

**2019-NCOV (SARS-COV-2) VACCINE****FIELD OF THE INVENTION**

The present invention relates to Coronavirus 2019-nCoV spike protein, polynucleotides  
5 encoding said spike protein, antibodies and vaccines for treatment or prevention of 2019-nCoV  
infection.

**BACKGROUND OF THE INVENTION**

Since 08 December 2019, several cases of pneumonia of unknown aetiology have been  
10 reported in Wuhan, Hubei province, China. Most patients worked at or lived around the local  
Huanan seafood wholesale market, where live animals were also on sale. In the early stages of  
this pneumonia, severe acute respiratory infection symptoms occurred, with some patients  
rapidly developing acute respiratory distress syndrome (ARDS), acute respiratory failure, and  
other serious complications. On 07 January 2020, a novel coronavirus was identified by the  
15 Chinese Center for Disease Control and Prevention (CDC) from the throat swab sample of a  
patient, and was subsequently named 2019-nCoV by WHO and has now been designated  
SARS-CoV-2.

Coronaviruses can cause multiple system infections in various animals and mainly  
respiratory tract infections in humans, such as severe acute respiratory syndrome (SARS) and  
20 Middle East respiratory syndrome (MERS). Most patients have mild symptoms and good  
prognosis.

So far, a few patients with 2019-nCoV have developed severe pneumonia, pulmonary  
oedema, ARDS, or multiple organ failure and have died. All costs of 2019-nCoV treatment are  
covered by medical insurance in China. At present, information regarding the epidemiology  
25 and clinical features of pneumonia caused by 2019-nCoV is scarce and no vaccine is available.  
Therefore, there is an ongoing need for the development of antigens which may be used in  
vaccines to prevent and treat 2019-nCoV infection. Further, there is a need to provide antigens  
that can be produced at scale inexpensively.

The present invention addresses one or more of the above needs by providing  
30 polynucleotides encoding 2019-nCoV antigens, particularly antigens from the spike protein of  
2019-nCoV, vectors comprising said polynucleotide, vectors encoding said antigens, and  
binding compounds (particularly antibodies and antibody-like molecules including aptamers  
and peptides) raised against the antigen, together with the use thereof (either alone or in  
combination) in the prevention or treatment of infection with 2019-nCoV. The polynucleotides

encoding the antigen are optimised for expression in host cells of interest. Antibodies and antibody-like molecules raised against the antigen may bind (e.g. specifically bind) to the antigen.

## 5 **SUMMARY OF THE INVENTION**

To-date, no vaccine has been developed for 2019-nCoV. The present inventors have developed polynucleotides encoding the 2019-nCoV spike protein, said polynucleotides being optimised for expression in commonly used expression systems. These polynucleotides provide increased level and duration of expression of the 2019-nCoV spike protein, making  
10 them advantageous for large-scale production of this antigen. Furthermore, the polynucleotides devised by the inventors encode the spike protein amino acid sequence in a form which retains the conformation of the native spike protein. Thus, the spike protein produced according to the present invention can give rise to an immunoprotective response, particularly through the production of neutralising antibodies.

15 Accordingly, the present invention provides an isolated polynucleotide encoding a spike protein from 2019-nCoV having at least 90% identity with SEQ ID NO: 1, or a fragment thereof that has a common antigenic cross-reactivity with said spike protein, wherein said polynucleotide is optimised for recombinant expression.

Said polynucleotide of may be optimised for expression in a host cell selected from:  
20 *Escherichia coli*; yeast, preferably *Komagataella* or *Saccharomyces*; and/or mammalian cells, preferably human cells. Optimisation may occur by omitting one or more cis-acting sequence motif, said one or more cis-acting sequence motif being independently selected from: an internal TATA-box; a chi-site; a ribosomal entry site; an AT-rich and/or GC-rich stretch of sequence; an RNA instability motif; a repeat sequence and/or an RNA secondary structure; a  
25 cryptic splice donor site; a cryptic splice acceptance site; and/or any combination of (a) to (i). Said polynucleotide may integrate into the host cell genome. Said polynucleotide may have a codon adaptation index (CAI) of at least about 0.80, preferably at least about 0.9, more preferably at least about 0.93. A polynucleotide of the invention may comprise or consist of a nucleic acid sequence having at least 90% identity to any one of SEQ ID NO: 2 to 8, 13, 14,  
30 26, 27, 28, 30 or 32. The polynucleotide of the invention typically encodes a spike protein, or fragment thereof which: (i) retains the conformational epitopes present in the native 2019-nCoV spike protein; (ii) results in the production of neutralising antibodies specific for the spike protein or fragment thereof when the nucleic acid or the encoded spike protein or fragment thereof is administered to a subject; and/or (iii) comprises or consists of receptor-

binding domain (RBD) of the 2019-nCoV spike protein, preferably having at least 90% identity with SEQ ID NO: 15.

The invention further provides an expression construct comprising polynucleotide of the invention, operably linked to a promoter.

5       The invention further provides a vaccine composition comprising a spike protein from 2019-nCoV having at least 90% identity with SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein, wherein optionally said fragment comprises or consists of receptor-binding domain (RBD) of the 2019-nCoV spike protein, preferably having at least 90% identity with SEQ ID NO: 15. Said vaccine typically results in  
10   the production of neutralising antibodies specific for the spike protein or fragment thereof when administered to a subject.

The invention also provides a viral vector, RNA vaccine or DNA plasmid that expresses a spike protein from 2019-nCoV having at least 90% identity with SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein, wherein  
15   optionally said fragment comprises or consists of receptor-binding domain (RBD) of the 2019-nCoV spike protein, preferably having at least 90% identity with SEQ ID NO: 15. Said viral vector, RNA vaccine or DNA plasmid may further encode a signal peptide. The signal peptide may direct secretion from human cells. The viral vector, RNA vaccine or DNA plasmid of the invention may further express one or more additional antigen or a fragment thereof, preferably  
20   one or more additional antigen from 2019-nCoV, or a fragment thereof. The spike protein or fragment thereof and the one or more additional antigen or fragment thereof may be expressed: as a fusion protein; or in separate viral vectors, RNA vaccines or DNA plasmids for use in combination. Said viral vector, RNA vaccine or DNA plasmid may comprise one or more polynucleotide or expression construct of the invention.

25       The invention also provides a fusion protein comprising a spike protein from 2019-nCoV having at least 90% identity with SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein, wherein optionally said fragment comprises or consists of receptor-binding domain (RBD) of the 2019-nCoV spike protein, preferably having at least 90% identity with SEQ ID NO: 15. Said VLP or fusion protein may  
30   further comprise: the Hepatitis B surface antigen (HBsAg), or a fragment thereof that has a common antigenic cross-reactivity with said HBsAg; the HPV 18 L1 protein, or a fragment thereof that has a common antigenic cross-reactivity with said HPV 18 L1 protein; the Hepatitis E P239 protein (HEV), or a fragment thereof that has a common antigenic cross-reactivity with said Hepatitis E P239 protein; and/or the HPV 16 L1 protein, or a fragment thereof that has a

common antigenic cross-reactivity with said HPV 16 L1 protein. The fusion protein may be encoded by a polynucleotide which comprises or consists of a nucleic acid sequence having at least 90% identity with any one of SEQ ID NO: 3, 5, 6, 8, 26, 27, 29,30, or 32; and/or may the fusion protein may comprise or consist of an amino acid sequence having at least 90% identity  
5 with any one of SEQ ID NO: 9, 10, 11, 12, 28,31, or 33.

The invention also provides a virus-like particle comprising a spike protein from 2019-nCoV having at least 90% identity with SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein, wherein optionally said fragment comprises or consists of receptor-binding domain (RBD) of the 2019-nCoV spike protein,  
10 preferably having at least 90% identity with SEQ ID NO: 15. Preferably the VLP comprises or consists of a fusion protein of the invention.

The invention also provides an antibody, or binding fragment thereof, that specifically binds to a 2091-nCoV spike protein antigen, or fragment thereof, as herein. Said antibody, or binding fragment thereof, may be a monoclonal or polyclonal antibody. Said antibody, or  
15 binding fragment thereof may be an Fab, F(ab')<sub>2</sub>, Fv, scFv, Fd or dAb.

The invention further provides an oligonucleotide aptamer that specifically binds to a 2019-nCoV spike protein or fragment thereof as defined herein.

The invention provides a vaccine composition comprising the viral vector, and/or RNA vaccine and/or DNA plasmid of the invention.

20 The invention also provides a polynucleotide of the invention, and/or the expression construct of the invention, and/or vaccine composition of the invention, and/or the viral vector and/or RNA vaccine and/or DNA plasmid of the invention and/or the virus-like particle of the invention, and/or the fusion protein of the invention, and/or the antibody of the invention and/or the aptamer of the invention for use in the treatment and/or prevention of 2019-nCoV infection.

25 The invention also provides the use of a polynucleotide of the invention, and/or the expression construct of the invention, and/or vaccine composition of the invention, and/or the viral vector and/or RNA vaccine and/or DNA plasmid of the invention and/or the virus-like particle of the invention, and/or the fusion protein of the invention, and/or the antibody of the invention and/or the aptamer of the invention in the manufacture of a medicament for the  
30 prevention and/or treatment of 2019-nCoV infection.

The invention also provides a method of producing a spike protein from 2019-nCoV having at least 90% identity with SEQ ID NO: 1, or a fragment thereof, comprising expressing a polynucleotide of the invention in a host cell, and optionally purifying the spike protein or

fragment. Said method may further comprise formulating said spike protein or fragment thereof with a pharmaceutically acceptable carrier or diluent.

### **DESCRIPTION OF FIGURES**

5

**Figure 1:** Schematic of the coronavirus's structure and the function of the structural proteins.

**Figure 2:** Tabulated results from ELISA reporting antibody titre at day0 and day14 following administration of 2019-nCoV spike protein and fusion proteins comprising 2019-nCoV spike protein produced according to the invention.

**Figure 3:** Western blots showing secreted protein HBSAg(EAAAK)<sub>3</sub>RBD released into culture medium after 40h (following concentration by centrifugal filtration). Left blot = HBSAg antibody. Right blot = 2019-nCoV antibody. For both blots, lane M = marker; lane 1 = HEV-GGGS-RBD #10B; lane 2 = HBSAg-EAAK-RBD #2A (293F secreted, total protein content by Bradford assay = 750µg/ml); lane 3 = HBSAg-EAAAK-RBD #2B (293F secreted, total protein content by Bradford assay = 500µg/ml); lane 4 = CoV-s #17 (293F surface bounded HisTag)

**Figure 4:** Graphs illustrating titres at d14 (A) and d42 (B) in mice inoculated with HBSAg(EAAAK)<sub>3</sub>RBD generated in HEK cells, either alone, or using aluminium hydroxide or Addavax™ adjuvants.

**Figure 5:** Western blots showing secreted protein HEV-(GGGS)<sub>3</sub>-RBD produced by *E. coli*. Left blot = #10A. Middle blot = #10B. Right blot = #10C. For all blots an anti-HEV mAB was used at a dilution of 1:4000.

**Figure 6:** Graphs illustrating titres at d14 (A) and d42 (B) in mice inoculated with HEV-GGGS-RBD generated in *E. coli*, either alone, or using aluminium hydroxide or Addavax™ adjuvants.

**Figure 7:** Western blot of recombinant HBSAg-(EAAAK)<sub>3</sub>-full-length 2019-nCoV spike protein fusion protein (HBSAg-(EAAAK)<sub>3</sub>-CoV-s) clone D8-SA01-01-01 (4x) and clone D8-SA01-02-01 (5X) expressed in HEK 293 cells.

30

### **DETAILED DESCRIPTION OF THE INVENTION**

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of

ordinary skill in the art. The meaning and scope of the terms should be clear; however, in the event of any latent ambiguity, definitions provided herein take precedent over any dictionary or extrinsic definition. It should be understood that this invention is not limited to the particular methodology, protocols, and reagents, etc., described herein and as such can vary. The terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention, which is defined solely by the claims. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. In this application, the use of "or" means "and/or" unless stated otherwise. Furthermore, the use of the term "including", as well as other forms, such as "includes" and "included", is not limiting.

The description of embodiments of the disclosure is not intended to be exhaustive or to limit the disclosure to the precise form disclosed. While specific embodiments of, and examples for, the disclosure are described herein for illustrative purposes, various equivalent modifications are possible within the scope of the disclosure, as those skilled in the relevant art will recognize. For example, while method steps or functions are presented in a given order, alternative embodiments may perform functions in a different order, or functions may be performed substantially concurrently. The teachings of the disclosure provided herein can be applied to other procedures or methods as appropriate. The various embodiments described herein can be combined to provide further embodiments. Aspects of the disclosure can be modified, if necessary, to employ the compositions, functions and concepts of the above references and application to provide yet further embodiments of the disclosure. Moreover, due to biological functional equivalency considerations, some changes can be made in protein structure without affecting the biological or chemical action in kind or amount. These and other changes can be made to the disclosure in light of the detailed description. All such modifications are intended to be included within the scope of the appended claims.

The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that such publications constitute prior art to the claims appended hereto.

### 30 **Coronaviruses**

Coronaviruses (CoVs) belong to the subfamily *Coronavirinae*, in the family *Coronaviridae* of the order *Nidovirales*. There are four genera: Alphacoronavirus, Betacoronavirus, Gammacoronavirus and Deltacoronavirus. Alphacoronaviruses and

Betacoronaviruses infect species of mammal, Gammacoronaviruses infect species of bird, and Deltacoronaviruses infect both species of mammals and birds.

CoVs are large enveloped single positive-sense RNA viruses. Mutation rates of RNA viruses are greater than DNA viruses, suggesting a more efficient adaptation process for survival.

CoVs have the largest genome among all RNA viruses, typically ranging from 27 to 32 kb. The CoV genome codes for at least four main structural proteins: spike (S), membrane (M), envelope (E), nucleocapsid (N) proteins and other accessory proteins which aid the replicative processes and facilitate entry into cells. Figure 1 summarises the coronavirus's structure and the function of the structural proteins. Briefly, the CoV genome is packed inside a helical capsid formed by the nucleocapsid and further surrounded by an envelope. Associated with the viral envelope are at least three structural proteins: the membrane and envelope proteins, which are involved in virus assembly, and the spike protein, which mediates virus entry into host cells. Some coronaviruses also encode an envelope-associated hemagglutinin-esterase protein (HE). The spike protein forms large protrusions from the virus surface, giving coronaviruses the appearance of having crowns, from which the name "Coronavirus" is derived. As well as mediating virus entry, the spike protein is a critical determinant of viral host range and tissue tropism and a major inducer of host immune responses.

2019-nCoV (officially named severe acute respiratory syndrome coronavirus 2, SARS-CoV-2, the two terms being used interchangeably herein) is the causative agent of coronavirus disease 2019 (COVID-19) and is contagious among humans. It is believed that 2019-nCoV originated in animals, with bats being a likely source given the genetic similarities of 2019-nCoV to SARS-CoV (79.5%) and bat coronaviruses (96%). Any disclosure herein in relation to CoVs also applies directly and without restriction to 2019-nCoV.

The CoV spike protein comprises three domains: (i) a large ectodomain; (ii) a transmembrane domain (which passes through the viral envelope in a single pass); and (iii) a short intracellular tail. The ectodomain consists of three receptor-binding subunits (3 x S1) and a trimeric stalk made of three membrane-fusion subunits (3 x S2). During virus entry, S1 binds to a receptor on the host cell surface for viral attachment, and S2 fuses the host and viral membranes, allowing viral genomes to enter host cells. Receptor binding and membrane fusion are the initial and critical steps in the coronavirus infection cycle. There is significant divergence in the receptors targeted by different CoVs.

The 2019-nCoV spike protein or immunogenic fragments thereof have therapeutic potential as antigens for vaccines against 2019-nCoV infection.

Accordingly, as described herein, the invention relates to a 2019-nCoV spike protein has at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity with SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein. Preferably the invention  
5 relates to a spike protein from 2019-nCoV has at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity with SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein. More preferably, the invention relates to a spike protein from 2019-nCoV having least 98%, at least 99% or more with SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with  
10 said spike protein. The spike protein from 2019-nCoV may comprise or consist of SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein.

According to the present invention, the 2019-nCoV spike protein or fragment thereof encoded by a polynucleotide of the invention maintains one or more conformational epitope  
15 present in native 2019-nCoV spike protein. As such, the 2019-nCoV spike protein or fragment thereof encoded by a polynucleotide of the invention is capable of giving rise to an immunoprotective effect. Typically said immunoprotective effect comprises the production of neutralising antibodies (nAb) which specifically bind to the one or more conformational epitope of the 2019-nCoV spike protein or fragment thereof encoded by a polynucleotide of  
20 the invention. A conformational epitope of a CoV spike protein has a specific three-dimensional structure that is found in the tertiary structure of the CoV spike protein. Said one or more conformational epitope is typically within the ectodomain of the spike protein. Preferably the 2019-nCoV spike protein or fragment thereof encoded by a polynucleotide of the invention retains all of the conformational epitopes present in native 2019-nCoV spike  
25 protein.

In some preferred embodiments, the invention relates to an immunogenic fragment of 2019-nCoV spike protein which is the receptor-binding domain (RBD) of the 2019-nCoV spike protein. This RBD is responsible for 2019-nCoV binding to a host cell and thus facilitates entry of 2019-nCoV particles into the host cell. The RBD corresponds to amino acid residues  
30 319 to 529 of SEQ ID NO: 1, as described herein is referred to as SEQ ID NO: 15. The RBD is encoded by bases corresponding to positions 955 to 1597 in the genome of the 2019-nCoV virus (Genbank Accession No. MN908947, version 3 of which (MN908947.3) was deposited 17 January 2020). Accordingly, as described herein, the invention relates to an RBD of the 2019-nCoV spike protein has at least 70%, at least 75%, at least 80%, at least 90%, at least

95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity with SEQ ID NO: 15. Preferably the invention relates to an RBD of the 2019-nCoV spike protein that has at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity with SEQ ID NO: 15. More preferably, the invention relates to an RBD of the 2019-nCoV spike protein having least 98%, at least 99% or more with SEQ ID NO: 15. The RBD of the 2019-nCoV spike protein may comprise or consist of SEQ ID NO: 15. Any and all disclosure herein relating to the 2019-nCoV spike protein (e.g. in relation to polynucleotides, viral vectors, DNA plasmids, RNA vaccines, virus-like particles (VLPs), fusion proteins, antibodies, compositions and pharmaceutical compositions, formulations and therapeutic indications) applies equally and without reservation to the RBD of the 2019-nCoV spike protein. References herein to RBD refer to the RBD of the 2019-nCoV spike protein.

CoVs are large enveloped single positive-sense RNA viruses. Mutation rates of RNA viruses are greater than DNA viruses, suggesting a more efficient adaptation process for survival. Thus, there is a risk that antigenic drift will also become a feature of the 2019-nCoV, or if 2019-nCoV becomes endemic in the population once the pandemic has subsided. Indeed, research to-date has already identified mutations within the receptor binding domain (RBD) of the spike protein of 2019-nCoV, particularly G476S and V483A/G, as well as a prevalent D614G mutation in the vicinity of the S1/S2 site (Saha *et al.*, ChemRxiv<sup>TM</sup> <http://doi.org/10.26434/chemrxiv.12320567.v1>), which the evidence suggests can enhance cell entry by the 2019-nCoV virion, and also broaden the host cell tropism. Other mutations reported in the 2019-nCoV spike protein include S943 (particularly S943P), L5 (particularly L5F), L8 (particularly L8F), V367 (particularly V367F), H49 (particularly H49Y), Y145 (particularly Y145H/del), Q239 (particularly Q239K), A831 (particularly A831V), D839 (particularly D839Y/N/E), and P1263 (particularly P1263L), or any combination thereof (Korber *et al.*, BioRxiv<sup>TM</sup> <https://doi.org/10.1101/2020.04.29.069054>).

Accordingly, the invention advantageously allow 2019-nCoV vaccine antigens to be modified if required to provide enhanced immunity against strains with mutated spike proteins as they arise. By way of non-limiting example, any 2019-nCoV spike protein or fragment thereof according to the invention may be modified (particularly by substitution) at position (i) D614, (ii) V483, (iii) G476, (iv) K417, (v), E484, (vi) N501, (vii) A570, and (viii) P681, or any combination of (including any two, any three, any four, any five, any six, any seven or all eight) of (i) to (viii). Alternatively or in addition, the 2019-nCoV spike proteins or fragments thereof may comprise deletion mutations, including deletions at one or more of amino acid residues 69, 70 and/or 144. As described herein, the positions of the mutations/modifications

typically corresponds to the numbering of amino acids in SEQ ID NO: 1 of the present invention.

Modification at position D614, particularly the D614G substitution, is preferred. In particular, any 2019-nCoV spike protein or fragment thereof according to the invention may  
5 comprise the following substitutions (i) G476S, (ii) V483A/G, (iii) D614G, (iv) K417N/T, (v), E484K, (vi) N501Y, (vii) A570D, and (viii) P681H, or any combination of (including any two, any three, any four, any five, any six, any seven or all eight) of (i) to (viii).

The invention also relates to 2019-nCoV spike proteins or fragments thereof from a variant 2019-nCoV. In particular, the invention may relate to 2019-nCoV spike proteins or  
10 fragments thereof from the B.1.1.7 strain (also known as 201/501Y.V1, which was first detected in the UK); the B.1.351 strain (also known as 20H/501.V2, which was first detected in South Africa), and/or the P1 strain (also known as 20J/501Y.V3, which was first detected in Japan and Brazil). The key mutations of the B.1.1.7 strain comprise deletion of residues 69/70 and 144Y, as well as N501Y, A570D, D614G and P681H substitutions. The key mutations of  
15 the B.1.351 strain comprise K417N, E484K, N501Y and D614G substitutions. The key mutations of the P.1 strain comprise E484K, K417N.T, N501Y and D614G.

All the disclosure herein in relation to polynucleotides, spike proteins and fragments thereof, VLPs, fusion proteins and DNA/RNA vaccine applies equally to different variants and strains of 2019-nCoV unless explicitly stated.

20

### **Polynucleotides**

The present invention provides a polynucleotide that encodes or expresses (the terms “encode” and “express” are used interchangeably herein) the protein or immunogenic fragment of the invention. The term polynucleotide encompasses both DNA and RNA sequences.  
25 Herein, the terms “nucleic acid”, “nucleic acid molecule” and “polynucleotide” are used interchangeably.

The invention provides an isolated polynucleotide encoding a spike protein from 2019-nCoV having at least 90% identity with SEQ ID NO: 1, or a fragment thereof that has a common antigenic cross-reactivity with said spike protein. For example, the polynucleotide  
30 may encode an RBD of the 2019-nCoV spike protein, preferably wherein said RBD has at least 90% identity with SEQ ID NO: 15. Exemplary polynucleotides encoding the RBD are shown in SEQ ID NO: 13, and the codon-optimised sequence of SEQ ID NO: 14.

The invention also encompasses polynucleotides encoding a variant spike protein from 2019-nCoV, as described above, or fragments thereof that have common antigenic cross-

reactivity with said variant spike protein. Said variant spike proteins typically have at least 90% identity with SEQ ID NO: 1.

A polynucleotide of the invention may be used for recombinant expression of the protein or immunogenic fragment of the invention (including in the form of a VLP or fusion protein), or as a DNA/RNA vaccine.

The present inventors are the first to provide improved polynucleotides encoding the 2019-nCoV spike protein or immunogenic fragments thereof. In particular, the present inventors have designed polynucleotides that are optimised for recombinant expression. A polynucleotide of the invention may be optimised for expression in one or more particular cell type, for example, eukaryotic cells (e.g. mammalian cells, yeast cells, insect cells or plants cells) or prokaryotic cells (bacterial cells). Typically the polynucleotides are optimised for expression in bacterial cells, yeast cells or mammalian cells. Preferably said polynucleotides are optimised for expression in *Escherichia coli* (for example, BL21(DE3), RV308(DE3), HMS174(DE3) or K12 strains), *Komagataella* (formally assigned as *Pichia*, particularly *Komagataella pastoris* or *Komagataella phaffii*), *Saccharomyces* (particularly *Saccharomyces cerevisiae*) or human cells (preferably 293 F cells, HEK 293 cells, HEK 293T cells or HeLa cells). Other cell types/expression systems of interest include *Pichia angusta*, *Hansenula polymorpha*, Chinese Hamster Ovary (CHO) cells and/or insect cell baculovirus-based expression systems.

The term “optimised” as used herein relates to optimisation for recombinant expression of the 2019-nCoV spike protein or immunogenic fragment thereof, and includes both codon optimisation and/or other modifications to the polynucleotide (both in terms of the nucleic acid sequence and other modifications) which increase the level and/or duration of expression of the 2019-nCoV spike protein from the polynucleotide within the host cell/organism, or which otherwise provide an advantage when expressing the 2019-nCoV spike protein, or fragment thereof, from a polynucleotide of the invention.

The term “codon optimised” refers to the replacement of at least one codon within a base polynucleotide sequence with a codon that is preferentially used by the host organism or cell in which the polynucleotide is to be expressed. Typically, the most frequently used codons in the host organism are used in the codon-optimised polynucleotide sequence. Methods of codon optimisation are well known in the art.

By way of non-limiting example, another form of polynucleotide optimisation are modifications which minimise RNA structure, as structures that involve or otherwise occlude the RBS and/or start codon in genes expressed in prokaryotes can impair expression.

Optimisation also encompasses modifications to the polynucleotide which optimise translation, either by increasing the rate of translation, or by balancing the rate of translation with the need to allow for efficient “self” or chaperone-aided protein folding, in which strategically placed slower codons or codon runs (e.g. at protein domain boundaries) could maximise folding efficiency whilst maintaining a high overall translation rate. Optimisation may also encompass the removal of deleterious motifs within the nucleic acid sequence of the polynucleotide. By way of non-limiting example, expressing a gene under control of a T7 promoter in *E. coli*, it is preferable to avoid both class I and II transcriptional termination sites. Shine–Dalgarno-like sequences within the coding sequence may cause incorrect downstream initiation or translational pauses in prokaryotic hosts. For expression in eukaryotic hosts/cells, potential splice signals, polyadenylation signals and other motifs affecting mRNA processing and stability may be removed. Other classes of deleterious motifs include sequences that promote ribosomal frameshifts and pauses. Any combination of modifications may be made to the polynucleotides of the invention to optimise expression in a host cell of interest.

Typically polynucleotides of the invention optimised for expression in bacterial cells, particularly *E. coli*, include cloned N-terminal and/or C-terminal deleted amino acids. Preferably about 1 to 20, more preferably about 1 to 15, most preferably about 5 to 10 cloned N-terminal and/or C-terminal deleted amino acids are included.

It will be understood by a skilled person that numerous different polynucleotides can encode the same polypeptide as a result of the degeneracy of the genetic code. It is also understood that skilled persons may, using routine techniques, make nucleotide substitutions that do not affect the polypeptide sequence encoded by the nucleic acid molecules to reflect the codon usage of any particular host organism in which the polypeptides are to be expressed. Therefore, unless otherwise specified, a “polynucleotide that encodes the protein or immunogenic fragment of the invention” includes all polynucleotide sequences that are degenerate versions of each other and that encode the same amino acid sequence.

A polynucleotide of the invention is typically designed so that it is capable of integrating into the genome of a host cell of interest. Different optimisation strategies may be used to facilitate integration depending on the desired host cell.

Typically a polynucleotide of the invention is optimised by the removal or omission of one or more cis-acting sequence motif (also referred to interchangeably as cis-acting elements or cis-acting regulatory elements). A cis-acting sequence motif is a sequence in the vicinity of the structural portion of a gene that is required for gene expression. Said one or more cis-acting sequence motif may be independently selected from: (a) an internal TATA-box; (b) a Chi-site;

(c) a ribosomal entry site; (d) an AT-rich and/or GC-rich stretch of sequence; (e) an RNA instability motif; (f) a repeat sequence and/or an RNA secondary structure; (g) a cryptic splice donor site; (h) a cryptic splice acceptance site; and/or (i) any combination of (a) to (i). These cis-acting sequence motifs are known in the art. By way of non-limiting example, regions of high GC content (e.g. above about 70%, preferably above about 80%) and/or low GC content (e.g. below about 40%, preferably below about 30%) are omitted. Preferably both regions of high and low GC content are omitted, in combination with the removal or omission of one or more other cis-acting sequence motif.

A polynucleotide of the invention may also be "codon optimised" as described herein. Codon optimisation preferably occurs in addition to the removal or omission of one or more cis-acting sequence motif as described herein.

The average GC content of a polynucleotide of the invention may also be modified to optimise expression of said polynucleotide. For example, the average GC content of a polynucleotide may be in the region of about 40% to about 60%, preferably about 40% to about 57%, more preferably about 45% to about 56%.

A polynucleotide of the invention typically has a codon adaptation index (CAI) of at least about 0.80, preferably at least about 0.9, more preferably at least about 0.91, at least about 0.92, at least about 0.93, at least about 0.94, at least about 0.95, at least about 0.96, at least about 0.97, at least about 0.98, at least about 0.99, or more up to about 1.0.

As a result of the optimising modifications, a polynucleotide of the invention may increase the expression of the encoded 2019-nCoV spike protein or fragment thereof by at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 100% or more compared with the corresponding non-optimised polynucleotide sequence. Preferably the expression level is increased by at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 100% or more, more preferably at least at least 70%, at least 80%, at least 90%, at least 100% or more compared with the corresponding non-optimised polynucleotide.

A polynucleotide of the invention may be capable of expression in the host cell for at least one week, at least two weeks, at least three weeks, at least one month, at least two months, at least three months, at least four months or more, preferably at least one month, at least two months, at least three months, at least four months or more.

The inventors have demonstrated that 2019-nCoV spike protein and fusion proteins comprising 2019-nCoV spike protein can be expressed at high levels in a variety of expression systems/host cells using their rationally designed optimised polynucleotides. Furthermore, the

present inventions have surprisingly demonstrated that 2019-nCoV spike protein and fusion proteins comprising 2019-nCoV spike protein can generate a strong antibody response in mice, which demonstrates their potential therapeutic utility. The inventors have exemplified their optimisation methodology by designing and generating optimised polynucleotides and fusion proteins as described in the Examples below.

Accordingly, a polynucleotide of the invention may comprise or consist of a nucleic acid sequence having at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity to any one of SEQ ID NOs: 2, 3, 4, 5, 6, 7 or 8. Preferably a polynucleotide of the invention may comprise or consist of a nucleic acid sequence having at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity to any one of SEQ ID NOs: 2, 3, 4, 5, 6, 7, 8, 13, 14, 26, 27, 29,30, or 32. More preferably, a polynucleotide of the invention may comprise or consist of a nucleic acid sequence having at least 98%, at least 99% or more identity to any one of SEQ ID NOs: 2, 3, 4, 5, 6, 7, 8, 13, 14, 26, 27, 29,30, or 32. A polynucleotide of the invention may comprise or consist of the nucleic acid sequence of any one of SEQ ID NOs: 2, 3, 4, 5, 6, 7, 8, 13, 14, 26, 27, 29, 30, or 32. In addition, the 5' cloning site, the 3' cloning site, or the 5' and 3' cloning sites identified in any of SEQ ID NOs: 2, 3, 4, 5, 6, 7, 8, 13, 14, 26, 27, 29, 30, or 32, or any variant thereof as described herein, may be deleted. Thus, the invention provides polynucleotides comprising or consisting of any one of SEQ ID NOs: 2, 3, 4, 5, 6, 7, 8, 13, 14, 26, 27, 29,30, or 32 but lacking the 5' cloning site, the 3' cloning site, or the 5' and 3' cloning sites identified in any of SEQ ID NOs: 2, 3, 4, 5, 6, 7, 8, 13, 14, 26, 27, 29,30, or 32. Alternatively, the 5' cloning site, the 3' cloning site, or the 5' and 3' cloning sites identified in any of SEQ ID NOs: 2, 3, 4, 5, 6, 7, 8, 13, 14, 26, 27, 29, 30, or 32, or any variant thereof as described herein, may be independently replaced with another appropriate cloning site. Suitable alternative cloning sites are well known in the art.

The invention particularly provides polynucleotides encoding an RBD of the 2019-nCoV spike protein. Accordingly, a polynucleotide of the invention may comprise or consist of a nucleic acid sequence having at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity to SEQ ID NO: 13, or to the codon-optimised sequence of SEQ ID NO: 14. Preferably a polynucleotide of the invention may comprise or consist of a nucleic acid sequence having at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity to SEQ ID NO: 13, or to the codon-optimised sequence of SEQ ID NO: 14. More preferably, a polynucleotide of the invention may comprise or consist of a nucleic acid sequence having at least 98%, at least 99%

or more identity to SEQ ID NO: 13, or to the codon-optimised sequence of SEQ ID NO: 14. A polynucleotide of the invention may comprise or consist of the nucleic acid sequence of SEQ ID NO: 13, or the codon-optimised sequence of SEQ ID NO: 14.

5 A polynucleotide of the invention typically encodes a 2019-nCoV spike protein, or an immunogenic fragment thereof which: (a) retains the conformational epitopes present in the native 2019-nCoV spike protein; and/or (b) results in the production of neutralising antibodies specific for the spike protein or fragment thereof when the nucleic acid or the encoded spike protein or fragment thereof is administered to a subject.

10 The polynucleotide of the invention typically expresses a spike protein from 2019-nCoV having at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity with SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein. Preferably a polynucleotide of the invention expresses a spike protein from 2019-nCoV having at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity with SEQ  
15 ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein. More preferably, a polynucleotide of the invention expresses a spike protein from 2019-nCoV having least 98%, at least 99% or more with SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein. A polynucleotide of the invention may express a spike protein from 2019-nCoV comprising or consisting of SEQ ID  
20 NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein.

A polynucleotide of the invention may be comprised in an expression construct to facilitate expression of the 2019-nCoV spike protein or fragment thereof. Accordingly, the invention further provides an expression construct comprising polynucleotide the invention.  
25 Typically, in such an expression construct a polynucleotide of the invention is operably linked to a suitable promoter. The polynucleotide may be linked to a suitable terminator sequence. The polynucleotide may be linked to both a promoter and terminator. Suitable promoter and terminator sequences are well known in the art.

The choice of promoter will depend on where the ultimate expression of the  
30 polynucleotide will take place. In general, constitutive promoters are preferred, but inducible promoters may likewise be used. The construct produced in this manner includes at least one part of a vector, in particular regulatory elements. The vector is preferably capable of expressing the nucleic acid in a given host cell. Any appropriate host cell may be used, such

as mammalian, bacterial, insect, yeast, and/or plant host cells. In addition, cell-free expression systems may be used. Such expression systems and host cells are standard in the art.

The 2019-nCoV spike protein or immunogenic fragment thereof encoded or expressed (the two terms are used interchangeably herein) by a polynucleotide of the invention typically  
5 retain the same binding affinity for its receptor as the native 2019-nCoV spike protein. In the context of the present invention, this may mean having a binding affinity for the 2019-nCoV spike protein receptor of at least 80%, at least 85%, at least 90%, at least 95%, at least 99% or more of that of the native 2019-nCoV spike protein. Preferably the 2019-nCoV spike protein or immunogenic fragment thereof expressed by a polynucleotide of the invention have a  
10 binding affinity for the 2019-nCoV spike protein of at least 90%, at least 95%, at least 99% or more of that of the native 2019-nCoV spike protein.

In some embodiments, the 2019-nCoV spike protein or immunogenic fragment thereof expressed by a polynucleotide of the invention have a binding affinity for the 2019-nCoV spike protein receptor greater than that of the full-length protein. For example, the 2019-nCoV spike  
15 protein or immunogenic fragment thereof expressed by a polynucleotide of the invention of the invention may have a binding affinity of at least 100%, at least 110%, at least 120%, or at least 150% or more of that of the native 2019-nCoV spike protein.

In other embodiments, the 2019-nCoV spike protein or immunogenic fragment thereof expressed by a polynucleotide of the invention may have a binding affinity for the 2019-nCoV  
20 spike protein receptor less than that of the native 2019-nCoV spike protein. For example, the 2019-nCoV spike protein or immunogenic fragment thereof expressed by a polynucleotide of the invention may have a binding affinity of less than 80%, less than 70%, less than 60%, less than 50% or less of that of the native 2019-nCoV spike protein.

The binding affinity of a 2019-nCoV spike protein or immunogenic fragment thereof  
25 expressed by a polynucleotide of the invention for its receptor may be quantified in terms of dissociation constant ( $K_d$ ).  $K_d$  may be determined using any appropriate technique, but SPR is generally preferred in the context of the present invention.

An immunogenic fragment of the 2019-nCoV spike protein expressed by a polynucleotide of the invention are typically greater than 200 amino acids in length. 2019-  
30 nCoV spike protein fragments of the present invention may comprise or consist of at least 200, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1100, or more amino acid residues in length. The fragments of the invention have a common antigenic cross-reactivity with the 2019-nCoV spike protein. In some preferred embodiments, the immunogenic fragment of the 2019-nCoV spike protein expressed

by a polynucleotide of the invention is an RBD of the 2019-nCoV spike protein as defined herein, preferably wherein said RBD has at least 90% identity with SEQ ID NO: 15.

The 2019-nCoV spike protein or immunogenic fragment thereof expressed by a polynucleotide of the invention may additionally comprise a leader sequence, for example to assist in the recombinant production and/or secretion of the 2019-nCoV spike protein or immunogenic fragment thereof. Any suitable leader sequence may be used, including conventional leader sequences known in the art. Suitable leader sequences include Bip leader sequences, which are commonly used in the art to aid secretion from insect cells and human tissue plasminogen activator leader sequence (tPA), which is routinely used in viral and DNA based vaccines and for protein vaccines to aid secretion from mammalian cell expression platforms.

The 2019-nCoV spike protein or immunogenic fragment thereof expressed by a polynucleotide of the invention may additionally comprise an N- or C-terminal tag, for example to assist in the recombinant production and/or purification of the 2019-nCoV spike protein or immunogenic fragment thereof. Any N- or C-terminal tag may be used, including conventional tags known in the art. Suitable tags sequences include C-terminal hexa-histidine tags and the “C-tag” (the four amino acids EPEA at the C-terminus), which are commonly used in the art to aid purification from heterologous expression systems, e.g. insect cells, mammalian cells, bacteria, or yeast. In other embodiments, the 2019-nCoV spike protein or immunogenic fragment thereof expressed by a polynucleotide of the invention are purified from heterologous expression systems without the need to use a purification tag.

The 2019-nCoV spike protein or immunogenic fragment thereof expressed by a polynucleotide of the invention may comprise a leader sequence and/or a tag as defined herein.

## **25 Viral Vectors, DNA Plasmids and RNA Vaccines**

The present invention also provides a vector: (a) comprising a polynucleotide of the invention; and/or (b) encoding a 2019-nCoV spike protein or immunogenic fragment thereof of the invention. The vector(s) may be present in the form of a vaccine composition or formulation.

30 The vector of the invention typically expresses a spike protein from 2019-nCoV having at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity with SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein. Preferably a vector of the invention expresses a spike protein from 2019-nCoV having at least 90%, at least 95%, at least

96%, at least 97%, at least 98%, at least 99% or more identity with SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein. More preferably, a vector of the invention expresses a spike protein from 2019-nCoV having least 98%, at least 99% or more with SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein. A vector of the invention may express a spike protein from 2019-nCoV comprising or consisting of SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein. In some preferred embodiments, the immunogenic fragment of the 2019-nCoV spike protein expressed by a vector of the invention is an RBD of the 2019-nCoV spike protein as defined herein, preferably wherein said RBD has at least 90% identity with SEQ ID NO: 15.

The vector of the invention may express a spike protein or immunogenic fragment thereof as defined herein which further comprises a signal peptide. Typically said signal peptide directs secretion of the 2019-nCoV spike protein or fragment thereof from a host cell of interest, such as a human cell, an *E. coli* cell or a yeast cell.

The vector of the invention may further express one or more additional antigen or a fragment thereof. The spike protein or fragment thereof and the one or more additional antigen or fragment thereof may be expressed as a fusion protein. Alternatively, separate vectors expressing the 2019-nCoV spike protein or fragment thereof and the one or more additional antigen or fragment thereof may be used. In such instances, said separate vectors may be used in combination, either sequentially or simultaneously. The one or more additional antigen may be the same antigen or a different antigen from 2019-nCoV, or a fragment thereof. More preferably, said one or more additional antigen is a different antigen from 2019-nCoV, such as an antigen from the 2019-CoV membrane protein or envelope protein.

The vector(s) of the invention may comprise any polynucleotide or expression construct as defined herein, or any combination thereof.

The vector(s) may be a viral vector. Such a viral vector may be an adenovirus (of a human serotype such as AdHu5, a simian serotype such as ChAd63, ChAdOX1 or ChAdOX2, or another form), an adeno-associated virus (AAV), or a poxvirus vector (such as a modified vaccinia Ankara (MVA)), or an adeno associated virus (AAV). ChAdOX1 and ChAdOX2 are disclosed in WO2012/172277 (herein incorporated by reference in its entirety). ChAdOX2 is a BAC-derived and E4 modified AdC68-based viral vector. Preferably said viral vector is an AAV vector adenovirus.

Viral vectors are usually non-replicating or replication impaired vectors, which means that the viral vector cannot replicate to any significant extent in normal cells (e.g. normal

human cells), as measured by conventional means – e.g. via measuring DNA synthesis and/or viral titre. Non-replicating or replication impaired vectors may have become so naturally (i.e. they have been isolated as such from nature) or artificially (e.g. by breeding in vitro or by genetic manipulation). There will generally be at least one cell-type in which the replication-impaired viral vector can be grown – for example, modified vaccinia Ankara (MVA) can be grown in CEF cells. By way of non-limiting example, the vector may be selected from a human or simian adenovirus or a poxvirus vector.

Typically, the viral vector is incapable of causing a significant infection in an animal subject, typically in a mammalian subject such as a human or other primate.

The vector(s) may be a DNA vector, such as a DNA plasmid. The vector(s) may be an RNA vector, such as a mRNA vector or a self-amplifying RNA vector. The DNA and/or RNA vector(s) of the invention may be capable of expression in eukaryotic and/or prokaryotic cells, particularly any host cell type described herein, or in a subject to be treated.

Typically the DNA and/or RNA vector(s) are capable of expression in a human, *E. coli* or yeast cell.

The present invention may be a phage vector, such as an AAV/phage hybrid vector as described in Hajitou et al., Cell 2006; 125(2) pp. 385-398; herein incorporated by reference.

The nucleic acid molecules of the invention may be made using any suitable process known in the art. Thus, the nucleic acid molecules may be made using chemical synthesis techniques. Alternatively, the nucleic acid molecules of the invention may be made using molecular biology techniques.

Vector(s) of the present invention may be designed *in silico*, and then synthesised by conventional polynucleotide synthesis techniques.

## **25 Virus-Like Particles**

Virus-like particles (VLPs) are particles which resemble viruses but do not contain viral nucleic acid and are therefore non-infectious. They commonly contain one or more virus capsid or envelope proteins which are capable of self-assembly to form the VLP. VLPs have been produced from components of a wide variety of virus families (Noad and Roy (2003), Trends in Microbiology, 11:438-444; Grgacic et al., (2006), Methods, 40:60-65). Some VLPs have been approved as therapeutic vaccines, for example Engerix-B (for hepatitis B), Cervarix and Gardasil (for human papilloma viruses).

Accordingly, the invention provides a VLP comprising a 2019-nCoV spike protein or immunogenic fragment thereof of the invention. A VLP of the invention typically comprises

a spike protein from 2019-nCoV having at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity with SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein. Preferably a VLP of the invention comprises a spike protein from 2019-nCoV having  
5 at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity with SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein. More preferably, a VLP of the invention comprises a spike protein from 2019-nCoV having least 98%, at least 99% or more with SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein. A VLP of the invention  
10 may comprise a spike protein from 2019-nCoV comprising or consisting of SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein. In some preferred embodiments, the immunogenic fragment of the 2019-nCoV spike protein comprised in a VLP of the invention is an RBD of the 2019-nCoV spike protein as defined herein, preferably wherein said RBD has at least 90% identity with SEQ ID NO: 15.

15 The skilled person will understand that VLPs can be synthesized through the individual expression of viral structural proteins, which can then self-assemble into the virus-like structure. Combinations of structural capsid proteins from different viruses can be used to create recombinant VLPs. In additions, antigens or immunogenic fragments thereof can be fused to the surface of VLPs. By way of non-limiting example, antigens or immunogenic  
20 fragments thereof of the invention may be coupled to a VLP using the SpyCatcher-SpyTag system (as described by Brune, Biswas, Howarth).

A VLP of the invention may comprise one or more additional protein antigen. The one or more additional antigen may be the same antigen or a different antigen from 2019-nCoV, or a fragment thereof. More preferably, said one or more additional antigen is a different antigen  
25 from 2019-nCoV, such as an antigen from the 2019-CoV membrane protein or envelope protein.

A VLP of the invention may comprise a fusion protein as described herein. A VLP of the invention may comprise a fusion protein of the 2019-nCoV spike protein or immunogenic fragment thereof with Hepatitis B surface antigen (HBsAg), human papillomavirus (HPV) 18  
30 L1 protein, HPV 16 L1 protein and/or Hepatitis E P239, preferably Hepatitis B surface antigen. Although these other viral proteins have been described in fusion proteins previous, to-date, there are no reports of fusion proteins being successfully generated comprising proteins of the size of 2019-nCoV spike protein. Furthermore, there are known limitations regarding the choice of expression systems for such fusion proteins. The present inventors have surprisingly

demonstrated that VLPs/fusion proteins comprising 2019-nCoV spike protein can be produced recombinantly in *E. coli*, yeast and human cells, and that these VLPs/fusion proteins can elicit an (immunoprotective) antibody response in animal models.

Thus, a VLP of the invention may be encoded by a polynucleotide which comprises or  
5 consists of a nucleic acid sequence having at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity to any one of SEQ ID NOs: 3, 5, 6 or 8. Preferably a VLP of the invention may be encoded by a polynucleotide which comprises or consists of a nucleic acid sequence having at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity to any one of  
10 SEQ ID NOs: 3, 5, 6 or 8. More preferably, a VLP of the invention may be encoded by a polynucleotide which comprises or consists of a nucleic acid sequence having at least 98%, at least 99% or more identity to any one of SEQ ID NOs: 3, 5, 6 or 8. A VLP of the invention may be encoded by a polynucleotide which comprises or consists of a nucleic acid sequence of any one of SEQ ID NOs: 3, 5, 6 or 8.

15 A VLP of the invention may be encoded by a polynucleotide which comprises or consists of a nucleic acid sequence having at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity to any one of SEQ ID NOs: 26, 27, 29,30, or 32. Preferably a VLP of the invention may be encoded by a polynucleotide which comprises or consists of a nucleic acid sequence having at least  
20 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity to any one of SEQ ID NOs: 26, 27, 29, 30, or 32. More preferably, a VLP of the invention may be encoded by a polynucleotide which comprises or consists of a nucleic acid sequence having at least 98%, at least 99% or more identity to any one of SEQ ID NOs: 26, 27, 29, 30, or 32. A VLP of the invention may be encoded by a polynucleotide which comprises or consists of the  
25 nucleic acid sequence of any one of SEQ ID NOs: 26, 27, 29, 30, or 32.

A VLP of the invention may comprise or consist of an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity to any one of SEQ ID NO: 9, 10, 11 or 12. Preferably a VLP of the invention may comprise or consist of an amino acid sequence having at least 90%,  
30 at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity to any one of SEQ ID NOs: 9, 10, 11 or 12. More preferably, a VLP of the invention may comprises or consists of an amino acid sequence having at least 98%, at least 99% or more identity to any one of SEQ ID NOs: 9, 10, 11 or 12. A VLP of the invention may comprise or consist of an amino acid sequence of any one of SEQ ID NOs: 9, 10, 11 or 12.

A VLP of the invention may comprise or consist of an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity to any one of SEQ ID NOs: 28, 31, or 33. Preferably a VLP of the invention may comprise or consist of an amino acid sequence having at least 90%,  
5 at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity to any one of SEQ ID NOs: 28, 31, or 33. More preferably, a VLP of the invention may comprises or consists of an amino acid sequence having at least 98%, at least 99% or more identity to any one of SEQ ID NOs: 28, 31, or 33. A VLP of the invention may comprise or consist of an amino acid sequence of any one of SEQ ID NOs: 28, 31, or 33.

10 The use of VLP may increase the efficacy of the immunoprotective response induced by the 2019-nCoV spike protein or immunogenic fragment and/or may increase the duration of the immunoprotective response as defined herein.

### **Fusion Proteins**

15 The invention further provides a fusion protein comprising a 2019-nCoV spike protein or immunogenic fragment thereof of the invention. A fusion protein of the invention typically comprises a spike protein from 2019-nCoV having at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity with SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with  
20 said spike protein. Preferably a fusion protein of the invention comprises a spike protein from 2019-nCoV having at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity with SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein. More preferably, a fusion protein of the invention comprises a spike protein from 2019-nCoV having least 98%, at least 99% or more with SEQ  
25 ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein. A fusion protein of the invention may comprise a spike protein from 2019-nCoV comprising or consisting of SEQ ID NO: 1, or a fragment thereof, that has a common antigenic cross-reactivity with said spike protein. In some preferred embodiments, the immunogenic fragment of the 2019-nCoV spike protein comprised in a fusion protein of the invention is an  
30 RBD of the 2019-nCoV spike protein as defined herein, preferably wherein said RBD has at least 90% identity with SEQ ID NO: 15.

A fusion protein of the invention typically also comprises a non-2019-nCoV domain or element, typically a non-2019-nCoV protein, polypeptide or peptide domain or element. A fusion protein of the invention may comprise the 2019-nCoV spike protein or immunogenic

fragment thereof and one or more of: Hepatitis B surface antigen (HBsAg); human papillomavirus (HPV) 18 L1 protein; HPV 16 L1 protein; and/or Hepatitis E P239 (HEV), preferably Hepatitis B surface antigen. As described above in the context of VLPs, the present inventors have surprisingly demonstrated that fusion proteins comprising 2019-nCoV spike  
5 protein can be produced recombinantly in *E. coli*, yeast and human cells, and that these fusion proteins can elicit an (immunoprotective) antibody response in animal models.

A fusion protein of the invention may be encoded by a polynucleotide which comprises or consists of a nucleic acid sequence having at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity to any  
10 one of SEQ ID NO: 3, 5, 6 or 8. Preferably a fusion protein of the invention may be encoded by a polynucleotide which comprises or consists of a nucleic acid sequence having at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity to any one of SEQ ID NOs: 3, 5, 6 or 8. More preferably, a fusion protein of the invention may be encoded by a polynucleotide which comprises or consists of a nucleic acid sequence having at  
15 least 98%, at least 99% or more identity to any one of SEQ ID NOs: 3, 5, 6 or 8. A VLP of the invention may be encoded by a polynucleotide which comprises or consists of a nucleic acid sequence of any one of SEQ ID NOs: 3, 5, 6 or 8.

A fusion protein of the invention may be encoded by a polynucleotide which comprises or consists of a nucleic acid sequence having at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity to any  
20 one of SEQ ID NOs: 26, 27, 29, 30, or 32. Preferably a fusion protein of the invention may be encoded by a polynucleotide which comprises or consists of a nucleic acid sequence having at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity to any one of SEQ ID NOs: 26, 27, 29, 30, or 32. More preferably, a fusion protein of the  
25 invention may be encoded by a polynucleotide which comprises or consists of a nucleic acid sequence having at least 98%, at least 99% or more identity to any one of SEQ ID NOs: 26, 27, 29, 30, or 32. A fusion protein of the invention may be encoded by a polynucleotide which comprises or consists of the nucleic acid sequence of any one of SEQ ID NOs: 26, 27, 29, 30, or 32.

30 A fusion protein of the invention may comprise or consist of an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity to any one of SEQ ID NO: 9, 10, 11 or 12. Preferably a fusion protein of the invention may comprise or consist of an amino acid sequence having at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more

identity to any one of SEQ ID NOs: 9, 10, 11 or 12. More preferably, a fusion protein of the invention may comprises or consists of an amino acid sequence having at least 98%, at least 99% or more identity to any one of SEQ ID NOs: 9, 10, 11 or 12. A fusion protein of the invention may comprise or consist of an amino acid sequence of any one of SEQ ID NOs: 9, 10, 11 or 12.

A fusion protein of the invention may comprise or consist of an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity to any one of SEQ ID NOs: 28, 31, or 33. Preferably a fusion protein of the invention may comprise or consist of an amino acid sequence having at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more identity to any one of SEQ ID NOs: 28, 31, or 33. More preferably, a fusion protein of the invention may comprise or consist of an amino acid sequence having at least 98%, at least 99% or more identity to any one of SEQ ID NOs: 28, 31, or 33. A fusion protein of the invention may comprise or consist of an amino acid sequence of any one of SEQ ID NOs: 28, 31, or 33.

A fusion protein of the invention may comprise a linker (also referred to interchangeably herein as a linker peptide, a spacer or a spacer peptide). A linker may be used to join two or more functional domains of a fusion protein of the invention. Typically, where a linker is present, it is used to join the 2019-nCoV spike protein or immunogenic fragment thereof domain of the fusion protein to the non-2019-nCoV spike protein domain of the fusion protein. Use of linkers in fusion proteins is routine in the art, and any conventional linker protein may be used in fusion proteins of the invention, provided that the resulting fusion protein retains the desired functional properties of the 2019-nCoV spike protein or immunogenic fragment thereof and the desired function properties of the non-2019-nCoV spike protein domain.

A linker may be a short peptide of up to about 30 amino acids, such as about 5-30 amino acids, about 5-25 amino acids, about 5-20 amino acids, about 10-20 amino acids, about 5-15 amino acids or about 10-15 amino acids in length. In some embodiments, the linker is about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19 or about 20 amino acids in length.

In some embodiments a rigid linker may be used in fusion proteins of the invention. Rigid linkers are conventionally used when it is necessary to keep a fixed distance between the different domains/portions of a fusion protein and to maintain their independent functions. Rigid linkers may also be used when the spatial separation of the fusion protein domains is critical to preserve the stability or bioactivity of the fusion proteins. An empirical rigid linker

with the sequence of A(EAAAK)<sub>n</sub>A (n = 2-5) (SEQ ID NO: 16) displayed  $\alpha$ -helical conformation, which is stabilized by Glu<sup>-</sup>-Lys<sup>+</sup> salt bridges. The length of linker may be adjusted by changing the copy number to achieve an optimal distance between domains. Helical linkers can also improve fusion protein folding and stability.

5           A non-limiting example of a rigid linker is EAAAKEAAKEAAAK (also referred to as (EAAAK)<sub>3</sub>, SEQ ID NO: 18), which may be encoded by the nucleic acid sequence (SEQ ID NO: 17). The length and structure of the linkers, controlling the distance between functional domains, affect the stability of the fusion protein. The stability and activity of the RBD fusion proteins of the invention are typically significantly improved after (EAAAK)<sub>3</sub> linker insertion,  
10 as shown in the Examples herein. Use of linkers, particularly (EAAAK)<sub>3</sub>, may also enhance yield of the recombinant fusion proteins of the invention.

Accordingly, rigid linkers, particularly (EAAAK)<sub>3</sub> (SEQ ID NO: 18), may be preferably used for expression of fusion proteins of the invention in mammalian cells, such as HEK 293 cells.

15           In some embodiments, flexible linkers may be used in fusion proteins of the invention. Flexible linkers are conventionally used when the joined domains require a certain degree of movement or interaction. Flexible linkers usually comprise or consist of small amino acid residues, such as glycine, threonine, arginine, serine, asparagine, glutamine, alanine, aspartic acid, proline, glutamic acid, lysine, leucine and/or valine, particularly glycine, serine, alanine,  
20 leucine and/or valine. Flexible linkers comprising or consisting of glycine, serine and/or alanine are preferred, with glycine and serine being particularly preferred. Accordingly, the most commonly used flexible linkers have sequences consisting primarily of stretches of Gly and Ser residues ("GS" linker), which comprise a sequence of (Gly-Gly-Gly-Gly-Ser)<sub>n</sub> (SEQ ID NO: 19). Non-limiting examples of GS linkers include GS5 or (GGGGS)<sub>1</sub> (SEQ ID NO:  
25 20); GS10 or (GGGGS)<sub>2</sub> (SEQ ID NO: 21); GS15 or (GGGGS)<sub>3</sub> (SEQ ID NO: 23); GS20 or (GGGGS)<sub>4</sub> (SEQ ID NO: 24); and GS25 or (GGGGS)<sub>5</sub> (SEQ ID NO: 25). Preferably, GS15 may be used, which may be encoded by (SEQ ID NO: 22). Flexible linkers may be preferably used for expression of fusion proteins of the invention in bacterial cells, such as *E. coli* cells.

Any appropriate linker, such as the exemplary linkers described herein may be used  
30 with any fusion protein of the invention (comprising any 2019-nCoV spike protein or immunogenic fragment domain and any non-2019-nCoV spike protein domain). By way of non-limiting example, a fusion protein of the invention may comprise or consist of HBSAg-(EAAAK)<sub>3</sub>-RBD (SEQ ID NO: 28), or a variant with at least 90% sequence identity thereto, which may be encoded by SEQ ID NO: 26 or 27, or a variant with at least 90% sequence

identity thereto. By way of a further non-limiting example, a fusion protein of the invention may comprise or consist of HBSAg-(EAAAK)<sub>3</sub>-full-length 2019-nCoV spike protein (SEQ ID NO: 33), or a variant with at least 90% sequence identity thereto, which may be encoded by SEQ ID NO: 32, or a variant with at least 90% sequence identity thereto. By way of further  
5 non-limiting example, a fusion protein of the invention may comprise or consist of HEV-GS15-RBD (SEQ ID NO: 31), or a variant with at least 90% sequence identity thereto, which may be encoded by (SEQ ID NO: 29 or 30), or a variant with at least 90% sequence identity thereto.

A fusion protein of the invention may preferably take the form of a VLP. Without being bound by theory, this is because HBSAg, HPV 18 L1 protein, HPB 16 L1 protein and Hepatitis  
10 E P239 protein are known to spontaneously form VLPs when expressed recombinantly, and this structure is retained when HBSAg, HPV 18 L1 protein, HPB 16 L1 protein and/or Hepatitis E P239 protein are present in fusion protein form combined with a 2019-nCoV spike protein of the invention (or immunogenic fragment thereof).

## 15 **Antibodies**

As described herein, the 2019-nCoV spike protein or fragment thereof encoded by a polynucleotide of the invention elicit the production of antibodies which specifically bind to the one or more conformational epitope found in native 2019-nCoV. Said antibodies are typically neutralising antibodies (nAb) as discussed below. These nAb are capable of  
20 mediating an immunoprotective effect against 2019-nCoV.

The term "antibody", as used herein, broadly refers to any immunoglobulin (Ig) molecule comprised of four polypeptide chains, two heavy (H) chains and two light (L) chains, or any functional fragment, mutant, variant, or derivation thereof, which retains the essential epitope binding features of an Ig molecule. Such mutant, variant, or derivative antibody entities  
25 are known in the art, non-limiting embodiments of which are discussed below.

In a full-length antibody, each heavy chain is comprised of a heavy chain variable region (abbreviated herein as VH) and a heavy chain constant region. The heavy chain constant region is comprised of three domains, CH1, CH2 and CH3. Each light chain is comprised of a light chain variable region (abbreviated herein as VL) and a light chain constant region. The  
30 light chain constant region is comprised of one domain, CL. The VH and VL regions can be further subdivided into regions of hypervariability, termed complementarity determining regions (CDR), interspersed with regions that are more conserved, termed framework regions (FR). Each VH and VL is composed of three CDRs and four FRs, arranged from amino-terminus to carboxy-terminus in the following order: FR1, CDR1, FR2, CDR2, FR3, CDR3,

FR4. Antibodies may be polyclonal (pAb) or monoclonal (mAb). When used therapeutically (i.e. to provide passive immunity), the administration of mAbs is preferred.

According to the invention, antibodies can be of any type (e.g., IgG, IgE, IgM, IgD, IgA and IgY), class (e.g., IgG1, IgG2, IgG3, IgG4, IgA1 and IgA2) or subclass and may be  
5 from any species (e.g., mouse, human, chicken, rat, rabbit, sheep, shark and camelid).

The term "antigen-binding fragment" of an antibody (or simply "binding fragment"), as used herein, refers to one or more fragments of an antibody that retain the ability to specifically bind to an antigen. It has been shown that the antigen-binding function of an antibody can be performed by one or more fragments of a full-length antibody. Single chain antibodies are also  
10 encompassed. Such antigen-binding fragments may also be bispecific, dual specific, or multi-specific, specifically binding to two or more different antigens. Thus, examples of binding fragments encompassed within the term "antigen-binding fragment" of an antibody include Fab, Fv, scFv, dAb, Fd, Fab' or F(ab')<sub>2</sub>, tandem scFv and diabodies.

Also encompassed are antibody constructs, defined as a polypeptide comprising one or  
15 more the antigen binding fragment of the invention linked to a linker polypeptide or an immunoglobulin constant domain. Linker polypeptides comprise two or more amino acid residues joined by peptide bonds and are used to link one or more antigen binding portions.

The term "antibody" as used herein may be a human antibody; defined as an antibody having variable and constant regions derived from human germline immunoglobulin  
20 sequences, but which may include amino acid residues not encoded by human germline immunoglobulin sequences (e.g., mutations introduced by random or site-specific mutagenesis in vitro or by somatic mutation in vivo), for example in the CDRs and in particular CDR3. Recombinant human antibodies are also encompassed.

An antibody of the invention may be a "chimeric antibody"; defined as an antibody  
25 which comprises heavy and light chain variable region sequences from one species and constant region sequences from another species. The present invention encompasses chimeric antibodies having, for example, murine heavy and light chain variable regions linked to human constant regions.

An antibody of the invention may be a "CDR-grafted antibody"; defined as an antibody  
30 which comprise heavy and light chain variable region sequences from one species but in which the sequences of one or more of the CDR regions of VH and/or VL are replaced with CDR sequences of another species, such as antibodies having murine heavy and light chain variable regions in which one or more of the murine CDRs (e.g., CDR3 or all three CDRs) has been replaced with human CDR sequences.

An antibody of the invention may be a "humanized antibody"; defined as an antibody which comprise heavy and light chain variable region sequences from a non-human species (e.g., a mouse) but in which at least a portion of the VH and/or VL sequence has been altered to be more "human-like", i.e., more similar to human germline variable sequences. One type  
5 of humanized antibody is a CDR-grafted antibody, in which human CDR sequences are introduced into non-human VH and VL sequences to replace the corresponding nonhuman CDR sequences.

The terms "Kabat numbering", "Kabat definitions" and "Kabat labelling" are used interchangeably herein. These terms, which are recognized in the art, refer to a system of  
10 numbering amino acid residues which are more variable (i.e. hypervariable) than other amino acid residues in the heavy and light chain variable regions of an antibody, or an antigen binding portion thereof (Kabat et al. (1971) Ann. NY Acad. Sci. 190:382-391 and Kabat, E.A., et al. (1991) Sequences of Proteins of Immunological Interest, Fifth Edition, U.S. Department of Health and Human Services, NIH Publication No. 91-3242).

15 Antibodies of the invention are not limited to a particular method of generation or production. Thus, the invention provides antibodies which have been manufactured from a hybridoma that secretes the antibody, as well as antibodies produced from a recombinantly produced cell that has been transformed or transfected with a polynucleotide or polynucleotides encoding the antibody. Such hybridomas, recombinantly produced cells, and polynucleotides  
20 form part of the invention.

An antibody, or antigen-binding fragment thereof, of the invention is selective or specific for the 2019-nCoV spike protein, or a particular epitope (preferably a conformational epitope) of the 2019-nCoV spike protein as described herein. By specific, it will be understood that the antibody binds to the molecule of interest, in this case the 2019-nCoV spike protein or  
25 fragment thereof, with no significant cross-reactivity to any other molecule, particularly any other protein. For example, a binding compound or antibody of the invention that is specific for a particular 2019-nCoV spike protein epitope of the invention will show no significant cross-reactivity with other 2019-nCoV spike protein epitopes. As another example, a binding compound or antibody of the invention that is specific for the 2019-nCoV spike protein will  
30 show no significant cross-reactivity with the 2019-nCoV membrane protein. Cross-reactivity may be assessed by any suitable method. Cross-reactivity of a binding compound (e.g. antibody) for a 2019-nCoV spike protein epitope with another 2019-nCoV spike protein epitope or a protein other than 2019-nCoV spike protein may be considered significant if the binding compound (e.g. antibody) binds to the other molecule at least 5%, 10%, 15%, 20%,

25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% or 100% as strongly as it binds to the 2019-nCoV spike protein epitope. A binding compound (e.g. antibody) that is specific for the 2019-nCoV spike protein or fragment thereof may bind to another molecule such as 2019-nCoV membrane protein at less than 90%, 85%, 80%, 75%,  
5 70%, 65%, 60%, 55%, 50%, 45%, 40%, 35%, 30%, 25% or 20% the strength that it binds to the 2019-nCoV spike protein epitope. Preferably, the binding compound (e.g. antibody) binds to the other molecule at less than 20%, less than 15%, less than 10% or less than 5%, less than 2% or less than 1% the strength that it binds to the 2019-nCoV spike protein epitope. Binding affinity may be quantified in any suitable way, e.g. by  $K_D$ .

10 The binding affinity of a 2019-nCoV spike protein antibody (preferably neutralising) of the invention for 2019-nCoV spike protein may be quantified in terms of dissociation constant ( $K_D$ ).  $K_D$  may be determined using any appropriate technique, but SPR is generally preferred in the context of the present invention. A 2019-nCoV spike protein antibody of the invention may bind to 2019-nCoV spike protein with a  $K_D$  of less than 1 $\mu$ M, less than 100nM,  
15 less than 50nM, less than 25nM, less than 10nM, less than 1nM, less than 900pM, less than 800pM, less than 700pM, less than 600pM, less than 500pM, less than 400pM, less than 300pM, less than 200pM, less than 100pM, less than 50pM, less than 25pM, less than 10pM, less than 5pM, or less. Typically a 2019-nCoV spike protein antibody of the invention binds to 2019-nCoV spike protein with a  $K_D$  of less 50nM, less than 10nM or less than 1nM.

20 Further antibodies which bind to the epitopes/antigens of the invention may be generated by producing variants of the antibodies of the invention. Such variants may have CDRs sharing a high level of identity with the CDRs of an antibody of the invention, for example may have CDRs each of which independently may differ by one or two amino acids from the antibody of the invention from which the variant antibody is derived, and wherein the  
25 variant retains the binding and functional properties of the antibody of the invention. Additionally, such antibodies may have one or more variations (e.g. a conservative amino acid substitution) in the framework regions. The variations in the amino acid sequences of the antibodies of the invention should maintain at least 75%, at least 80%, at least 85%, at least 90%, at least 95% and up to 99% sequence identity. Variants having at least 90% sequence  
30 identity with the antibodies of the invention, particularly the specific antibodies exemplified herein are specifically contemplated. Variation may or may not be limited to the framework regions and not present in the CDRs.

The term “neutralising antibody” is defined herein to mean an antibody which by itself (in the absence of any other 2019-nCoV spike protein antibody or other antibody against

another 2019-nCoV protein) has the ability to affect the function of the spike protein to which it binds. In particular, neutralising antibodies reduce the ability of 2019-nCoV viral particles expressing the spike protein from infecting a cell by neutralising or inhibiting the biological activity of the spike protein.

5           This neutralising activity may be quantified using any appropriate technique and measured in any appropriate units. This disclosure applies equally to neutralising antibodies of the invention (and to other binding compounds as described herein). For example, the effectiveness of a 2019-nCoV spike protein or immunogenic fragment thereof may be given in terms of their half maximal effective concentration (EC50), antibody titre stimulated (in terms  
10 of antibody units, AU) and/or EC50 in terms of AU. The latter of these gives an indication of the quality of the antibody response stimulated by the 2019-nCoV spike protein or immunogenic fragment thereof of the invention. Any appropriate technique may be used to determine the EC50, AU or EC50/AU. Conventional techniques are known in the art.

          The amount of antibody produced may be quantified using any appropriate method,  
15 with standard techniques being known in the art. For example, the amount of antibody produced may be measured by ELISA in terms of the serum IgG response induced by the 2019-nCoV spike protein or immunogenic fragment thereof of the invention. The amount of antibody produced may be given in terms of arbitrary antibody units (AU).

          The immune response (or immunogenicity) to a 2019-nCoV spike protein or  
20 immunogenic fragment thereof of the invention, particularly the antibody response, may be given as the half-maximal effective concentration in terms of the amount of antibody produced, i.e. EC50/AU. This gives an indication of the quality of the immune response generated to the 2019-nCoV spike protein or immunogenic fragment thereof. For example, a low EC50 (i.e. effective response) but a high number of antibody units generated is less effective (and gives a  
25 higher EC50/AU) than a low EC50 with a low number of antibody units. This value thus indicates the quality of the antibody response by representing the neutralising antibody activity (measured as the EC50) as a proportion of the total amount of anti-2019-nCoV spike protein or immunogenic fragment thereof IgG antibody produced (measured by ELISA in AU). A more effective vaccine thus induces the EC50 with less antibody (lower AU).

30           Typically, the neutralising 2019-nCoV spike protein antibodies of the invention reduce the infectivity of 2019-nCoV particles by at least at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90% or more.

          A 2019-nCoV spike protein or immunogenic fragment thereof of the invention may elicit an improved immune response, particularly an improved antibody response, compared

with the native 2019-nCoV spike protein or immunogenic fragment thereof or a 2019-nCoV spike protein or immunogenic fragment thereof produced by a non-optimised polynucleotide.

Any and all disclosure herein in relation to binding compounds preferably relates to antibodies as described herein.

5           Alternatively, other binding compounds, such as DNA oligonucleotide aptamers, RNA oligonucleotide aptamers, and other engineered biopolymers against 2019-nCoV spike protein or a fragment thereof, particularly an epitope within said 2019-nCoV spike protein or fragment, may also be able to replicate the activity of the antibodies and combinations thereof described here. Said alternative binding compounds may specifically bind to the protein or immunogenic  
10       fragment thereof of the invention.

Oligonucleotide aptamers may be identified or synthesised using well-established methods. The aptamer may further be optimised to render it suitable for therapeutic use, e.g. it may be conjugated to a monoclonal antibody to modify its pharmacokinetics and/or recruit Fc-dependent immune functions.

15

### **Compositions and Therapeutic Indications**

As described herein, the present inventors have demonstrated that immunisation with 2019-nCoV spike protein expressed by polynucleotides of the invention are able to generate a robust antibody response.

20           Accordingly, the present invention provides a polynucleotide, expression construct, viral vector, DNA plasmid or RNA vaccine which expresses a 2019-nCoV spike protein or immunogenic fragment thereof, or a 2019-nCoV spike protein or immunogenic fragment thereof, or a binding compound of the invention for use as a vaccine.

The invention also provides a vaccine composition comprising said polynucleotide,  
25       expression construct, viral vector, DNA plasmid or RNA vaccine which expresses a 2019-nCoV spike protein or immunogenic fragment thereof, or a 2019-nCoV spike protein or immunogenic fragment thereof, or a binding compound. The vaccine composition may optionally comprise a pharmaceutically acceptable excipient, diluent, carrier, propellant, salt and/or additive.

30           In some embodiments the vaccine composition comprises at least two different proteins or immunogenic fragments according to the invention, and/or at least two different polynucleotide molecules according to the invention. By way of non-limiting example, the vaccine composition may comprise a polynucleotide encoding a 2019-nCoV spike protein and a polynucleotide encoding a 2019-nCoV membrane protein.

The present invention also provides a method of stimulating or inducing an immune response in a subject using a polynucleotide, expression construct, viral vector, DNA plasmid or RNA vaccine which expresses a 2019-nCoV spike protein or immunogenic fragment thereof, or a 2019-nCoV spike protein or immunogenic fragment thereof, or a binding compound of the invention (as described above).

Said method of stimulating or inducing an immune response in a subject may comprise administering a polynucleotide, expression construct, viral vector, DNA plasmid or RNA vaccine which expresses a 2019-nCoV spike protein or immunogenic fragment thereof, or a 2019-nCoV spike protein or immunogenic fragment thereof, or a binding compound of the invention (as described above) to a subject.

In the context of the therapeutic uses and methods, a “subject” is any animal subject that would benefit from stimulation or induction of an immunoprotective response against 2019-nCoV. Typical animal subjects are mammals, such as primates, for example, humans.

Thus, the present invention provides a method for treating or preventing 2019-nCoV infection. Said method typically comprises the administration of a polynucleotide, expression construct, viral vector, DNA plasmid or RNA vaccine which expresses a 2019-nCoV spike protein or immunogenic fragment thereof, or a 2019-nCoV spike protein or immunogenic fragment thereof, a vaccine composition or a binding compound of the invention to a subject in need thereof.

The present invention also provides a polynucleotide, expression construct, viral vector, DNA plasmid or RNA vaccine which expresses a 2019-nCoV spike protein or immunogenic fragment thereof, or a 2019-nCoV spike protein or immunogenic fragment thereof, a vaccine composition or a binding compound of the invention for use in prevention or treatment of 2019-nCoV infection.

The present invention also provides the use of a polynucleotide, expression construct, viral vector, DNA plasmid or RNA vaccine which expresses a 2019-nCoV spike protein or immunogenic fragment thereof, or a 2019-nCoV spike protein or immunogenic fragment thereof, a vaccine composition or a binding compound of the invention for the manufacture of a medicament for the prevention or treatment of 2019-nCoV infection.

As used herein, the term “treatment” or “treating” embraces therapeutic or preventative/prophylactic measures, and includes post-infection therapy and amelioration of a 2019-nCoV infection.

As used herein, the term “preventing” includes preventing the initiation of infection by 2019-nCoV and/or reducing the severity or intensity of an infection by 2019-nCoV. The term

“preventing” includes inducing or providing protective immunity against infection by 2019-nCoV. Immunity to infection by a 2019-nCoV may be quantified using any appropriate technique, examples of which are known in the art.

A polynucleotide, expression construct, viral vector, DNA plasmid or RNA vaccine  
5 which expresses a 2019-nCoV spike protein or immunogenic fragment thereof, or a 2019-nCoV spike protein or immunogenic fragment thereof, a vaccine composition or a binding compound defined herein may be administered to a subject (typically a mammalian subject such as a human or other primate) already having a 2019-nCoV infection, a condition or symptoms associated with infection by 2019-nCoV, to treat or prevent infection by 2019-  
10 nCoV. For example, the subject may be suspected of having come in contact with 2019-nCoV, or has had known contact with 2019-nCoV, but is not yet showing symptoms of exposure.

When administered to a subject (e.g. a mammal such as a human or other primate) that already has a 2019-nCoV infection, or is showing symptoms associated with a 2019-nCoV infection, the polynucleotide, expression construct, viral vector, DNA plasmid or RNA vaccine  
15 which expresses a 2019-nCoV spike protein or immunogenic fragment thereof, or a 2019-nCoV spike protein or immunogenic fragment thereof, a vaccine composition or a binding compound as defined can cure, delay, reduce the severity of, or ameliorate one or more symptoms, and/or prolong the survival of a subject beyond that expected in the absence of such treatment.

Alternatively, a polynucleotide, expression construct, viral vector, DNA plasmid or RNA vaccine which expresses a 2019-nCoV spike protein or immunogenic fragment thereof, or a 2019-nCoV spike protein or immunogenic fragment thereof, a vaccine composition or a binding compound as defined herein may be administered to a subject (e.g. a mammal such as a human or other primate) who ultimately may be infected with 2019-nCoV, in order to prevent,  
20 cure, delay, reduce the severity of, or ameliorate one or more symptoms of said 2019-nCoV infection, or in order to prolong the survival of a subject beyond that expected in the absence of such treatment, or to help prevent that subject from transmitting a 2019-nCoV infection.

The treatments and preventative therapies of the present invention are applicable to a variety of different subjects of different ages. In the context of humans, the therapies are  
30 applicable to children (e.g. infants, children under 5 years old, older children or teenagers) and adults. In the context of other animal subjects (e.g. mammals such as primates), the therapies are applicable to immature subjects and mature/adult subjects. As used herein, the term “preventing” includes preventing the initiation of 2019-nCoV infection and/or reducing the severity or intensity of a 2019-nCoV infection. The term “preventing” includes inducing or

providing protective immunity against 2019-nCoV infection. Immunity to 2019-nCoV infection may be quantified using any appropriate technique, examples of which are known in the art.

As used, herein, a “vaccine” is a formulation that, when administered to an animal subject such as a mammal (e.g. a human or other primate) stimulates a protective immune response against 2019-nCoV infection. The immune response may be a humoral and/or cell-mediated immune response. A vaccine of the invention can be used, for example, to protect a subject from the effects of 2019-nCoV infection.

## 10 **Pharmaceutical Compositions and Formulations**

The term “vaccine” is herein used interchangeably with the terms “therapeutic/prophylactic composition”, “formulation” or “medicament”.

The vaccine of the invention (as defined above) can be combined or administered in addition to a pharmaceutically acceptable carrier. Alternatively or in addition the vaccine of the invention can further be combined with one or more of a salt, excipient, diluent, adjuvant, immunoregulatory agent and/or antimicrobial compound.

Pharmaceutically acceptable salts include acid addition salts formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or with organic acids such as acetic, oxalic, tartaric, maleic, and the like. Salts formed with the free carboxyl groups may also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

Administration of immunogenic compositions, therapeutic formulations, medicaments and prophylactic formulations (e.g. vaccines) is generally by conventional routes e.g. intravenous, subcutaneous, intraperitoneal, or mucosal routes. The administration may be by parenteral injection, for example, a subcutaneous, intradermal or intramuscular injection. Formulations comprising neutralizing antibodies may be particularly suited to administration intravenously, intramuscularly, intradermally, or subcutaneously.

Accordingly, immunogenic compositions, therapeutic formulations, medicaments and prophylactic formulations (e.g. vaccines) of the invention are typically prepared as injectables, either as liquid solutions or suspensions. Solid forms suitable for solution in, or suspension in, liquid prior to injection may alternatively be prepared. The preparation may also be emulsified, or the peptide encapsulated in liposomes or microcapsules.

The active immunogenic ingredients (such as the 2019-nCoV spike proteins, fragments thereof, nucleic acids encoding said spike proteins, expression vectors, viral vectors, DNA plasmids, RNA vaccines, fusion proteins and vaccine compositions) are often mixed with carriers, diluents, excipients or similar which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine.

Generally, the carrier, diluent, excipient or similar is a pharmaceutically-acceptable carrier. Non-limiting examples of pharmaceutically acceptable carriers include water, saline, and phosphate-buffered saline. In some embodiments, however, the composition is in lyophilized form, in which case it may include a stabilizer, such as BSA. In some embodiments, it may be desirable to formulate the composition with a preservative, such as thiomersal or sodium azide, to facilitate long term storage.

Examples of additional adjuvants which may be effective include but are not limited to: complete Freund's adjuvant (CFA), Incomplete Freund's adjuvant (IFA), Saponin, a purified extract fraction of Saponin such as Quil A, a derivative of Saponin such as QS-21, lipid particles based on Saponin such as ISCOM/ISCOMATRIX, *E. coli* heat labile toxin (LT) mutants such as LTK63 and/ or LTK72, aluminium hydroxide, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryl oxy)-ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted from bacteria, GLA-SE (GLA is a synthetic lipid A derivative, in GLA-SE this is formulated as an oil-in-water emulsion using squalene oil), monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2 % squalene/ Tween 80 emulsion, the MF59 formulation developed by Novartis, and the AS02, AS01, AS03 and AS04 adjuvant formulations developed by GSK Biologicals (Rixensart, Belgium). Preferred adjuvants include: aluminium hydroxide and aluminium phosphate gel; aluminium hydroxide and monophosphoryl lipid A (MPL); and 5% squalene (MF59). Other preferred adjuvants include alum or aluminium hydroxide and/or aluminium phosphate, GLA-SE and/or a squalene-based adjuvant such as MF59 or AddaVax™, or any combination thereof. Preferably a combination of aluminium hydroxide and/or aluminium phosphate, GLA-SE and AddaVax™ is used.

Examples of buffering agents include, but are not limited to, sodium succinate (pH 6.5), and phosphate buffered saline (PBS; pH 6.5 and 7.5).

Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations or formulations suitable for distribution as aerosols. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1%-2%.

Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders.

## DEFINITIONS

As used herein, the term "capable of" when used with a verb, encompasses or means the action of the corresponding verb. For example, "capable of interacting" also means interacting, "capable of cleaving" also means cleaves, "capable of binding" also means binds and "capable of specifically targeting..." also means specifically targets.

The term "variant", when used in relation to a protein, means a peptide or peptide fragment of the protein that contains one or more analogues of an amino acid (e.g. an unnatural amino acid), or a substituted linkage.

The term "derivative", when used in relation to a protein, means a protein that comprises the protein in question, and a further peptide sequence. The further peptide sequence should preferably not interfere with the basic folding and thus conformational structure of the original protein. Two or more peptides (or fragments, or variants) may be joined together to form a derivative. Alternatively, a peptide (or fragment, or variant) may be joined to an unrelated molecule (e.g. a second, unrelated peptide). Derivatives may be chemically synthesized, but will be typically prepared by recombinant nucleic acid methods. Additional components such as lipid, and/or polysaccharide, and/or polypeptide components may be included.

Reference to 2019-nCoV polynucleotides and/or proteins in the present specification embraces fragments and variants thereof. Variant 2019-nCoV spike proteins retain one or more conformational epitope of native spike protein and the ability to elicit the production of neutralising antibodies and/or an immunoprotective response. Variant 2019-nCoV spike protein polynucleotides of the invention encode such spike proteins. By way of example, a variant may have at least 80%, preferably at least 90%, more preferably at least 95%, and most

preferably at least 97 or at least 99% amino acid sequence homology with the reference sequence (e.g. a 2019-nCoV polynucleotide and/or protein of the invention, particularly any SEQ ID NO presented in the present specification which defines a 2019-nCoV polynucleotide and/or protein). Thus, a variant may include one or more analogues of a polynucleotide (e.g. an unnatural nucleic acid), or a substituted linkage. Also, by way of example, the term fragment, when used in relation to a 2019-nCoV polynucleotide and/or protein, means a polynucleotide having at least ten, preferably at least fifteen, more preferably at least twenty nucleic acid residues of the reference 2019-nCoV polynucleotide and/or protein. The term fragment also relates to the above-mentioned variants. Thus, by way of example, a fragment of a 2019-nCoV polynucleotide and/or protein of the present invention may comprise a nucleic acid sequence having at least 10, 20 or 30 nucleic acids, wherein the polynucleotide sequence has at least 80% sequence homology over a corresponding nucleic acid sequence (of contiguous) nucleic acids of the reference 2019-nCoV polynucleotide and/or protein sequence. These definitions of fragments and variants also apply to other polynucleotides of the invention. In the context of peptide sequences, the term fragment means a peptide having at least ten, preferably at least fifteen, more preferably at least twenty amino acid residues of the reference protein. The term fragment also relates to the above-mentioned variants. Thus, by way of example, a fragment may comprise an amino acid sequence having at least 10, 20 or 30 amino acids, wherein the amino acid sequence has at least 80% sequence homology over a corresponding amino acid sequence (of contiguous) amino acids of the reference sequence.

Some exemplified polynucleotides herein comprise additional motifs, particularly restriction enzyme sites, KOZAC sequences and/or motifs (e.g. tga taa) to drive protein secretion. One of skill in the art will appreciate that these additional sequences may be omitted. As such, the invention encompasses the exemplified nucleic acid sequences comprising the specified motifs, as well as the nucleic acid sequences lacking one or more of these motifs.

The terms "decrease", "reduced", "reduction", or "inhibit" are all used herein to mean a decrease by a statistically significant amount. The terms "reduce," "reduction" or "decrease" or "inhibit" typically means a decrease by at least 10% as compared to a reference level (e.g. the absence of a given treatment) and can include, for example, a decrease by at least about 10%, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99% , or more. As used herein, "reduction" or "inhibition" does not encompass a complete inhibition or reduction as

compared to a reference level. "Complete inhibition" is a 100% inhibition as compared to a reference level. A decrease can be preferably down to a level accepted as within the range of normal for an individual without a given disorder.

The terms "increased", "increase", "enhance", or "activate" are all used herein to mean an increase by a statically significant amount. The terms "increased", "increase", "enhance", or "activate" can mean an increase of at least 10% as compared to a reference level, for example an increase of at least about 20%, or at least about 30%, or at least about 40%, or at least about 50%, or at least about 60%, or at least about 70%, or at least about 80%, or at least about 90% or up to and including a 100% increase or any increase between 10-100% as compared to a reference level, or at least about a 2-fold, or at least about a 3-fold, or at least about a 4-fold, or at least about a 5-fold or at least about a 10-fold increase, or any increase between 2-fold and 10-fold or greater as compared to a reference level. In the context of a marker or symptom, an "increase" is a statistically significant increase in such level.

As used herein, a "subject" means a human or animal. Usually the animal is a vertebrate such as a primate, rodent, domestic animal or game animal. Primates include chimpanzees, cynomologous monkeys, spider monkeys, and macaques, e.g., Rhesus. Rodents include mice, rats, woodchucks, ferrets, rabbits and hamsters. Domestic and game animals include cows, horses, pigs, deer, bison, buffalo, feline species, e.g., domestic cat, canine species, e.g., dog, fox, wolf, avian species, e.g., chicken, emu, ostrich, and fish, e.g., trout, catfish and salmon. Preferably the subject is a mammal, e.g., a primate, e.g., a human. The terms, "individual," "patient" and "subject" are used interchangeably herein.

Preferably, the subject is a mammal. The mammal can be a human, non-human primate, mouse, rat, dog, cat, horse, or cow, but is not limited to these examples. Mammals other than humans can be advantageously used as subjects that represent animal models of pain. A subject can be male or female, adult or juvenile.

A subject can be one who has been previously diagnosed with or identified as suffering from or having a condition in need of treatment or one or more complications related to such a condition, and optionally, have already undergone treatment for a condition as defined herein or the one or more complications related to said condition. Alternatively, a subject can also be one who has not been previously diagnosed as having a condition as defined herein or one or more complications related to said condition. For example, a subject can be one who exhibits one or more risk factors for a condition or one or more complications related to said condition or a subject who does not exhibit risk factors.

A "subject in need" of treatment for a particular condition can be a subject having that condition, diagnosed as having that condition, or at risk of developing that condition.

As used herein, the terms "protein" and "polypeptide" are used interchangeably herein to designate a series of amino acid residues, connected to each other by peptide bonds between the alpha-amino and carboxyl groups of adjacent residues. The terms "protein", and "polypeptide" refer to a polymer of amino acids, including modified amino acids (e.g., phosphorylated, glycosylated, etc.) and amino acid analogues, regardless of its size or function. "Protein" and "polypeptide" are often used in reference to relatively large polypeptides, whereas the term "peptide" is often used in reference to small polypeptides, but usage of these terms in the art overlaps. The terms "protein" and "polypeptide" are used interchangeably herein when referring to a gene product and fragments thereof. Thus, exemplary polypeptides or proteins include gene products, naturally occurring proteins, homologs, orthologs, paralogs, fragments and other equivalents, variants, fragments, and analogs of the foregoing.

A polypeptide, e.g., a fusion polypeptide or portion thereof (e.g. a domain), can be a variant of a sequence described herein. Preferably, the variant is a conservative substitution variant. A "variant," as referred to herein, is a polypeptide substantially homologous to a native or reference polypeptide, but which has an amino acid sequence different from that of the native or reference polypeptide because of one or a plurality of deletions, insertions or substitutions. Polypeptide-encoding DNA sequences encompass sequences that comprise one or more additions, deletions, or substitutions of nucleotides when compared to a native or reference DNA sequence, but that encode a variant protein or fragment thereof that retains the relevant biological activity relative to the reference protein, e.g., at least 50% of the wildtype reference protein. As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters a single amino acid or a small percentage, (i.e. 5% or fewer, e.g. 4% or fewer, or 3% or fewer, or 1% or fewer) of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. It is contemplated that some changes can potentially improve the relevant activity, such that a variant, whether conservative or not, has more than 100% of the activity of wild-type, e.g. 110%, 125%, 150%, 175%, 200%, 500%, 1000% or more.

A given amino acid can be replaced by a residue having similar physiochemical characteristics, e.g., substituting one aliphatic residue for another (such as Ile, Val, Leu, or Ala for one another), or substitution of one polar residue for another (such as between Lys and Arg;

Glu and Asp; or Gln and Asn). Other such conservative substitutions, e.g., substitutions of entire regions having similar hydrophobicity characteristics, are known. Polypeptides comprising conservative amino acid substitutions can be tested in any one of the assays described herein to confirm that a desired activity of a native or reference polypeptide is retained. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles consistent with the disclosure. Typically conservative substitutions for one another include: 1) Alanine (A), Glycine (G); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); 7) Serine (S), Threonine (T); and 8) Cysteine (C), Methionine (M) (see, e.g., Creighton, *Proteins* (1984)).

Any cysteine residue not involved in maintaining the proper conformation of the polypeptide also can be substituted, generally with serine, to improve the oxidative stability of the molecule and prevent aberrant crosslinking. Conversely, cysteine bond(s) can be added to the polypeptide to improve its stability or facilitate oligomerization.

A polypeptide as described herein may comprise at least one peptide bond replacement. A single peptide bond or multiple peptide bonds, e.g. 2 bonds, 3 bonds, 4 bonds, 5 bonds, or 6 or more bonds, or all the peptide bonds can be replaced. An isolated peptide as described herein can comprise one type of peptide bond replacement or multiple types of peptide bond replacements, e.g. 2 types, 3 types, 4 types, 5 types, or more types of peptide bond replacements. Non-limiting examples of peptide bond replacements include urea, thiourea, carbamate, sulfonyl urea, trifluoroethylamine, ortho-(aminoalkyl)-phenylacetic acid, para-(aminoalkyl)-phenylacetic acid, meta-(aminoalkyl)-phenylacetic acid, thioamide, tetrazole, boronic ester, olefinic group, and derivatives thereof.

A polypeptide as described herein may comprise naturally occurring amino acids commonly found in polypeptides and/or proteins produced by living organisms, e.g. Ala (A), Val (V), Leu (L), Ile (I), Pro (P), Phe (F), Trp (W), Met (M), Gly (G), Ser (S), Thr (T), Cys (C), Tyr (Y), Asn (N), Gln (Q), Asp (D), Glu (E), Lys (K), Arg (R), and His (H). A polypeptide as described herein may comprise alternative amino acids. Non-limiting examples of alternative amino acids include D amino acids, beta-amino acids, homocysteine, phosphoserine, phosphothreonine, phosphotyrosine, hydroxyproline, gamma-carboxyglutamate; hippuric acid, octahydroindole-2-carboxylic acid, statine, 1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid, penicillamine (3-mercapto-D-valine), ornithine,

5 citruline, alpha-methyl-alanine, para-benzoylphenylalanine, paraaminophenylalanine, p-fluorophenylalanine, phenylglycine, propargylglycine, sarcosine, and tert-butylglycine), diaminobutyric acid, 7-hydroxy-tetrahydroisoquinoline carboxylic acid, naphthylalanine, biphenylalanine, cyclohexylalanine, amino-isobutyric acid, norvaline, norleucine, tert-leucine, 10 tetrahydroisoquinoline carboxylic acid, pipecolic acid, phenylglycine, homophenylalanine, cyclohexylglycine, dehydroleucine, 2,2-diethylglycine, 1-amino-1-cyclopentanecarboxylic acid, 1-amino-1-cyclohexanecarboxylic acid, amino-benzoic acid, amino-naphthoic acid, gamma-aminobutyric acid, difluorophenylalanine, nipecotic acid, alphaamino butyric acid, thienyl-alanine, t-butylglycine, trifluorovaline; hexafluoroleucine; fluorinated analogs; azide-10 modified amino acids; alkyne-modified amino acids; cyano-modified amino acids; and derivatives thereof.

A polypeptide may be modified, e.g. by addition of a moiety to one or more of the amino acids comprising the peptide. A polypeptide as described herein may comprise one or more moiety molecules, e.g. 1 or more moiety molecules per peptide, 2 or more moiety 15 molecules per peptide, 5 or more moiety molecules per peptide, 10 or more moiety molecules per peptide or more moiety molecules per peptide. A polypeptide as described herein may comprise one more types of modifications and/or moieties, e.g. 1 type of modification, 2 types of modifications, 3 types of modifications or more types of modifications. Non-limiting examples of modifications and/or moieties include PEGylation; glycosylation; HESylation; 20 ELPylation; lipidation; acetylation; amidation; end-capping modifications; cyano groups; phosphorylation; albumin, and cyclization.

Alterations of the original amino acid sequence can be accomplished by any of a number of techniques known to one of skill in the art. Amino acid substitutions can be introduced, for example, at particular locations by synthesizing oligonucleotides containing a 25 codon change in the nucleotide sequence encoding the amino acid to be changed, flanked by restriction sites permitting ligation to fragments of the original sequence. Following ligation, the resulting reconstructed sequence encodes an analogue having the desired amino acid insertion, substitution, or deletion. Alternatively, oligonucleotide-directed site-specific mutagenesis procedures can be employed to provide an altered nucleotide sequence having 30 particular codons altered according to the substitution, deletion, or insertion required. Techniques for making such alterations include those disclosed by Walder et al. (Gene 42:133, 1986); Bauer et al. (Gene 37:73, 1985); Craik (BioTechniques, January 1985, 12-19); Smith et al. (Genetic Engineering: Principles and Methods, Plenum Press, 1981); and U.S. Pat. Nos. 4,518,584 and 4,737,462, which are herein incorporated by reference in their entireties. A

polypeptide as described herein may be chemically synthesized and mutations can be incorporated as part of the chemical synthesis process.

As used herein, the terms "polynucleotides", "nucleic acid" and "nucleic acid sequence" refers to any molecule, preferably a polymeric molecule, incorporating units of ribonucleic acid, deoxyribonucleic acid or an analogue thereof. The nucleic acid can be either single-  
5 stranded or double-stranded. A single-stranded nucleic acid can be one nucleic acid strand of a denatured double- stranded DNA. Alternatively, it can be a single-stranded nucleic acid not derived from any double-stranded DNA. In one aspect, the nucleic acid can be DNA. In another aspect, the nucleic acid can be RNA. Suitable nucleic acid molecules are DNA, including  
10 genomic DNA or cDNA. Other suitable nucleic acid molecules are RNA, including mRNA.

As used herein the term "comprising" or "comprises" is used in reference to compositions, methods, and respective component(s) thereof, that are essential to the method or composition, yet open to the inclusion of unspecified elements, whether essential or not.

The term "consisting of" refers to compositions, methods, and respective components  
15 thereof as described herein, which are exclusive of any element not recited in that description of the invention.

As used herein the term "consisting essentially of" refers to those elements required for a given invention. The term permits the presence of elements that do not materially affect the basic and novel or functional characteristic(s) of that invention.

20

## SEQUENCE HOMOLOGY

Any of a variety of sequence alignment methods can be used to determine percent identity, including, without limitation, global methods, local methods and hybrid methods, such as, e.g., segment approach methods. Protocols to determine percent identity are routine  
25 procedures within the scope of one skilled in the art. Global methods align sequences from the beginning to the end of the molecule and determine the best alignment by adding up scores of individual residue pairs and by imposing gap penalties. Non-limiting methods include, e.g., CLUSTAL W, see, e.g., Julie D. Thompson et al., CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment Through Sequence Weighting, Position-Specific  
30 Gap Penalties and Weight Matrix Choice, 22(22) Nucleic Acids Research 4673-4680 (1994); and iterative refinement, see, e.g., Osamu Gotoh, Significant Improvement in Accuracy of Multiple Protein. Sequence Alignments by Iterative Refinement as Assessed by Reference to Structural Alignments, 264(4) J. Mol. Biol. 823-838 (1996). Local methods align sequences by identifying one or more conserved motifs shared by all of the input sequences. Non-limiting

methods include, e.g., Match-box, see, e.g., Eric Depiereux and Ernest Feytmans, Match-Box: A Fundamentally New Algorithm for the Simultaneous Alignment of Several Protein Sequences, 8(5) CABIOS 501 -509 (1992); Gibbs sampling, see, e.g., C. E. Lawrence et al., Detecting Subtle Sequence Signals: A Gibbs Sampling Strategy for Multiple Alignment, 5 262(5131 ) Science 208-214 (1993); Align-M, see, e.g., Ivo Van Walle et al., Align-M - A New Algorithm for Multiple Alignment of Highly Divergent Sequences, 20(9) Bioinformatics: 1428-1435 (2004).

Thus, percent sequence identity is determined by conventional methods. See, for example, Altschul et al., Bull. Math. Bio. 48: 603-16, 1986 and Henikoff and Henikoff, Proc. 10 Natl. Acad. Sci. USA 89:10915-19, 1992. Briefly, two amino acid sequences are aligned to optimize the alignment scores using a gap opening penalty of 10, a gap extension penalty of 1, and the "blosum 62" scoring matrix of Henikoff and Henikoff (ibid.) as shown below (amino acids are indicated by the standard one-letter codes).

15 Alignment score for determining sequence identity

#### BLOSUM62 table

|      | A  | R  | N  | D  | C  | Q  | E  | G  | H  | I  | L  | K  | M  | F  | P | S | T | W | Y | V |
|------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|---|---|---|---|---|---|
| 20 A | 4  |    |    |    |    |    |    |    |    |    |    |    |    |    |   |   |   |   |   |   |
| R    | -1 | 5  |    |    |    |    |    |    |    |    |    |    |    |    |   |   |   |   |   |   |
| N    | -2 | 0  | 6  |    |    |    |    |    |    |    |    |    |    |    |   |   |   |   |   |   |
| D    | -2 | -2 | 1  | 6  |    |    |    |    |    |    |    |    |    |    |   |   |   |   |   |   |
| C    | 0  | -3 | -3 | -3 | 9  |    |    |    |    |    |    |    |    |    |   |   |   |   |   |   |
| 25 Q | -1 | 1  | 0  | 0  | -3 | 5  |    |    |    |    |    |    |    |    |   |   |   |   |   |   |
| E    | -1 | 0  | 0  | 2  | -4 | 2  | 5  |    |    |    |    |    |    |    |   |   |   |   |   |   |
| G    | 0  | -2 | 0  | -1 | -3 | -2 | -2 | 6  |    |    |    |    |    |    |   |   |   |   |   |   |
| H    | -2 | 0  | 1  | -1 | -3 | 0  | 0  | -2 | 8  |    |    |    |    |    |   |   |   |   |   |   |
| I    | -1 | -3 | -3 | -3 | -1 | -3 | -3 | -4 | -3 | 4  |    |    |    |    |   |   |   |   |   |   |
| 30 L | -1 | -2 | -3 | -4 | -1 | -2 | -3 | -4 | -3 | 2  | 4  |    |    |    |   |   |   |   |   |   |
| K    | -1 | 2  | 0  | -1 | -3 | 1  | 1  | -2 | -1 | -3 | -2 | 5  |    |    |   |   |   |   |   |   |
| M    | -1 | -1 | -2 | -3 | -1 | 0  | -2 | -3 | -2 | 1  | 2  | -1 | 5  |    |   |   |   |   |   |   |
| F    | -2 | -3 | -3 | -3 | -2 | -3 | -3 | -3 | -1 | 0  | 0  | -3 | 0  | 6  |   |   |   |   |   |   |
| P    | -1 | -2 | -2 | -1 | -3 | -1 | -1 | -2 | -2 | -3 | -3 | -1 | -2 | -4 | 7 |   |   |   |   |   |

S 1 -1 1 0 -1 0 0 0 -1 -2 -2 0 -1 -2 -1 4  
 T 0 -1 0 -1 -1 -1 -1 -2 -2 -1 -1 -1 -2 -1 1 5  
 W -3 -3 -4 -4 -2 -2 -3 -2 -2 -3 -2 -3 -1 1 -4 -3 -2 11  
 Y -2 -2 -2 -3 -2 -1 -2 -3 2 -1 -1 -2 -1 3 -3 -2 -2 2 7  
 5 V 0 -3 -3 -3 -1 -2 -2 -3 -3 3 1 -2 1 -1 -2 -2 0 -3 -1 4

The percent identity is then calculated as:

$$\frac{\text{Total number of identical matches}}{\text{[length of the longer sequence plus the number of gaps introduced into the longer sequence in order to align the two sequences]}} \times 100$$

Substantially homologous polypeptides are characterized as having one or more amino acid substitutions, deletions or additions. These changes are preferably of a minor nature, that is conservative amino acid substitutions (see below) and other substitutions that do not significantly affect the folding or activity of the polypeptide; small deletions, typically of one to about 30 amino acids; and small amino- or carboxyl-terminal extensions, such as an amino-terminal methionine residue, a small linker peptide of up to about 20-25 residues, or an affinity tag.

#### Conservative amino acid substitutions

Basic: arginine  
           lysine  
 25           histidine  
 Acidic: glutamic acid  
           aspartic acid  
 Polar: glutamine  
           asparagine  
 30 Hydrophobic: leucine  
                   isoleucine  
                   valine  
 Aromatic: phenylalanine  
                   tryptophan

tyrosine  
Small: glycine  
alanine  
serine  
5 threonine  
methionine

In addition to the 20 standard amino acids, non-standard amino acids (such as 4-hydroxyproline, 6-N-methyl lysine, 2-aminoisobutyric acid, isovaline and a -methyl serine) may be substituted for amino acid residues of the polypeptides of the present invention. A limited number of non-conservative amino acids, amino acids that are not encoded by the genetic code, and unnatural amino acids may be substituted for clostridial polypeptide amino acid residues. The polypeptides of the present invention can also comprise non-naturally occurring amino acid residues.

Non-naturally occurring amino acids include, without limitation, trans-3-methylproline, 2,4-methano-proline, cis-4-hydroxyproline, trans-4-hydroxy-proline, N-methylglycine, allothreonine, methyl-threonine, hydroxy-ethylcysteine, hydroxyethylhomocysteine, nitroglutamine, homoglutamine, pipecolic acid, tert-leucine, norvaline, 2-azaphenylalanine, 3-azaphenyl-alanine, 4-azaphenyl-alanine, and 4-fluorophenylalanine. Several methods are known in the art for incorporating non-naturally occurring amino acid residues into proteins. For example, an *in vitro* system can be employed wherein nonsense mutations are suppressed using chemically aminoacylated suppressor tRNAs. Methods for synthesizing amino acids and aminoacylating tRNA are known in the art. Transcription and translation of plasmids containing nonsense mutations is carried out in a cell free system comprising an *E. coli* S30 extract and commercially available enzymes and other reagents. Proteins are purified by chromatography. See, for example, Robertson et al., *J. Am. Chem. Soc.* 113:2722, 1991; Ellman et al., *Methods Enzymol.* 202:301, 1991; Chung et al., *Science* 259:806-9, 1993; and Chung et al., *Proc. Natl. Acad. Sci. USA* 90: 10145-9, 1993). In a second method, translation is carried out in *Xenopus* oocytes by microinjection of mutated mRNA and chemically aminoacylated suppressor tRNAs (Turcatti et al., *J. Biol. Chem.* 271:19991-8, 1996). Within a third method, *E. coli* cells are cultured in the absence of a natural amino acid that is to be replaced (e.g., phenylalanine) and in the presence of the desired non-naturally occurring amino acid(s) (e.g., 2-azaphenylalanine, 3-azaphenylalanine, 4-azaphenylalanine, or 4-fluorophenylalanine). The non-naturally occurring amino acid is incorporated into the

polypeptide in place of its natural counterpart. See, Koide et al., Biochem. 33:7470-6, 1994. Naturally occurring amino acid residues can be converted to non-naturally occurring species by in vitro chemical modification. Chemical modification can be combined with site-directed mutagenesis to further expand the range of substitutions (Wynn and Richards, Protein Sci. 2:395-403, 1993).

A limited number of non-conservative amino acids, amino acids that are not encoded by the genetic code, non-naturally occurring amino acids, and unnatural amino acids may be substituted for amino acid residues of polypeptides of the present invention.

Essential amino acids in the polypeptides of the present invention can be identified according to procedures known in the art, such as site-directed mutagenesis or alanine scanning mutagenesis (Cunningham and Wells, Science 244: 1081-5, 1989). Sites of biological interaction can also be determined by physical analysis of structure, as determined by such techniques as nuclear magnetic resonance, crystallography, electron diffraction or photoaffinity labelling, in conjunction with mutation of putative contact site amino acids. See, for example, de Vos et al., Science 255:306-12, 1992; Smith et al., J. Mol. Biol. 224:899-904, 1992; Wlodaver et al., FEBS Lett. 309:59-64, 1992. The identities of essential amino acids can also be inferred from analysis of homologies with related components (e.g. the translocation or protease components) of the polypeptides of the present invention.

Multiple amino acid substitutions can be made and tested using known methods of mutagenesis and screening, such as those disclosed by Reidhaar-Olson and Sauer (Science 241:53-7, 1988) or Bowie and Sauer (Proc. Natl. Acad. Sci. USA 86:2152-6, 1989). Briefly, these authors disclose methods for simultaneously randomizing two or more positions in a polypeptide, selecting for functional polypeptide, and then sequencing the mutagenized polypeptides to determine the spectrum of allowable substitutions at each position. Other methods that can be used include phage display (e.g., Lowman et al., Biochem. 30: 10832-7, 1991; Ladner et al., U.S. Patent No. 5,223,409; Huse, WIPO Publication WO 92/06204) and region-directed mutagenesis (Derbyshire et al., Gene 46:145, 1986; Ner et al., DNA 7:127, 1988).

The following Examples illustrate the invention.

## **EXAMPLES**

The nucleic acid encoding the 2019-nCoV spike protein (S) has been modified for expression in various expression systems, including - *E. coli*, yeast and human cells to induce neutralizing antibodies response to protect against 2019-nCoV infection.

N and C terminal deleted amino acids (5-10) were cloned to express and refold the S protein in the native conformation of coronavirus in an *E. coli* system while others (Yeast and Human cells) as whole protein. S protein from coronavirus was also combined and express as fusion proteins with virus like particles of Hepatitis E (P239) and Human Papilloma virus like particle (18L1) in order to increase the efficacy and longevity of the immunoprotective response.

#### **Example 1: *E. coli* based 2019-nCoV spike protein expression**

The nucleic acid sequence of coronavirus S protein (Wuhan-Hu-1-CoV-S, GenBank: MN908947.3) was optimised for expression in *E. coli* using Geneart for codon optimisation and containing SacI and NotI single cloning sites to design the nucleic acid sequence of SEQ ID NO: 2.

The codon usage was adapted to the codon bias of *Escherichia coli* genes. In addition, regions of very high (> 80 %) or very low (< 30 %) GC content have been avoided where possible. Negative cis-acting sites (such as splice sites, TATA-boxes, etc.) which may negatively influence expression were eliminated wherever possible. GC content was adjusted (average GC content 45%) to prolong mRNA half life. Codon usage was adapted to the bias of *E. coli* resulting in a CAI (Codon Adaptation Index) value of 0.96. The optimized gene has been designed to allow high and stable expression rates in *E. coli*.

#### **Example 2: 2019-nCoV spike protein in Hepatitis E Virus-Like Particles**

The nucleic acid sequence of coronavirus S protein was optimised for expression in *E. coli* and used to generate Hepatitis E Virus-Like Particles comprising the S protein (Wuhan-Hu-1-HEV-CoV-S).

The gene synthesis was codon optimized for expression in *E. coli* by Geneart with SacI and NotI single cloning sites to design the nucleic acid sequence of SEQ ID NO: 3.

The codon usage was adapted to the codon bias of *Escherichia coli* genes. In addition, regions of very high (> 80 %) or very low (< 30 %) GC content have been avoided where possible. Negative cis-acting sites (such as splice sites, TATA-boxes, etc.) which may negatively influence expression were eliminated wherever possible. GC content (46%) was adjusted to prolong mRNA half life. Codon usage was adapted to the bias of *E. coli* resulting

in a CAI value of 0.96. The optimized gene has been designed to allow high and stable expression rates in *E. coli*.

**Example 3: *Komagataella pastoris* based 2019-nCoV spike protein expression**

5        The nucleic acid sequence of coronavirus S protein (Wuhan-Hu-1 CoV-S) was optimised for expression in *K. pastoris* using Geneart for codon optimisation and containing BstBI-NotI single cloning sites to design the nucleic acid sequence of SEQ ID NO: 4.

10        The codon usage was adapted to the codon bias of *K. pastoris* genes. In addition, regions of very high (> 80 %) or very low (< 30 %) GC content have been avoided where possible. Negative cis-acting sites (such as splice sites, TATA-boxes, etc.) which may negatively influence expression were eliminated wherever possible. GC content (48%) was adjusted to prolong mRNA half life. Codon usage was adapted to the bias of *K. pastoris* resulting in a CAI value of 0.84. The optimized gene has been designed to allow high and stable expression rates in *K. pastoris*.

15

**Example 4: expression of a fusion protein comprising 2019-nCoV spike protein in *K. pastoris***

20        The nucleic acid sequence of coronavirus S protein (Wuhan-Hu-1-HPV18 L1-CoV-S) was optimised for expression in *K. pastoris* as a fusion protein with HPV 18 L1. BstBI and NotI are single cloning sites to design the nucleic acid sequence of SEQ ID NO: 5

25        The codon usage was adapted to the codon bias of *K. pastoris* genes. In addition, regions of very high (> 80 %) or very low (< 30 %) GC content have been avoided where possible. Negative cis-acting sites (such as splice sites, TATA-boxes, etc.) which may negatively influence expression were eliminated wherever possible. GC content (48%) was adjusted to prolong mRNA half life. Codon usage was adapted to the bias of *K. pastoris* resulting in a CAI value of 0.84. The optimized gene has been designed to allow high and stable expression rates in *K. pastoris*.

30        **Example 5: expression of a fusion protein comprising 2019-nCoV spike protein in *K. pastoris***

      The nucleic acid sequence of coronavirus S protein (Wuhan-Hu-1-HPV16 L1-CoV-S) was optimised for expression in *K. pastoris* as a fusion protein with HPV 16 L1. BstBI and NotI are single cloning sites to design the nucleic acid sequence of SEQ ID NO: 6

The codon usage was adapted to the codon bias of *K. pastoris* genes. In addition, regions of very high (> 80 %) or very low (< 30 %) GC content have been avoided where possible. Negative cis-acting sites (such as splice sites, TATA-boxes, etc.) which may negatively influence expression were eliminated wherever possible. GC content (48%) was adjusted to prolong mRNA half life. Codon usage was adapted to the bias of *K. pastoris* resulting in a CAI value of **0.84**. The optimized gene has been designed to allow high and stable expression rates in *Pichia pastoris*

**Example 6: expression of 2019-nCoV spike protein in human 293 F cells**

The nucleic acid sequence of coronavirus S protein (Wuhan-Hu-1 CoV-S surface bounded protein, which is expressed and bound to the outer surface of 293 F cells) was optimised for expression in human 293 F cells using Geneart for codon optimisation and containing NheI-NotI single cloning sites to design the nucleic acid sequence of SEQ ID NO: 7.

The codon usage was adapted to the codon bias of *Homo sapiens* genes. In addition, regions of very high (> 80 %) or very low (< 30 %) GC content have been avoided where possible. Negative cis-acting sites (such as splice sites, TATA-boxes, etc.) which may negatively influence expression were eliminated wherever possible. GC content (56%) was adjusted to prolong mRNA half life. Codon usage was adapted to the bias of *Homo sapiens* resulting in a CAI value of 0.94. The optimized gene has been designed to allow high and stable expression rates in human cells.

**Example 7: expression of a fusion protein comprising 2019-nCoV spike protein in 293 F cells**

The nucleic acid sequence of coronavirus S protein (Wuhan-Hu-1-HBSAg-CoV-S) was optimised for expression in 293 F cells as a fusion protein with Hepatitis B surface antigen. The sequence contains NheI-NotI single cloning sites to design the nucleic acid sequence of SEQ ID NO: 8.

The codon usage was adapted to the codon bias of *Homo sapiens* genes. In addition, regions of very high (> 80 %) or very low (< 30 %) GC content have been avoided where possible. Negative cis-acting sites (such as splice sites, TATA-boxes, etc.) which may negatively influence expression were eliminated wherever possible. GC content (56%) was adjusted to prolong mRNA half life. Codon usage was adapted to the bias of *Homo sapiens*

resulting in a CAI\* (codon adaptation Index) value of **0.94**. The optimized gene has been designed to allow high and stable expression rates in human cells.

**Example 8: Adjuvination of 2019-nCoV spike protein using aluminium hydroxide and aluminium phosphate gel**

The 2019-nCoV spike protein and fusion proteins thereof of Examples 1, 3 and 6 were adsorbed in 0.5mg of aluminium hydroxide and aluminium phosphate gel for adjuvination.

**Example 9: Adjuvination of 2019-nCoV spike protein using monophosphoryl lipid and aluminium hydroxide**

The 2019-nCoV spike protein and fusion proteins thereof of Examples 1, 3 and 6 were mixed in monophosphoryl lipid (MPL) and aluminium hydroxide for adjuvination.

**Example 10: Adjuvination of 2019-nCoV spike protein using MF59**

The 2019-nCoV spike protein and fusion proteins thereof of Examples 1, 3 and 6 were mixed in MF59 (5% squalene) for adjuvination.

**Example 11: Generation of an antibody response in mice immunised with 2019-nCoV spike protein**

The immunogenicity of formulations comprising the 2019-CoV spike proteins and fusion proteins thereof of Examples 2, 4, 5 and 7 without adjuvants and the adjuvanted formulations of Examples 1, 3 and 6 were tested in BALB/c mice (5 mice per group). Vaccination was done on day 0 and 7 with 2µg antigen per administration. On day 0 and 14 serum samples were tested for sero negativity (day 0) and antibody response against S protein (day 14) by indirect ELISA. All formulations at day 14 induced high antibody titres while adjuvanted ones induced maximum response (Figure 2).

**Example 12: Expression of HBSAg-(EAAAK)<sub>3</sub>-RBD fusion protein in HEK cells (293F)**

The HBSAg-(EAAAK)<sub>3</sub>-RBD fusion protein of SEQ ID NO: 28 was expressed in HEK cells using the codon-optimised nucleic acid sequence of SEQ ID NO: 27. Briefly, the HBSAg-(EAAAK)<sub>3</sub>-RBD gene was optimised for human cell (293F) expression, cloned using *Nhe*-*Not*I into pcDNA3.1(+) and subject to clonal selection before transfection into 293F cells in suspension culture. The secreted HBSAg-(EAAAK)<sub>3</sub>-RBD fusion protein was harvested after 40h.

Western blotting using a polyclonal antibody against the 2019-nCoV spike protein (1:250) and a mouse monoclonal antibody against HBSAg (1:1,000) demonstrated strong secretion of the fusion protein by the HEK cells (Figure 3) in two separate cultures – #2A and #2B.

5

**Example 13: Generation of an antibody response in mice immunised with HBSAg-(EAAAK)<sub>3</sub>-RBD fusion in HEK cells**

The HBSAg-(EAAAK)<sub>3</sub>-RBD (#2A) from Example 12 was concentrated through 3K Amicon 15 ml centrifugal filter device spinning at 4000 rpm x 40 min.

10 Balb/c mice were then immunised using either (i) 50µg/dose (100µl); or (ii) a 50µg/dose (70µl), at a 1:1 v/v ratio with either aluminium hydroxide or Addavax™ (total volume administered = 140µl). Mice were primed with the vaccine on d0, followed by a first boost at d7, a second boost at d14 (with bleeding for analysis) and a third boost at d28. The final bleeding for analysis was conducted on d42. The experimental groups are shown in Table 1 below.

15

Table 1: Experimental Groups for mice immunised with HBSAg-(EAAAK)<sub>3</sub>-RBD

| GROUP | ANIMAL<br>N°<br>Balb/c | CAGE  | FORMULATION                                                                              | INJECTION<br>VOLUME |
|-------|------------------------|-------|------------------------------------------------------------------------------------------|---------------------|
| 8     | 3                      | 5     | Vaccine HBSAg-(EAAAK) <sub>3</sub> -RBD (293F)                                           | 100ml (ip)          |
| 9     | 2                      | 5 bis | Vaccine HBSAg-(EAAAK) <sub>3</sub> -RBD (293F)                                           | 100ml (ip)          |
| 10    | 3                      | 6     | Vaccine HBSAg-(EAAAK) <sub>3</sub> -RBD (293F)<br>with adjuvant Alu-280 (1:1 v/v ratio)  | 140ml (ip)          |
| 11    | 2                      | 6 bis | Vaccine HBSAg-(EAAAK) <sub>3</sub> -RBD (293F)<br>with adjuvant Alu-280 (1:1 v/v ratio)  | 140ml (ip)          |
| 12    | 3                      | 7     | Vaccine HBSAg-(EAAAK) <sub>3</sub> -RBD (293F)<br>with adjuvant Adda-Vax (1:1 v/v ratio) | 140ml (ip)          |
| 13    | 2                      | 7 bis | Vaccine HBSAg-(EAAAK) <sub>3</sub> -RBD (293F)<br>with adjuvant Adda-Vax (1:1 v/v ratio) | 140ml (ip)          |

Antibody titres were assessed at d14 and d42 using ELISA. As shown in Figure 4, VLPs of HBSAg-(EAAAK)<sub>3</sub>-RBD were able to achieve significant IgG titres at d14 and d42, with higher titres observed at d42. Administration with either adjuvant (aluminium hydroxide or AddaVax™) further increased the titre, again as shown in Figure 4.

- 5 Neutralisation assays were also conducted using mice immunized with VLPs of HBSAg-(EAAAK)<sub>3</sub>-RBD (alone or with aluminium hydroxide or AddaVax™). The results are shown in Table 2 below. The average neutralising titre was 1:1,200 – 1:2,700. The highest titre achieved was 1:5,120 and the lowest 1:640.

10 Table 2: Results of neutralisation assay using HBSAg-EAAAK<sub>3</sub>-RBD

| <u>TREATMENT</u>                               | <u>MOUSE</u>   | <u>TEST1</u> | <u>TEST2</u> |
|------------------------------------------------|----------------|--------------|--------------|
| <u>PBS</u>                                     | <u>1A</u>      | <u>5</u>     | <u>5</u>     |
|                                                | <u>1B</u>      | <u>5</u>     | <u>5</u>     |
|                                                | <u>1C</u>      | <u>5</u>     | <u>5</u>     |
|                                                | <u>1D</u>      | <u>5</u>     | <u>5</u>     |
| <u>HBSAg-(EAAAK)<sub>3</sub>-RBD</u>           | <u>5A</u>      | <u>640</u>   | <u>320</u>   |
|                                                | <u>5B</u>      | <u>1280</u>  | <u>1280</u>  |
|                                                | <u>5C</u>      | <u>5120</u>  | <u>5120</u>  |
|                                                | <u>5BIS A</u>  | <u>2560</u>  | <u>2560</u>  |
|                                                | <u>5BIS B</u>  | <u>640</u>   | <u>640</u>   |
| <u>HBSAg-(EAAAK)<sub>3</sub>-RBD + ALU</u>     | <u>6A</u>      | <u>1280</u>  | <u>2560</u>  |
|                                                | <u>6B</u>      | <u>2560</u>  | <u>2560</u>  |
|                                                | <u>6C</u>      | <u>2560</u>  | <u>2560</u>  |
|                                                | <u>6 BIS A</u> | <u>640</u>   | <u>640</u>   |
|                                                | <u>6 BIS B</u> | <u>2560</u>  | <u>5120</u>  |
| <u>HBSAg-(EAAAK)<sub>3</sub>-RBD + ADDAVAX</u> | <u>7A</u>      | <u>1280</u>  | <u>1280</u>  |
|                                                | <u>7B</u>      | <u>1280</u>  | <u>1280</u>  |
|                                                | <u>7C</u>      | <u>640</u>   | <u>640</u>   |
|                                                | <u>7 BIS A</u> | <u>1280</u>  | <u>1280</u>  |
|                                                | <u>7 BIS B</u> | <u>1280</u>  | <u>2560</u>  |

**Example 14: Expression of HEV-(GGGGS)<sub>3</sub>-RBD fusion in *E. coli***

The HEV-(GGGGS)<sub>3</sub>-RBD fusion protein of SEQ ID NO: 31 was expressed in *E. coli* using the nucleic acid sequence of SEQ ID NO: 30, which is codon-optimised for expression in *E. coli*. This nucleic acid sequence was cloned using *SacI-NotI* into pET26(+). A positive clone was selected (#10) and expressed in BL21. Three different positive colonies (#10A, #10B and #10C) were screened for protein expression, with HEV-RBD expression valuated using IPTG induction (6 hours) for all of #10A, #10B and #10C. Western blotting using an antibody against HEV was used to confirm fusion protein expression in all of #10A, #10B and #10C (Figure 5). Clone #10B was selected for future investigation and was purified using anion exchange chromatography using a 50mM phosphate buffer at pH 7.2 as the binding buffer (data not shown).

**Example 15: Generation of an antibody response in mice immunised with HEV-(GGGGS)<sub>3</sub>-RBD fusion protein generated in *E. coli***

Balb/c mice were immunised using HEV-(GGGGS)<sub>3</sub>-RBD (#10B) from Example 14 using 50µg/dose, at a 1:1 v/v ratio with either aluminium hydroxide or Addavax™ (total volume administered = 100µl). Mice were primed with the vaccine on d0, followed by a first boost at d7, a second boost at d14 (with bleeding for analysis) and a third boost at d28. The final bleeding for analysis was conducted on d42. The experimental groups are shown in Table 3 below.

20

Table 3: Experimental Groups for mice immunised with HEV-(GGGGS)<sub>3</sub>-RBD

| GROUP | ANIMAL<br>N°<br>Balb/c | CAGE  | FORMULATION                                                                               | INJECTION<br>VOLUME |
|-------|------------------------|-------|-------------------------------------------------------------------------------------------|---------------------|
| 1     | 4                      | 1     | PBS (negative control)                                                                    | 100µl (ip)          |
| 2     | 3                      | 2     | Vaccine HEV-(GGGGS) <sub>3</sub> -RBD (E. coli)                                           | 100µl (ip)          |
| 3     | 2                      | 2 bis | Vaccine HEV-(GGGGS) <sub>3</sub> -RBD (E. coli)                                           | 100µl (ip)          |
| 4     | 3                      | 3     | Vaccine HEV-(GGGGS) <sub>3</sub> -RBD (E. coli)<br>with adjuvant Alu-280 (1:1 v/v ratio)  | 100µl (ip)          |
| 5     | 2                      | 3 bis | Vaccine HEV-(GGGGS) <sub>3</sub> -RBD (E. coli)<br>with adjuvant Alu-280 (1:1 v/v ratio)  | 100µl (ip)          |
| 6     | 3                      | 4     | Vaccine HEV-(GGGGS) <sub>3</sub> -RBD (E. coli)<br>with adjuvant Adda-Vax (1:1 v/v ratio) | 100µl (ip)          |
| 7     | 2                      | 4 bis | Vaccine HEV-(GGGGS) <sub>3</sub> -RBD (E. coli)<br>with adjuvant Adda-Vax (1:1 v/v ratio) | 100µl (ip)          |

Antibody titres were assessed at d14 and d42 using ELISA. VLPs of HEV-(GGGGS)<sub>3</sub>-RBD were able to achieve significant neutralising titres at d14 and d42, with higher titres observed at d42. Administration with either adjuvant (aluminium hydroxide or AddaVax™) further increased the titre, as shown in Figure 6.

5

**Example 16: expression of a fusion protein comprising HBSAg-(EAAAK)<sub>3</sub>-full-length 2019-nCoV spike protein in HEK 293 cells**

The HBSAg-(EAAAK)<sub>3</sub>-full-length 2019-nCoV spike protein fusion protein (HBSAg-(EAAAK)<sub>3</sub>-CoV-s) of SEQ ID NO: 33 was expressed in HEK cells using the codon-optimised nucleic acid sequence of SEQ ID NO: 32. Briefly, HBSAg-(EAAAK)<sub>3</sub>-CoV-s was optimised for human cell (293F) expression and subject to clonal selection before transfection into 293F cells in suspension culture. The secreted HBSAg-(EAAAK)<sub>3</sub>-CoV-s was harvested after 40h.

Two different clones of the purified recombinant protein HBSAg-Co-V-s, D8-SA01-01-01 (4X) and the D8-SA01-02-01 (5x) were analysed by Western blot using a rabbit COVID-19 spike protein polyclonal antibody (My Biosource, MBS434243) used at a dilution 1:250. As shown in Figure 7, both clones produced sharp and highly expressed bands of the expected size, indicating strong expression of the fusion protein.

20 **SEQUENCE INFORMATION**

**SEQ ID NO: 1 – 2019-nCoV spike protein amino acid sequence**

MFVFLVLLPLVSSQCVNLTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVS GTNG  
 25 TKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLD SKTQSL L I V N N A T N V V I K V C E F Q F C N D P F L G V Y Y H K N N K  
 SWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEP  
 LVDLPIGINITRFQTL LALHRSYLT PGDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK  
 CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSP  
 30 T K L N D L C F T N V Y A D S F V I R G D E V R Q I A P G Q T G K I A D Y N Y K L P D D F T G C V I A W N S N N L D S K V G G N Y N  
 YLYRLFRKSNLKPFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYPYRVVLSFELLHAPATVC  
 GPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP  
 GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQ  
 TQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNISIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTEC  
 35 N L L L Q Y G S F C T Q L N R A L T G I A V E Q D K N T Q E V F A Q V K Q I Y K T P P I K D F G G F N F S Q I L P D P S K P S K R S F I E D L L F N K  
 VTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM  
 QMAYRFNGIGVTQNVLYENQKLIANQFN SAIGKIQDSLSSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAISS  
 VLNDILSRDLKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGHYLM

SFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDKAHFPREGVFSNGTHWFVTQRNFYEPQIIITDNTFVSGN  
CDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL  
QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

- 5 The RDB domain of the spike protein (residues 319 to 529) is underlined.

**SEQ ID NO: 2 – 2019-nCoV spike protein nucleic acid sequence – optimised for expression in *E. coli* and containing *SacI* and *NotI* single cloning sites. Described in Example 1**

10 GAGCTCatgt ttgtttttct ggttctgctg ccgctgggta gcagccagtg tgttaatctg  
accacacgta ccagctgcc tccggcataat accaatagct ttaccctggg tgtttattat  
ccggacaaag tttttcgtag cagcgttctg catagcacc aggacctgtt tctgccgttt  
tttagcaatg ttacctggtt tcatgccatt catgttagcg gcaccaatgg caccaaacgt  
15 tttgataatc cgtgtctgcc gtttaatgat ggtgtgtatt ttgcaagcac cgaaaaaagc  
aacattattc gcggttgat ttttgggtaca accctggata gcaaaaccca gagcctgctg  
attgttaata atgccaccaa tgtggtgatc aaagtgtgcg aatttcagtt ttgcaatgat  
ccgtttctgg gcgtgtatta ccacaaaaat aacaagagct ggatggaaag cgaatttcgt  
gtttatagca gcgccaataa ttgcaccttt gaatatgtta gccagccgtt tctgatggat  
20 ctggaaggta aacagggtaa ctttaaaaaac ctgcgcgagt tcgtgttcaa aaacatcgat  
ggttacttca aaatctatag caaacacacc ccgattaatc tggttcgtga tctgccgcag  
ggttttagcg cactggaacc gctggttgat ctgccaattg gtattaacat taccggtttt  
cagaccctgc tggcactgca tctagctat ctgacaccgg gtgatagcag cagcgggttg  
accgcaggcg cagcagcata ttatgttgggt tatctgcagc ctctgacctt tctgctgaaa  
25 tataacgaaa atggcacaat taccgatgcc gttgattgtg ccctggatcc gctgagcgaa  
accaaagtga cctgaaaag ctttaccgtt gagaaaggta tttatcagac cagcaatttt  
cgtgtgcagc cgaccgaaag cattgttcgt tttccgaata tcaccaatc gtgtccgttt  
ggcgaagttt ttaatgcaac ccgttttgcc agcgtttatg catggaatcg taaacgtatt  
agcaattgcg ttgccgatta tagcgttctg tataatagcg caagcttcag cacctttaa  
30 tgctatggtg ttagcccgac caaactgaat gatctgtgtt ttaccaatgt gtatgccgat  
agctttgtga ttctgtgtga tgaagttcgt cagattgcac cgggtcagac cggtaaaatt  
gcagattata actataaaact gccggatgat tttacgggtt gtgttattgc ctggaatagc  
aataatctgg acagcaaagt tgggtggcaac tataactatc tgtatcgctt gtttcgtaag  
35 agcaatctga aaccgtttga acgtgatatt agcaccgaga tttatcaggc aggtagcacc  
ccgtgtaatg gtgttgaagg ttttaattgc tattttccgc tgcagagcta tggttttcag  
ccgacaaatg gtgtgggtta tcagccgtat cgtgttgttg ttctgtcatt tgaactgctg  
catgcaccgg caaccgtttg tgggtccgaa aaaagtacca atctggtgaa aaataagtgc  
gtgaacttta actttaatgg tctgaccggc accggtgttc tgaccgaaag taacaaaaaa  
40 ttctgtccgt ttcagcagtt tggccgtgat attgcagata ccaccgatgc agttcgcat  
ccgcagacac tggaaattct ggatattacc ccgtgcagct ttggtggtgt ttcagttatt  
acaccgggta caaataaccag caatcaggtt gcagttctgt atcaggatgt taattgtacc  
gaagtcccg ttgcaattca tgcagatcag ctgaccccg cctggcgtgt gtatagcacc  
ggtagcaatg tgtttcagac acgtgcaggt tgtctgattg gtgcagaaca tgtgaataat  
45 agctatgaat gcgatattcc gattggtgcg ggtatttggt ccagctatca gaccagacc  
aatagtccgc gtcgtgcagc tagcgttgca agccagagca ttattgccta taccatgagc  
ctgggtgcag aaaatagcgt tgcctatagt aataacagca ttgccattcc gaccaacttt  
accattagcg ttaccaccga aattctgccg gttagcatga ccaaaaccag cgttgattgc  
accatgtata tttgtggtga tagtaccgaa tgtagcaatc tgctgctgca gtatggtagc  
50 ttttgcaccc agctgaatcg tgcactgacc ggtattgcag ttgaacagga taaaaacacg  
caagaagttt ttgcacaggt caagcagatc tataaaaccc ctccgattaa agattttggc  
ggtttcaatt ttagccagat cctgccggat ccgagcaaac cgagttaaag tagctttatt  
gaagatctgc tgttcaacaa agtgaccctg gcagatgcag gttttatcaa acagtatggt  
gattgcctgg gcgatattgc cgcacgtgat ctgatttggt cacagaaatt taacggcctg  
accgttctgc ctccgctgct gaccgatgaa atgattgcac agtataccag cgcactgctg  
gcaggcacca ttaccagtgg ttggaccttt ggtgccgtg ccgcactgca gattccgttt  
55 gcaatgcaga tggcatatcg ttttaatggt attggtgtta ccgagaacgt gctgtatgaa  
aaccagaaac tgattgcaaa ccagtttaat agcgcattg gcaaaattca ggatagcctg

```

5   agcagcaccg caagtgcact gggtaaactg caggacgttg ttaatcagaa tgcacaggca
    ctgaataccc tggttaaaca gctgagcagt aatttttggtg caatttcaag cgtgctgaac
    gatattctga gccgtctgga taaagttgaa gcagaagttc agattgatcg tctgattacc
    ggtcgtctgc aaagcctgca gacctatgtg acccagcagc tgattcgcgc agcagaaatt
10  cgtgcaagcg caaatctggc agccaccaa atgagcgaat gtgttctggg tcagagcaaa
    cgtgttgatt tttgcggcaa aggttatcac ctgatgagct ttccgcagag cgcaccgcat
    ggtgttgtgt ttctgcatgt tacctatgtt ccggcacaag aaaaaaactt tacaaccgct
    ccggcaattt gccatgatgg taaagcacat tttccgcgtg aagggtgttt tgttagtaat
    ggcacccatt ggtttgttac acagcgcaac ttttatgaac cgcagattat tacaaccgac
15  aacacctttg ttagcggtaa ctgtgatgtt gtgattggca ttgtgaataa caccgtttat
    gatccactgc agccggaact ggatagcttt aaagaagaac tggacaaata tttcaaaaac
    cacaccagtc cggatgttga tctgggtgat atttcaggtg ttaatgccag cgtggtgaac
    atccagaaag aaattgatcg cctgaatgaa gtggccaaaa atctgaatga aagcctgatt
    gatctgcaag aactggggaa atatgagcag tatatcaa atggcgtggta tatttggtcg
20  ggttttattg caggcctgat tgcaattgtt atggtgacca ttatgctgtg ttgtatgacc
    agctgttgta gctgtctgaa aggttggtgc agctgcggta gctgttgcaa atttgatgaa
    gatgatagcg aaccggtgct gaaaggtgtt aaactgcatt atacctaata aGCGGCGC

```

The 5' SacI single cloning site is single-underlined

20 The 3' NotI single cloning site is dash-underlined

The ATG start codon is in bold and italicised

The nucleic acid sequences of SEQ ID NO: 2 translates to give the native 2019-nCoV spike protein of SEQ ID NO: 1

25

**SEQ ID NO: 3 – nucleic acid encoding for fusion protein HEV-2019-nCoV spike protein– optimised for expression in *E. coli* and containing SacI and NotI single cloning sites. Described in Example 2**

```

30  gagctcATGA TTGCACTGAC CCTGTTTAAT CTGGCAGATA CCCTGTTAGG TGGTCTGCCG
    ACCGAACTGA TTAGCAGTGC CGGTGGTCAG CTGTTTATA GCCGTCCGGT GTTTAGCGCA
    AATGGTGAAC CGACCGTTAA ACTGTATACC AGCGTTGAAA ATGCACAGCA GGATAAAGGT
    ATTGCAATTC CGCATGATAT TGATCTGGGT GAAAGCCGTG TTGTGATTCA GGATTATGAT
35  AATCAGCATG AACAGGATCG TCCGACACCG AGTCCGGCAC CGAGCCGTCC GTTTAGCGTT
    CTGCGTGCAA ATGATGTTCT GTGGCTGAGC CTGACCGCAG CAGAATATGA TCAGAGCACC
    TATGGTAGCA GCACCGGTCC GGTTTATGTT AGCGATAGCG TTACCCTGGT TAATGTTGCA
    ACCGGTGAC AGGCAGTTGC ACGTAGCCTG GATTGGACCA AAGTGACCCT GGATGGTCGT
    CCGCTGAGCA CCATTGAGCA GTATAGCAAA ACCTTTTGTG TTCTGCCGCT GCGTGGTAAA
40  CTGAGCTTTT GGAAGCAGG CACCACCAA GCAGGTTATC CGTATAACTA TAATACCACC
    GCAAGCGATC AGCTGCTGGT TGAAAACGCA GCAGGTCATC GTGTTGCAAT TAGCACCTAT
    ACCACCAAGT TAGGTGCAGG TCCGGTTAGC ATTAGCGCAG TTGCAGTTCT GGCACCGCAT
    AGCGCAtttg tttttctggt tctgctgccg ctggttagca gccagtgtgt taatctgacc
    acacgtaccc agctgcctcc ggcataacc aatagcttta cccgtggtgt ttattatccg
    gacaaagttt ttcgtagcag cgttctgcat agcaccagc acctgtttct gccgtttttt
45  agcaatgtta cctggtttca tgcatcattc gttagcggca ccaatggcac caaacgtttt
    gataatcccg tgctgccgtt taatgatggt gtgtattttg caagcaccga aaaaagcaac
    attattcgcg gttggatttt tggtagaacc ctggatagca aaaccagag cctgctgatt
    gttaataatg ccaccaatgt ggtgatcaaa gtgtgcgaat ttcagttttg caatgatccg
    tttctgggcg tgtattacca caaaaataac aagagctgga tggaaagcga atttcgtgtt
50  tatagcagcg ccaataattg cacctttgaa tatgttagcc agccgtttct gatggatctg
    gaaggtaaac agggtaactt taaaaacctg cgcgagttcg tgttcaaaaa catcgatggt
    tacttcaaaa tctatagcaa acacaccccg attaatctgg ttcgtgatct gccgcagggt
    tttagcgcac tggaaaccgct ggttgatctg ccaattggta ttaacattac ccgttttcag
    accctgctgg cactgcatcg tagctatctg acaccgggtg atagcagcag cggttggacc

```

|    |             |             |             |             |             |                |
|----|-------------|-------------|-------------|-------------|-------------|----------------|
|    | gcaggcgcag  | cagcatatta  | tggttggttat | ctgcagcctc  | gtacctttct  | gctgaaatat     |
|    | aacgaaaatg  | gcacaattac  | cgatgccgtt  | gattgtgccc  | tggatccgct  | gagcgaaacc     |
|    | aaatgtaccc  | tgaaaagctt  | taccgttgag  | aaaggtat    | atcagaccag  | caattttcgt     |
| 5  | gtgcagccga  | cggaaagcat  | tggttcgtttt | ccgaatatca  | ccaatctgtg  | tccgttttggc    |
|    | gaagttttta  | atgcaaccgc  | ttttgccagc  | gtttatgcat  | ggaatcgtaa  | acgtattagc     |
|    | aattgcgttg  | ccgattatag  | cgttctgtat  | aatagcgcaa  | gcttcagcac  | ctttaaatgc     |
|    | tatggtgtta  | gcccgaacca  | actgaatgat  | ctgtgtttta  | ccaatgtgta  | tgccgatagc     |
|    | tttgtgattc  | gtggtgatga  | agttcgtcag  | attgcaccgg  | gtcagaccgg  | taaaattgca     |
| 10 | gattataact  | ataaactgcc  | ggatgatttt  | acgggttggt  | ttattgcctg  | gaatagcaat     |
|    | aatctggaca  | gcaaagttgg  | tggcaactat  | aactatctgt  | atcgctgtgt  | tcgtaagagc     |
|    | aatctgaaac  | cgtttgaacg  | tgatatttagc | accgagattt  | atcaggcagg  | tagcaccgcc     |
|    | tgtaatgggt  | ttgaagggtt  | taattgctat  | tttccgctgc  | agagctatgg  | ttttcagccg     |
|    | acaaatgggt  | tgggttatca  | gccgtatcgt  | gttgtgttgc  | tgtcatttga  | actgctgcat     |
|    | gcaccggcaa  | ccgtttgtgg  | tccgaaaaaa  | agtaccaatc  | tggtgaaaaa  | taagtgcgtg     |
| 15 | aactttaact  | ttaatggtct  | gaccggcacc  | ggtgttctga  | ccgaaagtaa  | caaaaaattc     |
|    | ctgccgtttc  | agcagtttgg  | ccgtgatatt  | gcagatacca  | ccgatgcagt  | tcgcgatccg     |
|    | cagacactgg  | aaattctgga  | tattaccccg  | tgcagctttg  | gtggtgtttc  | agttattaca     |
|    | ccgggtacaa  | ataccagcaa  | tcagggttgc  | gttctgtatc  | aggaatgtta  | ttgtaccgaa     |
|    | gttccggttg  | caattcatgc  | agatcagctg  | accccgacct  | ggcgtgtgta  | tagcaccggg     |
| 20 | agcaatgtgt  | ttcagacacg  | tgcaggttgt  | ctgattgggt  | cagaacatgt  | gaataatagc     |
|    | tatgaatgcg  | atattccgat  | tggtgcgggt  | atgtgtgcc   | gctatcagac  | ccagaccaat     |
|    | agtcgcgctc  | gtgcacgtag  | cgttgcaagc  | cagagcatta  | ttgcctatac  | catgagcctg     |
|    | ggtgcagaaa  | atagcgttgc  | ctatagtaat  | aacagcattg  | ccattccgac  | caactttacc     |
|    | attagcggtta | ccaccgaaat  | tctgccggtt  | agcatgacca  | aaaccagcgt  | tgattgcacc     |
| 25 | atgtatat    | gtggtgatag  | taccgaatgt  | agcaatctgc  | tgctgcagta  | tggttagcttt    |
|    | tgacccagc   | tgaatcgtgc  | actgaccggt  | attgcagttg  | aacaggataa  | aaacacgcaa     |
|    | gaagtttttg  | cacaggtcaa  | gcagatctat  | aaaaccctc   | cgattaaaga  | ttttggcggt     |
|    | ttcaatttta  | gccagatcct  | gccggatccg  | agcaaaccga  | gtaaacgtag  | ctttattgaa     |
| 30 | gatctgctgt  | tcaacaaagt  | gaccctggca  | gatgcagggt  | ttatcaaaac  | gtatggtgat     |
|    | tgctgtggcg  | atattgcccg  | acgtgatctg  | atgtgtgcac  | agaaatttaa  | cggcctgacc     |
|    | gttctgcctc  | cgctgctgac  | cgatgaaatg  | attgcacagt  | ataccagcgc  | actgctggca     |
|    | ggcaccatta  | ccagtgggtg  | gacctttggt  | gccggtgccg  | cactgcagat  | tccgtttgca     |
|    | atgcagatgg  | catatcggtt  | taatggattt  | ggtgttacct  | agaacgtgct  | gtatgaaaac     |
|    | cagaaaactga | ttgccaacca  | gttttaatagc | gccattggca  | aaattcagga  | tagcctgagc     |
| 35 | agcaccgcaa  | gtgcactggg  | taaactgcag  | gacgttggtta | atcagaatgc  | acaggcactg     |
|    | aataccctgg  | ttaaacagct  | gagcagtaat  | tttggtgcaa  | tttcaagcgt  | gctgaacgat     |
|    | attctgagcc  | gtctggataa  | agttgaagca  | gaagttcaga  | ttgatcgtct  | gattaccggg     |
|    | cgtctgcaaa  | gcctgcagac  | ctatgtgacc  | cagcagctga  | ttcgcgcagc  | agaaattcgt     |
|    | gcaagcgcaa  | atctggcagc  | cacaaaaatg  | agcgaatgtg  | ttctgggtca  | gagcaaacgt     |
| 40 | gttgattttt  | gcggcaaaag  | ttatcacctg  | atgagctttc  | cgcagagcgc  | accgcatggt     |
|    | gttgtgtttc  | tgcatgtttc  | ctatgttccg  | gcacaagaaa  | aaaactttac  | aaccgctccg     |
|    | gcaattttgc  | atgatggtaa  | agcacatttt  | ccgcgtgaag  | gtgtttttgt  | tagtaatggc     |
|    | acccattggt  | ttgttacaca  | gcgcaacttt  | tatgaaccgc  | agattattac  | aaccgacaac     |
|    | acctttgtta  | gcggtaactg  | tgatgtttgt  | attggcattg  | tgaataacac  | cgtttatgat     |
| 45 | ccactgcagc  | cggaactgga  | tagctttaaa  | gaagaactgg  | acaaatattt  | caaaaaccac     |
|    | accagtccgg  | atgttgatct  | gggtgatatt  | tcaggatatta | atgccagcgt  | ggtgaacatc     |
|    | cagaaagaaa  | ttgatcgctt  | gaatgaagtg  | gccaaaaatc  | tgaatgaaag  | cctgattgat     |
|    | ctgcaagaac  | tggggaaata  | tgagcagtat  | atcaaatggc  | cgtggtatat  | ttggctgggt     |
|    | tttattgcag  | gcctgattgc  | aattgttatg  | gtgaccatta  | tgctgtgttg  | tatgaccagc     |
| 50 | tggtgtagct  | gtctgaaagg  | ttgttgcagc  | tgccgtagct  | gttgcaaat   | tgatgaagat     |
|    | gatagcgaac  | cgggtgctgaa | aggtgttaaa  | ctgcattata  | cctaatagagc | <u>ggccgcg</u> |

The 5' SacI single cloning site is single-underlined

55 The HEV (p239 fragment) sequence is shown in capital letters

The 2019-nCoV spike protein encoding sequence is shown in lower case letters

The 3' NotI single cloning site is dash-underlined

SEQ ID NO: 4 – 2019-nCoV spike protein nucleic acid sequence – optimised for expression in *Komagataella pastoris* and containing BstB1 and NotI single cloning sites. Described in Example

3

5

10

15

20

25

30

35

40

45

50

**TTCGAAacga** tgttctgtgtt cttggctctg ttgccattgg tttcttccca gtgtgttaac  
ctgaccacta gaactcaatt gctccagcc tacaccaatt ccttcaccag aggtgtttac  
taccagaca aggtgttcag atcttccgtc ttgcaactca ctcaggactt gttcttgcca  
ttcttctcca acgttacctg gttccacgct attcacgttt ccggaactaa cggtagtaag  
agattcgaca acccagtcct gccattcaac gatggtgtct acttcgcttc taccgagaag  
tccaacatca tcagagggtg gatcttcggt actacccctg actctaagac tcagtccttg  
ctgatcggtta acaacgccac caacgttggt atcaagggtt gcgagttcca gttctgcaac  
gaccattct tgggtgtgta ctaccacaag aacaacaagt cttggatgga atccaggttc  
agagtttact cctccgcca caactgtacc ttcgagtaag tttccagcc attcttgatg  
gacttggagg gtaagcagg taacttcaag aacctgagag agttcgtttt caagaacatc  
gacggttact tcaagatcta ctccaagcac accccaatca acctggttag agatttgcca  
caaggtttct ccgctttgga gcttttggtt gacttgcca tcggtatcaa catcaccaga  
ttccagacct tgttggcctt gcacagatcc tacttgactc caggtgatcc ttcttccggt  
tggactgctg gtgctgctgc ttactatggt ggttacttgc agccaagaac ctctctgctg  
aagtacaacg agaacggaac tatcactgac gctgttgact gtgctttgga cccattgtct  
gagactaagt gcaccttgaa gtccttcacc gttgagaagg gtatctacca gacctccaac  
ttcagagttc agccaactga gtccatcgct agattcccaa acatcactaa cttgtgcca  
ttcgttgagg tgttcaacgc tactagattc gcttctggtt acgcttgga cagaagaga  
atctccaact gcgttgctga ctactccgtc ttgtacaact ctgcttcatt ctccacctc  
aagtgtacg gtgtttcccc aactaagttg aacgacctgt gtttactaa cgtctacgcc  
gactccttcg ttattagagg tgacgaggtt agacagatcg ctccagggtc aactggtaag  
atcgctgact acaactacaa gctgccagac gacttcaccg gttgtgttat tgcttggaa  
tccaacaacc tggactccaa ggttggtggt aactacaatt acctgtaccg tctgttcaga  
aagtccaact tgaagccatt cgagagagac atctccaccg agatctacca agctggttct  
actccatgta acggtgtcga gggtttcaac tgctacttcc cattgcaatc ctacggtttc  
caacctacca acggtgttg ataccagcca tacagagttg tcgttttgct cttcgagttg  
ttgcacgctc cagctactgt ttgtggtcca aagaagtcca ccaacttggt caagaacaa  
tgctgcaact ttaacttcaa cggcctgacc ggtactggtg ttttgactga atccaacaag  
aagtctctgc ctttccagca gttcggtaga gacattgctg aactactga cgcggttaga  
gatccacaga ctttgagat cttggacatc acccatggt ccttcggttg tgttccggt  
attaccctcg gaactaacac ctccaatcag gtcgctgtct tgtaccagga cgttaactgt  
actgaggttc cagttgctat ccacgctgac caattgactc caacttgag agtctactcc  
accggttcca acgttttcca aactagagcc ggttggttga tcggtgctga acacgtcaac  
aactcctacg agtgtgacat tccaattggt gctggtatct gtgcctccta ccaactcaa  
actaactccc caagaagggt tagatccgtt gcttcccaat ccattatcgc ttacaccatg  
tctttgggtg ccgagaactc tgttgcttac tctaacaact ctatcgctat cctaccaac  
ttcaccatct ccgttaccac tgagatcttg ccagtctcca tgaccaagac ttccggtgac  
tgtaccatgt acatctgttg tgactccact gagtgttcca acttgttgc gcaatacgg  
tccttctgca ccaggttgaa cagagctttg actggtattg ctgtcgagca agacaagaac  
actcaagagg ttttcgcca ggtgaagcag atctacaaga ctccacctat taaggacttc  
ggtggcttca acttctcca gatcttgcca gatccatcta agccctcaa gagatccttc  
attgaggacc tgcgtttcaa caaggttact ttggctgacg ccggtttcat caagcagtac  
ggtgattgct tgggtgacat tgcagctaga gacttgatct gtgccagaa gttcaacgg  
ttgaccgttt tgccacctt gttgaccgac gagatgatcg ctcagtacac ttctgctttg  
ttggccggta ctatcacttc tgggttgaca ttggagctg gtgccgcat gcaaattcca  
ttcgctatgc aaatggccta cagattcaac ggtatcggtg ttaccagaa cgtcctgtac  
gagaaccaga agcttatcgc caaccagttc aactccgcta tcggtgaagt tcaggactcc  
ttgtcctcta ctgcttctgc cttgggaaag ttgcaggatg ttgttaacca gaatgccag  
gctttgaaca ccctgggtta gcaactgtcc tctaacttcg gtgctatctc ctccggtttg

```

aacgacatct tgtcccgttt ggacaagggt gaggctgagg ttcagatcga cagattgatac
actggtagat tgcagtccct gcagacttac gttactcagc agttgattag agctgccgag
attagagcct ctgctaactt ggctgctact aagatgtccg agtgtgtttt gggtcagtc
5 aagagagttg acttctgcgg taagggttac cacctgatgt ctttcccaca atctgctcca
cacggtgtcg ttttcttgca cgttacttac gttccagctc aagagaagaa cttcactact
gctccagcca tttgtcacga tggtaaggct cactttcctc gtgagggtgt tttcgtttcc
aacggtactc actggttcgt caccagaga aacttttacg agccacagat catcaccacc
gacaacactt tcgtttctgg taactgtgac gtcgtcatcg gtatcgtgaa caacactgtc
10 tacgatccat tgcagccaga attggactcc ttcaaagagg aactggacaa gtactttaag
aaccacactt cccagacgt tgacctgggt gatatttccg gtattaacgc ctccgtgtgc
aacatccaaa aagagatcga ccgtttgaac gaggtcgcca agaactgaa cgagtccttg
attgacttgc aagagctggg caagtacgag cagtacatta agtggccatg gtacatttgg
ctgggtttca ttgctggttt gatcgccatc gttatggtca ccatcatgtt gtgctgtatg
acctcctggt gctcctggtt gaagggtgtg tgttcctgcg gttcctgttg taagtccgac
15 gaagatgact ccgagccagt cttgaagggt gttaagttgc actacactta aGCGGCCGC

```

The 5' BstBI single cloning site is single-underlined

The 3' NotI single cloning site is dash-underlined

Immediately following the 5' SacI is an ACG codon (needed for the coding sequence to be in frame  
 20 with the ATG start codon, which immediately follows the ACG). These two codons are shown in bold  
 and italicised.

The nucleic acid sequences of SEQ ID NO: 4 translates to give the native 2019-nCoV spike protein of  
 SEQ ID NO: 1

25

**SEQ ID NO: 5 – nucleic acid encoding for fusion protein HPV18L1/2019-nCoV spike protein–  
 optimised for expression in *K. pastoris* and containing BstBI and NotI single cloning sites.  
 Described in Example 4**

```

30 TTCGAAacgatggctctttggagaccatccgacaacactgtttacttgcc
accaccatccgttgctagagttgttaacactgacgactacgttactagaa
cttccatcttctaccacgctggttcttccagattgttgactgttggttaac
ccatacttcagagttccagctggaggtggttaacaagcaagacatccaaa
35 gttttccgcttaccagtacagagttttcagagttcagttgccagacccaa
acaagtttgattgccagacacttccatctacaacccagagactcagaga
cttgtttgggcttgctggtgttgaaatcggtagaggacagccattggg
tgttgggttgctggtcaccattctacaacaagttggacgacactgaat
cttctcacgctgctacttctaacgtttccgaggatgttagagacaacgtt
tccgttgactacaagcagactcagttgtgtatcttgggttgctccagc
40 tattggtgaacattgggctaagggtactgcttgtaagtcagaccattgt
ctcaggagattgtccaccattggagttgaagaacactgttttgaggac
ggtgatatggtgatactggttacggtgctatggacttctctactttgca
ggacactaagtgtgaagttccattggacatctgtcagtcocatctgtaagt
accagactacttgcaaatgtccgctgatccatacggtgactctatgttc
45 ttctgtttgagaagagagcagttgttcgctagacacttctggaacagagc
tgggtactatgggtgacactgttccacaatccttgtaacaaaggtactg
gaatgagagcttctcctggttcttgtgtttactctccatctccatccggt
tccattgttacttccgactccagttgttcaacaagccatactggttgca
taaggctcaaggtcacaacaacggtgtttgttggcacaaccagttgttcg
50 ttactgttgttgacactactagatccactaacttgactatctgtgcttcc
actcaatctccagttccaggacaatacagcgtactaagttcaagcagta

```

ctccagacacgttgaagagtacgacttgacgttcacatcttccagttgtgta  
ctatcactttgactgctgatgttatgtcctacatccactctatgaactcc  
tccatthttggaggattggaacttcggtgttccaccaccaactacttc  
attggttgacacttacagattcggttcagtcggttgctatcacttgcaaa  
5 aggacgctgctccagctgaaaacaaggacccatacgacaagttgaagttc  
tggaacggttgacttgaaagagaagttctccttggaacttggaaccaatccc  
attgggttagaaagtthtttggttcaggctggattgagaagaaagccaacta  
tcggtccaagaagagatcagctccatccgctactacttcatccaagcca  
gctaagagagtttagagtttagagctagaaagtTCGTGTTCTTGGTCCTGTT  
10 GCCATTGGTTTCTTCCCAGTGTGTTAACTTGACCACTAGAACTCAATTGC  
CTCCAGCCTACACCAATTCTTACCAGAGGTGTTTACTACCCAGACAAG  
GTGTTTCAGATCTTCCGTCTTGCCTCCACTCAGGACTTGTTCTTGGCATT  
CTTCTCCAACGTTACCTGGTTCACGCTATTACGTTTCCGGAACCTAACG  
GTACTAAGAGATTTCGACAACCCAGTCTGCCATTCAACGATGGTGTCTAC  
15 TTCGCTTCTACCGAGAAGTCCAACATCATCAGAGGTTGGATCTTCCGTAC  
TACCCTGGACTCTAAGACTCAGTCTTGTGATCGTTAACAACGCCACCA  
ACGTTGTCTCAAGGTTTTCGAGTTCCAGTTCTGCAACGACCCATTCTTG  
GGTGTGTACTACCACAGAACAACAAGTCTTGGATGGAATCCGAGTTCAG  
AGTTTACTCTCCGCCAACAACTGTACCTTCGAGTACGTTTCCCAGCCAT  
20 TCTTGATGGACTTGGAGGGTAAGCAGGGTAACCTCAAGAACCTGAGAGAG  
TTCGTTTTCTAAGAACATCGACGGTTACTTCAAGATCTACTCCAAGCACAC  
CCCAATCAACCTGGTTAGAGATTGTCACAAAGGTTTCTCCGCTTTGGAGC  
CTTGGTTGACTTGGCAATCGGTATCAACATCACCAGATTCCAGACCTTG  
TTGGCCTTGACAGATCCTTACTTGACTCCAGGTGATTCTTCTTCCGTTG  
25 GACTGCTGGTGTCTGCTTACTATGTTGGTTACTTGCAGCCAAGAACCT  
TCCTGCTGAAGTACAACGAGAACGGAACCTATCACTGACGCTGTTGACTGT  
GCTTTGGACCCATTGTCTGAGACTAAGTGCACCTTGAAGTCTTCCACCGT  
TGAGAAGGGTATCTACCAGACCTCCAACCTCAGAGTTCAGCCAACCTGAGT  
CCATCGTCAGATTCCCAACATCACTAACCTTGTCGCCATTTCGGTGAGGTG  
30 TTCAACGCTACTAGATTTCGCTTCTGTTTACGCCTGGAACAGAAAGAGAAT  
CTCCAACCTGCGTTGCTGACTACTCCGTCTGTACAACCTGCTTCATTCT  
CCACCTTCAAGTGCTACGGTGTTCCTCCCAACTAAGTTGAACGACCTGTGT  
TTCACTAACGTCTACGCCGACTCCTTCGTTATTAGAGGTGACGAGGTAG  
ACAGATCGCTCCAGGTCAAACCTGGTAAGATCGCTGACTACAACCTACAAGC  
35 TGCCAGACGACTTCAACCGTTGTGTTATTGCTTGGAACCTCAACAACCTG  
GACTCCAAGGTTGGTGGTAACTACAATTACCTGTACCGTCTGTTCAAGAA  
GTCCAACCTGAAGCCATTTCGAGAGAGACATCTCCACCGAGATCTACCAAG  
CTGGTTCTACTCCATGTAACGGTGTGAGGGTTTCAACTGCTACTTCCCA  
TTGCAATCCTACGGTTTCCAACCTACCAACGGTGTGGATACCAGCCATA  
40 CAGAGTTTCGCTTTTGTCTTTCGAGTTGTTGCACGCTCCAGCTACTGTTT  
GTGGTCCAAAGAAGTCCACCAACTTGGTCAAGAACAATGCGTCAACTTT  
AACTTCAACGGCTGACCGGTACTGGTGTGTTGACTGAATCCAACAAGAA  
GTTCTGCTTTCCAGCAGTTCCGTAGAGACATTGCTGACACTACTGACG  
CCGTTAGAGATCCACAGACTTTGGAGATCTTGACATCACCCCATGTTCC  
45 TTCGGTGGTGTTCGTTATTACCCCTGGAACCTAACACCTCCAATCAGGT  
CGCTGTCTGTACAGGACGTTAACTGTACTGAGGTTCCAGTTGCTATCC  
ACGCTGACCAATTGACTCCAACCTGGAGAGTCTACTCCACCGGTTCCAAC  
GTTTTCCAACCTAGAGCCGGTTGTTTGATCGGTGCTGAACACGTCAACAA  
CTCTACGAGTGTGACATTCCAATTGGTGTGCTGATCTGTGCTCCTACC  
50 AAACCTCAAACCTAACTCCCAAGAAGGGCTAGATCCGTTGCTTCCCAATCC  
ATTATCGCTTACACCATGTCTTTGGGTGCCGAGAACCTCTGTTGCCTACTC  
TAACAACCTCTATCGCTATCCCTACCAACTTACCATCTCCGTTACCACTG  
AGATCTTGCCAGTCTCCATGACCAAGACTTCCGTTGACTGTACCATGTAC  
ATCTGTGGTGACTCCACTGAGTGTTCACCTTGTGCTGCAATACGGTTC  
55 CTTCTGCACCCAGTTGAACAGAGCTTTGACTGGTATTGCTGTGAGCAAG  
ACAAGAACACTCAAGAGGTTTTCGCCCAGGTGAAGCAGATCTACAGACT  
CCACCTATTAAAGACTTCGGTGGCTTCAACTTCTCCAGATTTTGGCAGA  
TCCATCTAAGCCCTCCAAGAGATCCTTCAATTGAGACCTGCTGTTCAACA  
AGGTTACTTTGGCTGACGCCGGTTTCATCAAGCAGTACGGTGATTGCTTG  
60 GGTGACATTGCAGCTAGAGACTTGATCTGTGCCAGAAAGTTCAACGGTTT  
GACCGTTTTGCCACCTTTGTTGACCGACGAGATGATCGCTCAGTACACTT

CTGCTTTGTTGGCCGGTACTATCACTTCTGGTTGGACATTTGGAGCTGGT  
 GCCGCATTGCAAATTCATTGCTATGCAAATGGCCTACAGATTCAACGG  
 TATCGGTGTACCCAGAACGTCCTGTACGAGAACCAGAAGCTTATCGCCA  
 ACCAGTTCAACTCCGCTATCGGTAAGATTGAGGACTCCTTGTCTCTACT  
 5 GCTTCTGCCCTTGGGAAAGTTGCAGGATGTTGTTAACCAGAATGCCAGGC  
 TTTGAACACCCCTGGTTAAGCAACTGTCCTCTAACTTCGGTGCTATCTCCT  
 CCGTTTTGAACGACATCTTGTCCCGTTTGGACAAGGTTGAGGCTGAGGTT  
 CAGATCGACAGATTGATCACTGGTAGATTGCAGTCCCTGCAGACTTACGT  
 TACTCAGCAGTTGATTAGAGCTGCCGAGATTAGAGCCTCTGCTAACTTGG  
 10 CTGCTACTAAGATGTCCGAGTGTGTTTTGGGTGAGTCCAAGAGAGTTGAC  
 TTCTGCGGTAAAGGTTACCACCTGATGTCTTTCCACAATCTGCTCCACA  
 CGGTGTCGTTTTCTTGCACGTTACTTACGTTCCAGCTCAAGAGAAGAAGT  
 TCACTACTGCTCCAGCCATTTGTACGATGGTAAGGCTCACTTTCCTCGT  
 GAGGGTGTCTTCTGTTTCCAACGGTACTCACTGGTTCGTACCCAGAGAAA  
 15 CTTTTACGAGCCACAGATCATCACCACCGACAACACTTTCGTTTCTGGTA  
 ACTGTGACGTCGTATCGGTATCGTGAACAACACTGTCTACGATCCATTG  
 CAGCCAGAATTGGACTCCTTCAAAGAGGAAGTGGACAAGTACTTTAAGAA  
 CCACACTTCCCCAGACGTTGACCTGGGTGATATTTCCGGTATTAAACGCCT  
 CCGTTGTCAACATCCAAAAAGAGATCGACCGTTTGAACGAGGTCGCCAAG  
 20 AACTTGAACGAGTCTTGTGATTGACTTGCAAGAGCTGGGCAAGTACGAGCA  
 GTACATTAAGTGGCCATGGTACATTTGGCTGGGTTTATTGCTGGTTTGA  
 TCGCCATCGTTATGGTCACCATCATGTTGTGCTGTATGACCTCCTGTTGC  
 TCCTGTTTGAAGGTTGTTGTTCTGCGGTTCTGTTGTAAGTTCGACGA  
 AGATGACTCCGAGCCAGTCTTGAAGGGTGTAAAGTTGCACTACACTTAAG  
 25 **CGGCCGC**

The 5' BstBI single cloning site is single-underlined

The HPV18L1 sequence is shown in lower case letters

The 2019-nCoV spike protein encoding sequence is shown in capitalised letters

30 The 3' NotI single cloning site is dash-underlined

Immediately following the 5' BstBI is an ACG codon (needed for the coding sequence to be in frame with the ATG start codon, which immediately follows the ACG). These two codons are shown in bold and italicised.

35 **SEQ ID NO: 6 – nucleic acid encoding for fusion protein HPV16L1/2019-nCoV spike protein–  
 optimised for expression in *K. pastoris* and containing BstBI and NotI single cloning sites.  
 Described in Example 5**

40 **TTCGAAacgatgt**cctttgtggttgccatctgaagctactgtttacttgcc  
 accagttccagtttctaaagttgtttccactgacgaatacgttgctagaa  
 ctaacatctactaccacgctggtagttctagattgttggtgctgttggtcat  
 ccatacttcccaattaagaagccaaacaacaacaagattttggttccaaa  
 ggtttccggttgcaatacagagttttcagaatccatttgccagatccaa  
 45 acaagtttggtttccagatacttctttctacaaccagacactcaaaga  
 cttgtttgggcttggttggttggaagttggttagaggtcaaccattggg  
 tgttggtatttctggtcaccattgttgacaagttggacgatactgaaa  
 acgcttctgcttaacgctgctaacgctgggttgataacagagaatgtatt  
 tctatggactacaagcaaaactcaattgtgtttgattggttgtaagccacc  
 aattggtgaacattggggaaaggttctccatgtactaatgttgctgtta  
 50 accctggtgattgtccaccattggaattgattaacactgttattcaagac  
 ggtgatatggttgatactggtttcggtgctatggatttcactactttgca  
 agctaacaagtctgaagttccattggacatttgtaacttccatctgtaagt  
 acccagactacattaagatggtttctgaaccatacgggtgattcctttgttc

ttctacttgagaagagaacaaatgtttgttagacacttggtcaacagagc  
tggtgctgttggtgaaaacgttccagatgacttgtagacattaaggggtctg  
gttctactgctaacttggtctcttctaactactttccaactccatctggt  
tctatggttacttctgacgtcaaattttcaacaagccatactggttgca  
5 aagagcacaaggtcataacaacggatatttgttggggttaaccaattgttgcg  
ttactgttgttgacactactagatccactaacatgtccttgtgtgctgct  
atcttacttctgaaactacttacaagaacactaacttcaaagagtactt  
gagacacggagaagaatacgaacttgcaattcattttccaattgtgtaaga  
10 ttactttgactgctgacgttatgacttacattcactctatgaactctact  
atcttgggaagattggaacttcggattgcaaccaccaccaggtggtacttt  
ggaagatacttacagattcgttacttctcaagctattgcttgtcaaaagc  
atactccacctgctccaaaagaagatccattgaagaagtacactttctgg  
gaagttaacttgaaagaaaagtctctgctgatttggatcaattcccatt  
gggtagaaaagttttgttgcaagctggattgaaggctaaaccaaagtcca  
15 ctttgggaaagagaaaaggctactccaactacttcttacttctactact  
gctaagagaaaagaagagaaaattgtTCGTGTTCTTGGTCCTGTTGCCATT  
GGTTTCTTCCAGTGTGTTAACCTGACCACTAGAACTCAATTGCCCTCAG  
CCTACACCAATTCTTCCACAGAGGTGTTACTACCCAGACAAGGTGTTT  
AGATCTTCCGTCTTGCACTCCACTCAGGACTTGTCTTGCCATTCTTCTC  
20 CAACGTTACCTGGTTCCACGCTATTACGTTTCCGGAACCTAACGGTACTA  
AGAGATTCCGACAACCCAGTCTGCCATTCAACGATGGTGTCTACTTCGCT  
TCTACCGAGAAGTCCAACATCATCAGAGGTTGGATCTTCGGTACTACCTT  
GGACTCTAAGACTCAGTCTTGTGATCGTTAACAACGCCACCAACGTTG  
TCATCAAGGTTTGCGAGTTCCAGTTCTGCAACGACCCATTCTTGGGTGTG  
25 TACTACCACAAGAACAACAAGTCTTGGATGGAATCCGAGTTCAGAGTTTA  
CTCCTCCGCCAACAACTGTACCTTCGAGTACGTTTCCAGCCATTCTTGA  
TGGACTTGGAGGGTAAGCAGGGTAACCTCAAGAACCTGAGAGAGTTCGTT  
TTCAAGAACATCGACGGTTACTTCAAGATCTACTCCAAGCACACCCCAAT  
CAACCTGGTTAGAGATTTGCCACAAGGTTCTCCGCTTGGAGCCTTGG  
30 TTGACTTGCCAAATCGGTATCAACATCACCAGATTCCAGACCTTGTGGCC  
TTGCACAGATCCTACTTGACTCCAGGTGATTCTTCTCCGTTGGACTGC  
TGGTGCTGCTGCTTACTATGTTGGTTACTTGACGCCAAGAACCTTCTGC  
TGAAGTACAACGAGAACGGAACCTATCACTGACGCTGTTGACTGTGCTTG  
GACCCATTGTCTGAGACTAAGTGCACCTTGAAGTCTTACCGTTGAGAA  
35 GGGTATCTACAGACCTCAACTTCAGAGTTCAGCCAACTGAGTCCATCG  
TCAGATTCCCAAACATCACTAAGTGTGCCCATTGCGTGAGGTGTTCAAC  
GCTACTAGATTGCTTCTGTTTACGCCTGGAACAGAAAGAGAATCTCAA  
CTGCGTTGCTGACTACTCCGTCTTGTACAACCTCTGCTTCATTCTCCACCT  
TCAAGTGCTACGGTGTTCCTCCCAACTAAGTTGAACGACCTGTGTTTCACT  
40 AACGCTACGCCGACTCCTTCGTTATTAGAGGTGACGAGGTAGACAGAT  
CGCTCAGGTCAAACCTGGAAGATCGCTGACTACAACCTACAAGCTGCCAG  
ACGACTTCACCGGTTGTGTTATTGCTTGGAACTCCAACAACCTGGACTCC  
AAGGTTGGTGGTAACTACAATTACCTGTACCGTCTGTTGAGAAAGTCAA  
CTTGAAGCCATTGAGAGAGACATCTCCACCGAGATCTACCAAGCTGGTT  
45 CTACTCCATGTAACGGTGTGCGAGGTTTCAACTGCTACTTCCCATTGCAA  
TCCTACGGTTTCCAACCTACCAACGGTGTGGATACCAGCCATACAGAGT  
TGTCGTTTTGCTTTCGAGTTGTTGCACGCTCCAGCTACTGTTTGTGGTC  
CAAAGAAGTCCACCAACTTGGTCAAGAACAATGCGTCAACTTTAACTTC  
AACGGCCTGACCGGTACTGGTGTGTTTGTGACTGAATCCAACAAGAAGTCCCT  
50 GCCTTTCCAGCAGTTCCGTTAGAGACATTGCTGACACTACTGACGCCGTTA  
GAGATCCACAGACTTTGGAGATCTTGGACATCACCCCATGTTCTTCCGT  
GGTGTTCCTTATTACCCCTGGAACCTAACACCTCCAATCAGGTGCTGT  
CTTGTAACAGGACGTTAACTGTACTGAGGTTCCAGTTGCTATCCACGCTG  
ACCAATTGACTCCAACCTGGAGAGTCTACTCCACCGGTTCCAACGTTTTT  
55 CAACTAGAGCCGGTTGTTGATCGGTGCTGAACACGTCAACAACCTCCTA  
CGAGTGTGACATTCCAATTGGTGTGCTGCTGCTCCTACCAAACTC  
AACTAACTCCCAAGAAGGGCTAGATCCGTTGCTTCCCAATCCATTATC  
GCTTACACCATGTCTTTGGGTGCCGAGAACTCTGTTGCTTACTCTAACAA  
CTCTATCGCTATCCCTACCAACTTCACCATCTCCGTTACCACTGAGATCT  
60 TGCCAGTCTCCATGACCAAGACTTCCGTTGACTGTACCATGTACATCTGT  
GGTACTCCACTGAGTGTCCAACCTGTTGCTGCAATACGGTTCCTTCTG

CACCCAGTTGAACAGAGCTTTGACTGGTATTGCTGTGCGAGCAAGACAAGA  
 ACACTCAAGAGGTTTTTCGCCAGGTGAAGCAGATCTACAAGACTCCACCT  
 ATTAAGGACTTCGGTGGCTTCAACTTCTCCAGATTTTGCCAGATCCATC  
 5 TAAGCCCTCCAAGAGATCCTTCATTGAGGACCTGCTGTTCAACAAGGTTA  
 CTTTGGCTGACGCCGGTTTCATCAAGCAGTACGGTGATTGCTTGGGTGAC  
 ATTGACAGCTAGAGACTTGATCTGTGCCAGAAAGTTCAACGGTTTGACCGT  
 TTTGCCACCTTTGTTGACCGACGAGATGATCGCTCAGTACACTTCTGCTT  
 TGTTGGCCGGTACTATCACTTCTGGTTGGACATTTGGAGCTGGTGCCGCA  
 10 TTGCAAATTCATTTCGCTATGCAAATGGCTACAGATTC AACGGTATCGG  
 TGTTACCCAGAACGTCCTGTACGAGAACCAGAAGCTTATCGCCAACCACT  
 TCAACTCCGCTATCGGTAAGATT CAGGACTCCTTGTCTCTACTGCTTCT  
 GCCTTGGGAAAGTTGCAGGATGTTGTTAACCAGAATGCCAGGCTTTGAA  
 CACCCTGGTTAAGCAACTGTCCTCTAACTTCGGTGCTATCTCCTCCGTTT  
 15 TGAACGACATCTTGTCCCGTTTGGACAAGGTTGAGGCTGAGGTT CAGATC  
 GACAGATTGATCACTGGTAGATTGCAGTCCCTGCAGACTTACGTTACTCA  
 GCAGTTGATTAGAGCTGCCGAGATTAGAGCCTCTGCTAACTTGGCTGCTA  
 CTAAGATGTCGAGTGTGTTTGGGT CAGTCCAAGAGAGTTGACTTCTGC  
 GGTAAAGGTTACCACTGATGTCTTTCCCAACAATCTGCTCCACACGGTGT  
 CGTTTTCTTGACGTTACTTACGTTCCAGCTCAAGAGAAGAACTTCACTA  
 20 CTGCTCCAGCCATTTGTCACGATGGTAAGGCTCACTTCTCTCGTGAGGGT  
 GTTTTCGTTTCCAACGGTACTCACTGGTTCGTCAACCAGAGAACTTTTA  
 CGAGCCACAGATCATCACCACCGACAACACTTTCGTTTCTGGTAACTGTG  
 ACGTCCGTCATCGGTATCGTGAACAACACTGTCTACGATCCATTGCAGCCA  
 GAATTGGACTCCTTCAAAGAGGAACTGGACAAGTACTTTAAGAACCACAC  
 25 TTCCCCAGACGTTGACCTGGGTGATATTTCCGGTATTAACGCCTCCGTTG  
 TCAACATCCAAAAGAGATCGACCGTTTGAACGAGGTCGCCAAGAACTTG  
 AACGAGTCCTTGATTGACTTGCAAGAGCTGGGCAAGTACGAGCAGTACAT  
 TAAGTGGCCATGGTACATTTGGCTGGGTTTCATTGCTGGTTTGATCGCCA  
 TCGTTATGGTCACCATCATGTTGTGCTGTATGACCTCCTGTTGCTCCTGT  
 30 TTGAAGGGTTGTTGTTTCCTGCGGTTCCGTTGTAAGTTCGACGAAGATGA  
 CTCCGAGCCAGTCTTGAAGGGTGTAAAGTTGCACTACACTTAAG**CGGCCG**  
**C**

The 5' BstBI single cloning site is single-underlined

35 The HPV16L1 sequence is shown in lower case letters

The 2019-nCoV spike protein encoding sequence is shown in capitalised letters

The 3' NotI single cloning site is dash-underlined

Immediately following the 5' BstBI is an ACG codon (needed for the coding sequence to be in frame with the ATG start codon, which immediately follows the ACG). These two codons are shown in bold and italicised.

40

**SEQ ID NO: 7 – 2019-nCoV spike protein nucleic acid sequence – optimised for expression in humans (293F) and containing NheI and NotI single cloning sites. Described in Example 6**

45 **GCTAGCgaca** *tg*ttcgtggt tctgggtgctg ctgcctctgg tgtccagcca gtgtgtgaac  
 ctgaccacca gaacacagct gcctccagcc tacaccaata gcttcaccag gggcgtgtac  
 taccgccgaca aggtgttcag atctagcgtg ctgcacagca cccaggacct gtttctgccc  
 ttcttcagca acgtgacctg gttccacgcc atccacgtgt ccggcaccaa tggcaccaag  
 agattcgaca acccgtgct gcccttcaac gatgggggtg actttgccag caccgagaag  
 50 tccaacatca tcagaggctg gatcttcggc accacactgg acagcaagac ccagagcctg  
 ctgatcgtga acaacgccac caacgtggtc atcaaagtgt gcgagttcca gttctgcaac  
 gaccattcc tgggagtcta ctaccacaag aacaacaaga gctggatgga aagcgagttc

|    |             |             |            |            |            |                  |
|----|-------------|-------------|------------|------------|------------|------------------|
|    | cggggtgtaca | gcagcgccaa  | caactgcacc | ttcgagtagc | tgtcccagcc | tttctgatg        |
|    | gacctggaag  | gcaagcaggg  | caacttcaag | aacctgcgcg | agttcgtggt | caagaacatc       |
|    | gacggctact  | tcaagatcta  | cagcaagcac | acccctatca | acctcgtgcg | ggatctgcct       |
|    | cagggctttt  | ctgctctgga  | acctctggtg | gacctgccta | tcggcatcaa | catcaccogg       |
| 5  | tttcagaccc  | tgctggccct  | gcacagatct | tacctgacac | ctggcgatag | cagctctgga       |
|    | tggacagctg  | gcgcgctgc   | ctattatgtg | ggctacctgc | agcctcggac | cttctgctg        |
|    | aagtacaacg  | agaacggcac  | catcaccgac | gccgtggatt | gtgctctgga | tccctgagc        |
|    | gagacaaagt  | gcaccctgaa  | gtccttcacc | gtggaaaagg | gcatctacca | gaccagcaac       |
|    | ttcagagtgc  | agcccaccga  | gagcatcgtg | cggttcccca | atatcaccaa | tctgtgcccc       |
| 10 | ttcggcgagg  | tgttcaatgc  | cacaagatct | gccagcgtgt | acgcctggaa | ccggaagaga       |
|    | atcagcaact  | gcgtggccga  | ctacagcgtg | ctgtacaata | gcgccagctt | cagcaccttc       |
|    | aagtgtctacg | gcgtgtcccc  | taccaagctg | aacgacctgt | gcttcaccaa | tgtgtacgcc       |
|    | gacagcttcg  | tgatcagagg  | cgacgaagtt | cggcagatcg | ctcctggaca | gacaggcaag       |
|    | atcgccgatt  | acaactacaa  | gctgcccggc | gacttcaccg | gctgctgat  | cgcttggaa        |
| 15 | agcaacaacc  | tggactccaa  | agtgcggcgg | aactacaact | acctgtaccg | gctgttccgg       |
|    | aagtccaatc  | tgaagccctt  | cgagcgggac | atctccaccg | aaatctatca | ggccggcagc       |
|    | accccttgta  | acggcgtgga  | aggttccaac | tgctacttcc | caactgcagc | ctacggcttt       |
|    | cagcctacca  | atggcgtggg  | ctatcagccc | tatagagtgg | tggtgctgag | cttcgaactg       |
|    | ctgcatgccc  | ctgctaccgt  | gtgcggccct | aagaagtcta | ccaacctggt | caagaacaaa       |
| 20 | tgctggaact  | tcaacttcaa  | cggcctgacc | ggcacaggcg | tgctgacaga | gagcaacaag       |
|    | aagttcctgc  | ctttccagca  | gtttggccgg | gatatcgccg | ataccacaga | cgccgttaga       |
|    | gatccccaga  | caactggaaat | cctggacatc | accccatgca | gctttggcgg | agtgtctgtg       |
|    | atcaccctcg  | gcaccaatac  | cagcaatcag | gtggccgtgc | tgtatcagga | cgtgaactgt       |
|    | acagaggtgc  | ccgtggccat  | tcacgcggat | caactgacac | ccacttggag | agtgtactcc       |
| 25 | accggctcca  | acgtgttcca  | gactagagcc | ggatgtctga | tcggagccga | gcacgtgaac       |
|    | aatagctacg  | agtgcgacat  | ccccatcggc | gctggcatct | gtgccagcta | ccagacacag       |
|    | acaaatagcc  | ccagacgggc  | cagaagcgtg | gcctctcaga | gcatcattgc | ctacacaaatg      |
|    | agcctgggcg  | ccgagaattc  | tgtggcctac | agcaacaact | ctatcgctat | ccccaccaac       |
|    | ttcaccatca  | gcgtgaccac  | cgagatcctg | cctgtgtcca | tgaccaagac | cagcgtggac       |
| 30 | tgaccatgt   | acatctgcgg  | cgattccacc | gagtgacgca | acctgctgct | gcagtagcgc       |
|    | agcttctgca  | cccagctgaa  | tagagccctg | acagggatcg | ccgtggaaca | ggacaagaac       |
|    | acccaagagg  | tgctcgccca  | agtgaagcag | atctacaaga | cccctcctat | caaggacttc       |
|    | ggcggcttca  | atttcagcca  | gattctgccc | gatcctagca | agcccagcaa | gcggagcttt       |
|    | atcgaggacc  | tgctgttcaa  | caaagtgaac | ctggccgacg | ccggcttcat | caagcagtat       |
| 35 | ggcgattgcc  | tggcgacat   | tgccgccaga | gatctgattt | gcgccagaa  | gtttaacgga       |
|    | ctgacagtgc  | tgctcctct   | gctgaccgat | gagatgatcg | cccagtagac | atctgctctg       |
|    | ctggccggca  | caatcaccag  | cggatggaca | tttgagctg  | gcgcagccct | gcagatcccc       |
|    | tttgctatgc  | agatggccta  | ccggttcaac | ggcatcgag  | tgaccagaa  | tgtgctgtac       |
|    | gagaaccaga  | agctgatcgc  | caaccagttc | aacagcgcca | tcggcaagat | ccaggatagc       |
| 40 | ctgtctagca  | cagccagcgc  | tctgggcaaa | ctgcaggacg | tggtcaatca | gaacgctcag       |
|    | gccctgaaca  | ccctcgtgaa  | gcagctgagc | agcaatttcg | gcgccatcag | ctccgtgctg       |
|    | aacgatatcc  | tgagccggct  | ggataaggtg | gaagccgagg | tgagatcgca | cagactgatc       |
|    | acaggcagac  | tgagagccct  | ccagacatac | gtgacccagc | agctgatcag | agccgcagag       |
|    | attagagcct  | ctgccaatct  | ggccgccacc | aagatgtctg | agtgtgtgct | gggccagagc       |
| 45 | aagagagtgg  | atttctgcgg  | caagggctac | cacctgatga | gctttccaca | gtctgctcct       |
|    | cacggcgtgg  | tgcttctgca  | cgtgacctat | gtgcccgctc | aagagaagaa | cttcacaaca       |
|    | gcccctgcca  | tctgccacga  | cggaaaggcc | cattttccta | gagaaggcgt | gttctgttcc       |
|    | aacggcacc   | attggttctg  | gacacagcgg | aacttctacg | agccccagat | catcaccacc       |
|    | gacaacacct  | tcgtgtctgg  | caactgtgac | gtcgtgatcg | gcattgtgaa | caacaccgtg       |
| 50 | tacgaccttc  | tgagcccgca  | gctggacagc | ttcaaagagg | aactggacaa | gtactttaag       |
|    | aaccacacaa  | gcccgcagct  | ggacctgggc | gatattagcg | gcatcaatgc | ctccgtggtc       |
|    | aacatccaga  | aagagatcga  | ccggctgaac | gaggtggcca | agaatctgaa | cgagagcctg       |
|    | atcgacctgc  | aagaactggg  | gaagtacgag | cagtacatca | agtggccctg | gtacatctgg       |
|    | ctgggcttta  | tcgccggact  | gattgccatc | gtgatggcca | caatcatgct | gtgctgcatg       |
| 55 | accagctgct  | gtagctgcct  | gaagggtgct | tgagctgtg  | gcagctgtg  | caagttcgac       |
|    | gaggatgata  | gcgagcctgt  | gctgaagggc | gtgaaactgc | actacacc   | <u>GC_GGCCGC</u> |

The 5' NheI single cloning site is single-underlined

The 3' NotI single cloning site is dash-underlined

Immediately following the 5' NheI is an GAC codon (needed for the coding sequence to be in frame with the ATG start codon, which immediately follows the GAC). These two codons are shown in bold and italicised.

The nucleic acid sequences of SEQ ID NO: 7 translates to give the native 2019-nCoV spike protein of SEQ ID NO: 1

- 10 **SEQ ID NO: 8 – nucleic acid encoding for fusion protein HBSAg/2019-nCoV spike protein– optimised for expression in humans (293F) and containing NheI and NotI single cloning sites. Described in Example 7**

***GCTAGC***GACatgaactttctgggcggtacgacagtatgccttggacaaaattcacaatctccgacgtctaatacac  
15 tccccatacaagttgtccaccgacttgcgccggtataggtggatgtgtctcagacgattcataatctttctcttc  
attcttcttctgtgcctgatattcttctgtggtccttctggattaccaggaatgcttcccggtgtgctcctctgatt  
cctggttcatccactacatctacgggtccctgtagaacatgcaccacacctgcacagggcacctccatgtatccg  
tcatgctgtgcaagaaaccatcagatggtaactgcacgtgcataccgatccccctcatcatggcggttgggaaa  
tttctgtgggagtgggcctcagcccggttttccTTCGTGTTTCTGGTGCTGCTGCCCTCTGGTGTCAGCCAGTGT  
20 GTGAACCTGACCACAGAACACAGCTGCCTCCAGCCTACACCAATAGCTTCACCAGGGGCGTGTACTACCCGAC  
AAGGTGTTCAAGTCTAGCGTGCTGCACAGCACCCAGGACCTGTTTCTGCCCTTCTTCAGCAACGTGACCTGGTTC  
CACGCCATCCACGTGTCCGGCACCAATGGCACCAAGAGATTTCGACAAACCCCGTGCTGCCCTTCAACGATGGGGTG  
TACTTTGCCAGCACCGAGAAGTCCAACATCATCAGAGGCTGGATCTTCGGCACCACTGGACAGCAAGACCCAG  
AGCCTGCTGATCGTGAACAACGCCACCAACGTGGTCATCAAAGTGTGCGAGTTCAGTTCTGCAACGACCCATTCT  
25 CTGGGAGTCTACTACCACAAGAACAACAAGAGCTGGATGGAAAGCGAGTTCCGGGTGTACAGCAGCGCCAACAAC  
TGCACCTTCGAGTACGTGTCCAGCCTTTCCTGATGGACCTGGAAGGCAAGCAGGGCAACTTCAAGAACCTGCGC  
GAGTTCGTGTTCAAGAACATCGACGGCTACTTCAAGATCTACAGCAAGCACACCCCTATCAACCTCGTGCGGGAT  
CTGCCTCAGGGCTTTTCTGCTCTGGAACCTCTGGTGGACCTGCCATCGGCATCAACATCACCCGGTTTCAGACC  
CTGCTGGCCCTGCACAGATCTTACCTGACACCTGGCGATAGCAGCTCTGGATGGACAGCTGGCGCCGCTGCCAT  
30 TATGTGGGCTACCTGCAGCCTCGGACCTTCTGCTGAAGTACAACGAGAACGGCACCATCACCGACGCCGTGGAT  
TGTGCTCTGGATCCCCTGAGCGAGACAAAGTGCACCTGAAGTCCTTCACCGTGAAAAGGGCATCTACCAGACC  
AGCAACTTCAGAGTGCAGCCACCGAGAGCATCGTGCGGTTCCTCAATATCACCAATCTGTGCCCTTCGGCGAG  
GTGTTCAATGCCACAAGATTTGCCAGCGGTACGCCTGGAACCGGAAGAGAATCAGCAACTGCGTGGCCGACTAC  
AGCGTGCTGTACAATAGCGCCAGCTTCAGCACCTTCAAGTGCTACGGCGGTGCCCTACCAAGCTGAACGACCTG  
35 TGCTTCACCAATGTGTACGCCGACAGCTTCGTGATCAGAGGCGACGAAGTTCGGCAGATCGCTCCTGGACAGACA  
GGCAAGATCGCCGATTACAACTACAAGCTGCCCCGACGACTTCACCGGCTGCGTGATCGCCTGGAATAGCAACAAC  
CTGGACTCCAAAGTCGGCGGCAACTACAACCTGTACCGGCTGTTCCGGAAGTCCAATCTGAAGCCCTTCGAG  
CGGGACATCTCCACCGAAATCTATCAGGCCGGCAGACCCCTTGTAACGGCGTGGAAGGCTTCAACTGCTACTTC  
CCACTGCAGTCTACGGCTTTCAGCCTACCAATGGCGTGGGCTATCAGCCCTATAGAGTGGTGGTGTGAGCTTC  
40 GAACTGCTGCATGCCCTGTACCGTGTGCGGCCCTAAGAAGTCTACCAACCTGGTCAAGAACAATGCGTGAAC  
TTCAACTTCAACGGCCTGACCGGCACAGGCGTGCTGACAGAGAGCAACAAGAAGTTCCTGCCCTTCCAGCAGTTT  
GGCCGGGATATCGCCGATACCACAGACGCCGTTAGAGATCCCCAGACACTGGAAATCCTGGACATACCCCATGC

AGCTTTGGCGGAGTGTCTGTGATCACCCCTGGCACCAATACCAGCAATCAGGTGGCCGTGCTGTATCAGGACGTG  
 AACTGTACAGAGGTGCCCCGTGGCCATTACGCGCATCAACTGACACCCACTTGGAGAGTGTACTCCACCGGCTCC  
 AACGTGTTCCAGACTAGAGCCGGATGTCTGATCGGAGCCGAGCACGTGAACAATAGCTACGAGTGCGACATCCCC  
 ATCGGCGCTGGCATCTGTGCCAGCTACCAGACACAGACAAATAGCCCCAGACGGGCCAGAAGCGTGGCCTCTCAG  
 5 AGCATCATTGCCTACACAATGAGCCTGGGCGCCGAGAATTCTGTGGCCTACAGCAACAACCTCTATCGCTATCCCC  
 ACCAACTTCACCATCAGCGTGACCACCGAGATCCTGCCTGTGTCCATGACCAAGACCAGCGTGGACTGCACCATG  
 TACATCTGCGGCGATTCCACCGAGTGCGCAACCTGCTGCTGCGTACGGCAGCTTCTGCACCCAGCTGAATAGA  
 GCCCTGACAGGGATCGCCGTGGAACAGGACAAGAACACCCAAGAGGTGTTGCCCCAAGTGAAGCAGATCTACAAG  
 ACCCCTCCTATCAAGGACTTCGGCGGCTTCAATTTAGCCAGATTCTGCCCGATCCTAGCAAGCCCAGCAAGCGG  
 10 AGCTTTATCGAGGACCTGCTGTTCAACAAAGTGACACTGGCCGACGCCGGCTTCATCAAGCAGTATGGCGATTGC  
 CTGGGCGACATTGCCGCCAGAGATCTGATTTGCGCCGAGAAGTTTAAACGGACTGACAGTGCTGCCTCCTCTGCTG  
 ACCGATGAGATGATCGCCAGTACACATCTGCTCTGCTGGCCGGCACAATCACCAGCGGATGGACATTTGGAGCT  
 GGCGCAGCCCTGCAGATCCCCTTTGCTATGCAGATGGCCTACCGGTTCAACGGCATCGGAGTGACCCAGAATGTG  
 CTGTACGAGAACCAGAAGCTGATCGCCAACAGTTCAACAGCGCCATCGGCAAGATCCAGGATAGCCTGTCTAGC  
 15 ACAGCCAGCGCTCTGGGCAAACCTGCAGGACGTGGTCAATCAGAACGCTCAGGCCCTGAACACCCTCGTGAAGCAG  
 CTGAGCAGCAATTTGCGCGCCATCAGCTCCGTGCTGAACGATATCCTGAGCCGGCTGGATAAGGTGGAAGCCGAG  
 GTGCAGATCGACAGACTGATCAGGCGAGCTGCAGAGCCTCCAGACATACGTGACCCAGCAGCTGATCAGAGCC  
 GCCGAGATTAGAGCCTCTGCCAATCTGGCCGCCACCAAGATGTCTGAGTGTGTGCTGGGCCAGAGCAAGAGAGTG  
 GATTTCTGCGGCAAGGGCTACCACCTGATGAGCTTTCCACAGTCTGCTCCTCACGGCGTGGTGTCTGCACGTG  
 20 ACCTATGTGCCCGCTCAAGAGAAGAACTTCAACAGCCCCCTGCCATCTGCCACGACGGAAGGCCCATTTTCTCT  
 AGAGAAGGCGTGTTCTGTGTCCAACGGCACCCATTGGTTCTGTGACACAGCGGAACCTTCTACGAGCCCCAGATCATC  
 ACCACCGACAACACCTTCGTGTCTGGCAACTGTGACGTCGTGATCGGCATTGTGAACAACACCGTGTACGACCCT  
 CTGCAGCCCCGAGCTGGACAGCTTCAAAGAGGAAGTGGACAAGTACTTTAAGAACCACACAAGCCCCGACGTGGAC  
 CTGGGCGATATTAGCGGCATCAATGCCCTCCGTGGTCAACATCCAGAAAGAGATCGACCGGCTGAACGAGGTGGCC  
 25 AAGAATCTGAACGAGAGCCTGATCGACCTGCAAGAACTGGGGAAGTACGAGCAGTACATCAAGTGGCCCTGGTAC  
 ATCTGGCTGGGCTTTATCGCCGGACTGATTGCCATCGTGATGGTCAACATCATGCTGTGCTGCATGACCAGCTGC  
 TGTAGCTGCCTGAAGGGCTGTTGCAGCTGTGGCAGCTGCTGCAAGTTCGACGAGGATGATAGCGAGCCTGTGCTG  
 AAGGGCGTGAAACTGCACTACACCGCGGCCGC

30 The 5' NheI single cloning site is single-underlined

The HSBAG sequence is shown in lower case letters

The 2019-nCoV spike protein encoding sequence is shown in capitalised letters

The 3' NotI single cloning site is dash-underlined

Immediately following the 5' NheI is an GAC codon (needed for the coding sequence to be in frame  
 35 with the ATG start codon, which immediately follows the GAC). These two codons are shown in bold  
 and italicised.

**SEQ ID NO: 9 – amino acid sequence corresponding to SEQ ID NO: 3**

(fusion protein HEV-2019-nCoV spike protein– optimised for expression in *E. coli* and containing  
 40 **SacI** and **NotI** single cloning sites. Described in Example 2)

MIALTLFNLADTLLGGLPTELISSAGGQLFYSRPVVSANGEPTVKLYTSVENAQQDKGIAIPHDIDLGESRVVIQ  
 DYDNQHEQDRPTPSPAPSRPFVLRANDVLWLSLTAAEYDQSTYGSSTGPVYVSDSVTLVNVATGAQAVARSLDW

TKVTLDGRPLSTIQQYSKTFVFLPLRGKLSFWEAGTTKAGYPYNYNTTASDQLLVENAAGHRVAISTYTTSLGAG  
 PVSISAVAVLAPHSAFVFLVLLPLVSSQCVNLTTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSN  
 VTWFHAIHVS GTNGTKRFDNPVLPFNDGVYFAS TEKSN IIRGWI FGTTLD SKTQSL L I VNNATNVVIK VCE FQFC  
 5 ND PFLGVY YHKNNKSWMESEFRVYSSANNCTFEYVSQPF LMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPIN  
 LVRDLPQGFSALEPLVDLPIGINITRFQTL LALHRSYLT PGDSSSGWTAGAAAYVGYLQPRTFLLKY NENG TIT  
 DAVDCALDPLSETKCTLKSF TVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGVEFNATRFASVYAWNRRKISNC  
 VADYSVLYNSASFSTFKCYGVSPTKLN DLCTNVYADSFVIRGDEV RQIAPGQTGKIADYNYKLPDDFTGCVIAW  
 NSNNLDSKVGGNYNLYRLFRKSNLKPFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV  
 10 VLSFELLHAPATVCGPKKSTNLVKNKCVNFENGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILD  
 ITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTRVYSTGSNVFQTRAGCLIGA EHVNSYE  
 CDIPIGAGICASYQTQTN SPRRARSVASQSIIAYTMSLGAENSVAYSNN SIAIPTNFTISVTTEILPVSMTKTSV  
 DCTMYICGDSTEC SNLL LQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSK  
 PSKRSFIEDLLFNKVT LADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMI AQYTSALLAGTITSGW  
 TFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFN SAIGKIQDSL S STASALGKLQDVVNQNAQALNT  
 15 LVKQLSSNFGAIS SVLNDILSR LDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQ  
 SKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAI CHDGKAHFPREGVFVSNGTHWFVTQRNFYE  
 PQIITTDNTFVSGNCDVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRL  
 NEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCMTSCC SCLKGCCSCGSCCKFDEDD S  
 20 EPVLKGVKLHYT

**SEQ ID NO: 10 – amino acid sequence corresponding to SEQ ID NO: 5**

**(fusion protein HPV18L1/2019-nCoV spike protein– optimised for expression in *K. pastoris* and containing BstB1 and NotI single cloning sites. Described in Example 4)**

25 MALWRPSDNTVYLP PPSVARVVNTDDYVTRTSIFYHAGSSRLLTVGNPYFRVPAGGGNKQDIPKVSAYQYRVFRV  
 QLPDPNKFGLPDTSIYNPETQRLVWACAGVEIGRGQPLGVGLSGHPFYNKLDDTES SHAATS NVSEDVRDNVSVD  
 YKQTQLCILGCAPAIGEHWAKGTACKSRPLSQGDCCPPELKN TVLEDGDMVD TGYGAMDFSTLQDTKCEVPLDIC  
 30 QSICKYPDY LQMSADPYGDSMFFCLRREQLFARHFWRAGTMGDTV PQSLYIKGTGMRASPGSCVYSPSPSGSIV  
 TSDSQLFNKPYWLHKAQGHNNGV CWHNQLFVTVD TTRSTNL TICASTQSPVPGQYDATKFKQYSRHVEEYDLQF  
 IFQLCTITLTADVMSYIHS MNSSI LEDWNFGVPPPTTSLVD TYRFVQSVAITCQKDAAPAENKDPYDKLKFWNV  
 DLKEKFSLDLDQYPLGRKFLVQAGLRKPTIGPRKRSAPSATTSSKPAKRVRVRARKFVFLVLLPLVSSQCVNL T  
 TRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVS GTNGTKRFDNPVLPFNDGVYFAS  
 35 TEKSN IIRGWI FGTTLD SKTQSL L I VNNATNVVIK VCE FQFCND PFLGVY YHKNNKSWMESEFRVYSSANNCTFE  
 YVSQPF LMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTL LAL  
 HRSYLT PGDSSSGWTAGAAAYVGYLQPRTFLLKY NENG TITDAVDCALDPLSETKCTLKSF TVEKGIYQTSNFR  
 VQPTESIVRFPNITNLCPFGVEFNATRFASVYAWNRRKISNCVADYSVLYNSASFSTFKCYGVSPTKLN DLCTN  
 VYADSFVIRGDEV RQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERDIS  
 40 TEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVNFEN  
 GLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTE  
 VPVAIHADQLTPTRVYSTGSNVFQTRAGCLIGA EHVNSYECDIPIGAGICASYQTQTN SPRRARSVASQSIIA  
 YTMSLGAENSVAYSNN SIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTEC SNLL LQYGSFCTQLNRALTG  
 45 IAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVT LADAGFIKQYGDCLGDI  
 AARDLICAQKFNGLTVLPPLLTDEMI AQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYEN  
 QKLIANQFN SAIGKIQDSL S STASALGKLQDVVNQNAQALNTLVKQLSSNFGAIS SVLNDILSR LDKVEAEVQID  
 RLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYV  
 PAQEKNFTTAPAI CHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVIGIVNNTVYDPLQPE  
 50 LDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWL  
 GFIAGLIAIVMVTIMLCMTSCC SCLKGCCSCGSCCKFDEDDSEPV LKGVKLHYT

**SEQ ID NO: 11 – amino acid sequence corresponding to SEQ ID NO: 6**

**(fusion protein HPV16L1/2019-nCoV spike protein– optimised for expression in *K. pastoris* and containing BstB1 and NotI single cloning sites. Described in Example 5)**

55 MSLWLPSEATVYLP PVPVSKVVSTDEYVARTNIYYHAGTSRLLAVGHPYFPIKKPNNNKILVPKVSGLQYRVFRI  
 HLPDPNKFGFPDTSFYNPDTQRLVWACVGVVEVGRGQPLGVGISGHPLLNKLDDETENASAYAANAGVDNRECI SMD

YKQTQLCLIGCKPPIGEHWGKSGPCTNVAVNP GD C P P L E L I N T V I Q D G M V D T G F G A M D F T T L Q A N K S E V P L D I C  
 T S I C K Y P D Y I K M V S E P Y G D S L F F Y L R R E Q M F V R H L F N R A G A V G E N V P D D L Y I K G S G S T A N L A S S N Y F P T P S G S M V  
 T S D A Q I F N K P Y W L Q R A Q G H N N G I C W G N Q L F V T V V D T T R S T N M S L C A A I S T S E T T Y K N T N F K E Y L R H G E E Y D L Q F I  
 F Q L C K I T L T A D V M T Y I H S M N S T I L E D W N F L Q P P P G G T L E D T Y R F V T S Q A I A C Q K H T P P A P K E D P L K K Y T F W E V N  
 5 L K E K F S A D L D Q F P L G R K F L L Q A G L K A K P K F T L G K R K A T P T T S S T S T A K R K R K L F V F L V L L P L V S S Q C V N L T T R  
 T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L F L P F F S N V T W F H A I H V S G T N G T K R F D N P V L P F N D G V Y F A S T E  
 K S N I I R G W I F G T T L D S K T Q S L L I V N N A T N V V I K V C E F Q F C N D P F L G V Y Y H K N N K S W M E S E F R V Y S S A N N C T F E Y V  
 S Q P F L M D L E G K Q G N F K N L R E F V F K N I D G Y F K I Y S K H T P I N L V R D L P Q G F S A L E P L V D L P I G I N I T R F Q T L L A L H R  
 S Y L T P G D S S S G W T A G A A Y Y V G Y L Q P R T F L L K Y N E N G T I T D A V D C A L D P L S E T K C T L K S F T V E K G I Y Q T S N F R V Q  
 10 P T E S I V R F P N I T N L C P F G E V F N A T R F A S V Y A W N R K R I S N C V A D Y S V L Y N S A S F S T F K C Y G V S P T K L N D L C F T N V Y  
 A D S F V I R G D E V R Q I A P G Q T G K I A D Y N Y K L P D D F T G C V I A W N S N N L D S K V G G N Y N Y L R L F R K S N L K P F E R D I S T E  
 I Y Q A G S T P C N G V E G F N C Y F P L Q S Y G F Q P T N G V G Y Q P Y R V V L S F E L L H A P A T V C G P K K S T N L V K N C V N F N F N G L  
 T G T G V L T E S N K K F L P F Q Q F G R D I A D T T D A V R D P Q T L E I L D I T P C S F G G V S V I T P G T N T S N Q V A V L Y Q D V N C T E V P  
 V A I H A D Q L T P T W R V Y S T G S N V F Q T R A G C L I G A E H V N N S Y E C D I P I G A G I C A S Y Q T Q T N S P R R A R S V A S Q S I I A Y T  
 15 M S L G A E N S V A Y S N N S I A I P T N F T I S V T T E I L P V S M T K T S V D C T M Y I C G D S T E C S N L L L Q Y G S F C T Q L N R A L T G I A  
 V E Q D K N T Q E V F A Q V K Q I Y K T P P I K D F G G F N F S Q I L P D P S K P S K R S F I E D L L F N K V T L A D A G F I K Q Y G D C L G D I A A  
 R D L I C A Q K F N G L T V L P P L L T D E M I A Q Y T S A L L A G T I T S G W T F G A G A A L Q I P F A M Q M A Y R F N G I G V T Q N V L Y E N Q K  
 L I A N Q F N S A I G K I Q D S L S S T A S A L G K L Q D V V N Q N A Q A L N T L V K Q L S S N F G A I S S V L N D I L S R L D K V E A E V Q I D R L  
 I T G R L Q S L Q T Y V T Q Q L I R A A E I R A S A N L A A T K M S E C V L G Q S K R V D F C G K G Y H L M S F P Q S A P H G V V F L H V T Y V P A Q  
 20 E K N F T T A P A I C H D G K A H F P R E G V F V S N G T H W F V T Q R N F Y E P Q I I T T D N T F V S G N C D V V I G I V N N T V Y D P L Q P E L D  
 S F K E E L D K Y F K N H T S P D V D L G D I S G I N A S V V N I Q K E I D R L N E V A K N L N E S L I D L Q E L G K Y E Q Y I K W P W Y I W L G F I  
 A G L I A I V M V T I M L C C M T S C C S C L K G C C S C G S C K F D E D D S E P V L K G V K L H Y T

- 25 **SEQ ID NO: 12 – amino acid sequence corresponding to SEQ ID NO: 8  
 (fusion protein HBSAg/2019-nCoV spike protein– optimised for expression in humans (293F) and  
 containing NheI and NotI single cloning sites. Described in Example 7)**

M N F L G G T T V C L G Q N S Q S P T S N H S P T S C P P T C P G Y R W M C L R R F I I F L F I L L L C L I F L L V L L D Y Q G M L P V C P L I P G S  
 30 S T T S T G P C R T C T T P A Q G T S M Y P S C C C T K P S D G N C T C I P I P S S W A F G K F L W E W A S A R F S F V F L V L L P L V S S Q C V N L  
 T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L F L P F F S N V T W F H A I H V S G T N G T K R F D N P V L P F N D G V Y F A  
 S T E K S N I I R G W I F G T T L D S K T Q S L L I V N N A T N V V I K V C E F Q F C N D P F L G V Y Y H K N N K S W M E S E F R V Y S S A N N C T F  
 E Y V S Q P F L M D L E G K Q G N F K N L R E F V F K N I D G Y F K I Y S K H T P I N L V R D L P Q G F S A L E P L V D L P I G I N I T R F Q T L L A  
 L H R S Y L T P G D S S S G W T A G A A Y Y V G Y L Q P R T F L L K Y N E N G T I T D A V D C A L D P L S E T K C T L K S F T V E K G I Y Q T S N F  
 35 R V Q P T E S I V R F P N I T N L C P F G E V F N A T R F A S V Y A W N R K R I S N C V A D Y S V L Y N S A S F S T F K C Y G V S P T K L N D L C F T  
 N V Y A D S F V I R G D E V R Q I A P G Q T G K I A D Y N Y K L P D D F T G C V I A W N S N N L D S K V G G N Y N Y L R L F R K S N L K P F E R D I  
 S T E I Y Q A G S T P C N G V E G F N C Y F P L Q S Y G F Q P T N G V G Y Q P Y R V V L S F E L L H A P A T V C G P K K S T N L V K N C V N F N F  
 N G L T G T G V L T E S N K K F L P F Q Q F G R D I A D T T D A V R D P Q T L E I L D I T P C S F G G V S V I T P G T N T S N Q V A V L Y Q D V N C T  
 E V P V A I H A D Q L T P T W R V Y S T G S N V F Q T R A G C L I G A E H V N N S Y E C D I P I G A G I C A S Y Q T Q T N S P R R A R S V A S Q S I I  
 40 A Y T M S L G A E N S V A Y S N N S I A I P T N F T I S V T T E I L P V S M T K T S V D C T M Y I C G D S T E C S N L L L Q Y G S F C T Q L N R A L T  
 G I A V E Q D K N T Q E V F A Q V K Q I Y K T P P I K D F G G F N F S Q I L P D P S K P S K R S F I E D L L F N K V T L A D A G F I K Q Y G D C L G D  
 I A A R D L I C A Q K F N G L T V L P P L L T D E M I A Q Y T S A L L A G T I T S G W T F G A G A A L Q I P F A M Q M A Y R F N G I G V T Q N V L Y E  
 N Q K L I A N Q F N S A I G K I Q D S L S S T A S A L G K L Q D V V N Q N A Q A L N T L V K Q L S S N F G A I S S V L N D I L S R L D K V E A E V Q I  
 D R L I T G R L Q S L Q T Y V T Q Q L I R A A E I R A S A N L A A T K M S E C V L G Q S K R V D F C G K G Y H L M S F P Q S A P H G V V F L H V T Y V  
 45 P A Q E K N F T T A P A I C H D G K A H F P R E G V F V S N G T H W F V T Q R N F Y E P Q I I T T D N T F V S G N C D V V I G I V N N T V Y D P L Q P  
 E L D S F K E E L D K Y F K N H T S P D V D L G D I S G I N A S V V N I Q K E I D R L N E V A K N L N E S L I D L Q E L G K Y E Q Y I K W P W Y I W L  
 G F I A G L I A I V M V T I M L C C M T S C C S C L K G C C S C G S C K F D E D D S E P V L K G V K L H Y T A A

- 50 **SEQ ID NO: 13 – RBD 2019-nCoV spike protein nucleic acid sequence**

G C T A G C G A C g c c a c c A T G A G A G T C C A A C C A A C A G A A T C T A T T G T T A G A T T  
 T C C T A A T A T T A C A A A C T T G T G C C C T T T T G G T G A A G T T T T T A A C G C C A C C A  
 G A T T T G C A T C T G T T T A T G C T T G G A A C A G G A A G A G A A T C A G C A A C T G T G T T  
 G C T G A T T A T T C T G T C C T A T A T A A T T C C G C A T C A T T T T C C A C T T T T A A G T G  
 55 T T A T G G A G T G T C T C T A C T A A A T T A A A T G A T C T C T G C T T T A C T A A T G T C T  
 A T G C A G A T T C A T T T G T A A T T A G A G G T G A T G A A G T C A G A C A A A T C G C T C C A

- GGGCAAACCTGGAAAGATTGCTGATTATAATTATAAAATTACCAGATGATTT  
TACAGGCTGCGTTATAGCTTGGAATTCTAACAATCTTGATTCTAAGGTTG  
GTGGTAATTATAATTACCTGTATAGATTGTTTAGGAAGTCTAATCTCAAA  
CCTTTTGAGAGAGATATTTCAACTGAAATCTATCAGGCCGGTAGCACACC  
5 TTGTAATGGTGTGGAAGGTTTTAATTGTTACTTTCCTTTACAATCATATG  
GTTTCCAACCCACTAATGGTGTGTTGTTACCAACCATACAGAGTAGTAGTA  
CTTTCTTTTGAACCTTCTACATGCACCAGCAACTGTTTGTGGACCTAAAAA  
GtgataaGCGGCCGC
- 10 KOZAC sequence added (gcc acc, underlined) before the starting ATG (**bold**).  
Secreted form tga taa added (double underlined) before NotI – This tga taa sequence is a “two  
stop codon” motif that interrupts protein synthesis, facilitating secretion into the extracellular  
medium (also included in other sequences, as described below).  
Unique Restriction sites have been added respectively at 5’ end NheI and at the 3’ end, NotI  
15 (dash underlined)

**SEQ ID NO: 14 – RBD 2019-nCoV spike protein nucleic acid sequence - human codon  
optimized for 293F (HEK) cell expression.**

- 20 GCTAGCGACgccacc**ATG**AGAGTGCAGCCTACAGAGTCTATCGTGCGGTTCCCCAA  
CATCACCAATCTGTGCCCTTTCGGCGAGGTGTTCAACGCCACAAGATTTGCCAGC  
GTGTACGCCTGGAACCGGAAGAGAATCAGCAACTGCGTGGCCGACTACAGCGTG  
CTGTACAATAGCGCCAGCTTCAGCACCTTCAAGTGCTACGGCGTGTCCCCTACCA  
AGCTGAACGACCTGTGCTTCACCAATGTGTACGCCGACAGCTTCGTGATCAGAGG  
25 CGACGAAGTTCGGCAGATCGCTCCTGGACAGACAGGCAAGATCGCCGATTACAA  
CTACAAGCTGCCCCGACGACTTCACCGGCTGCGTGATCGCCTGGAATAGCAACAA  
CCTGGACAGCAAAGTCGGCGGCAACTACAACCTACCTGTACCGGCTGTTCCGGAA  
GTCCAACCTGAAGCCTTTCGAGCGGGACATCAGCACCGAGATCTATCAGGCCGG  
CAGCACCCCTTGTAATGGCGTGGAAGGCTTCAACTGCTACTTCCCCTGCAGTCC  
30 TACGGCTTCCAGCCTACAAACGGCGTGGGCTACCAGCCTTATAGAGTGGTGGTGC  
TGAGCTTCGAACCTGCTGCATGCCCCCTGCTACAGTGTGCGGCCCAAGAAGtgata  
aGCGGCCGC

- KOZAC sequence added (gcc acc, underlined) before the starting ATG (**bold**).  
35 Secreted form tga taa added (double underlined) before NotI  
Unique Restriction sites have been added respectively at 5’ end NheI and at the 3’ end, NotI  
(dash underlined)

- 40 **SEQ ID NO: 15 – RBD 2019-nCoV spike protein amino acid sequence corresponding to SEQ ID  
NOs: 13 and 14**

- MRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLVNSASF  
TFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRIAPGQTGKIADYNYKLPDDFTGC  
VIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCY  
45 FPLQSYGFQPTNGVGYQPYRVVLSFELLHAPATVCGPKK

**SEQ ID NO: 16 – rigid EAAAK linker consensus amino acid sequence**

A(EAAK)<sub>n</sub>A (n = 2-5)

**SEQ ID NO: 17 – rigid (EAAAK)<sub>3</sub> linker nucleic acid sequence**

GAA GCC GCC GCT AAA GAG GCC GCT GCC AAA GAA GCT GCT GCT AAG

**SEQ ID NO: 18 – rigid (EAAK)<sub>3</sub> linker amino acid sequence**

EAAAKEAAAKEAAAK

**SEQ ID NO: 19 – flexible GS<sub>n</sub> linker consensus amino acid sequence**

(Gly-Gly-Gly-Gly-Ser)<sub>n</sub> (n=1-6)

**SEQ ID NO: 20 – flexible GS5 ((GGGGS)<sub>1</sub>) linker amino acid sequence**

GGGGS

**SEQ ID NO: 21 – flexible GS10 ((GGGGS)<sub>2</sub>)linker amino acid sequence**

GGGGSGGGGS

**SEQ ID NO: 22 – flexible GS15 linker nucleic acid sequence**

GGT GGT GGT GGT AGC GGT GGT GGC GGT TCA GGT GGC GGT GGT TCA

**SEQ ID NO: 23 – flexible GS15 ((GGGGS)<sub>3</sub>) linker amino acid sequence**

GGGGS GGGGS GGGGS

**SEQ ID NO: 24 – flexible GS20 ((GGGGS)<sub>4</sub>) linker amino acid sequence**

GGGGSGGGGSGGGGSGGGGSG

**SEQ ID NO: 25 – flexible GS25 ((GGGGS)<sub>5</sub>) linker amino acid sequence**

GGGGS GGGGS GGGGS GGGGS GGGGS

**SEQ ID NO: 26 – HBSAg-(EAAAK)<sub>3</sub>-RBD nucleic acid sequence**

GCTAGCGACgccacc**AT**GATTGCACTGACCCTGTTTAATCTGGCAGATAC  
CCTGTTAGGTGGTCTGCCGACCGAACTGATTAGCAGTGCCGGTGGTCAGC  
TGTTTTATAGCCGTCCGGTTGTTAGCGCAAATGGTGAACCGACCGTTAAA  
CTGTATACCAGCGTTGAAAATGCACAGCAGGATAAAGGTATTGCAATTCC  
GCATGATATTGATCTGGGTGAAAGCCGTGTTGTGATTTCAGGATTATGATA  
ATCAGCATGAACAGGATCGTCCGACACCGAGTCCGGCACCAGCCGTCCG  
TTTAGCGTTCTGCGTGCAAATGATGTTCTGTGGCTGAGCCTGACCGCAGC  
AGAATATGATCAGAGCACCTATGGTAGCAGCACCGGTCCGGTTTATGTTA

GCGATAGCGTTACCCTGGTTAATGTTGCAACCGGTGCACAGGCAGTTGCA  
 CGTAGCCTGGATTGGACCAAAGTGACCTGGATGGTCGTCCGCTGAGCAC  
 CATTACAGCAGTATAGCAAAACCTTTTTTGTCTGCCGCTGCGTGGTAAAC  
 TGAGCTTTTGGGAAGCAGGCACCACCAAAGCAGGTTATCCGTATAACTAT  
 5 AATACCACCGCAAGCGATCAGCTGCTGGTTGAAAACGCAGCAGGTCATCG  
 TGGTGAATTAGCACCTATACCACCAGTTTAGGTGCAGGTCCGGTTAGCA  
 TTAGCGCAGTTGCAGTTCTGGCACCGCATTACAGCCgaagcagccgctaaa  
gaagcagccgctaaagaagcagccgctaaaAGAGTCCAACCAACAGAATC  
 TATTGTTAGATTTCTTAATATTACAAACTTGTGCCCTTTTGGTGAAGTTT  
 10 TTAACGCCACCAGATTTGCATCTGTTTATGCTTGGAACAGGAAGAGAATC  
 AGCAACTGTGTTGCTGATTATTCTGTCTATATAATTCCGCATCATTTTC  
 CACTTTTAAAGTGTATGGAGTGTCTCCTACTAAATTAAATGATCTCTGCT  
 TTTACTAATGTCTATGCAGATTCATTTGTAATTAGAGGTGATGAAGTCAGA  
 CAAATCGCTCCAGGGCAAACCTGGAAAGATTGCTGATTATAATTATAAATT  
 15 ACCAGATGATTTTACAGGCTGCGTTATAGCTTGGAATTCTAACAATCTTG  
 ATTCTAAGGTTGGTGGTAATTATAATTACCTGTATAGATTGTTTAGGAAG  
 TCTAATCTCAAACCTTTTGAGAGAGATATTCAACTGAAATCTATCAGGC  
 CGGTAGCACACCTTGTAATGGTGTGAAAGTTTTAATTGTTACTTTCCTT  
 TACAATCATATGGTTTCCAACCCACTAATGGTGTGTTACCAACCATAC  
 20 AGAGTAGTAGTACTTTCTTTTGAACTTCTACATGCACCAGCAACTGTTTG  
 TGGACCTAAAAAGtgataaGCGGCCGC

KOZAC sequence added (gcc acc, underlined) before the starting ATG (**bold**).

Secreted form tga taa added (double underlined) before NotI

25 Unique Restriction sites have been added respectively at 5' end NheI and at the 3' end, NotI (dash underlined)

The bold and dotted underlined sequence corresponds to the (**EAAAK**)<sub>3</sub> linker.

30 **SEQ ID NO: 27 – HBSAg-(EAAAK)<sub>3</sub>-RBD nucleic acid sequence human codon optimised for 293f (HEK) cell expression**

GCTAGCGACgccacc**ATGA**ATTTTCTCGGCGGCACAACAGTGTGCCTGGGCCAGAA  
 TAGCCAGTCTCCTACCAGCAATCACAGCCCCACCAGCTGTCCTCCAACCTGTCCT  
 35 GGCTACAGATGGATGTGCCTGCGGCGGTTTCATCATCTTTCTGTTTCATCCTGCTGCT  
 GTGCCTGATCTTCCTGCTGGTGTGCTGGATTACCAGGGAATGCTGCCTGTGTGT  
 CCTCTGATCCCTGGCAGCAGCACACAAGCACAGGCCCTTGCAGAACCTGCACA  
 ACACCAGCTCAGGGCACCAGCATGTACCCTAGCTGCTGTTGTACCAAGCCTAGCG  
 ACGGCAACTGCACATGCATCCCCATTCTAGCAGCTGGGCCTTCGGCAAGTTTCT  
 40 GTGGGAATGGGCCAGCGCCAGATTTTCCGAAGCCGCCGCTAAAGAGGCCGCTGC  
 CAAAGAAGCTGCTGCTAAGAGAGTGCAGCCCACC**AGTCTATCGTGC****GGTTCC**  
**CCAACATCACCAATCTGTGCCCTTT**CGGCGAGGTGTTCAACGCCACAAGATTT  
 GCCAGCGTGTACGCCTGGAACCGGAAGAGAATCAGCAACTGCGTGGCCGACTAC  
 AGCGTGCTGTACAATAGCGCCAGCTTCAGCACCTTCAAGTGCTACGGCGTGTCCC  
 45 CTACCAAGCTGAACGACCTGTGCTTCACCAATGTGTACGCCGACAGCTTCGTGAT  
 CAGAGGCGACGAAGTTCGGCAGATCGCTCCTGGACAGACAGGCAAGATCGCCGA  
 TTACAACCTACAAGCTGCCCGACGACTTCACCGGCTGCGTGATCGCCTGGAATAGC  
 AACAACCTGGACAGCAAAGTCGGCGGCAACTACAACCTACCTGTACCGGCTGTTC  
 CGGAAGTCCAACCTGAAGCCTTTTCGAGCGGGACATCAGCACCGAAATCTACCAG  
 50 GCCGGCAGCACCCCTTGTAATGGCGTGGAAGGCTTCAACTGCTACTTCCCACTGC

AGTCCTACGGCTTCCAGCCTACAAACGGCGTGGGCTACCAGCCTTATAGAGTGGT  
GGTGCTGAGCTTCGAACTGCTGCATGCCCCTGCTACAGTGTGCGGCCCAAGAAG  
tgataaGCGGCCGC

- 5 KOZAC sequence added (gcc acc, underlined) before the starting ATG (**bold**).  
Secreted form tga taa added (double underlined) before NotI  
Unique Restriction sites have been added respectively at 5' end NheI and at the 3' end, NotI  
(dash underlined)  
The bold and dotted underlined sequence corresponds to the (**EAAAK**)<sub>3</sub> linker.

10

**SEQ ID NO: 28 – HBSAg-(EAAAK)<sub>3</sub>-RBD amino acid sequence corresponding to SEQ ID NOs:  
26 AND 27**

- MNFLGGTTVCLGQNSQSPTSNSHPTSCPPTCPGYRWMCLRRFIIFLLCLIFLLVLL  
15 DYQGMLPVCPLIPGSSTTSTGPCRTCTTPAQGTSMYPSCCCTKPSDGNCTCIPISSWA  
FGKFLWEWASARFSEAAAKEAAAKEAAAKRVQPTESIVRFPNITNLCPFGEVFNATR  
FASVYAWNRRKRISNCVADYSVLYNSASFSTFKCYGVSP TKLNDLCFTNVYADSFVIR  
GDEV RQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKS  
NLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSEFEL  
20 LHAPATVCGPKK

The (EAAAK)<sub>3</sub> linker is underlined.

- 25 **SEQ ID NO: 29– HEV-GS15-RBD nucleic acid sequence**

- GAGCTC**ATG**ATTGCACTGACCCTGTTTAATCTGGCAGATACCCTGCTGGG  
TGGTCTGCCGACCGAACTGATTAGCAGTGCCGGTGGTCAGCTGTTTTATA  
GCCGTCCGGTTGTTAGCGCAAATGGTGAACCGACCGTTAACTGTATACC  
30 AGCGTTGAAAATGCACAGCAGGATAAAGGTATTGCAATTCCGCATGATAT  
TGATCTGGGTGAAAGCCGTGTTGTGATTCAGGATTATGATAATCAGCATG  
AACAGGATCGTCCGACCCGAGTCCGGCACCGAGCCGTCCGTTAGCGTT  
CTGCGTGCAATGATGTTCTGTGGCTGAGCCTGACCGCAGCAGAATATGA  
TCAGAGCACCTATGGTAGCAGCACCGGTCCGGTTTATGTTAGCGATAGCG  
35 TTACCCTGGTTAATGTTGCAACCGGTGCACAGGCAGTTGCACGTAGCCTG  
GATTGGACCAAAGTGACCCTGGATGGTCCGCTGAGCACCATTCAGCA  
GTATAGCAAAACCTTTTTTGTCTGCCGCTGCGTGGTAACTGAGCTTTT  
GGGAAGCAGGCACCACCAAGCAGGTTATCCGTATAACTATAATACCACC  
GCAAGCGATCAGCTGCTGGTTGAAAACGCAGCAGGTCATCGTGTGCAAT  
40 TAGCACCTATACCACAGTCTGGGTGCAGGTCCGGTTAGCATTAGCGCAG  
TTGCAGTTCTGGCACCGCATAGCGCAggtggaggaggttctggaggcggt  
ggaagtgggtggcgaggttagcAGAggtccaaccaacagaatctattgttag  
atttcctaataattacaaacttgtgcccttttgggtgaagtttttaacgccca

ccagatttgcattctgtttatgcttggaaacaggaagagaatcagcaactgt  
gttgctgattattctgtcctatataaattccgcatcattttccacttttaa  
gtgttatggagtgtctcctactaaattaaatgatctctgctttactaatg  
tctatgcagattcattttgtaattagaggtgatgaagtcagacaaatcgct  
5 ccaggggcaaactggaaagattgctgattataattataaattaccagatga  
ttttacaggctgcgttatagcttggaaattctaacaatcttgattctaagg  
ttggtggtaattataaattacctgtatagattgttttaggaagtctaattctc  
aaaccttttgagagagatatttcaactgaaatctatcaggccggtagcac  
accttgtaatggtgttggaaggttttaattgttacttttctttacaatcat  
10 atggtttccaaccactaatggtgttggttaccaaccatacagagtagta  
gtactttcttttgaacttctacatgcaccagcaactgtttgtggacctaa  
aaagtgaataaGCGGCCGC

starting ATG (bold)

15 Unique Restriction sites have been added respectively at 5' end, SacI and at the 3' end, NotI  
(dash underlined)

Secreted form tga taa added (double underlined) before NotI

The bold and dotted underlined sequence corresponds to the GS15 linker.

20

**SEQ ID NO: 30 – HEV-GS15-RBD nucleic acid sequence optimized for *E.coli* expression**

25 GAGCTCATGATTGCACTGACCCTGTTTAATCTGGCAGATACCCCTGTTAGGTGGT  
TGCCGACCGAACTGATTAGCAGTGCCGGTGGTCAGCTGTTTTATAGCCGTCCGGT  
TGTTAGCGCAAATGGTGAACCGACCGTTAAACTGTATACCAGCGTTGAAAATGC  
ACAGCAGGATAAAGGTATTGCAATTCGCGCATGATATTGATCTGGGTGAAAGCCG  
TGTGTGATTACAGGATTATGATAATCAGCATGAACAGGATCGTCCGACACCGAGT  
CCGGCACCGAGCCGTCCGTTTAGCGTTCTGCGTGCAAATGATGTTCTGTGGCTGA  
GCCTGACCGCAGCAGAATATGATCAGAGCACCTATGGTAGCAGCACCGGTCCGG  
30 TTTATGTTAGCGATAGCGTTACCCTGGTTAATGTTGCAACCGGTGCACAGGCAGT  
TGCACGTAGCCTGGATTGGACCAAAGTGACCCTGGATGGTCGTCCGCTGAGCAC  
CATTCAGCAGTATAGCAAAACCTTTTTTGTCTGCCGCTGCGTGGTAAACTGAGC  
TTTTGGGAAGCAGGCACCACCAAAGCAGGTTATCCGTATAACTATAATACCACC  
GCAAGCGATCAGCTGCTGGTTGAAAACGCAGCAGGTCATCGTGTGCAATTAGC  
35 ACCTATAACCACAGTTTAGGTGCAGGTCCGGTTAGCATTAGCGCAGTTGCAGTTC  
TGGCACCGCATTACAGCCGGTGGTGGTGGTAGCGGTGGTGGCGGTTACAGGTGG  
CGGTGGTTACAGTGTTACAGCCGACAGAAAGCATTGTTCGTTTTCCGAATATCAC  
CAATCTGTGTCCGTTTGGCGAAGTTTTTAATGCAACCCGTTTTGCAAGCGTTTATG  
CCTGGAATCGTAAACGTATTAGCAATTGCGTTGCCGATTATAGCGTGCTGTATAA  
40 TAGCGCAAGCTTTAGCACCTTTAAATGCTATGGTGTTAGCCCGACCAAACCTGAAT  
GATCTGTGTTTTACCAATGTGTATGCCGATAGCTTTGTGATTCTGTGGTGATGAAGT  
TCGTCAGATTGCACCGGGTCAGACCGGTAAAATTGCAGATTATAACTACAACT  
GCCGGATGATTTTACGGGTTGTGTTATTGCATGGAATAGCAATAACCTGGATAGC  
AAAGTTGGTGGCAACTATAACTATCTGTATCGCCTGTTTCGTAAGAGCAATCTGA  
45 AACCGTTTGAACGTGATATTAGCACCGAAATTTATCAGGCAGGTAGCACCCCGTG  
CAATGGTGTTGAAGGTTTTAATTGTTATTTCCGCTGCAGAGCTATGGTTTTACGC  
CTACCAATGGTGTGGGTTATCAGCCGTATCGTGTTGTTGTTCTGTCAATTTGAACGT  
CTGCATGCACCGGCAACCGTTTTGTGGTCCGAAAAAAtgataaGCGGCCGC

starting ATG (bold)

Secreted form tga taa added (double underlined) before NotI

Unique Restriction sites have been added respectively at 5' end, SacI and at the 3' end, NotI

5 (dash underlined)

The bold and dotted underlined sequence corresponds to the GS15 linker.

**SEQ ID NO: 31 – HEV-GS15-RBD amino acid sequence corresponding to SEQ ID NOs: 29 AND**

10 **30**

MIALTLFNLADTLLGGLPTELISSAGGQLFYSRPVVSANGEPTVKLYTSV  
 ENAQQDKGIAIPHDIDLGESRVVIQDYDNQHEQDRPTSPAPSRPFSLR  
 ANDVLWLSLTAAEYDQSTYGSSTGPVYVSDSVTLNVNATGAQAVARSL  
 15 DWTKVTLTGDRPLSTIQQYSKTFVPLRGKLSFWEAGTTKAGYPYNYNT  
 TASDQLLVENAAGHRVAISTYTTSLGAGPVSISAVAVLAPHSAGGGGSG  
GGGSGGGGSRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNKRIS  
 NCVADYSVLVNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEV  
 20 QIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFR  
 KSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQP  
 YRVVLSFELLHAPATVCGPKK

The GS15 linker is underlined

25 **SEQ ID NO: 32 – HBSAg-(EAAAK)<sub>3</sub>-full-length 2019-nCoV spike protein nucleic acid sequence**  
**human codon optimised for 293f (HEK) cell expression**

AAGCTTGCCgccaacc**ATG**GAGAACATCACATCAGGATTCCTAGGACCCCTGCTCGTGTTACA  
 GGCGGGGTTTTCTTGTGACAAGAATCCTCACAAATACCACAGAGTCTAGACTCGTGGTGGA  
 30 CTTCTCTCAATTTTCTAGGGGATCACCCGTGTGTCTGGGCCAAAATTCGCAGTCCCCAACC  
 TCCAATCACTCACCAACCTCTTGCTCCTCAATTTGTCCTGGCTATCGCTGGATGTGTCTGCG  
 GCGTTTTATCATATTCCTCTTCATCCTGCTGCTATGCCTCATCTTCTTGTGGTTCTTCTGG  
 ACTACCAGGGTATGTTGCCCGTTTGTCTCTAATTCCAGGATCAACAACCTACCAACACGGGA  
 CCATGCAAGACCTGCACGACTCCTGCTCAAGGAACTCTATGTTCCCTCTTGTGCTGTAC  
 35 AAAACCTACCGACGGAACCTGCACTTGTATTTCCCATCCCATCATCCTGGGCTTTTCGCAAAAT  
 ACCTATGGGAGTGGGCCTCAGTCCGTTTCTCCTGGCTCAGTTTACTAGTGCCATTTGTTTCAG  
 TGGTTCGTAGGGCTTTCCCCCACTGTTTGGCTTTCCGCTATATGGATGATGTGGTATTGGGG  
 GCCAAGTCTGTACAGCATCGTGAGTCCCTTTATACCTCTATTACCAATTTTCTTTTGTCTTT  
 GGGTATACATTGAGGCTGCCGCAAAGGAAGCCGCAGCTAAAGAGGCAGCTGCCAAGTTCTGTG  
 40 TTCTTGGTTCTGTGCTGCCCTGGTGTCTAGCCAGTGCGTGAACCTGACCACCAGAACACAGCT  
 GCCTCCAGCCTACACCAACAGCTTCACCAGAGGCGTGTACTACCCGACAAGGTGTTCCGGT  
 CCTCCGTGCTGCATTCTACCCAGGACCTGTTCTCCTGCTTTCTTCTCCAACGTGACCTGGTTC  
 CACGCCATCCATGTGTCTGGCACCAACGGCACCAAGAGATTCGACAACCCCGTGCTGCCTTT  
 CAACGACGGGGTGTACTTTGCCTCCACCGAGAAGTCCAACATCATCAGAGGCTGGATCTTCG  
 45 GCACAACCTGGACAGCAAGACCCAGAGCCTGCTGATCGTGAACAACGCCACCAACGTGGTC  
 ATCAAAGTGTGCGAGTTCCAGTTCTGCAACGACCCCTTCCTGGGCGTCTACTACCACAAGAA  
 CAACAAGTCCTGGATGGAATCCGAGTTCCGGGTGTACTCCTCCGCCAACAACTGCACCTTCG

AATACGTGTCCCAGCCTTTCTGATGGACCTGGAAGGCAAGCAGGGCAACTTCAAGAACCTG  
CGCGAGTTTCGTGTTCAAGAACATCGACGGCTACTTCAAGATCTACTCCAAGCACACCCCTAT  
CAACCTCGTGC GG GATCTGCCTCAGGGCTTCTCTGCTCTGGAACCCCTGGTGGATCTGCCCA  
TCGGCATCAACATCACCCGGTTTTCAGACCCCTGCTGGCCCTGCACCGGTCTTATTTGACCCCT  
5 GGCGACTCCTCTTCTGGCTGGACTGCTGGCGCCGCTGCTTACTATGTGGGCTACCTGCAGCC  
TCGGACCTTTCTGCTGAAGTACAACGAGAATGGCACCATCACCGACGCCGTGGACTGTGCTC  
TGGATCCTCTGTCCGAGACAAAGTGCACCCCTGAAGTCCTTCACCGTGGAAAAGGGCATCTAC  
CAGACCTCCAACCTCCGGGTGCAGCCACCGAGTCTATCGTGCGGTTCCCTAACATCACCAA  
CCTGTGTCTTTTCGGCGAGGTGTTCAATGCCACCAGATTTCGCCTCTGTGTACGCCTGGAACC  
10 GGAAGCGGATCTCTAACTGCGTGGCCGACTACAGCGTGTGTACAACCTCCGCCTCCTTCAGC  
ACCTTCAAGTGCTACGGCGTGTCCCTACAAAGCTGAACGACCTGTGCTTCACAAACGTGTA  
CGCCGACAGCTTCGTGATCCGGGGAGATGAAGTGC GGCAGATCGCTCCTGGACAGACCCGGCA  
AGATCGCCGATTACAAC TACAAGCTGCCCGACGACTTCACCGGCTGTGTGATCGCTTGGAAC  
TCCAACAACCTGGACTCCAAGTCCGGCGGCAACTACAAC TACCTGTACCGGCTGTTCCGGAA  
15 GTCTAACCTGAAGCCTTTTCGAGCGGGACATCAGCACCGAGATCTACCAGGCTGGCAGCACCC  
CTTGTAACGGCGTGGGCTATCAGCCCTACAGAGTGGTGGTGTGTCTTTCGAGCTGCTGCATGC  
ACCAATGGCGTGGGCTATCAGCCCTACAGAGTGGTGGTGTGTCTTTCGAGCTGCTGCATGC  
TCCTGCTACCGTGTGCGGCCCTAAGAAATCTACCAACCTGGTCAAGAACAATGCGTGAAC T  
TCAACTTCAACGGCCTGACCGGCACCGGCGTGTGTGACAGAGTCCAACAAGAAGTTCTGCCA  
20 TTCCAGCAGTTCGGCCGGGATATCGCCGATACACAGATGCCGTCAGGGACCCCTCAGACACT  
GGAAATCCTGGACATCACCCCTTGCTCCTTCGGCGGAGTGTCTGTGATCACCCAGGCACCA  
ACACCTCTAACCAGGTGGCCGTGTGTATCAGGACGTGAAC TGTACCGAGGTGCCCCGTGGCT  
ATCCATGCCGATCAGCTGACCCCTACATGGCGCGTGTACTCCACCGGCTCTAACGTGTTCCA  
GACAAGAGCTGGCTGTCTGATCGGCGCTGAGCACGTGAACAATTCCTACGAGTGCGACATCC  
25 CCATCGGAGCCGGAATCTGCGCCTCTTATCAGACCCAGACCAACTCTCCCAGACGGGCCAGA  
TCTGTGGCCAGCCAGTCTATCATTTGCTTACACCATGAGCCTGGGCGCCGAGAACTCTGTGGC  
CTACAGCAACAAC TCTATCGCTATCCCCACCAACTTCACCATCTCCGTGACCACAGAGATCC  
TGCCAGTGTCCATGACCAAGACCAGCGTGGACTGCACCATGTACATCTGCGGCGACTCTACC  
GAGTGTCTCAAACCTGCTGCTCCAGTACGGCTCCTTCTGCACCCAGCTGAATAGAGCCCTGAC  
30 CGGAATCGCCGTGGAACAGGACAAGAACACCCAAGAGGTGTTTCGCCCAAGTGAAGCAGATCT  
ACAAGACCCCTCCTATCAAGGACTTCGGCGGCTTCAATTTCTCCCAGATTCTGCCCCGATCCT  
AGCAAGCCCTCCAAGCGGTCTTTTCATCGAGGACCTGCTGTTCAACAAAGTGACACTGGCCGA  
CGCCGGCTTCATCAAGCAGTACGGCGACTGTCTGGGCGACATTGCCGCTAGGGATCTGATCT  
GCGCCCAGAAGTTTAACGGACTGACAGTGTGCCTCCTCTGCTGACCGATGAGATGATCGCC  
35 CAGTACACCTCCGCACTGCTGGCTGGCACAATCACCTCTGGATGGACATTTGGCGCTGGCGC  
TGCTCTGCAAATCCCATTCTGCTATGCAAATGGCCTACCGGTTCAACGGCATCGGCGTGACCC  
AGAATGTGCTGTACGAGAACCAGAAGCTGATCGCCAACCAGTTC AACAGCGCCATCGGAAAG  
ATCCAGGACAGCCTGTCCAGCACCGCTTCTGCCCTGGGAAAGCTGCAGGATGTGGTCAACCA  
GAACGCTCAGGCCCTGAACACCCTCGTGAAGCAGCTGTCTAGCAACTTCGGCGCCATCTCCT  
40 CTGTGCTGAACGATATCCTGAGCCGGCTGGACAAGGTGGAAGCCGAGGTGCAGATCGACAGA  
CTGATCACCGGACGGCTGCAGTCCCTGCAGACCTATGTTACCCAGCAGCTGATCCGGGCTGC  
CGAGATTAGAGCCTCTGCCAATCTGGCCGCAACCAAGATGTCTGAGTGTGTGCTGGGACAGT  
CCAAGAGAGTGGACTTCTGCGGCAAGGGCTACCACCTGATGAGCTTCCCTCAGTCTGCTCCT  
CACGGCGTGGTGTCTTCTGCACGTGACCTACGTGCCCGCTCAAGAGAAGAACTTTACCACCGC  
45 TCCTGCCATCTGCCACGACGGCAAGGCTCACTTTCTAGAGAAGGCGTGTTCGTGTCTAACG  
GCACCCATTGGTTCGTGACACAGCGGAACTTCTACGAGCCCCAGATCATCACCACCGACAAC  
ACCTTCGTGTCCGGCAACTGCGACGTGTGATCGGAATTGTGAACAATACCGTGTACGACCC  
TCTGCAGCCCCGAGCTGGACTCCTTCAAAGAGGAACTGGACAAGTACTTTAAGAACCACACAA  
GCCCCGACGTGGACCTGGGAGACATCTCTGGCATCAACGCCTCCGTGGTCAACATCCAGAAA  
50 GAGATCGACCGCTGAACGAGGTGGCCAAGAATCTGAACGAGTCCCTGATCGACCTGCAAGA  
ACTGGGGAAGTACGAGCAGTACATCAAGTGGCCCTGGTACATCTGGCTGGGCTTTATCGCTG

GCCTGATCGCTATCGTGATGGTCACAATCATGCTGTGCTGTATGACCTCCTGTTGCTCCTGC  
CTGAAGGGCTGCTGCTCTTGCGGCTCTTGCTGCAAGTTCGACGAGGACGACTCTGAGCCCGT  
GCTGAAAGGCGTGAAGCTGCACTATACCTGATGACTCGAG

- 5 KOZAC sequence added (gcc acc, underlined) before the starting ATG (**bold**).  
The bold and dotted underlined sequence corresponds to the (**EAAAK**)<sub>3</sub> linker.

- 10 **SEQ ID NO: 33 – HBSAg-(EAAAK)<sub>3</sub>-full-length 2019-nCoV spike protein amino acid sequence**  
**corresponding to SEQ ID NO: 32**

MENITSGFLGPLLVLQAGFFLLTRILTIPQSLDSWWTSLNFLGGSPVCLGQNSQSPTSNNHSP  
TSCPPICPGYRWMCLRRFIIIFLFIILLCLIFLLVLLDYQGMLPVCPLIPGSTTTNTGPCKTC  
TTPAQGNSMFPSCCCTKPTDGNCTCIPSSWAFAYLWELWASVRFWSLLVPFVQWFWGL  
15 SPTVWLSAIWMMWYWGPSLYSIVSPFIFLLPIFFCLWVYIEAAAKEAAAKEAAAKFVFLVLL  
PLVSSQCVNLTTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHV  
SGTNGTKRFDNPFVLPFNDGVYFASTEKSNIRGWIFGTTLDSTQSLIVNNATNVVIKVCE  
FQFCNDPFLGVYHKNKSWMESEFRVYSSANNCTFEYVSQPFMDLEGKQGNFKNLREFVF  
KNIDGYFKIYSKHTPINLVRDLPOGFSALEPLVDLPIGINITRFQTLALHRSYLTTPGDSSS  
20 GWTAGAAAYYVGYLQPRFTLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNF  
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLNSASFSTFKCY  
GVSPKTLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLD  
SKVGGNYNYLYRLFRKSNLKPFFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVG  
YQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFG  
25 RDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQ  
LTPTWRVYSTGSNVFQTRAGCLIGAHEVNNSEYCDIPIGAGICASYQTQTNSPRRARVASQ  
SIIAYTMSLGAENSVAYSNNIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECNL  
LLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSK  
RSFIEDLLFNKVTADAGFIKQYGDCLGDIARDLICAQKFNGLTVLPPLLTDEMIAQYTS  
30 LLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSL  
SSTASALGKLQDVVNQNAQALNTLVKQLSSNFAGAISSVLNDILSRDKVEAEVQIDRLITGR  
LQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVF  
LHVITYVPAQEKNFTTAPAICHGDKAHFPREGVVFVSNNGTHWFVTQRNFYEPQIIITDNTFVSG  
NCDVIVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVNNIQKEIDRL  
35 NEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCC  
SCGSCCKFDEDDSEPVLKGVKLHYT

The (EAAAK)<sub>3</sub> linker is underlined