFNAVDCMSDFMSEIKCKTQSI  
 PPTGVYELNGYTVQPVADVRRKPDLPNCNIEAWLNDKSVPSPLNWERKTFSNCFNMSSLSMFI  
 45 QADSFTCNNIDAAKIYGMCFSSITIDKFAPNRRKVDLQGLNLGYLQSSNYRIDTTATSCQLYYN  
 LPAANVSVSRENPSTWKNRFGFIEDSVFVPQPTGVFTNHSVVYAQHCFKAPKNFCPCSSCSCPGK  
 NNGIGTCPAGTNSLTCDNLCTLDPITLKAPDITYKCPQSKSLVGIGEHCSGLAVKSDYCGNNSCTC  
 QPQAFLGWSADSCLQGDKNIFANFILHDVNNGLTCTSDLQKANTEIELGVCVNYDLYGISGQGI  
 FVEVNATYYNSWQNLLYDSNGNLYGFRDYITNRTFMIHSCYSGRVSAAYHANSSEPALLFRNIKC  
 50 NYVFNNLSLTRLQLPINYSFDSYLGCVVNAYNSTAISVQTCDLTVGSGYCVDYSKNGGSGSAITTG  
 YRFTNFEPFTVNSVNDLSLEPVGGLYEIQIPSEFTIGNMEEFIQTSSPKVTIDCAAFVCGDYAACK

5 LQLVEYGSFCDNINAILTEVNELLDTTQLQVANSLMNGVTLSTKLKDGVNFNVDINFSFVLGCL  
GSECSKASSRSAIEDLLFDKVKLSDVGFVEAYNNCTGGAEIRDLCVQSYKGIKVLPLLSNQI  
SGYTLAATSASLFPPWTAAAGVPFYLVNQYRINGLGVTMDVLSQNQKLIANAFNNALHAIQQGFD  
ATNSALVKIQAVVNANAEALNNLLQQLSNRFGAISASLQEILSRLDPPAEAEQIDRLINGRLTAL  
NAYVSQQLSDSTLVKFSAAQAMEKVNCEVKSQSSRINFCGNGNHIISLVQNAPYGLYFIHFNYVP  
TKYVTAKVSPGLCIAGNRGIAPKSGYFVNVNNTWMTGSGYYYPEPITENNVVVMSTCAVNYTKA  
PYVMLNTSIPNLPDFKEELDQWFKNQTSVAPDLSLDYINVTFLDLQVEMNRLQEAIKVLNHSYIN  
LKDIGTYEYYVKWPWYVWLLICLAGVAMLVLLFFICCCTGCGTSCFKKCGGCCDDYTGYQELVIK  
TSHDD

10 **PDI-OC43-H5iTMCT-AA (SEQ ID NO: 143)**

MAKNVAIFGLLFSLLVLVPSQIFAVIGDLNCTLDPRLKGSFNNRDTGPPSISIDTVDVTNGLGTY  
YVLDRVYLNNTLFLNGYYPTSGSTYRNMALKGTDLLSTLWFKPPFLSDFINGIFAKVKNTKVFKD  
GVMYSEFPAITIGSTFVNNTSYSVVVQPTINSTQDGVNKLQGLLEVSVQCYNMCEYPHTICHPNL  
15 GNHFKEWLHYDTGVVSCLYKRNFTYDVNATYLYFHFYQEGGTFYAYFTDTGFVTKFLFNVYLGMA  
LSHYVMPLTLCIRRPKDGFSLEYWVTPLTPRQYLLAFNQDGIIFNAVDCMSDFMSEIKCKTQSI  
PPTGVYELNGYTVQPVADVRRKPDLPNCNIEAWLNDKSVSPNLNWERKTFSNCFNMSSSLMSFI  
QADSFTCNNIDAAKIYGMCFSSITIDKFAIPNRRKVDLQGLNLGYLQSSNYRIDTTATSCQLYYN  
LPAANVSVSRENPSSTWNKRFGFIEDSVFVPQPTGVFTNHSVVYAQHCFAKPNFCPCSSCSCPGK  
20 NNGIGTCPAGTNSLTCDNLCTLDPITLKAPDITYKCPQSKSLVGIGEHCSGLAVKSDYCGNNSCTC  
QPQAFGLWSADSCLQGDKNIFANFILHDVNGLTCTDLDQKANTEIELGVCVNYDLYGISGQGI  
FVEVNATYNSWQNLLYDSNGNLYGFRDYITNRTFMIHSCYSGRVSAAYHANSSEPALLFRNIKC  
NYVFNNSLTRQLQPINYSFDSYLGCVVNAYNSTAISVQTCDLTVGSGYCVDSKNGGSGSAITTG  
YRFTNFEPFTVNSVNDLSLEPVGGLYEIQIPSEFTIGNMEEFIQTSSPKVTIDCAAFVCGDYAACK  
25 LQLVEYGSFCDNINAILTEVNELLDTTQLQVANSLMNGVTLSTKLKDGVNFNVDINFSFVLGCL  
GSECSKASSRSAIEDLLFDKVKLSDVGFVEAYNNCTGGAEIRDLCVQSYKGIKVLPLLSNQI  
SGYTLAATSASLFPPWTAAAGVPFYLVNQYRINGLGVTMDVLSQNQKLIANAFNNALHAIQQGFD  
ATNSALVKIQAVVNANAEALNNLLQQLSNRFGAISASLQEILSRLDPPAEAEQIDRLINGRLTAL  
NAYVSQQLSDSTLVKFSAAQAMEKVNCEVKSQSSRINFCGNGNHIISLVQNAPYGLYFIHFNYVP  
TKYVTAKVSPGLCIAGNRGIAPKSGYFVNVNNTWMTGSGYYYPEPITENNVVVMSTCAVNYTKA  
30 PYVMLNTSIPNLPDFKEELDQWFKNQTSVAPDLSLDYINVTFLDLQVEMNRLQEAIKVLNHSYIN  
LKDIGTYEYYVKWPWYQILSIYSTVASSLALAIMMAGLSLWMCNNGSLQCRICI

**PDI-OC43-H5iCT-AA (SEQ ID NO: 144)**

MAKNVAIFGLLFSLLVLVPSQIFAVIGDLNCTLDPRLKGSFNNRDTGPPSISIDTVDVTNGLGTY  
35 YVLDRVYLNNTLFLNGYYPTSGSTYRNMALKGTDLLSTLWFKPPFLSDFINGIFAKVKNTKVFKD  
GVMYSEFPAITIGSTFVNNTSYSVVVQPTINSTQDGVNKLQGLLEVSVQCYNMCEYPHTICHPNL  
GNHFKEWLHYDTGVVSCLYKRNFTYDVNATYLYFHFYQEGGTFYAYFTDTGFVTKFLFNVYLGMA  
LSHYVMPLTLCIRRPKDGFSLEYWVTPLTPRQYLLAFNQDGIIFNAVDCMSDFMSEIKCKTQSI  
40 PPTGVYELNGYTVQPVADVRRKPDLPNCNIEAWLNDKSVSPNLNWERKTFSNCFNMSSSLMSFI  
QADSFTCNNIDAAKIYGMCFSSITIDKFAIPNRRKVDLQGLNLGYLQSSNYRIDTTATSCQLYYN  
LPAANVSVSRENPSSTWNKRFGFIEDSVFVPQPTGVFTNHSVVYAQHCFAKPNFCPCSSCSCPGK  
NNGIGTCPAGTNSLTCDNLCTLDPITLKAPDITYKCPQSKSLVGIGEHCSGLAVKSDYCGNNSCTC  
QPQAFGLWSADSCLQGDKNIFANFILHDVNGLTCTDLDQKANTEIELGVCVNYDLYGISGQGI  
FVEVNATYNSWQNLLYDSNGNLYGFRDYITNRTFMIHSCYSGRVSAAYHANSSEPALLFRNIKC  
45 NYVFNNSLTRQLQPINYSFDSYLGCVVNAYNSTAISVQTCDLTVGSGYCVDSKNGGSGSAITTG  
YRFTNFEPFTVNSVNDLSLEPVGGLYEIQIPSEFTIGNMEEFIQTSSPKVTIDCAAFVCGDYAACK  
LQLVEYGSFCDNINAILTEVNELLDTTQLQVANSLMNGVTLSTKLKDGVNFNVDINFSFVLGCL  
GSECSKASSRSAIEDLLFDKVKLSDVGFVEAYNNCTGGAEIRDLCVQSYKGIKVLPLLSNQI

5 SGYTLAATSASLFPPWTAAGVPFYLVNQYRINGLVGTMVLSQNQKLIANAFNNALHAIQQGFD  
 ATNSALVKIQAVVNANAEALNNLLQQLSNRFGAISASLQEILSRDPPEAEAQIDRLINGRLTAL  
 NAYVSQQLSDSTLVKFSAAQAMEKVNCEVKSQSSRINFCGNGNHIISLVQNAPYGLYFIHFNYVP  
 TKYVTAKVSPGLCIAGNRGIAPKSGYFVNVNNTWMTGSGYYPPEPITENNVVMSTCAVNYTKA  
 PYVMLNTSIPNLPDFKEELDQWFKNQTSVAPDLSLDYINVTFLDLQVEMNRLQEAIKVLNHSYIN  
 LKDIGTYEYYVKWPWYVWLLICLAGVAMLVLLFFISLWMCNGLQCRICI

#### PDI-OC43-H5iCT(V4)-AA (SEQ ID NO: 145)

10 MAKNVAIFGLLFSLVLVPSQIFAVIGDLNCTLDPRKGSFNNRDTGPPSISIDTVDVNTGLGTY  
 YVLDRVYLNNTLFLNGYYPTSGSTYRNMALKGTDLSTLWFKPPFLSDFINGIFAKVKNTKVFKD  
 GVMYSEFFAITIGSTFVNTSYSVVVQPTINSTQDGVNKLQGLLEVSVCQYNMCEYPHTICHPNL  
 GNHFKEWLHYDTGVVSCLYKRNFTYDVNATYLYFHFYQEGGTFYAYFTDTGFVTKFLFNVYLGMA  
 LSHYYVMPLTCIRRPKDGFSLEYWVTPLTPRQYLLAFNQDGIIFNAVDCMSDFMSEIKCKTQSIA  
 15 PPTGVYELNGYTVQPVADVRRKPDLPNCNIEAWLNDKSVSPNLWERKTFSNCFNMSSLSMFI  
 QADSFTCNNIDAAKIYGMCFSSITIDKFAIPNRRKVDLQGLNLGYLQSSNYRIDTTATSCQLYYN  
 LPAANVSVSRENFPSTWNRKFGFIEDSVFVPQPTGVFTNHSVVYAQHCFAKPNFCPCSSCSCPGK  
 NNGIGTCPAGTNSLTCDNLCTLDPITLKAPDXYKCPQSKSLVGIGEHCSGLAVKSDYCGNNSCTC  
 20 QPQAFGLWSADSCLQGDKNIFANFILHDVNNGLTCTSDLQKANTEIELGVCVNYDLYGISGQGI  
 FVEVNATYINSWQNLLYDSNGNLYGFRDYITNRTFMIHSCYSGRVSAAYHANSSEFALLFRNIKC  
 NYVFNNSLTRQLQPINYSFDSYLGCVVNAYNSTAISVQTCDLTVGSGYCVDSKNGGSGSAITTG  
 YRFTNFEPFTVNSVNDLSLEPVGGLEYEQIPSEFTIGNMEEFIQTSSPKVTIDCAAFVCGDYAACK  
 LQLEVEYGSFCDNINAILTEVNELLDTTQLQVANSLMNGVTLSTKLKDGVNFNVDINFSPLGCL  
 GSECSKASSRSAIEDLLFDKVKLSDVGFVEAYNNCTGGAEIRDLCVQSYKGIKVLPLLSNQI  
 25 SGYTLAATSASLFPPWTAAGVPFYLVNQYRINGLVGTMVLSQNQKLIANAFNNALHAIQQGFD  
 ATNSALVKIQAVVNANAEALNNLLQQLSNRFGAISASLQEILSRDPPEAEAQIDRLINGRLTAL  
 NAYVSQQLSDSTLVKFSAAQAMEKVNCEVKSQSSRINFCGNGNHIISLVQNAPYGLYFIHFNYVP  
 TKYVTAKVSPGLCIAGNRGIAPKSGYFVNVNNTWMTGSGYYPPEPITENNVVMSTCAVNYTKA  
 PYVMLNTSIPNLPDFKEELDQWFKNQTSVAPDLSLDYINVTFLDLQVEMNRLQEAIKVLNHSYIN  
 30 LKDIGTYEYYVKWPWYVWLLICLAGVAMLVLLFFICCSNGLQCRICI

#### PDI-OC43-H1cCT-AA (SEQ ID NO: 146)

35 MAKNVAIFGLLFSLVLVPSQIFAVIGDLNCTLDPRKGSFNNRDTGPPSISIDTVDVNTGLGTY  
 YVLDRVYLNNTLFLNGYYPTSGSTYRNMALKGTDLSTLWFKPPFLSDFINGIFAKVKNTKVFKD  
 GVMYSEFFAITIGSTFVNTSYSVVVQPTINSTQDGVNKLQGLLEVSVCQYNMCEYPHTICHPNL  
 GNHFKEWLHYDTGVVSCLYKRNFTYDVNATYLYFHFYQEGGTFYAYFTDTGFVTKFLFNVYLGMA  
 LSHYYVMPLTCIRRPKDGFSLEYWVTPLTPRQYLLAFNQDGIIFNAVDCMSDFMSEIKCKTQSIA  
 40 PPTGVYELNGYTVQPVADVRRKPDLPNCNIEAWLNDKSVSPNLWERKTFSNCFNMSSLSMFI  
 QADSFTCNNIDAAKIYGMCFSSITIDKFAIPNRRKVDLQGLNLGYLQSSNYRIDTTATSCQLYYN  
 LPAANVSVSRENFPSTWNRKFGFIEDSVFVPQPTGVFTNHSVVYAQHCFAKPNFCPCSSCSCPGK  
 NNGIGTCPAGTNSLTCDNLCTLDPITLKAPDXYKCPQSKSLVGIGEHCSGLAVKSDYCGNNSCTC  
 QPQAFGLWSADSCLQGDKNIFANFILHDVNNGLTCTSDLQKANTEIELGVCVNYDLYGISGQGI  
 FVEVNATYINSWQNLLYDSNGNLYGFRDYITNRTFMIHSCYSGRVSAAYHANSSEFALLFRNIKC  
 NYVFNNSLTRQLQPINYSFDSYLGCVVNAYNSTAISVQTCDLTVGSGYCVDSKNGGSGSAITTG  
 YRFTNFEPFTVNSVNDLSLEPVGGLEYEQIPSEFTIGNMEEFIQTSSPKVTIDCAAFVCGDYAACK  
 45 LQLEVEYGSFCDNINAILTEVNELLDTTQLQVANSLMNGVTLSTKLKDGVNFNVDINFSPLGCL  
 GSECSKASSRSAIEDLLFDKVKLSDVGFVEAYNNCTGGAEIRDLCVQSYKGIKVLPLLSNQI  
 SGYTLAATSASLFPPWTAAGVPFYLVNQYRINGLVGTMVLSQNQKLIANAFNNALHAIQQGFD  
 ATNSALVKIQAVVNANAEALNNLLQQLSNRFGAISASLQEILSRDPPEAEAQIDRLINGRLTAL

NAYVSQQLSDSTLVKFSAQAAMEKVNCEVKSQSSRINFCGNGNHIISLVQNAPYGLYFIHFNYVP  
TKYVTAKVSPGLCIAGNRGIAPKSGYFVNVNNTWMTGSGYYYPEPITENNVVMSTCAVNYTKA  
PYVMLNTSIPNLPDFKEELDQWFKNQTSVAPDLSLDYINVTFLDLQVEMNRLQEAIKVLNHSYIN  
LKDIGTYEYYVKWPWYVWLLICLAGVAMLVLLFFISFWMCSNGSLQCRICI

5

## IF(CoV229EwtCT).r (SEQ ID NO: 147)

ACGACACGACTAAGGCCTTCACTGTATGTGGATCTTTTCGACATCGTA

## PDI-229E -wtTMCT-DNA (SEQ ID NO: 148)

10 atggcgaaaaacgttgcgatttttcggcttattgttttctcttcttgtgttggttccttctcagat  
cttcgcgCAAACGACTAATGGGCTGAACACCAGTTACAGCGTCTGTAACGGCTGCGTCGGATATA  
GCGAGAACGTGTTTCGCAGTGGAAAGTGGGGGTACATTCCCTTCCGACTTCGCTTTCAATAACTGG  
TTTCTCCTGACTAACACAAGCTCCGTCGTGGATGGCGTGGTCAGGTCCCTTCAGCCTCTTCTCCT  
15 GAATTGCCTGTGGTCTGTGTCCGGGTTAAGATTCACTACAGGCTTCGTATACTTCAACGGGACGG  
GCCGGGGGATTGCAAGGGCTTCTCCTCCGACGTGCTGTGATGTCAGATGTCGTTACAATCTGAAC  
TTCGAAGAGAACTTACGGCGGGGGACAATCCTGTTCAAACATCATATGGCGTAGTCGTATTTTA  
CTGCACCAATAATACCCTGGTGAGTGGGGACGCCCATATTCCCTTCGGAACAGTGCTGGGTAACT  
TTTACTGTTTTGTCAACACTACGATCGGAAACGAAACCACTAGCGCCTTTGTGCGAGCTCTGCCA  
20 AAAACAGTTAGGGAGTTCGTGATCTCTCGGACCGGTCACCTCTATATCAACGGCTACCGTTATTT  
TACTTTGGGCAACGTCGAAGCCGTCAATTTTAAATGTGACAACCTGCAGAGACAACCTGACTTTTGCA  
CTGTGGCTCTCGCCAGTTATGCCGATGTGCTGGTGAATGTAAGTCAAACGTCAATTGCCAACATC  
ATCTATGTAACTCAGTAATCAACCGGCTCCGCTGTGACCAACTCTCATTCGACGTCCCCGACGG  
ATTCTATTCACGAGCCCGATTGAGAGCGTGAACCTGCCAGTTTCCATCGTATCCCTCCCAGTTT  
25 ACCACAAGCACACTTTTATCGTTCTCTACGTAGATTTTAAACCCAGTCAGGAGGAGGGAAATGC  
TTCAACTGCTACCCGGCTGGCGTGAACATCACCTTGGCCAATTTTAAATGAACTAAAGGGCCCCCT  
TTGCGTGGATACGTACACTTTACCACAAAGTATGTTGCAGTCTATGCTAACGTCGGCAGGTGGT  
CAGCGTCCATTAACACAGGCAATTGCCGTTCTCTTTTCGGGAAAGTGAAACAACCTTCGTGAAGTTT  
GGAAGTGTGTGCTTCAGTTTGAAAGACATTCCGGGCGGCTGCGCCATGCCATTGTGGCTAATTG  
GGCTTATTCCAAGTACTACACCATTTGGCTCTCTCTACGTTAGCTGGAGCGACGGTGACGGTATAA  
30 CGGGCGTACCACAACCGGTGGAAGGGGTGAGCTCTTTCATGAATGTCACTCTGGACAAGTGTACC  
AAATATAATATATACGATGTGAGTGGAGTGGGCGTTATACGCGTGTCTAACGACACCTTTCTAAA  
CGGCATAACCTACACAAGCACGTGAGGCAATCTGTTAGGTTTTTAAAGACGTCACTAAAGGCACTA  
TATATAGCATCACCCCATGCAACCCACCTGATCAATTAGTCGTATATCAGCAAGCTGTTGTGGGT  
GCTATGCTGTGCAAAAACTTACACAGCTACGGGTTCTCCAATGTGGTGGAACCTGCCAAATTTCTT  
35 TTACGCTAGCAATGGCACATATAACTGTACTGACGCGCTCTTGACTTACAGTTTATTCCGAGTGT  
GCGCGGACGGCAGCATTATCGCCGTGCAGCCGGCCAATGTCAGCTATGATTCCGTTTCCGCCATC  
GTGACAGCCAACCTTGTGATTCCCTCTAACTGGACAACGTCTGTCCAAGTCAATATCTGCAGAT  
CACCTCAACCCCATAGTAGTCGATTGCTCAACCTACGTCTGCAACGGTAATGTCAGATGTGTGTCG  
AGCTGCTCAAGCAGTACACCTCCGCTGTAAGACTATTGAGGATGCATTAAGAAAATAGTGCAAGA  
40 TTGGAAGCGCCGATGTGTGCGAAATGCTAACCTTCGATAAGAAGGCATTACACTGGCGAACGT  
AAGCTCTTTTCGGCGATTACAACCTGTCTTCGGTAATCCCTAGCTTGCCACATCCGGCTCTCGGG  
TGCGGGGGCGGAGCGCTATCGAGGACATTTTATTCTCGAAACTGGTTACATCTGGGCTCGGAACT  
GTGGACGCCGATTACAAGAAGTGCACCAAGGGCCTAAGCATCGCCGACCTCGCTGTGCTCAGTA  
CTACAACCGGAATTATGGTGCTGCCAGGTGTGCTGACGCAGAGCGGATGGCTATGTATACCGGCA  
45 GTCTCATTTGGCGGGATTGCGTTGGGCGGCCTGACGTCCGCTGTCTCCATCCCTTTCTCTGGCT  
ATACAAGCCCGACTGAATTATGTGGCCCTGCAGACTGATGTCTGCAAGAAAATCAGAAGATTCT  
TGCCGCCAGCTTCAACAAGGCCATGACTAATATTGTGGATGCGTTTACCGGAGTGAATGACGCCA  
TACCCCAAACGTCCCAAGCCCTGCAGACAGTCGCCACGGCGTTAAACAAAATCCAGGATGTAGTG

AATCAGCAAGGGAACAGCTTGAATCACCTGACGTCCCAGTTAAGACAGAACTTTCAGGCAATCAG  
 TAGCTCAATCCAGGCTATCTACGATCGATTAGATCCTCCTCAGGCAGATCAGCAGGTGGATCGGC  
 TCATCACCGGCCGCTCGCGGCATTGAATGTTTTTCGTAAGTCATACCTTGACCAAGTACACGGAG  
 GTGAGGGCCAGTCGCCAGCTGGCTCAGCAAAAAGTGAATGAGTGTGTGAAATCACAGAGCAAACG  
 5 GTACGGGTTTTGTGGAAATGGGACGCACATCTTTAGCATCGTTAATGCTGCCCCGAAGGGTTAG  
 TCTTCCTGCACACTGTGCTCCTTCCCTACCCAGTATAAAGATGTCGAAGCATGGTCTGGGCTCTGT  
 GTCGATGGAACAAACGGTTATGTCTTCGACAGCCAAACCTCGCTCTCTATAAAGAAGGGAATTA  
 CTATAGGATCACCTCAAGAATCATGTTTCGAGCCCAGGATACCAACAATGGCCGATTTTGTGCAGA  
 TTGAAAATTGTAACGTGACCTTTGTGAATATCAGTCGATCCGAGCTTCAAACGATTGTTCTTGAG  
 10 TACATCGACGTGAATAAAACTCTACAAGAGCTGTCCTATAAACTGCCTAATTATACCGTGCCTGA  
 CCTTGATGTCGAGCAATACAACCAGACTATTCTGAACCTGACATCGGAAATCTCTACATTGGAGA  
 ATAAAAGCGCCGAGCTCAATTACACAGTGCAGAAGCTGCAGACCCGTGATCGACAATATTAACAGC  
 ACTCTTGTGGACTTAAAGTGGCTGAACCGTGTGGAGACTTACATCAAGTGGCCCTGGTGGGTGTG  
 GCTCTGTATTTCCGTGGTCTTATATTTGTTGTAAGTATGCTGCTCCTGTGCTGTTGCTCAACCG  
 15 GGTGCTGCGGTTTTTTCTCCTGTTTCGCCTCATCCATCCGTGGCTGTTGTGAGAGCACTAAACTG  
 CCATATTACGATGTCGAAAAGATCCACATACAGTGA

## PDI-229E -H5iTMCT-DNA (SEQ ID NO: 149)

atggcgaaaaacggttgcgattttcggcttattgttttctcttcttgtgttggttccttctcagat  
 20 cttcgcgCAAACGACTAATGGGCTGAACACCAGTTACAGCGTCTGTAACGGCTGCGTCGGATATA  
 GCGAGAACGTGTTTCGAGTGGAAAGTGGGGGTACATTCCTTCCGACTTCGCTTTCAATAACTGG  
 TTTCTCCTGACTAACACAAGCTCCGTCTGGATGGCGTGGTCAGGTCCTTTCAGCCTCTTCTCCT  
 GAATTGCCTGTGGTCTGTGTCCGGGTTAAGATTCACTACAGGCTTCGTATACTTCAACGGGACGG  
 25 GCCGGGGGATTGCAAGGGCTTCTCCTCCGACGTGCTGTCAGATGTGATCCGTTACAATCTGAAC  
 TTCGAAGAGAACTTACGGCGGGGGACAATCCTGTTCAAACATCATATGGCGTAGTCGTATTTTA  
 CTGACCAATAATACCCTGGTGAGTGGGGACGCCATATTCCTTCGGAACAGTGCTGGGTAACT  
 TTTACTGTTTTGTCAACACTACGATCGGAAACGAAACCACTAGCGCCTTTGTTCGGAGCTCTGCCA  
 AAAACAGTTAGGGAGTTTCGTGATCTCTCGGACCGGTCACTTCTATATCAACGGCTACCGTTATTT  
 TACTTTGGGCAACGTGCAAGCCGTCAATTTAATGTGACAACTGCAGAGCAACTGACTTTTGCA  
 30 TGTGGCTCTCGCCAGTTATGCCGATGTGCTGGTGAATGTAAGTCAAACGTCGAATTGCCAACATC  
 ATCTATTGTAACCTCAGTAATCAACCGGCTCCGCTGTGACCAACTCTCATTCGACGTCCCCGACGG  
 ATTCTATTCCACGAGCCCGATTACAGAGCGTGGAACTGCCAGTTTCCATCGTATCCCTCCCAGTTT  
 ACCACAAGCACACTTTTATCGTTCTCTACGTAGATTTTAAACCCCACTCAGGAGGAGGGAAATGC  
 TTCAACTGCTACCCGGCTGGCGTGAACATCACCTTGGCCAATTTTAAATGAACTAAAGGGCCCT  
 35 TTGCGTGGATACGTACACTTTACCACAAAGTATGTTGCAGTCTATGCTAACGTCCGCAGGTGGT  
 CAGCGTCCATTAACACAGGCAATTGCCCGTTCTCTTTCGGGAAAGTGAACAACCTTCGTGAAGTTT  
 GGAAGTGTGCTTCAGTTTGAAAGACATTCGCGGCGGCTGCGCCATGCCTATTGTGGCTAATTG  
 GGCTTATTCCAAGTACTACACCATTTGGCTCTCTCTACGTTAGCTGGAGCGACGGTGACGGTATAA  
 CGGGCGTACCACAACCGGTGGAAGGGGTGAGTCTTTTCATGAATGTCACTCTGGACAAGTGTACC  
 40 AAATATAATATATACGATGTGAGTGGAGTGGCGTTATACGCGTGTCTAACGACACCTTTCTAAA  
 CGGCATAACCTACACAAGCACGTACAGCAATCTGTTAGGTTTTAAAGACGTCACTAAAGGCACTA  
 TATATAGCATCACCCCATGCAACCCACCTGATCAATTAGTCGTATATCAGCAAGCTGTTGTGGGT  
 GCTATGCTGTCAGAAAACCTTACCAGCTACGGGTTCTCCAATGTGGTGGAACTGCCCAAATCTTT  
 TTACGCTAGCAATGGCACATATAACTGTACTGACGCCGTCTTGACTTACAGTTCATTTCGGAGTGT  
 45 GCGCGGACGGCAGCATTATCGCCGTGCAGCCGGCCAATGTCAGCTATGATTCCGTTTCCGCCATC  
 GTGACAGCCAACTTGTGATTCCTCTAACTGGACAACGTCTGTCCAAGTCGAATATCTGCAGAT  
 CACCTCAACCCCATAGTAGTCGATTGCTCAACCTACGTCTGCAACGGTAATGTCAGATGTGTGCG  
 AGCTGCTCAAGCAGTACACCTCCGCCTGTAAGACTATTGAGGATGCATTAAGAAAATAGTGCAAGA  
 TTGGAAGCGCCGATGTGTGCGAAATGCTAACCTTCGATAAGAAGGCATTACACTGGCGAACGT  
 50 AAGCTCTTTCGGCGATTACAACCTGTCTTCGGTAATCCCTAGCTTGCCACATCCGGCTCTCGGG

5 TGGCGGGGCGGAGCGCTATCGAGGACATTTTATTCTCGAACTGGTTACATCTGGGCTCGGAACT  
GTGGACGCCGATTACAAGAAGTGCACCAAGGGCCTAAGCATCGCCGACCTCGCCTGTGCTCAGTA  
CTACAACGGAATTATGGTGCTGCCAGGTGTGCTGACGCAGAGCGGATGGCTATGTATACCGGCA  
GTCTCATTTGGCGGGATTGCGTTGGGCGGCCTGACGTCCGCTGTCTCCATCCCTTTCTCTCTGGCT  
10 ATACAAGCCCGACTGAATTATGTGGCCCTGCAGACTGATGTCTTGAAGAAAAATCAGAAGATTCT  
TGCCGCCAGCTTCAACAAGGCCATGACTAATATTGTGGATGCGTTTACCGGAGTGAATGACGCCA  
TACCCCAAACGTCCCAAGCCCTGCAGACAGTCGCCACGGCGTTAAACAAAATCCAGGATGTAGTG  
AATCAGCAAGGGAACAGCTTGAATCACCTGACGTCCAGTTAAGACAGAACTTTCAGGCAATCAG  
TAGCTCAATCCAGGCTATCTACGATCGATTAGATCCCTCCTCAGGCAGATCAGCAGGTGGATCGGC  
15 TCATCACCGGCCGCTCGCGGCATTGAATGTTTTTCGTAAGTCATACCTTGACCAAGTACACGGAG  
GTGAGGGCCAGTCGCCAGCTGGCTCAGCAAAAAGTGAATGAGTGTGTGAAATCACAGAGCAAAACG  
GTACGGGTTTTGTGGAATGGGACGCACATCTTTAGCATCGTTAATGCTGCCCCGAAGGGTTAG  
TCTTCTGCACACTGTGCTCCTTCTACCCAGTATAAAGATGTCGAAGCATGGTCTGGGCTCTGT  
GTGATGGAACATAACGGTTATGTCTTTCGACAGCCAAACCTCGCTCTCTATAAAGAAGGGAATTA  
20 CTATAGGATCACCTCAAGAATCATGTTTCGAGCCAGGATACCAACAATGGCCGATTTTGTGCAGA  
TTGAAAATTGTAACGTGACCTTTGTGAATATCAGTCGATCCGAGCTTCAAACGATTGTTCTGAG  
TACATCGACGTGAATAAACTCTACAAGAGCTGTCTATAAACTGCCTAATTATACCGTGCCTGA  
CCTTGTAGTCGAGCAATACAACCAGACTATTCTGAACCTGACATCGGAAATCTCTACATTGGAGA  
ATAAAGCGCCGAGCTCAATTACACAGTGCAGAAGCTGCAGACCCTGATCGACAATATTAACAGC  
ACTCTGTGGACTTAAAGTGGCTGAACCGTGTGGAGACTTACATCAAGTGGCCCTGGTGGcaaat  
25 actgtcaattttattcaacagtgggcgagttccctagcactggcaatcatgatggctggtctatcct  
tatggatgtgctccaatggatcggttacaatgcagaatttgcattTGA

## PDI-229E -H5iCT-DNA (SEQ ID NO: 150)

25 atggcgaaaaacggttgcgattttctcggttattgttttctcttcttctgtgttggttccttctcagat  
cttcgcgCAAAACGACTAATGGGCTGAACACCAGTTACAGCGTCTGTAACGGCTGCGTCGGATATA  
GCGAGAACGTGTTTCGAGTGGAAAGTGGGGGTACATTCCCTCCGACTTCGCTTTCAATAACTGG  
TTTCTCTGACTAACACAAGCTCCGTCGTGGATGGCGTGGTCAGGTCCCTTTCAGCCTCTTCTCTCT  
30 GAATTGCCTGTGGTCTGTGTCCGGTTAAGATTACTACAGGCTTCGTATACTTCAACGGGACGG  
GCCGGGGGATTGCAAGGGCTTCTCCTCCGACGTGCTGTGATGTCAGATGTGATCCGTTTACAATCTGAAC  
TTCGAAGAGAACTTACGGCGGGGACAATCCTGTTCAAACATCATATGGCGTAGTCGTATTTTAA  
CTGCACCAATAATACCCTGGTGAGTGGGACGCCATATTCCCTTCGGAACAGTGTGCGGTAACCT  
TTTACTGTTTTGTCAACACTACGATCGGAAACGAAACCACTAGCGCCTTTGTGCGAGCTCTGCCA  
35 AAAACAGTTAGGGAGTTCGTGATCTCTCGGACCGGTCACTTCTATATCAACGGCTACCGTTATTT  
TACTTTGGGCAACGTGCAAGCCGTCAATTTTAAATGTGACAACCTGCAGAGACAACCTGACTTTTGCA  
CTGTGGCTCTCGCCAGTTATGCCGATGTGCTGGTGAATGTAAGTCAAACGTCAATTGCCAACATC  
ATCTATTGTAACCTCAGTAATCAACCGGCTCCGCTGTGACCAACTCTCATTCGACGTCCCCGACGG  
ATTCTATTCCACGAGCCCGATTACAGAGCGTGGAACTGCCAGTTTCCATCGTATCCCTCCCAGTTT  
40 ACCACAAGCACACTTTTATCGTTCTCTACGTAGATTTTAAACCCAGTCAGGAGGAGGGAAATGC  
TTCAACTGCTACCCGGCTGGCGTGAACATCACCTTGGCCAATTTTAAATGAACTAAAGGGCCCCCT  
TTGCGTGGATACGTACACTTTACCACAAAGTATGTTGCAGTCTATGCTAACGTCCGCAGGTGGT  
CAGCGTCCATTAACACAGGCAATTGCCCGTTCTCTTTTCGGGAAAGTGAACAACTTCGTGAAGTTT  
GGAAGTGTGTGCTTCAGTTTGAAAGACATTCCGGGCGGCTGCGCCATGCCATTGTGGCTAATTG  
GGCTTATTCCAAGTACTACACCATTTGGCTCTCTCTACGTTAGCTGGAGCGACGGTGACGGTATAA  
45 CGGGCGTACCACAACCGGTGGAAGGGGTGAGCTCTTTCATGAATGTCACTCTGGACAAGTGTACC  
AAATATAATATATACGATGTGAGTGGAGTGGGCGTTATACGCGTGTCTAACGACACCTTTCTAAA  
CGGCATAACCTACACAAGCACGTGAGGCAATCTGTTAGGTTTTTAAAGACGTCACTAAAGGCACTA  
TATATAGCATCACCCCATGCAACCCACCTGATCAATTAGTCGTATATCAGCAAGCTGTTGTGGGT  
GCTATGCTGTGAGAAAATTCACAGCTACGGGTTCTCCAATGTGGTGGAAGTGCCTAAATCTT  
50 TTACGCTAGCAATGGCACATATAACTGTACTGACGCCGTCTTGACTTACAGTTTATTTCGGAGTGT

5 GCGCGGACGGCAGCATTATCGCCGTGCAGCCGGCCAATGTCAGCTATGATTCCGTTTCCGCCATC  
GTGACAGCCAACTTGTGCGATTCCCTCTAACTGGACAACGTCTGTCCAAGTCGAATATCTGCAGAT  
CACCTCAACCCCATAGTAGTCGATTGCTCAACCTACGTCTGCAACGGTAATGTCAGATGTGTGCG  
AGCTGCTCAAGCAGTACACCTCCGCCTGTAAGACTATTGAGGATGCATTAAGAAATAGTGCAAGA  
10 TTGGAAGCGCCGATGTGTGCGGAAATGCTAACCTTCGATAAGAAGGCATTACACTGGCGAACGT  
AAGCTCTTTCGGCGATTACAACCTGTCTTCGGTAATCCCTAGCTTGCCACATCCGGCTCTCGGG  
TGGCGGGGCGGAGCGCTATCGAGGACATTTTATTCTCGAACTGGTTACATCTGGGCTCGGAACT  
GTGGACGCCGATTACAAGAAGTGCACCAAGGGCCTAAGCATCGCCGACCTCGCCTGTGCTCAGTA  
15 CTACAACGGAATTATGGTGCTGCCAGGTGTCGCTGACGCAGAGCGGATGGCTATGTATACCGGCA  
GTCTCATTTGGCGGGATTGCGTTGGGCGGCCTGACGTCCGCTGTCTCCATCCCTTTCTCTCTGGCT  
ATACAAGCCCGACTGAATTATGTGGCCCTGCAGACTGATGTCTGCAAGAAAATCAGAAGATTCT  
TGCCGCCAGCTTCAACAAGGCCATGACTAATATTGTGGATGCGTTTACCGGAGTGAATGACGCCA  
TCACCCAAACGTCCCAAGCCCTGCAGACAGTCGCCACGGCGTTAAACAAAATCCAGGATGTAGTG  
AATCAGCAAGGGAACAGCTTGAATCACCTGACGTCCCAGTTAAGACAGAACTTTCAGGCAATCAG  
20 TAGCTCAATCCAGGCTATCTACGATCGATTAGATCCTCCTCAGGCAGATCAGCAGGTGGATCGGC  
TCATCACCGGCCCGCTCGCGGCATTGAATGTTTTTCGTAAGTCATACCTTGACCAAGTACACGGAG  
GTGAGGGCCAGTCGCCAGCTGGCTCAGCAAAAAGTGAATGAGTGTGTGAAATCACAGAGCAAACG  
GTACGGGTTTTGTGGAATGGGACGCACATCTTTAGCATCGTTAATGCTGCCCCGAAGGGTTAG  
TCTTCCTGCACACTGTGCTCCTTCTACCCAGTATAAAGATGTCGAAGCATGGTCTGGGCTCTGT  
25 GTCGATGGAACCTAACGGTTATGTCCTTCGACAGCCAAACCTCGCTCTCTATAAAGAAGGGAATTA  
CTATAGGATCACCTCAAGAATCATGTTTCGAGCCAGGATACCAACAATGGCCGATTTTGTGCAGA  
TTGAAAATTGTAACGTGACCTTTGTGAATATCAGTCGATCCGAGCTTCAAACGATTGTTTCTGAG  
TACATCGACGTGAATAAACTCTACAAGAGCTGTCTATAAACTGCCTAATTATACCGTGCCTGA  
CCTTGTAGTCGAGCAATACAACCAGACTATTCTGAACCTGACATCGGAAATCTCTACATTGGAGA  
30 ATAAAAGCGCCGAGCTCAATTACACAGTGCAGAAGCTGCAGACCCTGATCGACAATATTAACAGC  
ACTCTTGTGGACTTAAAGTGGCTGAACCGTGTGGAGACTTACATCAAGTGGCCCTGGTGGGTGTG  
GCTCTGTATTTCCGTGGTCCTTATATTTGTTGTAAGTATGCTGCTCCTGtctttatggatgtgct  
ccaatggatcggttacaatgcagaatttgcattTGA

30 PDI-229E -H5iCT(V4)-DNA (SEQ ID NO: 151)

atggcgaaaaacggttgcgattttcggcttattgttttctcttcttgtgttggttccttctcagat  
cttcgcgCAAAACGACTAATGGGCTGAACACCAGTTACAGCGTCTGTAACGGCTGCGTCGGATATA  
35 GCGAGAACGTGTTGCGAGTGGAAGTGGGGGTACATTCCCTCCGACTTCGCTTTCAATAACTGG  
TTTCTCCTGACTAACACAAGCTCCGTGCTGGATGGCGTGGTCAGGTCCTTTCAGCCTCTTCTCCT  
GAATTGCCTGTGGTCTGTGTCCGGGTTAAGATTCACCTACAGGCTTCGTATACTTCAACGGGACGG  
GCCGGGGGATTGCAAGGGCTTCTCCTCCGACGTGCTGTGATGTCAGATGTGATCCGTTACAATCTGAAC  
TTCGAAGAGAACTTACGGCGGGGGACAATCCTGTTCAAAACATCATATGGCGTAGTCGTATTTTA  
CTGCACCAATAATACCCTGGTGAGTGGGACGCCATATTCCTTCGGAACAGTGCTGGGTAACCT  
40 TTTACTGTTTTGTCAACACTACGATCGGAACGAAACCACTAGCGCCTTTGTGCGAGCTCTGCCA  
AAAACAGTTAGGGAGTTCTGTGATCTCTCGGACCGGTCACCTCTATATCAACGGCTACCGTTATTT  
TACTTTGGGCAACGTGCAAGCCGTCAATTTTAATGTGACAACCTGCAGAGACAACCTGACTTTTGCA  
CTGTGGCTCTCGCCAGTTATGCCGATGTGCTGGTGAATGTAAGTCAAACGTCAATTGCCAACATC  
ATCTATTGTAACCTCAGTAATCAACCGGCTCCGCTGTGACCAACTCTCATTCGACGTCCCCGACGG  
ATTCTATTCCACGAGCCCGATTACAGAGCGTGAACCTGCCAGTTTCCATCGTATCCCTCCCAGTTT  
45 ACCACAAGCACACTTTTATCGTTCTCTACGTAGATTTTAAACCCAGTCAGGAGGAGGGAAATGC  
TTCAACTGCTACCCGGCTGGCGTGAACATCACCTTGGCCAATTTTAATGAACTAAAGGGCCCT  
TTGCGTGATACGTACACTTTACCACAAAGTATGTTGCAGTCTATGCTAACGTCGGCAGGTGGT  
CAGCGTCCATTAAACACAGGCAATTGCCCGTTCTCTTTCGGGAAAGTGAACAACTTCGTGAAGTTT  
GGAAGTGTGTGCTTCAGTTTGAAGACATTCCGGGCGGCTGCGCCATGCCATATTGTGGCTAATTG  
50 GGCTTATTCCAAGTACTACACCATTGGCTCTCTCTACGTTAGCTGGAGCGACGGTGACGGTATAA

CGGGCGTACCACAACCGGTGGAAGGGGTGAGCTCTTTTCATGAATGTCACTCTGGACAAGTGTACC  
AAATATAATATATACGATGTGAGTGGAGTGGGCGTTATACGCGTGTCTAACGACACCTTTCTAAA  
CGGCATAACCTACACAAGCACGTGAGGCAATCTGTTAGGTTTTAAAGACGTCACTAAAGGCACTA  
TATATAGCATCACCCCATGCAACCCACCTGATCAATTAGTCGTATATCAGCAAGCTGTTGTGGGT  
5 GCTATGCTGTGCAGAAAACTTCACCAGCTACGGGTTCTCCAATGTGGTGGAAGTGCCCAAATTCTT  
TTACGCTAGCAATGGCACATATAACTGTACTGACGCCGTCTTGACTTACAGTTTCATTTCGGAGTGT  
GCGCGGACGGCAGCATTATCGCCGTGCAGCCGGCCAATGTCAGCTATGATTCCGTTTCCGCCATC  
GTGACAGCCAACCTTGTGATTCCCTCTAACTGGACAACGTCTGTCCAAGTGAATATCTGCAGAT  
CACCTCAACCCCCATAGTAGTCGATTGCTCAACCTACGTCTGCAACGGTAATGTCAGATGTGTCTG  
10 AGCTGCTCAAGCAGTACACCTCCGCCTGTAAGACTATTGAGGATGCATTAAGAAATAGTGCAAGA  
TTGGAAGCGCCGATGTGTGCGAAATGCTAACCTTCGATAAGAAGGCATTACACTGGCGAACGT  
AAGCTCTTTCGGCGATTACAACCTGTCTTCGGTAATCCCTAGCTTGCCACATCCGGCTCTCGGG  
TGGCGGGGCGGAGCGCTATCGAGGACATTTTATTCTCGAACTGGTTACATCTGGGCTCGGAACT  
GTGGACGCCGATTACAAGAAGTGCACCAAGGGCCTAAGCATCGCCGACCTCGCCTGTGCTCAGTA  
15 CTACAACGGAATTATGGTGCTGCCAGGTGTGCTGACGCAGAGCGGATGGCTATGTATACCGGCA  
GTCTCATTTGGCGGGATTGCGTTGGGCGGCCTGACGTCCGCTGTCTCCATCCCTTTCTCTGGCT  
ATACAAGCCCGACTGAATTATGTGGCCCTGCAGACTGATGTCTTGCAGAAAAATCAGAAGATTCT  
TGCCGCCAGCTTCAACAAGGCCATGACTAATATTGTGGATGCGTTTACCGGAGTGAATGACGCCA  
TCACCCAAACGTCCCAAGCCCTGCAGACAGTCGCCACGGCGTTAAACAAAATCCAGGATGTAGTG  
20 AATCAGCAAGGGAACAGCTTGAATCACCTGACGTCCAGTTAAGACAGAACTTTCAGGCAATCAG  
TAGCTCAATCCAGGCTATCTACGATCGATTAGATCCTCCTCAGGCAGATCAGCAGGTGGATCGGC  
TCATCACCGGCCGCTCGCGGCATTGAATGTTTTTCGTAAGTCATACCTTGACCAAGTACACGGAG  
GTGAGGGCCAGTCGCCAGCTGGCTCAGCAAAAAGTGAATGAGTGTGTGAAATCACAGAGCAAACG  
GTACGGGTTTTGTGGAATGGGACGCACATCTTTAGCATCGTTAATGCTGCCCCGAAGGGTTAG  
25 TCTTCTGCACACTGTGCTCCTTCTACCCAGTATAAAGATGTGCAAGCATGGTCTGGGCTCTGT  
GTCGATGGAACAAACGGTTATGTCTTCGACAGCCAAACCTCGCTCTCTATAAAGAAGGGAATTA  
CTATAGGATCACCTCAAGAATCATGTTTCGAGCCAGGATACCAACATGGCCGATTTTGTGCAGA  
TTGAAAATTGTAACGTGACCTTTGTGAATATCAGTCGATCCGAGCTTCAAACGATTGTTTCTGAG  
TACATCGACGTGAATAAACTCTACAAGAGCTGTCTTATAAACTGCCTAATTATACCGTGCCTGA  
30 CCTTGTAGTCGAGCAATACAACCAGACTATTCTGAACCTGACATCGGAAATCTCTACATTGGAGA  
ATAAAAGCGCCGAGCTCAATTACACAGTGCAGAAGCTGCAGACCCTGATCGACAATATTAACAGC  
ACTCTTGTGACTTAAAGTGGCTGAACCGTGTGGAGACTTACATCAAGTGGCCCTGGTGGGTGTG  
GCTCTGTATTTCCGTGGTCTTATATTTGTTGTAAGTATGCTGCTCCTGtgctgctccaatggat  
cgttacaatgcagaattttgcattTGA  
35

**PDI-229E -HlcCT-DNA (SEQ ID NO: 152)**

atggcgaaaaacggttgcgattttctggcttattgttttctcttcttgtgttggttccttctcagat  
cttcgqgCAAAACGACTAATGGGCTGAACACCAGTTACAGCGTCTGTAACGGCTGCGTCGGATATA  
GCGAGAACGTGTTTCGAGTGGAAAGTGGGGGTACATTCCCTCCGACTTCGCTTTCAATAACTGG  
40 TTTCTCCTGACTAACACAAGCTCCGTCGTGGATGGCGTGGTCAGGTCCCTTCAGCCTCTTCTCCT  
GAATTGCCTGTGGTCTGTGTCCGGTTAAGATTCACTACAGGCTTCGTATACTTCAACGGGACGG  
GCCGGGGGATTGCAAGGGCTTCTCCTCCGACGTGCTGTGAGATGTGATCCGTTACAATCTGAAC  
TTCGAAGAGAACTTACGGCGGGGGACAATCCTGTTCAAACATCATATGGCGTAGTCGTATTTTA  
CTGCACCAATAATAACCTGGTGAGTGGGGACGCCATATTCCCTTCGGAACAGTGTGGGTAAC  
45 TTTACTGTTTTGTCAACACTACGATCGGAACGAAACCACTAGCGCCTTTGTGCGAGCTCTGCCA  
AAAACAGTTAGGGAGTTTCGTGATCTCTCGGACCGGTCACTTCTATATCAACGGCTACCGTTATTT  
TACTTTGGGCAACGTGCAAGCCGTCAATTTTAAATGTGACAACCTGCAGAGACAACCTGACTTTTGCA  
CTGTGGCTCTCGCCAGTTATGCCGATGTGCTGGTGAATGTAAGTCAAACGTCAATTGCCAACATC  
ATCTATTGTAACCTCAGTAATCAACCGGCTCCGCTGTGACCAACTCTCATTCGACGTCCCCGACGG  
50 ATTCTATTCCAGAGCCCGATTACAGAGCGTGAAGTGGCAAGTTCCATCGTATCCCTCCCAGTTT

ACCACAAGCACACTTTTATCGTTCTCTACGTAGATTTTAAACCCAGTCAGGAGGAGGGAAATGC  
 TTCAACTGCTACCCGGCTGGCGTGAACATCACCTTGGCCAATTTTAATGAACTAAAGGGCCCCCT  
 TTGCGTGGATACGTACACTTTACCACAAAGTATGTTGCAGTCTATGCTAACGTCGGCAGGTGGT  
 CAGCGTCCATTAACACAGGCAATTGCCCGTTCTCTTTCCGGGAAAGTGAACTTTCGTGAAGTTT  
 5 GGAAGTGTGTGCTTCAGTTTGAAAGACATTCGGGGCGGCTGCGCCATGCCATTGTGGCTAATTG  
 GGCTTATTCCAAGTACTACACCATTTGGCTCTCTCTACGTTAGCTGGAGCGACGGTGACGGTATAA  
 CGGGCGTACCACAACCGGTGGAAGGGGTGAGCTCTTTCATGAATGTCACTCTGGACAAGTGTACC  
 AAATATAATATATACGATGTGAGTGGAGTGGGCGTTATACGCGTGTCTAACGACACCTTTCTAAA  
 CGGCATAACCTACACAAGCACGTCAGGCAATCTGTTAGGTTTTAAAGACGTCACCTAAAGGCACCTA  
 10 TATATAGCATCACCCCATGCAACCCACCTGATCAATTAGTCGTATATCAGCAAGCTGTTGTGGGT  
 GCTATGCTGTGAGAAACTTACCAGCTACGGGTTCTCCAATGTGGTGGAACTGCCCAAATTCTT  
 TTACGCTAGCAATGGCACATATAACTGTACTGACGCCGTCTTGACTTACAGTTTATTTCGGAGTGT  
 GCGCGGACGGCAGCATTATCGCCGTGACGCCGGCCAATGTCAGCTATGATTCCGTTTCCGCCATC  
 GTGACAGCCAACCTTGTGATTCCCTCTAACTGGACAACGTCTGTCCAAGTGAATATCTGCAGAT  
 15 CACCTCAACCCCATAGTAGTCGATTGCTCAACCTACGTCTGCAACGGTAATGTCAGATGTGTGCG  
 AGCTGCTCAAGCAGTACACCTCCGCTGTAAAGACTATTGAGGATGCATTAAGAAATAGTGCAAGA  
 TTGGAAGCGCCGATGTGTGCGAAATGCTAACCTTCGATAAGAAGGCATTACACTGGCGAACGT  
 AAGCTCTTTCGGCGATTACAACCTGTCTTCGGTAATCCCTAGCTTGCCCATCCGGCTCTCGGG  
 TGGCGGGGCGGAGCGCTATCGAGGACATTTTATTCTCGAACTGGTTACATCTGGGCTCGGAACT  
 20 GTGGACGCCGATTACAAGAAGTGCACCAAGGGCCTAAGCATCGCCGACCTCGCCTGTGCTCAGTA  
 CTACAACGGAATTATGGTGCTGCCAGGTGTGCTGACGCAGAGCGGATGCTATGTATACCGGCA  
 GTCTCATTGGCGGGATTGCGTTGGGCGGCCTGACGTCCGCTGTCTCCATCCCTTTCTCTGCGCT  
 ATACAAGCCCGACTGAATTATGTGGCCCTGCAGACTGATGTCTGCAAGAAAATCAGAAGATTCT  
 TGCCGCCAGCTTCAACAAGGCCATGACTAATATTGTGGATGCGTTTACCGGAGTGAATGACGCCA  
 25 TCACCCAAACGTCCCAAGCCCTGCAGACAGTCGCCACGGCGTTAAACAAAATCCAGGATGTAGTG  
 AATCAGCAAGGGAACAGCTTGAATCACCTGACGTCCAGTTAAGACAGAACTTTCAGGCAATCAG  
 TAGCTCAATCCAGGCTATCTACGATCGATTAGATCCTCCTCAGGCAGATCAGCAGGTGGATCGGC  
 TCATCACCGGCCGCTCGCGGCATTGAATGTTTTCGTAAGTCATACCTTGACCAAGTACACGGAG  
 GTGAGGGCCAGTCGCCAGCTGGCTCAGCAAAAAGTGAATGAGTGTGTGAAATCACAGAGCAAACG  
 30 GTACGGGTTTTGTGGAATGGGACGCACATCTTTAGCATCGTTAATGCTGCCCCGAAGGGTTAG  
 TCTTCTGACACTGTGCTCCTTCCCTACCCAGTATAAAGATGTGGAAGCATGGTCTGGGCTCTGT  
 GTCGATGGAATAACGGTTATGTCCTTCGACAGCCAAACCTCGCTCTATAAAGAAGGGAATTA  
 CTATAGGATCACCTCAAGAATCATGTTGAGCCAGGATACCAACATGGCCGATTTTGTGACAGA  
 TTGAAAATTGTAAAGTGAACCTTTGTGAATATCAGTCGATCCGAGCTTCAAACGATTGTTTCTGAG  
 35 TACATCGACGTGAATAAACTCTACAAGAGCTGTCTATAAACTGCCTAATTATACCGTGCCTGA  
 CCTTGTAGTCGAGCAATACAACCAGACTATTCTGAACCTGACATCGGAAATCTCTACATTGGAGA  
 ATAAAAGCGCCGAGCTCAATTACACAGTGCAGAAGCTGCAGACCCTGATCGACAATATTAACAGC  
 ACTCTTGTGGACTTAAAGTGGCTGAACCGTGTGGAGACTTACATCAAGTGCCCTGGTGGGTGTG  
 40 GCTCTGTATTTCCGTGGTCTTATATTTGTTGTAAGTATGCTGCTCCTGagcttctggtatgtgct  
 ctaatgggtctctacagtgtagaatatgtattTGA

# PDI-229E -wtTMCT-AA (SEQ ID NO: 153)

MAKNVAIFGLLFSLVLVPSQIFAQTTNGLNTSY SVCNGCVGYSENVFAVESGGYIPSDFAFNW  
 FLLTNTSSVVDGVVRSFQPLLLNCLWSVSGLRFTTG FVYFNGTGRGDCKGFSSDVLSDVIRYNLN  
 45 FEENLRRGTILFKTSYGVVVFYCTNNTLVSGDAHIPFGTVLGNFYCFVNTTIGNETTSFVVGALP  
 KTVREFVISRTGHFYINGYRYFTLGNVEAVNFNVTTAETD FCTVALASYADVLVNVQS IANI  
 IYCN SVINRLRCDQLSFDVPDGFYSTSPIQSVELPVSIVSLPVYHKHFTIVLYVDFK PQSGGGKC  
 ENCYPAGVNITLANFNETKGPLCVDTSHFTTKYVAVYANVGRWSASINTGNCPFSFGKVNNFVKF  
 GSVCFSLKDIPGGCAMP I VANWAYS KYTYTIGSLYVSWSDGDGITGVPQ PVEGVSSFMNVTLDKCT  
 50 KYNIYDVSGVG VIRVSNDFLNGITYTSTSGNLLGFKDVTKG TIYSITPCNPPDQLVVYQQAVVG

5 AMLSENFTSYGFSNVVELPKFFYASNGTYNCTDAVLTYSSFGVCADGSI IAVQPANVSYDSVSAI  
 VTANLSIPSNWTTSVQVEYLQITSTPIVDCSTYVCNGNVRCVELLKQYTSACKTIEDALRNSAR  
 LESADVSEMLTFDKKAFTLANVSSFGDYNLSSVIPSLPTSGSRVAGRSAIEDILFSKLVTSGLGT  
 VDADYKKCTKGLSIADLACAQYYNGIMVLPGVADAERMAMYTGSLIGGIALGGLTSAVSIPFSLA  
 IQARLNYVALQTDVLQENQKILAASFNKAMTNIVDAFTGVNDAITQTSQALQTVATALNKIQDQV  
 10 NQQGNSLNHLTSQLRQNFQAISSSIQAIYDRLDPPQADQQVDRLITGRLAALNVFVSHTLTKEYTE  
 VRASRQLAQQKVNCEVKSKRYGFCGNGTHIFSI VNAAPEGLVFLHTVLLPTQYKDVEAWSGLC  
 VDGNGYVLRQPNLALYKEGNYYRITSRIMFEPRIPTMAFVQIENCNVT FVNISRSELQTI VPE  
 YIDVNKTLQELSYKLPNYTPDLVVEQYNQTI LNLTSEISTLENKSAELNYTVQKLQTLIDNINS  
 TLVDLKWLN RVETIYIKWPWWVWLCISVVLIFVVSMLLLCCCSTGCCGFSCFASSIRGCCESTKL  
 PYYDVEKIHQ

**PDI-229E -H5iTMCT-AA (SEQ ID NO: 154)**

15 MAKNVAIFGLLFSLLVLPVPSQIFAQTTNGLNTSY SVCNGCVGYSENVFAVESGGYIPSDFAFNNW  
 FLLTNTSSVVDGTVRSFQPLLLNCLWSVSGLRFTTG FVYFNGTGRGDCKGFSSDVLSDVIRYNLN  
 FEENLRRGTILFKTSYGVVVFYCTNNTLVSGDAHIFPGTVLGNFYCFVNTTIGNETTSAFVGALP  
 KTVREFVISRTGHFYINGYRYFTLGNVEAVNFNVTTAETTD FCTVALASYADVLVNVSQTSIANI  
 IYCNSVINRLRCDQLSFDVPDGFYSTSPIQSVELPVSIVSLPVYHKHTFIVLYVDFKPKQSGGGKC  
 20 FNCYPAGVNITLANFNETKGPLCVDTSHTTCKYVAVYANVGRWSASINTGNCFPSFGKVN NFVKF  
 GSVCFSLKDIPGGCAMP I VANWAYSKYYTIGSLYVSWSDGDGITGVQPVEGVSSFMNVTLDKCT  
 KYNIYDVSGVG VIRVSN DTF L NGITYTSTSGNLLGFKDVT KGTIYSITPCNPPDQLVVYQQAVVG  
 AMLSENFTSYGFSNVVELPKFFYASNGTYNCTDAVLTYSSFGVCADGSI IAVQPANVSYDSVSAI  
 VTANLSIPSNWTTSVQVEYLQITSTPIVDCSTYVCNGNVRCVELLKQYTSACKTIEDALRNSAR  
 LESADVSEMLTFDKKAFTLANVSSFGDYNLSSVIPSLPTSGSRVAGRSAIEDILFSKLVTSGLGT  
 25 VDADYKKCTKGLSIADLACAQYYNGIMVLPGVADAERMAMYTGSLIGGIALGGLTSAVSIPFSLA  
 IQARLNYVALQTDVLQENQKILAASFNKAMTNIVDAFTGVNDAITQTSQALQTVATALNKIQDQV  
 NQQGNSLNHLTSQLRQNFQAISSSIQAIYDRLDPPQADQQVDRLITGRLAALNVFVSHTLTKEYTE  
 VRASRQLAQQKVNCEVKSKRYGFCGNGTHIFSI VNAAPEGLVFLHTVLLPTQYKDVEAWSGLC  
 VDGNGYVLRQPNLALYKEGNYYRITSRIMFEPRIPTMAFVQIENCNVT FVNISRSELQTI VPE  
 30 YIDVNKTLQELSYKLPNYTPDLVVEQYNQTI LNLTSEISTLENKSAELNYTVQKLQTLIDNINS  
 TLVDLKWLN RVETIYIKWPWWQILSIYSTVASSLALAIMMAGLSLWMC SNGSLQCRICI

**PDI-229E -H5iCT-AA (SEQ ID NO: 155)**

35 MAKNVAIFGLLFSLLVLPVPSQIFAQTTNGLNTSY SVCNGCVGYSENVFAVESGGYIPSDFAFNNW  
 FLLTNTSSVVDGTVRSFQPLLLNCLWSVSGLRFTTG FVYFNGTGRGDCKGFSSDVLSDVIRYNLN  
 FEENLRRGTILFKTSYGVVVFYCTNNTLVSGDAHIFPGTVLGNFYCFVNTTIGNETTSAFVGALP  
 KTVREFVISRTGHFYINGYRYFTLGNVEAVNFNVTTAETTD FCTVALASYADVLVNVSQTSIANI  
 IYCNSVINRLRCDQLSFDVPDGFYSTSPIQSVELPVSIVSLPVYHKHTFIVLYVDFKPKQSGGGKC  
 40 FNCYPAGVNITLANFNETKGPLCVDTSHTTCKYVAVYANVGRWSASINTGNCFPSFGKVN NFVKF  
 GSVCFSLKDIPGGCAMP I VANWAYSKYYTIGSLYVSWSDGDGITGVQPVEGVSSFMNVTLDKCT  
 KYNIYDVSGVG VIRVSN DTF L NGITYTSTSGNLLGFKDVT KGTIYSITPCNPPDQLVVYQQAVVG  
 AMLSENFTSYGFSNVVELPKFFYASNGTYNCTDAVLTYSSFGVCADGSI IAVQPANVSYDSVSAI  
 VTANLSIPSNWTTSVQVEYLQITSTPIVDCSTYVCNGNVRCVELLKQYTSACKTIEDALRNSAR  
 LESADVSEMLTFDKKAFTLANVSSFGDYNLSSVIPSLPTSGSRVAGRSAIEDILFSKLVTSGLGT  
 45 VDADYKKCTKGLSIADLACAQYYNGIMVLPGVADAERMAMYTGSLIGGIALGGLTSAVSIPFSLA  
 IQARLNYVALQTDVLQENQKILAASFNKAMTNIVDAFTGVNDAITQTSQALQTVATALNKIQDQV  
 NQQGNSLNHLTSQLRQNFQAISSSIQAIYDRLDPPQADQQVDRLITGRLAALNVFVSHTLTKEYTE  
 VRASRQLAQQKVNCEVKSKRYGFCGNGTHIFSI VNAAPEGLVFLHTVLLPTQYKDVEAWSGLC  
 VDGNGYVLRQPNLALYKEGNYYRITSRIMFEPRIPTMAFVQIENCNVT FVNISRSELQTI VPE

YIDVNKTLQELSYKLPNYTVPDLVVEQYNQITILNLTSEISTLENKSAELNYTVQKLQTLIDNINS  
TLVDLKWLNVRVETIYKWPWWVWLCISVVLI FVVSMLLLSLWMCNGLQCRICI

**PDI-229E -H5iCT(V4)-AA (SEQ ID NO: 156)**

5 MAKNVAIFGLLFSLLVLPVPSQIFAQTTNGLNTSYSVCNCGCVGYSENVFAVESGGYIPSDFAFNNW  
FLLTNTSSVVDGTVRSFQPLLLNCLWSVSGLRFTTGFBVYFNGTGRGDCKGFSSDVLSDVIRYNLN  
FEENLRRTILEFKTSYGVVVFYCTNNTLVSGDAHIPFGTVLGNFYCFVNTTIGNETTSFVVGALP  
KTVREFVISRTGHFYINGYRYFTLGNVEAVNFNVTTAETTDCTVALASYADVLVNVQSISIANI  
10 IYCNVSVINRLRCDQLSFDVPDGFYSTSPIQSVELPVSIVSLPVYHKHTFIVLYVDFKPKQSGGGKC  
FNCYPAGVNITLANFNETKGPLCVDTSHTTQYVAVYANVGRWSASINTGNCPFSFGKVNNVFKF  
GSVCFSKLDIPGGCAMPIVANWAYSKYYTIGSLYVSWSDGDGITGVPQPEGVSSFMNVTLDKCT  
KYNIDVSGVGVIRVSNDFLNGITYTSTSGNLLGFKDVTGKTIYSITPCNPPDQLVVYQQAVVG  
AMLSENFYSYGFSSNVVELPKFFYASNGTYNCTDAVLTYSSFGVCADGSI IAVQPANVSYSVSAI  
15 VTANLSIPSNWTTSVQVEYLQITSTPIVVDCCSTYVCNGNVRCVELLKQYTSACKTIEDALRNSAR  
LESADVSEMLTFDKKAFTLANVSSFGDYNLSSVIPSLPTSGSRVAGRSAIEDILFSKLVTSGLGT  
VDADYKKCTKGLSIADLACAQYNGIMVLPGVADAERMAMYTGSLIGGIALGGLTSAVSI PFSLA  
IQARLNYVALQTDVLQENQKILAASFNKAMTNIVDAFTGVNDAITQTSQALQTVATALNKIQDQV  
NQQGNLSNLHTSQLRQNFQAISSSIQAIYDRLDPPQADQQVDRLITGRLAALNVFVSHTLTQYTE  
VRASRLAQQKVNCEVKSKRYGFCGNTHIFSI VNAAPEGLVFLHTVLLPTQYKDVEAWSGLC  
20 VDGTYGVLRQPNLALYKEGNYRITSRIMFEPRIPTMAADFVQIENCNVTFVNISRSELQTI VPE  
YIDVNKTLQELSYKLPNYTVPDLVVEQYNQITILNLTSEISTLENKSAELNYTVQKLQTLIDNINS  
TLVDLKWLNVRVETIYKWPWWVWLCISVVLI FVVSMLLLCCSNGLQCRICI

**PDI-229E-H1cCT-AA (SEQ ID NO: 157)**

25 MAKNVAIFGLLFSLLVLPVPSQIFAQTTNGLNTSYSVCNCGCVGYSENVFAVESGGYIPSDFAFNNW  
FLLTNTSSVVDGTVRSFQPLLLNCLWSVSGLRFTTGFBVYFNGTGRGDCKGFSSDVLSDVIRYNLN  
FEENLRRTILEFKTSYGVVVFYCTNNTLVSGDAHIPFGTVLGNFYCFVNTTIGNETTSFVVGALP  
KTVREFVISRTGHFYINGYRYFTLGNVEAVNFNVTTAETTDCTVALASYADVLVNVQSISIANI  
IYCNVSVINRLRCDQLSFDVPDGFYSTSPIQSVELPVSIVSLPVYHKHTFIVLYVDFKPKQSGGGKC  
30 FNCYPAGVNITLANFNETKGPLCVDTSHTTQYVAVYANVGRWSASINTGNCPFSFGKVNNVFKF  
GSVCFSKLDIPGGCAMPIVANWAYSKYYTIGSLYVSWSDGDGITGVPQPEGVSSFMNVTLDKCT  
KYNIDVSGVGVIRVSNDFLNGITYTSTSGNLLGFKDVTGKTIYSITPCNPPDQLVVYQQAVVG  
AMLSENFYSYGFSSNVVELPKFFYASNGTYNCTDAVLTYSSFGVCADGSI IAVQPANVSYSVSAI  
VTANLSIPSNWTTSVQVEYLQITSTPIVVDCCSTYVCNGNVRCVELLKQYTSACKTIEDALRNSAR  
35 LESADVSEMLTFDKKAFTLANVSSFGDYNLSSVIPSLPTSGSRVAGRSAIEDILFSKLVTSGLGT  
VDADYKKCTKGLSIADLACAQYNGIMVLPGVADAERMAMYTGSLIGGIALGGLTSAVSI PFSLA  
IQARLNYVALQTDVLQENQKILAASFNKAMTNIVDAFTGVNDAITQTSQALQTVATALNKIQDQV  
NQQGNLSNLHTSQLRQNFQAISSSIQAIYDRLDPPQADQQVDRLITGRLAALNVFVSHTLTQYTE  
VRASRLAQQKVNCEVKSKRYGFCGNTHIFSI VNAAPEGLVFLHTVLLPTQYKDVEAWSGLC  
40 VDGTYGVLRQPNLALYKEGNYRITSRIMFEPRIPTMAADFVQIENCNVTFVNISRSELQTI VPE  
YIDVNKTLQELSYKLPNYTVPDLVVEQYNQITILNLTSEISTLENKSAELNYTVQKLQTLIDNINS  
TLVDLKWLNVRVETIYKWPWWVWLCISVVLI FVVSMLLLSFWMCNGLQCRICI

**Native OC43-CoV S protein wtTM/CT AA (AVR40344) (SEQ ID NO: 158)**

45 MFLILLISLPTAFVIGDLNCTLDPRLKGSFNNRDTGPPSISIDTVDVNTGLGTYVLDREVYLN  
TLFLNGYYPTSGSTYRNMAKGTDLSTLWFKPPFLSDFINGIFAKVKNTKVFKDGVMYSEFP  
PAI

5 TIGSTFVNNTSYSVVVQPRRTINSTQDGVNKLQGLLEVSVQCYNMCEYPHTICHPNLGNHFKEWLWHY  
 DTGVVSCLYKRNFYTDVNATYLYFHFYQEGGTFYAYFTDTGFVTKFLFNVYLGMAISHYYVMPLT  
 CIRRPKDGFSLEYWVTPPTPRQYLLAFNQDGIIFNAVDCMSDFMSEIKCKTQSIAPPTGVYELNG  
 YTVQPVADVRRKPDLPNCNIEAWLNDKSVPSPLNWERKTFSNCFNMSSLMSTIQADSFCTCNNI  
 10 DAAKIYGMCFSSITIDKFAIPNRRKVDLQGLNLGYLQSSNYRIDTTATSCQLYNNLPAANVSISR  
 FNPSTWNKRFGFIEDSVFVPQPTGVFTNHSVVYAQHCFAKAPKNFCPCSSCSCPGKNNIGTCTPAG  
 TNSLTCDNLCTLDPITLKAPDTYKCPQSKSLVGIGEHCSGLAVKSDYCGNNSCTCQPPQAFGLWSA  
 DSCIQGDKNIFANFILHDVNNGLTCTSTDLQKANTEIELGVCVNYDLYGISGQGI FVEVNATYYN  
 15 SWQNLLYDSNGNLYGFRDYITNRTFMHSCYSGRVSAAYHANSSEPALLFRNICKNYVFNNSLTR  
 QLQPINYSFDSYLGCVVNAYNSTAISVQTCDLTVGSGYCVDYSKNRRSRRAITTG YRFTNFEPFT  
 VNSVNDLSLEPVGGLYEIQIPSEFTIGNMEEFIQTSSPKVTIDCAAFVCGDYAACKLQLVEYGSFC  
 DNINAILTEVNELLDTTQLQVANSLMNGVTLSTKLKDGVNFNVDINFPVLGCLGSECSKASSR  
 SAIEDLLFDKVKLSDVGFVEAYNNCTGGAEIRDLCVQSYKGIKVLPLLSNQISGYTLAATSA  
 20 SLFPPWTAAAGVPFYLNVQYRINGLGVTMDVLSQNQKLIANAFNNALHAIQQGFATNSALVKIQ  
 AVVNANAEALNNLLQQLSNRFGAISASLQEILSRDLAEAEAQIDRLINGRLTALNAYVSQQQLSD  
 STLVKFSAAQAMEKVNCEVKSQSSRINFCCGNNGHIIISLVQNAPYGLYFIHFNYVPTKYVTAKVSP  
 GLCIAGNRGIAPKSGYFVNVTNTWMTYTGSGYYPPEPITENNVVMSTCAVNYTKAPYVMLNTSIP  
 NLPDFKEELDQWFKNQTSVAPDLSLDYINVTFLDLQVEMNRLQEAIKVLNHSYINLKDIGTYEYY  
 VKWPWYVWLLICLAGVAMLVLLFFICCTGCGTSCFKKCGGCCDDYTG YQELVIKTSDDD

Native 229E S protein wtTM/CT AA (P15423) (SEQ ID NO: 159)

25 MFVLLVAYALLHIAGCQTTNGLNTSYSVCNGCVGYSENVFAVESGGYIPSDFAFNNWFLLTNTSS  
 VVDGVVRSFQPLLLNCLWSVSGLRFTTG FVYFNGTGRGDCKGFSDDLSDVIRYNLNFEEENLRG  
 TILFKTSYGVVVFYCTNNTLVSGDAHIFPGTVLGNFYCFVNTTIGNETTSAFVFGALPKTVREFVI  
 30 SRTHFYINGYRYFTLGNVEAVNFNVTAEITDFCTVALASYADVLVNVSQTSIANI IYCNSVIN  
 RLRCQQLSFDVPDGFYSTSPIQSVELPVSIVSLVPVYHKHTFIVLYVDFKPPQSGGGKCFNCYPAGV  
 NITLANFNETKGPLCVDTSHTFTKYVAVYANVGRWSASINTGNCPFSFGKVN NFVKFGSVC FSLK  
 DIPGGCAMPIVANWAYSKYYTIGSLYVSWSDGDGITGVPQPVGVSSFMNVTLDKCTKYNIYDVS  
 35 GVG VIRVSNDTFLNGITYTSTSGNLLGFKDVTKGTIYSITPCNPPDQLVVYQAVVGAMLSNF  
 SYGFSNVVELPKFFYASNGTYNCTDAVLTYSSFGVCADGSI IAVQPRNVSYDSVSAIVTANLSIP  
 SNWTTSVQVEYLQITSTPIVVDCTSYVCNGNVRCVELLKQYTSACKTIEDALRNSARLESADVSE  
 MLTFDKKAFTLANVSSFGDYNLSSVIPSLPTSGSRVAGRSAIEDILFSLKLVTSGLGTVDADYKKC  
 TKGLSIADLACAQYNGIMVLPGVADAERMAMYTGSLIGGIALGGLTSAVSIPFSLAIQARLNYV  
 40 ALQTDVLQENQKILAASFNKAMTNIVDAFTGVNDAITQTSQALQTVATALNKIQDVVNQQGNSLN  
 HLTSQLRQNFQAISSSIQAIYDRLDTIQADQQVDRLITGRLAALNVFVSHTLTKYTEVRASRQLA  
 QQKVNCEVKSQSKRYGFCGNGTHIFSI VNAPEGLVFLHTVLLPTQYKDVEAWSGLCVDGTNGYV  
 LRQPNLALYKEGNYRITSRIMFEPRIPTMAADFVQIENCNVTFVNI SRSELQTI VPEYIDVNKTL  
 QELSYKLPNYTVPDLVVEQYNQITILNLTSEISTLENKSAELNYTVQKLQTLIDNINSTLVDLKWL  
 NRVETYIKWPWWVWLCISVVLI FVVSMLLLCCCSTGCCGFFSCFASSIRGCCSTKLPYYDVEKI  
 HIQ

Native OC43-CoV S protein wtTM/CT AA (AVR40344) without signal peptide (SEQ ID NO: 160)

45 VIGDLNCTLDPRLKGSFNNRDTGPPSISIDTVDVNTGLGTYVLDREVYLNNTLFLNGYYPTSGST  
 YRNMALKGTDLSTLWFKPPFLSDFINGIFAKVKNTKVKFDGVMYSEFPAITIGSTFVNNTSYSVV  
 VQPRRTINSTQDGVNKLQGLLEVSVQCYNMCEYPHTICHPNLGNHFKEWLWHYDTGVVSCLYKRNF  
 YDVNATYLYFHFYQEGGTFYAYFTDTGFVTKFLFNVYLGMAISHYYVMPLT CIRRPKDGFSLEYW  
 VTPPTPRQYLLAFNQDGIIFNAVDCMSDFMSEIKCKTQSIAPPTGVYELNGYTVQPVADVRRKP

5 DLPNCNIEAWLNDKSVPSPLNWERKTFNSCNFMSSLSMFIQADSFTCNNIDA AKIYGMCFSSIT  
 IDKFAIPNRRKVDLQLGNLGYLQSSNYRIDTTATSCQLYYNLPAA NVSVSRFNPSTWNKRFGFIE  
 DSVFVPQPTGVFTNHSV VYAQHCFKAPKNFCPCSSCSCPGKNNIGITCPAGTNSLTCDNLCTLDP  
 10 I TLKAPD TYKCPQSKSLV GIGEHCSGLAVKSDYCGNNSCTCQPQAF LGWSAD SCLQGDKCNIFAN  
 FILHDVNNGLTCSTDLQKANTEIELGVCVNYDLYGISGQGI FVEVNATYYNSWQNLLYDSNGNLY  
 GFRDYITNRTFMIHSCYSGRVSAAYHANSSEPALLFRNI KCNYVFNNSLTRQLQPINYSFDSYLG  
 CVVNAYNSTAISVQTCDLTVGSGYCVDY SKNRRSRRAITTG YRFTNFEPFTVNSVND SLEPVGGL  
 YEIQIPSEFTIGNMEEFIQTSSPKVTIDCAAFVCGDYAACKLQLVEYGS FCDNINAILTEVNELL  
 15 DTTQLQVANSLMNGVTLS TKLKDG VNFNVDDINFSPVLGCLGSECSKASSRSAIEDLLDFDKVKLS  
 DVGFVEAYNNCTGGAEIRD LICVQSYKG I KVL PPL LSENQISGYTLAATSASLFPFWTAAAGVPF  
 YLNVQYRINGLGVTMDVLSQNQKLIANAFNNALHAIQQGF DATNSALVKIQAVVNANAEALNNLL  
 QQLSNRFGAISASLQEILSRLDALEAEAQIDRLINGRLTALNAYVSQQLSDSTLVKFSAQAAMEK  
 VNECVKSQSSRINFCGNGNHII SLVQNAPYGLYFIHFNYVPTKYVTAKVSPGLCIAGNRGIAPKS  
 GYFVN VNNTWMYTGS GYYP EPITENN VVMSTCAVNYTKAPYVMLNTSIPNLDPFKEELDQWFK  
 20 NQTSVAPDLSLDYINVTFLDLQVEMNRLQEAIKVLNHSYINLKDIGTYEYYVKWPWYVWLLICLA  
 GVAMLVLLFFICCTGCGTSCFKKCGGCCDDYTG YQELVIKTS HDD

**Native 229E S protein wtTM/CT AA (P15423) without signal peptide (SEQ ID NO: 161)**

20 QTTNGLNTSY SVCNGCVGYSENVFAVESGGYIPSDFAFNNWFLLTNTSSVVDG VVRSFQPLLLNC  
 LWSVSGLRFTTG FVYFNGTGRGDCKGFSSDVLSDVIRYNLNFEENLRRTILFKTSYGVVVFYCT  
 NNTLVSGDAHIPFGTVLGNFYCFVNTTIGNETTSAFVGALPKTVREFVISRTGHFYINGYRYFTL  
 GNVEAVNFNVTTAETTD FCTVALASYADVLVNVSQTSIANIIYCNSVINRLRCDQLSFDVPDGFY  
 25 STSPIQSVELPVSIVSLPVYHKHTFIVLYVDFKPKQSGGGKCFNCYPAGVNITLANFNETKGPLCV  
 DTSHTFTKYVAVYANVGRWSASINTGNCPFSFGKVN NFVKFGSVCFSLKDI PGGCAMPIVANWAY  
 SKYYTIGSLYVSWSDGDGITGVPQ PVEGVSSFMNVTL DKCTKYNIYDVSGVG VIRVSNDTFLNGI  
 TYTSTSGNLLGFKDVT KGTIYSITPCNPPDQLV VYQQA VVGAM LSENFTSYGFSNVVELPKFFYA  
 SNGTYNCTDAVLTYSSFGVCADGSI IAVQPRNVSYDSVSAIVTANLSIPS NWTTSVQVEYLQITS  
 30 TPIVVD CSTYVCNGNVRCVELLKQYTSACKTIEDALRNSARLESADVSEMLTFDKKAFTLANVSS  
 FGDYNLSSVIPSLPTSGSRVAGRSAIEDILFSKLVT SGLGTVDADYKKCTKGLSIADLACAQYYN  
 GIMVLPGVADAERMAMYTGSLIGGIALGGLTS AVSIPFSLAIQARLNYVALQTDVLQENQKILAA  
 SFNKAMTNIVDAFTGVNDAITQTSQALQT VATALNKIQDVVNQQGNSLNHLTSQLRQNFQAIS  
 35 IQAIYDRLDTIQADQQVDRLITGR LAALNVFVSHTLT KYTEVRASRQLAQQKVNECVKSQSKRYG  
 FCGNGTHIFSI VNAAPEGLVFLHTVLLPTQYKDVEAWSGLCVDGTNGYVLRQPNLALYKEGNYR  
 ITSRI MFEPRIPTMADFVQIENCNVT FVNISRSELQTIPEYIDVNKTLQELSYKLPNYTV PDLV  
 VEQYNQTI LNLTSEISTLENKSAELNYTVQKLQTLIDNINSTLVDLKWLNRVETIYIKWPWWVWLC  
 ISVVLIFVVSMLLLCCCSTGCCGFFSCF ASSIRGCC ESTKLPYYDVEKIHIQ

**TMCT region of modified PDI-OC43-COV wtTMCT-AA (SEQ ID NO: 162)**

WYVWLLICLAGVAMLVLLFFICCTGCGTSCFKKCGGCCDDYTG YQELVIKTS HDD

**40 TMCT region of modified PDI-OC43-COV H5iTMCT-AA (SEQ ID NO: 163)**

WYQILSIYSTVASSLALAIMMAGLSLWMCSNGSLQCRICI

**TMCT region of modified PDI-OC43-COV H5iCT-AA (SEQ ID NO: 164)**

WYVWLLICLAGVAMLVLLFFISLWMCSNGSLQCRICI

TMCT region of modified PDI-OC43-COV H5iCT(V4)-AA (SEQ ID NO: 165)

WYVWLLICLAGVAMLVLLFFICCSNGSLQCRICI

TMCT region of modified PDI-OC43-COV H1cCT-AA (SEQ ID NO: 166)

WYVWLLICLAGVAMLVLLFFISFWMCSNGSLQCRICI

5 TMCT region of modified PDI-229E-wtTMCT-AA (SEQ ID NO: 167)

WWVWLCISVVLIFVVSMLLLCCCSTGCCGFFSCFASSIRGCCSTKLPYYDVEKIHQ

TMCT region of modified PDI-229E-H5iTMCT-AA (SEQ ID NO: 168)

WWQILSIYSTVASSLALAIMMAGLSLWMCSNGSLQCRICI

TMCT region of modified PDI-229E-H5iCT-AA (SEQ ID NO: 169)

10 WWVWLCISVVLIFVVSMLLLSLWMCSNGSLQCRICI

TMCT region of modified PDI-229E-H5iCT(V4)-AA (SEQ ID NO: 170)

WWVWLCISVVLIFVVSMLLLCCSNGSLQCRICI

TMCT region of modified PDI-229E-H1cCT-AA (SEQ ID NO: 171)

WWVWLCISVVLIFVVSMLLLSFWMCNSNGSLQCRICI

15 TM/CT Region of Modified OC43-CoV S protein with intervening peptide sequence X<sub>n</sub> (SEQ ID NO: 172)

WYVWLLICLAGVAMLVLLFFI - (X)<sub>n</sub> - CSNGSXXCXICI

TM/CT Region of Modified OC43-CoV S protein with intervening peptide sequence X<sub>n</sub> (SEQ ID NO: 173)

20 WWVWLCISVVLIFVVSMLLL - (X)<sub>n</sub> - CSNGSXXCXICI

[00343] All citations are hereby incorporated by reference.

[00344] The present invention has been described with regard to one or more  
embodiments. However, it will be apparent to persons skilled in the art that a number of  
variations and modifications can be made without departing from the scope of the  
invention as defined in the claims.

25

## WHAT IS CLAIMED IS:

1. A modified coronavirus S-protein comprising, in series,
  - an ectodomain derived from a coronavirus S-protein,
  - a transmembrane and cytosolic tail domain (TMCT), wherein the TMCT is a chimeric TMCT, comprising:
    - a transmembrane domain (TM), wherein the TM or a portion of the TM is derived from a coronavirus S-protein and
    - a cytosolic tail (CT), wherein the CT or a portion of the CT is derived from an influenza hemagglutinin (HA) protein.
2. The modified coronavirus S-protein of claim 1, wherein the TM is directly fused to the CT.
3. The modified coronavirus S-protein of claim 1, wherein the TM is a chimeric TM comprising a N-terminal sequence derived from the coronavirus S-protein TM and a C-terminal sequence derived from the influenza HA protein TM.
4. The modified coronavirus S-protein of claim 3, wherein the chimeric TM comprises a N-terminal sequence derived from the coronavirus S-protein TM comprising at least 20 amino acids corresponding to amino acids 1-20 of SEQ ID NO: 18 or SEQ ID NO: 169, or at least 21 amino acids corresponding to amino acids 1-21 of SEQ ID NO: 118 or 164, or at least 22 amino acids corresponding to amino acids 1-22 of SEQ ID NO: 123 and one or more than one amino acid from the C-terminal end of the influenza HA protein TM.
5. The modified coronavirus S-protein of claim 4, wherein the one or more than one amino acid from the C-terminal end of the influenza HA protein TM are selected from AGL or conserved substitution of AGL, MAGL or conserved substitution of MAGL.

6. The modified coronavirus S-protein of claim 1, wherein the CT is a chimeric CT comprising a N-terminal sequence derived from the coronavirus S-protein CT and a C-terminal sequence derived from the influenza HA protein CT.
- 5 7. The modified coronavirus S-protein of claim 6, wherein the chimeric CT comprises a C-terminal sequence derived from the influenza HA protein CT comprising amino acids corresponding to amino acids 27-37 of SEQ ID NO: 18, 126, 128, 129, 130 or 131; or amino acids corresponding to amino acids 27-36 of SEQ ID NO: 127 and one or more than one amino acid from the N-terminal end of the coronavirus S-protein CT.
- 10 8. The modified coronavirus S-protein of claim 7, wherein the one or more than one amino acid from the N-terminal end of the coronavirus S-protein CT are selected from C or a conserved substitution of C, CC or a conserved substitution of CC, or CCM or a or a conserved substitution of CCM.
- 15 9. The modified coronavirus S-protein of claim 3, wherein the chimeric TM comprises amino acids corresponding to amino acids 1-20 of SEQ ID NO: 18 or SEQ ID NO: 169, or to amino acids 1-21 of SEQ ID NO: 118 or SEQ ID NO: 164, or amino acids 1-22 of SEQ ID NO: 123.
- 20 10. The modified coronavirus S-protein of claim 4, wherein the chimeric CT comprises amino acids corresponding to amino acids of 27-37 of SEQ ID NO: 18, 126, 128, 129, 130 or 131; or amino acids 27-36 of SEQ ID NO: 127.
- 25 11. The modified coronavirus S-protein of claim 1, wherein the chimeric TMCT comprises a chimeric TM comprising amino acids corresponding to amino acids 1-20 of SEQ ID NO: 18 or SEQ ID NO: 169, or to amino acids 1-21 of SEQ ID NO: 118 or SEQ ID NO: 164, or amino acids 1-22 of SEQ ID NO: 123 and, a chimeric CT comprising amino acids corresponding to amino acids 27-37 of SEQ ID NO: 18, 126, 128, 129, 130 or 131; amino acids 27-36 of SEQ ID NO: 127 or a combination thereof.
12. The modified S-protein of any one of claims 1-11, wherein the S-protein comprises one or more than one amino acid substitution when compared to a wild-type coronavirus S-protein amino acid sequence.

13. The modified S-protein of claim 12, wherein the one or more than one amino acid substitution comprises:

- i) substitutions that restricts the processing at a cleavage site between S1 and S2 subunit,
- ii) substitution of one or more than one amino acid to one or more than one proline, or
- iii) substitutions that restrict the processing at the cleavage site between the S1 and the S2 subunit and substitution of one or more than one amino acid to one or more than one proline.

14. The modified S-protein of claim 12 or 13, wherein the one or more than one amino acid substitution correspond to amino acids at positions 971 and 972, when compared to reference amino acid sequence of SEQ ID NO: 2.

15. The modified S-protein of claim 12 or 13, wherein the one or more than one amino acid substitution correspond to amino acids at positions 802, 927, 971 and 972, when compared to reference amino acid sequence of SEQ ID NO: 2.

16. The modified S-protein of claim 14 or 15, wherein the modified S-protein further comprises one or more than one amino acid substitution corresponding to amino acids at positions 667, 668, 670, or a combination thereof, when compared to reference amino acid sequence of SEQ ID NO: 2.

17. The modified S-protein of claim 16, wherein the substitution of the amino acid corresponding to the amino acid at position 667 of SEQ ID NO: 2 is to glycine or a conserved substitution of glycine, the substitution of the amino acid corresponding to position 668 of SEQ ID NO: 2 is to a serine or a conserved substitution of serine, the substitution of the amino acid corresponding to position 670 of SEQ ID NO: 2 is to a serine or a conserved substitution of serine, the substitution of the amino acid corresponding to the amino acid at positions 971 and 972 of SEQ ID NO: 2 is to a proline or a conserved substitution of proline.

18. The modified S-protein of claim 16, wherein the substitution of the amino acid corresponding to the amino acid at position 667 of SEQ ID NO: 2 is to glycine or a conserved substitution of glycine, the substitution of the amino acid corresponding to

5 position 668 of SEQ ID NO: 2 is to a serine or a conserved substitution of serine, the substitution of the amino acid corresponding to position 670 of SEQ ID NO: 2 is to a serine or a conserved substitution of serine, the substitution of the amino acid corresponding to the amino acid at positions 802, 927, 971 and 972 of SEQ ID NO: 2 is to a proline or a conserved substitution of proline.

19. The modified S-protein of any one of claims 14-18, wherein the modified S-protein further comprises an amino acid substitution corresponding to amino acid at position 923, when compared to reference amino acid sequence of SEQ ID NO: 2.

10 20. The modified S-protein of claim 19, wherein the substitution of the amino acid corresponding to the amino acid at position 667 of SEQ ID NO: 2 is to glycine or a conserved substitution of glycine, the substitution of the amino acid corresponding to position 668 of SEQ ID NO: 2 is to a serine or a conserved substitution of serine, the substitution of the amino acid corresponding to position 670 of SEQ ID NO: 2 is to a serine or a conserved substitution of serine, the substitution of the amino acid  
15 corresponding to the amino acid at positions 802, 927, 971 and 972 of SEQ ID NO: 2 is to a proline or a conserved substitution of proline and the substitution of the amino acid corresponding to position 923 of SEQ ID NO: 2 is to phenylalanine or a conserved substitution of phenylalanine.

20 21. The modified S-protein of any one of claims 12-20, wherein the one or more than one substitution maintains the S-protein in a pre-fusion state or produces are higher yield of the modified S-protein when expressed in a host or host cell, when compared to the yield of a corresponding S-protein without the one or more than one substitutions expressed in the host or host cell.

25 22. The modified S-protein of any one of claims 1-21, wherein the S-protein comprises an S1 subunit and an S2 subunit, wherein the S2 subunit comprises the chimeric TMCT.

23. The modified S-protein of any one of claims 1-22, wherein the influenza HA protein is derived from influenza type B or influenza subtype H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, or H16.

24. The modified S-protein of any one of claims 1-23, wherein the coronavirus S-protein is derived from SARS-CoV-2, SARS-CoV-1, MERS-CoV, OC43-CoV or 229E-CoV.
- 5 25. The modified S-protein of any one of claims 1-23, wherein the ectodomain is derived from SARS-CoV-2, SARS-CoV-1, MERS-CoV, OC43-CoV or 229E-CoV, the TM or the portion of the TM is derived from SARS-CoV-2, SARS-CoV-1, MERS-CoV, OC43-CoV or 229E-CoV or wherein the ectodomain and the TM or the portion of the TM are derived from SARS-CoV-2, SARS-CoV-1, MERS-CoV, OC43-CoV or 229E-CoV.
- 10 26. The modified S-protein of any one of claims 1-25, wherein the TMCT comprises a sequence having about 80% to about 100% identity with the sequence of SEQ ID NO: 18, 19, 37, 38, 39, 64, 126, 127, 128, 129, 130, 131, 118, 119, 120, 123, 124, 125, 134, 135, 164, 165, 166, 169, 170, 171, 172 or 173.
27. The modified S-protein of any one of claims 1-26, wherein the modified S-protein is produced as a precursor protein, the precursor protein comprising the modified S-protein and a signal peptide.
- 15 28. The modified S-protein of claim 27, wherein the signal peptide is native or non-native to the modified S-protein.
29. The modified S-protein of claim 27 or 28, wherein the signal peptide is derived from the signal peptide of protein disulfide-isomerase (PDI).
- 20 30. The modified S-protein of any one of claims 1-29, wherein the modified S-protein comprises plant specific N-glycans.
- 25 31. The modified S-protein of claim 1, wherein the amino acid sequence of the S-protein comprises from 80% to 100% identity with amino acids of SEQ ID NO: 5, 21, 30, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 95, 96, 97, 108, 109, 110, 144, 145, 146, 155, 156 or 157, or with amino acids 24-1259 of SEQ ID NO: 47 amino acids 25-1259 of SEQ ID NO: 48, amino acids 25-1259 of SEQ ID NO: 49, amino acids 25-1259 of SEQ ID NO: 50, amino acids 25-1259 of SEQ ID NO: 51, amino acids 25-1259 of SEQ ID NO: 52, amino acids 25-1259 of SEQ ID NO: 53, amino acids 25-1259 of SEQ

ID NO: 54, amino acids 25-1259 of SEQ ID NO: 55, amino acids 25-1259 of SEQ ID  
 NO: 56, amino acids 25-1259 of SEQ ID NO: 57, amino acids 25-1259 of SEQ ID NO:  
 58, amino acids 25-1262 of SEQ ID NO: 59, amino acids 25-1261 of SEQ ID NO: 60,  
 amino acids 25-1258 of SEQ ID NO: 61, or amino acids 25-1256 of SEQ ID NO: 62,  
 5 amino acids 25-1243 of SEQ ID NO: 95, amino acids 25-1240 of SEQ ID NO: 96, amino  
 acids 25-1243 of SEQ ID NO: 97, amino acids 25-1341 of SEQ ID NO: 108, amino acids  
 25-1338 of SEQ ID NO: 109, amino acids 25-1341 of SEQ ID NO: 110, or amino acids  
 25-1351 of SEQ ID NO: 144, amino acids 25- 1348 of SEQ ID NO: 145, amino acids 25-  
 1351 of SEQ ID NO: 146, amino acids 25- 1159 of SEQ ID NO: 155, amino acids 25-  
 10 1156 of SEQ ID NO: 156, or amino acids 25- 1159 of SEQ ID NO: 157.

32. The modified S-protein of any one of claims 1-30, wherein the CT or portion of the  
 CT comprises from 80% to 100% identity with the sequence of SEQ ID NO: 15, or with  
 amino acids 35-50 of SEQ ID NO 6, 8, 7, 9, 10, 12, 13 or 14, , or with amino acids 34-49  
 of SEQ ID NO 11, or with amino acids 553-568 of SEQ ID NO:3 or with amino acids 22-  
 15 37 of SEQ ID NO:18, or with amino acids 21-40 of SEQ ID NO: 19, or with amino acids  
 21-39 of SEQ ID NO: 37, or with amino acids 25-36 of SEQ ID NO: 38 or with amino  
 acids 24-34 of SEQ ID NO: 39, or amino acids 22-37 of SEQ ID NO: 126, 128, 129, 130  
 or 131 or amino acids 22-36 of SEQ ID NO: 127

33. The modified S-protein of claim 27, wherein the precursor protein comprises from  
 20 80% to 100% identity with amino acids 1-1234 of SEQ ID 1, or with amino acids 1-1234  
 of SEQ ID NO: 5, with amino acids 1-1243 of SEQ ID NO: 30, with amino acids 1-1227  
 of SEQ ID NO: 95 , with amino acids 1-1325 of SEQ ID NO: 108, with amino acids 1-  
 1216 of SEQ ID NO: 112, with amino acids 1-1318 of SEQ ID NO: 113, with amino  
 acids 1-1335 of SEQ ID NO: 144, with amino acids 1-1143 of SEQ ID NO: 155, with  
 25 amino acids 1-1325 of SEQ ID NO: 158 , with amino acids 1-1135 of SEQ ID NO: 159,  
 and wherein the amino acid sequence of the CT comprises from 80% to 100% identity  
 with the sequence of SEQ ID NO: 15, or with amino acids 35-50 of SEQ ID NO 6, 8, 7,  
 9, 10, 12, 13 or 14, or with amino acids 34-49 of SEQ ID NO 11, or with amino acids  
 553-568 of SEQ ID NO:3.

34. A nucleic acid comprising a nucleotide sequence encoding the modified S protein of any one of claims 1-33.
35. A trimer comprising the modified S-protein of any one of claims 1-33.
- 5 36. A virus like particle (VLP) comprising the modified S-protein of any one of claims 1-33 or the trimer of claim 35.
37. The VLP of claim 36, wherein the VLP further comprises plant lipids.
38. A composition comprising an effective dose of the modified S-protein of any one of claims 1-33, the trimer of claim 35 or the VLP of any one of claims 36-37 and a pharmaceutically acceptable carrier, adjuvant, vehicle or excipient.
- 10 39. A vaccine for inducing an immune response, the vaccine comprising an effective dose of the modified S-protein of any one of claims 1-33, the trimer of claim 35, the VLP of any one of claims 36-37, or the composition of claim 38.
40. The vaccine of claim 39, wherein the vaccine is a multivalent vaccine, comprising a mixture of VLP.
- 15 41. A method for inducing immunity to a Coronavirus infection in a subject, the method comprising administering the composition of claim 38 or the vaccine of claim 39 or 40 to the subject.
42. The method of claim 41, wherein the composition or vaccine is administered once to the subject.
- 20 43. The method of claim 41, wherein the composition or vaccine is administered multiple times to the subject.
44. The method of claim 41, wherein the composition or vaccine is administered as an initial dose and one or more than one subsequent doses are administered between 1 day and 6 months from the administration of the initial dose.

45. An antibody or antibody fragment prepared using the composition of claim 38 or the vaccine of claim 39 or 40.

46. A non-human host or host cell comprising the modified S-protein of any one of claims 1-33, the trimer of claim 35, the nucleic acid of claims 34 or the VLP of claims 36 or 37.

47. A method of producing a modified S-protein in a non-human host or host cell comprising:

a) introducing the nucleic acid of claim 34 into the non-human host or host cell, or providing the non-human host or host cell comprising the nucleic acid of claim 34, and

b) incubating the non-human host or host cell under conditions that permit the expression of the nucleic acid, thereby producing the modified S-protein.

48. A method of producing a virus like particle (VLP) in a non-human host or host cell comprising:

a) introducing the nucleic acid of claim 34 into the non-human host or host cell, or providing the non-human host or host cell comprising the nucleic acid of claim 34, and

b) incubating the non-human host or host cell under conditions that permit the expression of the nucleic acid, thereby producing the VLP.

49. The method of claim 47 or 48, the method further comprising step c) of harvesting the non-human host or host cell.

50. A method of increasing yield of production of a modified Coronavirus S-protein in a non-human host or host cell comprising:

a) introducing the nucleic acid of claims 34 into the host or host cell; or providing a non-human host or host cell comprising the nucleic acid of claim 34; and

b) incubating the non-human host or host cell under conditions that permit expression of the modified S-protein encoded by the nucleic acid, thereby producing modified S-protein at a higher yield compared to a host or host cell expressing the unmodified S-protein.

5 51. A method of increasing yield of production of virus like particle (VLP) in a non-human host or host cell comprising:

10 a) introducing into a non-human host or host cell a nucleic acid encoding a modified coronavirus S-protein comprising, in series, an ectodomain derived from a coronavirus S-protein, a transmembrane and cytosolic tail domain (TMCT), wherein the TMCT is a chimeric TMCT, comprising: a transmembrane domain (TM), wherein the TM or a portion of the TM is derived from a coronavirus S-protein and a cytosolic tail (CT), wherein the CT or a portion of the CT is derived from an influenza hemagglutinin (HA) protein, or

15 providing a non-human host or host cell comprising the nucleic acid encoding the modified coronavirus S-protein comprising, in series, an ectodomain derived from a coronavirus S-protein, a transmembrane and cytosolic tail domain (TMCT), wherein the TMCT is a chimeric TMCT, comprising: a transmembrane domain (TM), wherein the TM or a portion of the TM is derived from a coronavirus S-protein and a cytosolic tail (CT), wherein the CT or a portion of the CT is derived from an influenza hemagglutinin (HA) protein; and

20

b) incubating the non-human host or host cell under conditions that permit expression of the modified S-protein encoded by the nucleic acid, thereby producing VLP comprising modified S-protein at a higher yield compared to the yield of VLP in a host of host cell that expresses unmodified S protein.

25 52. A method of producing a virus like particle (VLP) in a non-human host or host cell comprising:

a) introducing into the a non-human host or host cell a nucleic acid encoding a modified coronavirus S-protein comprising, in series, an ectodomain derived

5 from a coronavirus S-protein, a transmembrane and cytosolic tail domain (TMCT), wherein the TMCT is a chimeric TMCT, comprising: a transmembrane domain (TM), wherein the TM or a portion of the TM is derived from a coronavirus S-protein and a cytosolic tail (CT), wherein the CT or a portion of the CT is derived from an influenza hemagglutinin (HA) protein; or

10 providing a non-human host or host cell comprising the nucleic acid encoding the modified coronavirus S-protein comprising, in series, an ectodomain derived from a coronavirus S-protein, a transmembrane and cytosolic tail domain (TMCT), wherein the TMCT is a chimeric TMCT, comprising: a transmembrane domain (TM), wherein the TM or a portion of the TM is derived from a coronavirus S-protein and a cytosolic tail (CT), wherein the CT or a portion of the CT is derived from an influenza hemagglutinin (HA) protein; and

b) incubating the non-human host or host cell under conditions that permit the expression of the nucleic acid, thereby producing the VLP.

15 53. The method of any one of claims 48, 51 or 52, wherein the VLP is further extracted and purified from the non-human host or host cell.

54. A VLP produced by the method of claim 48, 51 or 52.

55. A modified S-protein produced by the method of claim 47 or 50.

20 56. The method of any one of claims 47-55, or the non-human host or host cell of claim 46, wherein the non-human host or host cell comprises a plant, portion of a plant, a plant cell, a fungi, a fungi cell, an insect, an insect cell, an animal or an animal cell.

25 57. A composition comprising a virus-like particles (VLP), the VLP comprising a modified coronavirus S-protein comprising, in series, an ectodomain derived from a coronavirus S-protein, a transmembrane and cytosolic tail domain (TMCT), wherein the TMCT is a chimeric TMCT, comprising: a transmembrane domain (TM), wherein the TM or a portion of the TM is derived from a coronavirus S-protein and a cytosolic tail (CT), wherein the CT or a portion of the CT is derived from an influenza hemagglutinin

(HA) protein and wherein the S-protein comprises substitutions at positions 667, 668, 670, 971 and 972 when compared to reference amino acid sequence of SEQ ID NO: 2.

5 58. A composition comprising a virus-like particle (VLP), the VLP comprising a modified coronavirus S-protein comprising, in series, an ectodomain derived from a coronavirus S-protein, a transmembrane and cytosolic tail domain (TMCT), wherein the TMCT is a chimeric TMCT, comprising: a transmembrane domain (TM), wherein the TM or a portion of the TM is derived from a coronavirus S-protein and a cytosolic tail (CT), wherein the CT or a portion of the CT is derived from an influenza hemagglutinin (HA) protein and wherein the S-protein comprises a glycine substitution at position 667,  
10 a serine substitution at position 668, a serine substitution at position 670, a proline substitution at position 971 and a proline substitution at position 972, the position corresponding to reference amino acid sequence of SEQ ID NO: 2.

15 59. A composition comprising a virus-like particle (VLP), the VLP comprising a modified coronavirus S-protein, the modified S-protein comprising the sequence of SEQ ID NO: 21 or amino acids 25-1259 of SEQ ID NO: 51.

20 60. A composition comprising a virus-like particle (VLP), the VLP comprising modified coronavirus S-protein comprising, in series, an ectodomain derived from a coronavirus S-protein, a transmembrane and cytosolic tail domain (TMCT), wherein the TMCT is a chimeric TMCT, comprising: a transmembrane domain (TM), wherein the TM or a portion of the TM is derived from a coronavirus S-protein and a cytosolic tail (CT), wherein the CT or a portion of the CT is derived from an influenza hemagglutinin (HA) protein and wherein the S-protein comprises substitutions at positions 667, 668, 670, 802, 923, 927, 971 and 972 when compared to reference amino acid sequence of SEQ ID NO: 2.

25