CORONAVIRUS POLYPEPTIDE

Field of the Invention

The invention provides polypeptides and compositions of said polypeptides and/or their encoding polynucleotides for the prophylactic vaccination and/or therapeutic treatment of coronavirus infections, as well as methods for the manufacture of a polypeptide vaccine and the use of polypeptides and/or their encoding polynucleotides in treating, preventing, and/or diagnosing coronavirus infection.

Background of the Invention

Following the emergence of SARS-CoV-2 in late 2019 and the resulting global pandemic, there is an urgent need for many and varied vaccines against SARS-CoV-2 and treatments for COVID-19, and particularly for vaccines that can accommodate novel mutations and strains as they emerge. Even more so, there is pressing need for broad-acting vaccines and therapies which can protect against multiple members of the *Coronaviridae* family, which is recognised as high-risk for emergent zoonoses and future epidemics.

Coronaviruses are enveloped viruses with a linear, single-stranded positive-sense RNA genome encoding four or five structural proteins, and four to eight non-structural/accessory proteins. Of the structural proteins, all coronaviruses comprise a nucleocapsid protein (N), small envelope protein (E), membrane protein (M), and spike protein (S). Some coronaviruses (including some betacoronaviruses, but not SARS-CoV-2, SARS-CoV, or MERS-CoV) further encode a hemagglutinin-esterase (HE) surface protein. The N protein complexes with the RNA genome to form a helical nucleocapsid which is surrounded by a membranous envelope in which M, E, and S are embedded transmembrane. Vaccination efforts to produce neutralising antibodies focus primarily on structural proteins exposed on the surface, and coronavirus vaccine development to date has focussed particularly on the coronavirus S protein as the major transmembrane protein and antigenic target.

Spike 'S' glycoproteins protrude as trimers with stalk-and-club morphology from the coronavirus envelope and mediate viral-host binding and viral entry. Each monomer comprises a S1 (representing the club portion) and S2 (the stalk) subunit. In severe acute respiratory syndrome-related coronaviruses, including SARS-CoV and SARS-CoV-2, the S1 subunit of S mediates viral binding to host receptors, for example human angiotensin-converting enzyme 2 ('ACE2'), via a receptor binding domain ('RBD'). Upon RBD binding to the host receptor, host protease activity cleaves the S1 subunit from the transmembrane S2

subunit. The S2 subunit undergoes a conformational change and a fusion protein contained within the S2 inserts into the host cell membrane. In a further conformation change, two hydrophobic heptad repeat ('HR') regions within S2 associate to form a six-helix bundle, bringing the viral envelope into close proximity with the host cell membrane and allowing the viral and host cell plasma membranes to fuse. Thus, the nucleocapsid and its encapsulated viral material is subsumed into the host cell.

Both cellular and humoral responses are crucial to the adaptive immune response for host clearance of a virus. Following initial exposure to the virus' molecular signature, either through initial infection or through vaccination, an immune memory is formed in the host which allows cellular and humoral responses to be mounted more rapidly upon re-exposure to the molecular signature.

Humoral antibody responses are stimulated in response to extracellular three-dimensional molecular epitopes. Antibody responses are known to play a key role in SARS-CoV-2 immunity and much of the attention in vaccine development in response to the SARS-CoV-2 pandemic to date relates to neutralising antibody responses to full-length S protein delivered via viral vector, DNA, RNA, recombinant protein, and protein subunit platform technologies (Thanh Le *et al.*, 2020).

T cell responses have been postulated to play a significant role in the immune clearance of, and immune memory to, coronaviruses including SARS-CoV-2, and cross-reactive T cell responses have been documented (Grifoni *et al.*, 2020; Altmann & Boyton, 2020). In contrast to antibody responses, cellular responses including T cell responses are stimulated in response to contiguous two-dimensional fragments of pathogenic proteins and polymers. Techniques for stimulation of specific T cell responses are less well characterised and lie generally within the field of therapeutic (rather than prophylactic) vaccination e.g. for the treatment of cancers.

The present invention provides a polypeptide which is able to effectively stimulate neutralising antibodies against key coronavirus epitopes. As described herein, the polypeptide also generates effective T cell responses, may be readily adapted to represent novel strains and/or to incorporate multiple coronavirus or viral variant sequences, can motivate immune responses across a broad range of individuals regardless of inherent antibody epitope specificity bias, B- or T-cell receptor polymorphism, or HLA-type, and is deliverable in native protein form without the need for challenging and burdensome vector or liposomal delivery. The polypeptide uses key epitopes of coronavirus surface proteins in an overlapping and

intracellularly cleavable arrangement to preferentially stimulate neutralising antibody production and, at the same time, stimulate the T cell response. Surprisingly, we have found that this is effective in preventing coronavirus infection, and in treating an existing coronavirus infection.

Summary of the Invention

The present invention provides a polypeptide for the vaccination against and/or treatment of a coronavirus infection.

In a first aspect, a polypeptide for vaccination against and/or treatment of a coronavirus infection comprising two or more peptide fragments comprising sequences derived from coronavirus S protein, and one or more exogenous protease cleavage site sequences located between each of the two or more peptide fragments, is provided.

In some embodiments, the coronavirus is a betacoronavirus, optionally a severe acute respiratory syndrome-related coronavirus, optionally SARS-CoV-2. In some embodiments, the coronavirus is a human coronavirus.

In some embodiments, at least two of the two or more peptide fragments comprise at least one overlapping sequence. In some embodiments, at least two of the two or more peptide fragments comprise sequences derived from the S1 and/or S2 subunit of the S protein. In some embodiments, at least one of the two or more peptide fragments comprises a sequence derived from the receptor binding domain (RBD), optionally the receptor binding motif (RBM), of the S1 subunit.

In some embodiments, the polypeptide comprises three or more peptide fragments each comprising sequences derived from the S protein of the coronavirus, wherein at least two of the three or more peptide fragments comprise sequences derived from the same S protein subunit, and wherein at least two of the three or more peptide fragments comprise at least one overlapping sequence. In some embodiments the three or more peptide fragments comprise at least one peptide fragment comprising a sequence derived from the S1 subunit, and at least one peptide fragment comprising a sequence derived from the S2 subunit.

In some embodiments, at least one peptide fragment comprises one or more antibody epitope(s) of a coronavirus S protein, optionally a SARS-CoV-2 S protein. In some embodiments the antibody epitope is a linear antibody epitope. In some embodiments the

antibody epitope is a neutralising epitope. In some embodiments the linear antibody epitope a neutralising linear antibody epitope.

In some embodiments, the polypeptide, when administered to a subject, stimulates production of anti-coronavirus antibodies in said subject. In some embodiments the anti-coronavirus antibodies are neutralising antibodies.

In some embodiments at least one peptide fragment comprises one or more CD4+ T cell epitope(s) and/or one or more CD8+ T cell epitope(s) of a coronavirus S protein. In some embodiments at least one peptide fragment comprises one or more CD4+ T cell epitope(s) and/or one or more CD8+ T cell epitope(s) of a SARS-CoV-2 S protein.

In some embodiments, the polypeptide, when administered to a subject or cell sample, stimulates T cell responses specific to the coronavirus in said subject or cell sample.

In some embodiments the polypeptide has five or more peptide fragments. In some embodiments the polypeptide has ten or more peptide fragments.

In some embodiments the one or more exogenous protease cleavage site sequence is a cathepsin protease cleavage sequence. In some embodiments the one or more exogenous protease cleavage site sequence is a cathepsin S cleavage sequence. In some embodiments the one or more exogenous protease cleavage site sequence is LRMK (SEQ ID NO.: 34).

In some embodiments, each of the overlapping sequences of the polypeptide of the invention are between 2 and 31 amino acids long. In some embodiments, each of the overlapping sequences of the polypeptide are at least 8 amino acids long.

In some embodiments, the peptide fragments comprise at least one of the epitopes: SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, and/or SEQ ID NO: 33.

In some embodiments, at least one peptide fragment of the polypeptide of the invention comprises a sequence with at least 90% sequence identity to a sequence selected from the group: SEQ ID NO: 1; SEQ ID NO: 2; SEQ ID NO: 3; SEQ ID NO: 4; SEQ ID NO: 5; SEQ ID NO: 6; SEQ ID NO: 7; SEQ ID NO: 8; SEQ ID NO: 9; SEQ ID NO: 10; SEQ ID NO: 11; SEQ ID NO: 12; and, optionally, wherein the polypeptide is immunostimulatory.

In a second aspect, herein is provided a formulation for administration to a subject comprising the polypeptide of the invention. In some embodiments the formulation is a SARS-CoV-2 vaccine formulation. In some embodiments, the formulation additionally comprises the RBD, RBM, HR1 and/or HR2 of a coronavirus. In some embodiments the coronavirus is SARS-CoV-2.

In a third aspect, herein is provided a formulation for administration to a subject comprising the polypeptide of the invention and the RBD, RBM, HR1 and/or HR2 of a coronavirus. In some embodiments the formulation is a SARS-CoV-2 vaccine formulation. In some embodiments the coronavirus is SARS-CoV-2.

In a fourth aspect, a polynucleotide encoding a polypeptide of the invention is provided.

In a fifth aspect, a composition comprising a delivery vehicle and the polynucleotide of the invention is provided. In some embodiments the composition additionally comprises one or more polynucleotide(s) encoding the RBM, RBD, HR1, and/or HR2 of a coronavirus. In some embodiments the coronavirus is SARS-CoV-2. In some embodiments the composition further comprises a pharmaceutically acceptable carrier.

In a further aspect, a method for manufacturing a vaccine comprising a polypeptide of the invention is provided, comprising expressing the polynucleotide of the invention in one or more cell *in vitro* to produce a polypeptide of the invention; and isolating said polypeptide. In some embodiments the method additionally comprises the steps of: expressing a polynucleotide encoding at least the RBM of a coronavirus in one or more cells in vitro to produce a protein sequence comprising at least the RBM; isolating the protein sequence; and combining the polypeptide and the protein sequence into a formulation for administration. In some embodiments the coronavirus is SARS-CoV-2.

In a further aspect, a method for vaccination against and/or treatment of a coronavirus infection is provided comprising administering, to a subject in need thereof, the polypeptide of the invention. In some embodiments the method additionally comprises administering, to the subject, the RBM, RBD, HR1, and/or HR2 of a coronavirus. In some embodiments the coronavirus is SARS-CoV-2. In some embodiments the polypeptide of the invention is administered to a human at any dose between 5 $\mu g.kg^{-1}$ and 5 $mg.kg^{-1}$. In some embodiments, the RBM, RBD, HR1 and/or HR2 is administered to a human at any dose between 5 $\mu g.kg^{-1}$ and 5 $mg.kg^{-1}$.

In a further aspect a method for vaccination against and/or treatment of a coronavirus infection comprising administering, to a subject in need thereof, the polynucleotide of the invention is provided. In some embodiments the polynucleotide of the invention is administered to a human at any dose between 5 $\mu g.kg^{-1}$ and 5 $mg.kg^{-}1$. In an aspect a method for vaccination against and/or treatment of a coronavirus infection comprising administering, to a subject in need thereof, the composition of the invention is provided. In some embodiments the composition is administered to a human at any dose between 5 $\mu g.kg^{-1}$ and 5 $mg.kg^{-}1$. In some embodiments the coronavirus is the coronavirus is a betacoronavirus. In some embodiments the coronavirus is a severe acute respiratory syndrome-related coronavirus. In some embodiments the coronavirus is SARS-CoV-2.

In a further aspect, a method for the diagnosis of a coronavirus infection is provided, comprising administering a polypeptide of the invention to a patient and detecting resulting T cell stimulation. In some embodiments, the step of detecting T cell stimulation comprises detecting the production of cytokines.

It will be apparent to the skilled person that any of the above embodiments may also apply, as appropriate, to the following aspects.

In an aspect, a polypeptide comprising two or more peptide fragments which comprise sequences derived from coronavirus M, S, E, and/or N proteins and at least one overlapping sequence, said polypeptide also comprising one or more protease cleavage site sequence(s) located between the peptide fragments, is provided.

In some embodiments, the coronavirus is a betacoronavirus, optionally a human betacoronavirus. In some embodiments, the betacoronavirus is a severe acute respiratory syndrome-related coronavirus, optionally a human severe acute respiratory syndrome-related coronavirus. In some embodiments, the severe acute respiratory syndrome-related coronavirus is SARS-CoV-2.

In some embodiments, at least two of the two or more peptide fragments comprise sequences derived from the S protein, preferably the S1 and/or S2 subunit of the S protein. In one embodiment, at least one of the two or more peptide fragments comprises a sequence derived from the receptor binding domain of the S1 subunit. In some embodiments, at least one of the two or more peptide fragments comprises a sequence derived from the HR2 and/or HR1 domain of the S2 subunit.

In some embodiments, the polypeptide comprises three or more peptide fragments comprising sequences derived from the S protein of SARS-CoV-2, wherein at least two of the peptide fragments comprise sequences derived from the same S protein subunit and comprise at least one overlapping sequence. In some embodiments, the three or more peptide fragments comprise at least one peptide fragment comprising a sequence derived from the S1 subunit, and at least one peptide fragment comprising a sequence derived from the S2 subunit.

In some embodiments, the one or more protease cleavage site sequence is a cathepsin protease cleavage sequence, preferably cathepsin S, more preferably an LRMK protease cleavage sequence.

In an aspect, a composition comprising a delivery vehicle and a polypeptide of the invention and/or a polynucleotide of the invention is provided.

In a further aspect, a composition comprising a pharmaceutically acceptable carrier and a polypeptide of the present invention and/or a polynucleotide of the invention and/or a delivery vehicle is provided.

In a further aspect, a method for the vaccination against and/or treatment of a coronavirus infection is provided, comprising administering to a subject a polypeptide of the invention in a delivery vehicle and/or pharmaceutically acceptable carrier. In some embodiments, the polypeptide is administered to a human at a dose between 5 $\mu g.kg^{-1}$ and 5 $mg.kg^{-1}$. In some embodiments, the coronavirus is a betacoronavirus, preferably a severe acute respiratory syndrome-related coronavirus, more preferably SARS-CoV-2.

In a further aspect, a method for the vaccination and/or treatment of a coronavirus infection is provided, comprising administering to a subject a polynucleotide of the invention in a delivery vehicle and/or pharmaceutically acceptable carrier. In some embodiments, the polynucleotide is administered to a human at a dose between 5 $\mu g.kg^{-1}$ and 5 $mg.kg^{-1}$. In some embodiments, the coronavirus is a betacoronavirus, preferably a severe acute respiratory syndrome-related coronavirus, more preferably SARS-CoV-2.

Description of Figures

Aspects and embodiments of the invention will now be described with reference to the Figures and Examples in which:

Figure 1. Plasmid map of constructed plasmid pET30a.

Figure 2. Electrophoretic analysis of the vector plasmid pET30a, demonstrating successful insertion of the ROP gene into *E.coli*.

Figure 3. SDS-PAGE analysis of:
(a) The induction rate of ROP-COVS. Lane 1 is before induction; Lane 2 is 4 hours after induction with IPTG to final concentration 0.2 mM; Lane M is the molecular weight marker (14.4-94.0 kDa). ROP-COVS is effectively induced by IPTG.
(b) Purification of ROP-COVS. Lane 1 is sample before purification; Lane 2 is the flow-through; Lane 3 is eluted with 48 mM Imidazole; Lane 4 is eluted with 78 mM Imidazole; Lane 5 & 6 is eluted with 105 mM Imidazole; Lane 7 is eluted with 138 mM Imidazole; Lane M is the molecular weight marker (14.4-94.0 kDa). Lanes 6 to 8 have over 95% purity.
(c) Refolding of ROP-COVS.

Figure 4. ELISA antibody titration analysis of the sera of mice following vaccination with ROP-COVS or a control, as indicated.

Figure 5. An illustrative schematic of one embodiment of a polypeptide of the invention.

Figure 6. Serum neutralisation data derived via surrogate RBD-ACE2 ELISA neutralisation assay.

Figure 7. Purified IgG neutralisation data derived via surrogate RBD-ACE2 ELISA neutralisation assay.

Detailed Description of the Invention

In a first aspect the invention provides a polypeptide for vaccination against and/or treatment of a coronavirus infection comprising two or more peptide fragments comprising sequences derived from coronavirus S protein, and one or more exogenous protease cleavage site sequences located between each of the two or more peptide fragments.

In an aspect the invention provides a polypeptide for vaccination against and/or treatment of a coronavirus infection, comprising two or more peptide fragments, wherein the two or more peptide fragments comprise sequences derived from at least one of the M, S, E, or N

coronavirus proteins, or any combination thereof, wherein at least two of the two or more peptide fragments comprise sequences from the same coronavirus protein and comprise at least one overlapping sequence, further comprising one or more protease cleavage site sequences located between each of the two or more peptide fragments.

The present invention may be better understood with use of the following definitions.

"Recombinant" as used herein refers to any polymer, optionally a polypeptide, which is non-naturally occurring or artificially constructed, having been manufactured by gene recombination techniques in a bacterium (for example, but not limited to, an *E.coli* bacterium).

"Polypeptide" as used herein refers to a linear chain of amino acids linked by means of peptide bonds which is longer than a 'peptide' or 'peptide fragment', as used herein.

"Peptide" as used herein refers to a linear chain of more than one amino acids linked by means of peptide bonds which is shorter than a 'polypeptide' as used herein.

"Peptide fragment" as used herein refers to an amino acid chain (a "peptide") which is a piece of a larger polypeptide. In other words, two or more peptide fragments, if fragments of the same larger polypeptide, can together form all or part of the primary sequence of the larger polypeptide. In this case, the larger polypeptide may be the recombinant polypeptide of the present invention.

"Protein" as used herein refers to a molecular entity composed primarily of one or more peptides and/or polypeptides (usually, but not essentially, having more 100 amino acids) and which has folded into, or presents as, a 3-dimensional conformation.

"Vaccine" as used herein refers to a substance capable of generating a protective immune memory against a pathogen in a subject, wherein said subject is an animal and optionally a human. Said protective immune memory may amount to full immunity and/or a reduction in severity or symptoms of the disease associated with said pathogen upon subsequent infection.

"Coronavirus" as used herein refers to a member of the *Coronaviridae* family as defined by the Coronavirus Study Group, a working group of the International Committee on Taxonomy of Viruses and as used in the Coronaviridae Study Group of the International Committee on Taxonomy of Viruses (2020) https://dx.doi.org/10.1038%2Fs41564-020-0695-z.

"Betacoronavirus" as used herein refers to a member of the *Betacoronavirus* genus as defined by the Coronavirus Study Group, a working group of the International Committee on Taxonomy of Viruses and as used in Coronaviridae Study Group of the International Committee on Taxonomy of Viruses (2020) https://dx.doi.org/10.1038%2Fs41564-020-0695-z. Amongst subspecies grouped within the *Betacoronavirus* genus are SARS-CoV, SARS-CoV-2, and MERS-CoV.

"Severe acute respiratory syndrome-related coronavirus" as used herein refers to a member of the *Severe acute respiratory syndrome-related coronavirus* species as defined by the Coronavirus Study Group, a working group of the International Committee on Taxonomy of Viruses and as used in Coronaviridae Study Group of the International Committee on Taxonomy of Viruses (2020) https://dx.doi.org/10.1038%2Fs41564-020-0695-z. Amongst other subspecies grouped within the *Severe acute respiratory syndrome-related coronavirus* species are SARS-CoV, SARS-CoV-2, SARSr-CoV BtKY72, SARSr-CoV RaTG13, SARS-CoV PC4-227, SARS-CoVGZ-02, Bat SARS CoVRf1/2004, Civet SARS CoVSz3/2003.

"Epitope" as used herein refers to a portion of a peptide fragment, peptide, polypeptide, protein, glycoprotein, lipoprotein, carbohydrate, lipid, or otherwise which is recognised by the adaptive immune system, and particularly by antibodies, B cells, and/or T cells, via receptor binding interactions.

"LRMK" as used herein refers to the Leu-Arg-Met-Lys amino acid sequence SEQ ID NO: 34, being a cleavage site recognised by *inter alia* Cathepsin S. In some embodiments, a cleavable linker is provided and in some further embodiments, that linker is LRMK.

"Exogenous" as used herein means artificially introduced. It may also mean not present in the native sequence, for example the wild type (including any variants), at least in the location at which it is now artificially introduced. For example, a polypeptide may comprise two sequences which are contiguous in a native protein, and which are separated by an exogenous protease cleavage site i.e. a cleavage site which is not present in the contiguous native sequence. As another example, in the context of a polypeptide comprising peptide fragments comprising sequences derived from a SARS-CoV-2 S protein and comprising an exogenous protease cleavage site between each peptide fragment, the exogenous protease cleavage site is a cleavage site that has been artificially introduced or which is not natively found in the SARS-CoV-2 S protein at the location within the S protein amino acid sequence at which it is now located.

"Overlap" as used herein refers to a portion or 'sub-sequence' of an amino acid sequence which is the same, or substantially the same, in two different amino acid sequences or peptides or peptide fragments, preferably in such a way that the sub-sequence at the C-terminal end of one amino acid sequence or peptide or peptide fragment is the same as or substantially similar to the sub-sequence at the N-terminal end of another amino acid sequence or peptide or peptide fragment, and/or vice versa. Overlap may or may not be reflected in the polynucleotide sequences which encode said amino acid sequences. It will be clear to the skilled reader that 'peptide fragments which overlap' therefore means 'peptide fragments having at least one overlap'.

"Identity" as used herein is the degree of similarity between two sequences, in other words the degree to which two sequences match one another in terms of residues, as determined by

comparing two or more polypeptide or polynucleotide sequences. Identity can be determined using the degree of similarity of two sequences to provide a measurement of the extent to which the two sequences match. Numerous programs are well known by the skilled person for comparing polypeptide or polynucleotide sequences, for example (but not limited to) the various BLAST and CLUSTAL programs. Percentage identity can be used to quantify sequence identity. To calculate percentage identity, two sequences (polypeptide or nucleotide) are optimally aligned (i.e. positioned such that the two sequences have the highest number of identical residues at each corresponding position and therefore have the highest percentage identity) and the amino acid or nucleic acid residue at each position is compared with the corresponding amino acid or nucleic acid at that position. In some instances, optimal sequence alignment can be achieved by inserting space(s) in a sequence to best fit it to a second sequence. The number of identical amino acid residues or nucleotides provides the percentage identity, e.g. if 9 residues of a 10 residue long sequence are identical between the two sequences being compared then the percentage identity is 90%. Percentage identity is generally calculated along the full length of the two sequences being compared.

"Variant" as used in the context of a peptide, polypeptide, and/or protein herein refers to a peptide, polypeptide, and/or protein which has an amino acid sequence with at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or at least 99% identity, optionally 60-100%, 65-100%, 70-100%, 75-100%, 80-100%, 85-100%, 90-100%, 91-100%, 92-100%, 93-100%, 94-100%, 95-100%, 96-100%, 97-100%, 98-100% identity, to the wild-type peptide, polypeptide, and/or protein. Where the variant differs from the wildtype, this may be due to substitution of amino acids within the sequence, and/or due to the addition or loss of amino acids from either or both ends of, or even internally within, a sequence. "Variant" may also be used in the context of a virus, (herein "a viral variant") to refer to a virus possessing one or more mutations in its genome sequence

"Neutralising" as used herein refers to antibodies which can inhibit or block a key component of the viral replication cycle. Viral replication may thereby be lessened and/or prevented.

"Broad-acting" as used herein refers to a vaccine, therapeutic or antibody which is effective against multiple different viral species, sub-species, and/or viral variants. As an illustrative example, a broad-acting coronavirus vaccine may be effective at preventing infection across sub-species e.g. may prevent infection with SARS-CoV-2 and with SARS-CoV; in another illustrative example, a broad-acting coronavirus vaccine may be effective at preventing infection across species e.g. may prevent infection with SARS-CoV-2, SARS-CoV, MERS, HKU1, and OC43, amongst others.

"Derived from" herein and throughout means 'identical to or substantially similar to a portion of'. A peptide fragment having a sequence derived from a coronavirus protein is a peptide fragment containing an amino acid sequence which is identical to, or substantially similar to,

a contiguous portion of the amino acid sequence of said coronavirus protein. 'Substantially similar' herein and throughout means that the amino acid sequence has at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or at least 99% identity, optionally 70-100%, 75-100%, 80-100%, 85-100%, 90-100%, 91-100%, 92-100%, 93-100%, 94-100%, 95-100%, 96-100%, 97-100%, 98-100% identity to the reference coronavirus protein sequence, the reference SEQ ID NO, or a contiguous portion or sub-sequence thereof as will be apparent from the context. 'At least' herein and throughout means, in some embodiments, the recited percentage up to and including 100%. For example, 'at least 75%' can mean, in some embodiments, '75% to 100%'. Frequently, the nucleic acid sequence of a peptide fragment having a sequence derived from a coronavirus protein will differ from the coronavirus protein nucleic acid sequence to a greater degree than will the amino acid sequence of the peptide fragment from the coronavirus protein amino acid sequence. This is due to reasons of preparation and optimisation of expression of the polypeptide, for example codon optimisation. For the avoidance of doubt, it is the amino acid sequence of a peptide fragment which is derived from – in that it has at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or at least 99% identity to a contiguous portion of – the amino acid sequence of a coronavirus protein. The nucleic acid sequences may differ to a greater extent and may have a lower sequence identity due to the inherent redundancy of the genetic code for amino acids.

'At least' herein and throughout means, in some embodiments, the recited number of peptide fragments up to and including the total number of peptide fragments present in the polypeptide. For example, in a polypeptide with 14 peptide fragments 'at least two peptide fragments' would mean, in some embodiments, 'two to 14 peptide fragments', or any number in between.

Where a dosage is expressed in '$\mu g.kg^{-1}$', this is intended to mean the mass of the agent in micrograms per mass of the subject in kilograms. It will be clear to the skilled reader, therefore, that $mg.kg^{-1}$ means the mass of the agent in milligrams per mass of the subject in kilograms. Said agent may be the polypeptide, or corresponding polynucleotide, as provided herein, and may be provided to a mammalian subject, preferably a human.

In an embodiment, the invention provides a polypeptide for vaccination against and/or treatment of a coronavirus infection comprising two or more peptide fragment(s), at least one of which comprises one or more linear antibody epitope(s) of a coronavirus S protein, and comprising a protease cleavage site sequence located between each peptide fragment. In another embodiment, the invention provides a polypeptide for vaccination against and/or treatment of a coronavirus infection comprising two or more peptide fragment(s), at least one of which comprises one or more linear antibody epitope(s) of a coronavirus S protein, and

comprising an exogenous (i.e. artificially introduced) protease cleavage site sequence located between each peptide fragment. Said exogenous cleavage site constitutes a difference and is useful because it allows peptide fragments to be liberated in a desired manner. In some embodiments the exogenous protease cleavage site sequence is for an intracellular protease and thereby allows peptide fragments to be liberated from the polypeptide intracellularly. In some embodiments, the coronavirus S protein is a SARS-CoV-2 S protein. In some embodiments at least one linear antibody epitope is a neutralising epitope. In some embodiments, the two or more peptide fragments comprise amino acid sequences which overlap. This may allow that, for example, while one or more linear antibody epitope sequence is comprised wholly within one peptide fragment, other peptide fragment(s) comprise partial sequences of said linear antibody epitope. In some embodiments, at least one peptide fragment comprises one or more CD4+ and/or CD8+ T cell epitope(s) of a coronavirus protein, optionally a coronavirus S protein, optionally the S protein of SARS-CoV-2. A peptide fragment may comprise one epitope only (whether a linear antibody epitope or CD4+ or CD8+ T cell epitope). Equally, a peptide fragment may comprise some or all of two epitopes, for example some or all of a linear antibody epitope and some or all of a T cell epitope. A peptide fragment may comprise no epitope.

In an embodiment, the invention provides peptide fragments which overlap and which comprise sequences derived from coronavirus surface proteins which have been fused together in such a way to form a recombinant polypeptide which generates neutralising antibodies against said surface proteins, and from which the peptide fragments can be liberated intracellularly to stimulate $CD4^+$ and $CD8^+$ T cell responses. Said neutralising antibodies generated from the polypeptide of the present invention may block viral binding to host receptors and/or block viral fusion. In some embodiments, the polypeptide of the present invention stimulates antibodies against a coronavirus S protein. In some embodiments, the polypeptide of the present invention stimulates antibodies against the SARS-CoV-2 S protein. In some embodiments, the polypeptide of the present invention stimulates antibodies against the S1 and/or S2 subunits of the SARS-CoV-2 S protein, optionally against key functional regions thereof, e.g. the RBD of S1, and/or the HR1 and/or HR2 of S2.

In some embodiments, the polypeptide may be a recombinant polypeptide. In some embodiments, the polypeptide may be a Recombinant Overlapping Peptide ('ROP').

In one aspect, herein is provided a polypeptide comprising two or more peptide fragments, wherein each of the two or more peptide fragments comprise amino acid sequences derived from one or more structural protein(s) encoded by coronavirus sub-genomic RNA, preferably

the membrane protein ('M') and/or spike protein ('S') and/or small envelope protein ('E') and/or nucleocapsid protein ('N') of a coronavirus, and further comprising one or more protease cleavage site sequences located between each of the two or more peptide fragments. The skilled person could also envisage that the hemagglutinin-esterase protein ('HE'), which is present in some but not all coronaviruses, could be used. In some embodiments, the protease cleavage site sequence is exogenous.

In some embodiments, at least two of the two or more peptide fragments comprise sequences which derive from the same protein, e.g. the M, S, E, N, or HE protein, and said sequences of said at least two peptide fragments overlap – i.e. said at least two of the two or more peptide fragments comprise an 'overlapping sequence'. In some embodiments, at least two of the two or more peptide fragments comprise sequences which derive from the same protein, optionally the M, S, E, N or HE protein, and said sequences of said at least two peptide fragments overlap. In some embodiments, at least two of the two or more peptide fragments comprise sequences which derive from the same protein, optionally the M, S, E, N or HE protein, and said at least two of the two or more peptide fragments comprise an overlapping sequence.

An "overlapping sequence" is a portion or sub-sequence of an amino acid sequence which is present in two or more peptide fragments of the polypeptide of the present invention. In one embodiment, the C-terminal end of one peptide fragment of the present invention comprises an amino acid sequence which is the same or substantially similar to the amino acid sequence at the N-terminal end of another peptide fragment of the polypeptide of the invention, wherein said sequence which is the same or substantially similar is the 'overlapping sequence'. As an illustrative example, the first peptide fragment may comprise residues 1 to 10 of the amino acid sequence of the coronavirus protein from which the fragment is derived, and the second peptide fragment may comprise residues 5 to 15 of said protein sequence, thus the residues corresponding to residues 5 to 10 of the coronavirus protein sequence from which each peptide fragment is derived are the 'overlapping sequence', and the overlapping sequence is 5 amino acids long; i.e. the degree of overlap is 5 amino acids. Polypeptides comprising these overlapping sequences may be referred to as recombinant overlapping polypeptides ('ROPs'). In some embodiments, where sequences overlap, said sequences are derived from the same coronavirus protein. In some embodiments, where sequences overlap, said sequences are derived from the same coronavirus protein subunit. In some embodiments, where sequences overlap, said sequences are derived from contiguous portions of a coronavirus protein. In some embodiments, where sequences overlap, said sequences are derived from contiguous sequences of a coronavirus protein.

In some embodiments, the polypeptide may comprise multiple overlapping sequences. As an illustrative example, the first peptide fragment may comprise residues 1 to 10 of the amino acid sequence of the coronavirus protein from which the fragment is derived, the second peptide fragment may comprise residues 5 to 15, and the third peptide fragment may comprise 11 to 20. Thus, in the illustrative example, there are two overlapping sequences in the polypeptide, specifically residues 5 to 10 in the first and second sequence, and 11 to 15 in the second and third sequence. In addition, or alternatively, there may be one or more overlapping sequences, but not all the peptide fragments need contain an overlapping sequence. As an illustrative example, the first and second peptide fragments may contain an overlapping sequence defined by residues 5 to 10, but the third peptide fragment may comprise residues 16 to 25, and thus not overlap with either. In some embodiments the polypeptide may comprise 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or more overlapping sequences. Any number of overlapping sequences may be present and this is only limited by the number and size of the peptide fragments of the polypeptide.

A 'sequence which overlaps' or 'a peptide fragment which overlaps' is a sequence or peptide fragment having an overlap, as defined herein, with another sequence or peptide fragment comprised within the polypeptide of the present invention. A 'sequence which overlaps' comprises an overlapping sequence.

In one embodiment, the polypeptide of the invention comprises peptide fragments comprising a sequence which overlaps with that of one other peptide fragment within the polypeptide – for example, by means of its N-terminal sequence or its C-terminal sequence. In another embodiment, the polypeptide of the invention comprises peptide fragments comprising a sequence which overlaps with those of two other peptide fragments within the polypeptide – for example, by means of its N-terminal sequence and its C-terminal sequence. In some embodiments, the polypeptide of the invention additionally comprises one or more peptide fragment(s) which comprise a sequence which does not overlap with the sequence of any other peptide fragment contained within the polypeptide.

In some embodiments the polypeptide may comprise 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or more overlapping sequences.

In some embodiments, the overlapping sequence is 2 to 40 amino acids long, i.e. the degree of overlap between two sequences which overlap is 2 to 40 amino acids. In some embodiments, the overlapping sequence is 2 to 31 amino acids long, i.e. the degree of overlap between two sequences which overlap is 2 to 31 amino acids. In other embodiments, the

overlapping sequence is 4 to 30 amino acids long, i.e. the degree of overlap between two sequences which overlap is 4 to 30 amino acids. In other embodiments, the overlapping sequence is 6 to 20 amino acids long, i.e. the degree of overlap between two sequences which overlap is 6 to 20 amino acids. In preferred embodiments, the overlapping sequence is 8 to 17 amino acids long, i.e. the degree of overlap between two sequences which overlap is 8 to 17 amino acids. In some embodiments, the overlapping sequence is 8, 9, 10 or 11 amino acids long, i.e. the degree of overlap is 8, 9, 10, or 11 amino acids. In some embodiments, the overlapping sequence is 12 amino acids long, i.e. the degree of overlap is 12 amino acids. In other embodiments, the overlapping sequence is 13 amino acids, 14 amino acids, 15, 16, or 17 amino acids long, i.e. the degree of overlap is 13, 14, 15, 16, or 17 amino acids. In a most preferred embodiment, the overlapping sequence – or degree of overlap – is at least 8 amino acids for the generation of a cytotoxic T lymphocyte ('CTL') ($CD8^+$ T cell) response and/or at least 12 amino acids for the generation of a T helper cell ($CD4^+$ T cell) response.

Any one peptide fragment may be 2 to 55 amino acids in length, more preferably 8 to 50 amino acids in length, more preferably 12 to 45 amino acids, more preferably 20 to 40 amino acids in length. In a preferred embodiment, each peptide fragment is 25 to 40 amino acids long, more preferably 28 to 38 amino acids long, even more preferably 29 to 37 amino acids long. In preferred embodiments, each peptide fragment is 29, 30, 31, 32, 33, 34, 35, 36, or 37 amino acids in length.

In all embodiments of the invention, peptide fragments are linked together to form the polypeptide by means of at least one protease cleavage site sequence located between each linearly adjacent peptide fragment. 'Linearly adjacent' is taken here to mean peptide fragments which are immediately sequential in terms of secondary structure or amino acid sequence. In some embodiments, short linkers and/or cleavage sites, such as protease cleavage sites, can be used to separate linearly adjacent peptide fragments. In some embodiments, the protease cleavage site sequence is exogenous. Accordingly, one or more exogenous protease cleavage site sequences may separate each peptide fragment. Peptide fragments are connected by means of one or more exogenous protease cleavage site sequences. In one embodiment of the invention, two or more peptide fragments are linked together in tandem to form the polypeptide by means of at least one exogenous protease cleavage site sequence located between each linearly adjacent peptide fragment. In another embodiment, three or more peptide fragments are linked together in tandem to form the polypeptide by means of at least one exogenous protease cleavage site sequence located between each linearly adjacent peptide fragment. In another embodiment, 4 to 30, 5 to 20 peptide fragments, more preferably 10 to 15, 11 to 14, 12, or 13 peptide fragments are linked together in tandem to form the

polypeptide by means of at least one exogenous protease cleavage site sequence located between each linearly adjacent peptide fragment.

It is an object of the present invention to provide a polypeptide for the therapeutic and/or prophylactic treatment of a coronavirus infection. In some embodiments, the coronavirus is a member of the *Coronaviridae* family as defined by the Coronavirus Study Group of the International Committee on Taxonomy of Viruses, more preferably a member of the *Betacoronavirus* genus, and more preferably a member of the *severe acute respiratory syndrome-related* species (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020, DOI: 10.1038%2Fs41564-020-0695-z). In a preferred embodiment, the coronavirus infection is a SARS-CoV-2 infection. In a preferred embodiment, the polypeptide of the invention is for vaccination against and/or treatment of a SARS-CoV-2 infection.

Said prophylactic and/or therapeutic polypeptide may be provided to a mammalian subject, optionally a feline, a rodent, a primate, preferably a human. Accordingly, in some embodiments the invention provides a polypeptide for vaccination against and/or treatment of a human coronavirus infection, preferably wherein the coronavirus is a member of the *Coronaviridae* family as defined by the Coronavirus Study Group of the International Committee on Taxonomy of Viruses, more preferably a member of the *Betacoronavirus* genus, and more preferably a member of the *severe acute respiratory syndrome-related* species, and further wherein the infection is in a human. In a preferred embodiment, the coronavirus infection is a human SARS-CoV-2 infection. In a preferred embodiment, the polypeptide of the invention is for vaccination against and/or treatment of a SARS-CoV-2 infection in a human.

High degrees of homology are seen across members of the severe acute respiratory syndrome-related species. For example, the spike 'S' protein of SARS-CoV-2 has 76 % overall sequence identity to the S protein of SARS-CoV; the S1 subunit of the S protein of SARS-CoV-2 has 64 % identity to the S1 subunit of SARS-CoV; the receptor binding domain ('RBD') of the S1 subunit of SARS-CoV-2 has 74 % identity to the RBD of SARS-CoV; the S2 subunit of the S protein of SARS-CoV-2 has 90 % identity to the S2 subunit of SARS-CoV; the heptad-repeat 2 '(HR2') region of the S2 subunit of the S protein of SARS-CoV-2 has 100 % identity to the HR2 region of SARS-CoV (Jaimes *et al.*, 2020); the heptad-repeat 1 ('HR1') region of the S2 subunit of the S protein of SARS-CoV-2 has 93 % identity to the HR1 region of SARS-CoV (Xia, S., et al. (2020)). The peptide fragments of the present invention may derive from any, or multiple, coronaviruses. Due to high homology in the S protein and its component structures, and the ability to incorporate multiple peptide fragments into the present invention,

optionally from multiple coronaviruses and/or strains, in some embodiments the anti-coronavirus antibodies raised by the polypeptide of the present invention are cross-reactive against various coronaviruses and/or strains, i.e. are broad-acting.

In one embodiment, the polypeptide of the invention comprises between 2 and 30 peptide fragments derived from one or more structural protein(s) encoded by coronavirus sub-genomic RNA, preferably the M, S, E, and/or N protein of a coronavirus, optionally the M, S, E, and/or N protein of a severe acute respiratory syndrome-related coronavirus, optionally the M, S, E, and/or N protein of the SARS-CoV-2 virus. In another embodiment, the polypeptide comprises between 3 and 25 peptide fragments derived from the M, S, E, and/or N protein of a coronavirus, optionally the M, S, E, and/or N protein of a severe acute respiratory syndrome-related coronavirus, optionally the M, S, E, and/or N protein of the SARS-CoV-2 virus. In another embodiment, the polypeptide comprises 5 to 20 peptide fragments, more preferably the polypeptide comprises 10 to 15, 11 to 14, 12, or 13 peptide fragments derived from the M, S, E, and/or N protein of a coronavirus, optionally the M, S, E, and/or N protein of a severe acute respiratory syndrome-related coronavirus, optionally the M, S, E, and/or N protein of the SARS-CoV-2 virus. The skilled person will appreciate that 'derived from' carries the meaning outlined above.

In a preferred embodiment, the polypeptide of the invention comprises peptide fragments derived from the S protein of a coronavirus, optionally a betacoronavirus, optionally a severe acute respiratory syndrome-related coronavirus, optionally SARS-CoV-2, and raises antibody and T cell responses against the S protein. The skilled person will appreciate that 'derived from' carries the meaning outlined above.

In one embodiment of the invention, the polypeptide comprises two or more peptide fragments, at least one – optionally more than one – of which comprises a sequence derived from the S protein of a severe acute respiratory syndrome-related coronavirus, optionally SARS-CoV-2. In an embodiment of the invention, at least one – optionally more than one – of said two or more peptide fragments comprises a sequence derived from the S1 subunit. In an embodiment, at least one – optionally more than one – of said two or more peptide fragments comprises a sequence derived from the RBD of the S1 subunit. In some embodiments, at least one – optionally more than one – of said two or more peptide fragments comprises a sequence derived from the receptor binding motif ('RBM') of the RBD. In another preferred embodiment of the invention, at least one – optionally more than one – of said two or more peptide fragments comprises a sequence derived from the S2 subunit. In an embodiment, at least one – optionally more than one – of said two or more peptide fragments comprises a sequence

derived from the heptad repeat 2 ('HR2') domain of the S2 subunit. In another embodiment, at least one – optionally more than one – of said two or more peptide fragments comprises a sequence derived from the heptad repeat 1 ('HR1') domain of the S2 subunit. The skilled person will appreciate that 'derived from' carries the meaning outlined above.

It will be understood that there exist multiple viral variants and/or strains of *Severe acute respiratory syndrome-related coronavirus* subspecies such as SARS-CoV-2 (e.g. B.1.1.7, B.1.351, P.1, B.1.427, B.1.429), and that new viral variants will continue to emerge. The peptide fragments of the present invention may derive from any, or multiple, of such viral variant strains. The amino acid sequence of the peptide fragments can be readily adjusted to represent new mutations and variants in order to provide immune protection to a subject receiving the fusion protein of the present invention against emergent viral variant strains.

In an embodiment of the invention, the polypeptide comprises three or more peptide fragments. At least two of said three or more peptide fragments comprise sequences which overlap.

In one embodiment, at least one, at least two, or at least three of said three or more peptide fragments comprises a sequence derived from the S1 subunit. In a further embodiment, at least one, at least two, or at least three of said peptide fragments comprises a sequence derived from the RBD. In a further embodiment, at least one, at least two, or at least three of said peptide fragments comprises a sequence derived from the receptor binding motif ('RBM') of the RBD. The skilled person will appreciate that 'derived from' carries the meaning outlined above.

The RBD is a domain of the S1 subunit of S proteins which binds to a host receptor. The RBD of SARS-CoV-2 binds strongly to angiotensin-converting enzyme 2 (ACE2) of at least humans and bats (Tai, W., et al. (2020)). The RBD of SARS-CoV binds ACE2. The RBD of MERS-CoV binds dipeptidyl peptidase 4 (DPP4). The RBD of SARS-CoV-2 may be represented as SEQ ID NOs: 15 or 16 and in some embodiments, the RBD comprises residues 318 to 541 of SARS-CoV-2 S proteins (Yi, C., et al. (2020)). In other embodiments, the RBD may comprise residues 319 to 529, 331-524, or 336-516 of SARS-CoV-2 S proteins (Shang, J., *et al.* (2020); Tai, W., et al. (2020); Lan, J., *et al.* (2020)). The RBD of SARS-CoV may comprise residues 306-527 and/or 318-510 of SARS-CoV S proteins; the RBD of MER-CoV S may comprise residues 377-588 of MERS-CoV S proteins (Yi, C., et al. (2020); Tai, W., et al. (2020)). The skilled person will appreciate that the boundaries of the RBD as defined by residue numbers

may vary slightly and as seen above. Thus, the RBD may comprise the amino acid sequences having residues defined above or variants thereof.

The RBM is a motif of the S1 subunit of S proteins, and within the RBD, which binds to a host receptor. The RBM of SARS-CoV-2 may be represented as SEQ ID NO: 17 and, in some embodiments, the RBM of SARS-CoV-2 comprises residues 438-506 of SARS-CoV-2 S proteins (Lan, J., (2020)). The skilled person will appreciate that the boundaries of the RBM as defined by residue numbers may vary slightly. Thus, the RBM may comprise the amino acid sequences having residues defined above or variants thereof.

The RBD and RBM are key functional regions of the S1 subunit of the coronavirus S protein, being essential for coronavirus-host receptor binding. Antibodies which bind to, and are directed against, key functional regions are able to block, interfere with, or prevent the viral function of said regions, sterically or otherwise. By providing one or more sequences of the RBD, optionally further the RBM, the polypeptide of the present invention stimulates the generation of neutralising and/or broad-acting antibodies against the RBD, and optionally the RBM, which preclude binding of viral S1 to host receptors.

In one embodiment, the polypeptide stimulates antibody production, the antibodies being specific to the relevant coronavirus protein. In another embodiment, the polypeptide stimulates the production of neutralising antibodies, optionally against the S protein. In a preferred embodiment, the polypeptide stimulates the production of neutralising antibodies against the S1 subunit, optionally against the RBD, further optionally against the RBM.

In an embodiment of the invention, at least two or at least three of said three or more peptide fragments comprise a sequence derived from the S1 subunit, optionally wherein these sequences overlap. In a further embodiment, at least two or at least three of said peptide fragments comprise a sequence derived from the RBD, optionally wherein these sequences overlap. In a further embodiment, at least two or at least three of said peptide fragments comprise a sequence derived from the RBM of the RBD, optionally wherein these sequences overlap.

In another embodiment, at least one, at least two, or at least three of said three or more peptide fragments comprise a sequence derived from the S2 subunit, optionally wherein these sequences overlap. In an embodiment, at least one, at least two, or at least three of said peptide fragments comprise a sequence derived from the HR1 region, optionally wherein these sequences overlap. In another embodiment, at least one, at least two, or at least three

of said peptide fragments comprise a sequence derived from the HR2 region, optionally wherein these sequences overlap. In yet another embodiment, at least one peptide fragment is derived from the HR1 region and at least one peptide fragment is derived from the HR2 region, optionally wherein these sequences overlap. In an embodiment, at least one, at least two, or at least three of said peptide fragments comprise a sequence derived from a region adjacent the HR1 and/or HR2 regions, optionally from the intervening region between HR1 and HR2, optionally wherein these sequences overlap. The skilled person will appreciate that 'derived from' carries the meaning outlined above.

HR1 is a heptad repeat which forms a 6-helical bundle (6HB) with the HR2 heptad repeat which brings the viral envelope into close proximity with host cell membranes for fusion. HR1 may be represented as SEQ ID NO: 35 and in some embodiments comprises residues 910 to 988 of SARS-CoV-2 S proteins. In other embodiments, HR1 may comprise residues 912 to 984 or 920 to 970 of SARS-CoV-2 S proteins (Xia, S., et al. (2020)). The HR1 of SARS-CoV may comprise residues 902 to 952 of SARS-CoV S proteins. The skilled person will appreciate that the boundaries of the HR1 as defined by residue numbers may vary slightly. Thus, the HR1 may comprise the amino acid sequences having residues defined above or variants thereof.

HR2 is a heptad repeat which forms a 6-helical bundle (6HB) with the HR2 heptad repeat which brings the viral envelope into close proximity with host cell membranes for fusion. HR2 may be represented as SEQ ID NO: 19 and in some embodiments comprises residues 1159-1211 of SARS-CoV-2 S proteins. In other embodiments, HR2 may comprise residues 1163-1202 of SARS-CoV-2 S proteins (Xia, S., et al. (2020)). The HR2 of SARS-CoV may comprise residues 1145-1184 of SARS-CoV S proteins. The skilled person will appreciate that the boundaries of the HR2 as defined by residue numbers may vary slightly. Thus, the HR2 may comprise the amino acid sequences having residues defined above or variants thereof.

The HR1 and HR2 regions are key functional regions of the S2 subunit of the coronavirus S protein, being essential for fusion of the viral envelope with the host cell membrane. Antibodies which bind to or close to, i.e. are directed against, key functional regions are able to block, interfere with, or prevent the viral function of said regions, sterically or otherwise. By providing one or more sequences of HR1 and/or HR2, the polypeptide of the present invention stimulates the generation of neutralising and/or broad-acting antibodies against HR1 and/or HR2 which preclude viral entry into host cells. This is distinct from any use of isolated amino acid sequences of HR1 and/or HR2 (whether in native or stapled form) to directly inhibit coronavirus entry via direct binding of said HR1 and/or HR2 sequences to the coronavirus'

own HR1 and/or HR2 and subsequent preclusion of formation of the coronavirus-associated HR1-HR2 6HB, as described in e.g. CN111560054 and CN111732637.

In a preferred embodiment, the polypeptide stimulates the production of neutralising antibodies against the S2 subunit, further optionally against the HR1 and/or HR2 regions.

In a preferred embodiment of the invention, the polypeptide comprises three or more peptide fragments, wherein at least two of said three or more peptide fragments comprise sequences which overlap, and further wherein at least one of said three or more peptide fragments comprises a sequence derived from the S1 subunit of the S protein of a severe acute respiratory syndrome-related coronavirus, optionally SARS-CoV-2, and at least one of said three or more peptide fragments comprises a sequence derived from the S2 subunit of the S protein of a severe acute respiratory syndrome-related coronavirus, optionally SARS-CoV-2. In a preferred embodiment, the at least one peptide fragment comprising a sequence derived from S1 comprises a sequence derived from the RBD of S1, optionally the RBM, and the at least one peptide fragment comprising a sequence derived from S2 comprises a sequence derived from HR1 and/or HR2. In one embodiment a polypeptide comprises at least two peptide fragments each comprising a sequence derived from S1 and at least one peptide fragment comprising a sequence derived from S2, wherein the at least two peptide fragments comprising sequences derived from S1 comprise an overlapping sequence. In another embodiment, a polypeptide comprises at least two peptide fragments each comprising a sequence derived from S2 and at least one peptide fragment comprising a sequence derived from S1, wherein the at least two peptide fragments comprising sequences derived from S1 comprise an overlapping sequence. The skilled person will appreciate that 'derived from' carries the meaning outlined above.

In embodiments in which the polypeptide comprises peptide fragments comprising sequences derived from S1 and comprises peptide fragments comprising sequences derived from S2, in some embodiments two or more of the sequences derived from S1 overlap. In other embodiments two or more of the sequences derived from S2 overlap. In other embodiments two or more of the sequences derived from S1 overlap and two or more of the sequences derived from S2 overlap.

For example, in some embodiments the polypeptide comprises four peptide fragments, three of which comprise sequences derived from S1 and one of which comprises a sequence derived from S2. In one such embodiment all three of said three sequences derived from S1 overlap. In another embodiment two of said three sequences derived from S1 overlap.

Equally, for example, in embodiments in which the polypeptide comprises four peptide fragments, three of which comprise sequences derived from S2 and one of which comprises a sequence derived from S1, in one such embodiment all three of said three sequences derived from S2 overlap. In another embodiment two of said three sequences derived from S2 overlap.

In a further example, in some embodiments the polypeptide comprises four peptide fragments, two of which comprise sequences derived from S1 and two of which comprise sequences derived from S2. In one embodiment none of the sequences overlap. In another embodiment said sequences derived from S1 overlap. In another embodiment, said sequences derived from S2 overlap. In another embodiment said sequences derived from S1 overlap and said sequences derived from S2 overlap.

The skilled person will appreciate that the logic in the above examples can apply to a polypeptide having any number of peptide fragments, wherein some comprise sequences derived from S1 and some comprise sequences derived from S2.

In one embodiment, the polypeptide comprises 5 or more peptide fragments. In another embodiment, the polypeptide comprises 10 or more peptide fragments. In another embodiment, the polypeptide comprises 12 peptide fragments derived from the S protein of a severe acute respiratory syndrome-related coronavirus, preferably SARS-CoV-2. In one embodiment, the polypeptide comprises 11 peptide fragments derived from the S protein of a severe acute respiratory syndrome-related coronavirus, preferably SARS-CoV-2. In one embodiment, the polypeptide comprises 13 peptide fragments derived from the S protein of a severe acute respiratory syndrome-related coronavirus, preferably SARS-CoV-2. In one embodiment, the polypeptide comprises 14, 15, or 16 peptide fragments derived from the S protein of a severe acute respiratory syndrome-related coronavirus, preferably SARS-CoV-2.

In some embodiments, the polypeptide comprises 8 to 18 peptide fragments, 9 to 17 peptide fragments, 10 peptide fragments, 11 peptide fragments, 12 peptide fragments, 13 peptide fragments, 14, 15, or 16 peptide fragments, each of which comprises a sequence derived from the S1 subunit or the S2 subunit of the S protein of a severe acute respiratory syndrome-related coronavirus, optionally SARS-CoV-2. Preferably, said polypeptide comprises both peptide fragments comprising sequences derived from S1 and peptide fragments comprising sequences derived from S2. In some embodiments, said sequences derived from the S1 subunit overlap, and said sequences derived from the S2 subunit overlap.

In one embodiment, the polypeptide comprises at least three peptide fragments comprising a sequence derived from the RBD. In another embodiment, at least 4 peptide fragments comprise a sequence derived from the RBD. In another embodiment, at least 5, 6, or 7 peptide fragments comprise a sequence derived from the RBD. In some embodiments, 8, 9, 10, 11, or 12 peptide fragments comprise a sequence derived from the RBD. In some embodiments, 9 peptide fragments comprise a sequence derived from the RBD. In some embodiments, more than 12 peptide fragments comprise a sequence derived from the RBD. Optionally, at least one, at least two, or at least three of said peptide fragments comprising a sequence derived from the RBD comprise a sequence derived from the RBM. In some embodiments, the polypeptide comprises 9 peptide fragments which comprise a sequence derived from the RBD, wherein five of said peptide fragments comprising a sequence derived from the RBD comprise a sequence derived from the RBM.

In some embodiments, the polypeptide comprises at least one peptide fragment comprising a sequence derived from HR1. In another embodiment, at least 2 peptide fragments comprise a sequence derived from HR1. In another embodiment, 3 peptide fragments comprise a sequence derived from HR1. In some embodiments, 4, 5, 6, or more peptide fragments comprise a sequence derived from HR1.

In one embodiment, the polypeptide comprises at least one peptide fragment comprising a sequence derived from HR2. In another embodiment, at least 2 peptide fragments comprise a sequence derived from HR2. In another embodiment, 3 peptide fragments comprise a sequence derived from HR2. In some embodiments, 4, 5, 6, or more peptide fragments comprise a sequence derived from HR2.

The SARS-CoV-2 NTD of S1 may comprise residues 13 to 303 of the S protein. The SARS-CoV-2 RBD may comprise residues 318 to 541 of the S protein. Alternatively, the SARS-CoV-2 RBD may comprise residues 319 to 529 (SEQ ID NO: 16) of the S protein. The SARS-CoV-2 RBM may comprise residues 438 to 506 (SEQ ID NO: 17) of the S protein. The SARS-CoV-2 fusion peptide may comprise residues 816 to 855 of the S protein. The SARS-CoV-2 HR1 may comprise residues 910 to 988 (SEQ ID NO: 35) of the S protein. The SARS-CoV-2 HR2 may comprise residues 1159 to 1211 (SEQ ID NO: 19) of the S protein. The region of the SARS-CoV-2 S2 intervening between HR1 and HR2 may comprise residues 970 to 1163 of the S protein.

In some embodiments, the polypeptide of the present invention comprises peptide fragments which tile the whole or one or more contiguous portion(s) of one or more key functional

region(s) of a coronavirus S protein, optionally the SARS-CoV-2 S protein. To 'tile' or means to cover (or represent) a sequence in an overlapping manner. For example, three peptide fragments comprising amino acids 1 to 10, 6 to 18, and 13 to 21 of a hypothetical protein may be said to tile amino acids 1 to 21 of said hypothetical protein. In this way, key epitopes (neutralising linear antibody epitopes and/or T cell epitopes) contained within said key functional region(s) are represented in the peptide fragments, and may be represented more than once (i.e. in more than one peptide fragment) in full form, partial form, or both. In some embodiments, the polypeptide comprises peptide fragments which tile the RBM. In some embodiments, the polypeptide comprises at least two peptide fragments which tile the whole RBM. In some embodiments, at least three or at least four peptide fragments tile the whole RBM. In some embodiments, five peptide fragments tile the whole RBM.

In some embodiments, the polypeptide comprises peptide fragments which tile the whole or a contiguous portion of the RBD. In some embodiments, the polypeptide comprises at least two, at least three, or at least four peptide fragments which tile a contiguous portion of the RBD, optionally at least five peptide fragments which tile a contiguous portion of the RBD, optionally at least six, at least seven, at least eight peptide fragments which tile a contiguous portion of the RBD. In some embodiments, the polypeptide comprises nine peptide fragments which tile a contiguous portion of the RBD. Optionally, at least one, at least two, or at least three of said peptide fragments tiling a contiguous portion of the RBD tile the whole or a portion of the RBM. In some embodiments, the polypeptide comprises nine peptide fragments which tile a contiguous portion of the RBD, wherein five of said peptide fragments tiling a contiguous portion of the RBD tile the whole RBM.

In some embodiments, the polypeptide comprises peptide fragments which tile the whole or a contiguous portion of the NTD of S1. In some embodiments, the polypeptide comprises peptide fragments which tile the whole or a contiguous portion of HR1 and/or HR2 and/or the intervening region between HR1 and HR2. In some embodiments, at least one, at least two, or at least three peptide fragments tile a contiguous portion of HR1. In some embodiments at least one, at least two, or at least three peptide fragments tile a contiguous portion of HR2. In some embodiments, the polypeptide comprises peptide fragments which tile the whole or a contiguous portion of the S2 fusion peptide.

In some embodiments, the polypeptide comprises nine peptide fragments which tile a contiguous portion of the RBD, wherein five of said peptide fragments tiling a contiguous portion of the RBD tile the whole RBM, and further comprises three peptide fragments which tile a contiguous portion of HR2.

Said peptide fragments which tile the whole or one or more contiguous portion(s) of one or more key functional region(s) may be arranged sequentially within the amino acid sequence of the polypeptide, or may be arranged non-sequentially.

Antibodies raised against a whole viral protein will differ amongst members of a population (Watson, Glanville & Marasco, 2017). In some embodiments, the polypeptide of the invention provides multiple different sequences of the RBD which overlap, optionally further the RBM, so the polypeptide of the present invention is effective at generating neutralising antibodies in a larger proportion of the vaccinated population. In some embodiments, the multiple and overlapping peptide fragments of the present invention compensate this population-wide variation; they provide alternative epitopes for immune recognition via the ability to tile, or provide greater coverage of, key functional regions such as the RBD, optionally the RBM.

In one embodiment, at least one, at least two, or at least three of three or more peptide fragments comprise antibody epitopes, preferably neutralising antibody epitopes.

In some embodiments, the polypeptide of the present invention generates antibodies against those regions of a coronavirus protein against which it is targeted, i.e. those sequences of the coronavirus protein which are comprised within the peptide fragments of the polypeptide. In some embodiments, the polypeptide comprises one, two, three, or more than three antibody epitopes. In some embodiments, the polypeptide comprises two antibody epitopes. In some embodiments, the polypeptide comprises three or more antibody epitopes.

Preferably, the antibody epitopes are linear antibody epitopes (also known as linear B cell epitopes) of a coronavirus S protein, optionally the SARS-CoV-2 S protein. In some embodiments, the full sequence of each antibody epitope is represented once within the polypeptide, optionally by means of multiple sequences which overlap. In some embodiments, the full sequence of each antibody epitope is represented more than once, optionally by means of multiple sequences which overlap, within the polypeptide of the present invention. In some embodiments, a peptide fragment comprises a full antibody epitope sequence. In some embodiments, one, two, three, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 peptide fragments comprise a full antibody epitope sequence. In some embodiments, at least two peptide fragments comprise a full antibody epitope sequence. In embodiments, at least three peptide fragments comprise a full antibody epitope sequence. In some embodiments, at least four peptide fragments comprise a full antibody epitope sequence. In some embodiments, one, two, three, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 peptide fragments comprise at least part of a linear antibody epitope sequence. In some embodiments, at least three peptide fragments

comprise at least part of a linear antibody epitope sequence. In some embodiments, at least 4, 5, 6, 7, or 8 peptide fragments comprise at least part of an antibody epitope sequence In some embodiments, 9 peptide fragments comprise at least part of an antibody epitope sequence. Since in some embodiments the peptide fragments comprise sequences which overlap, in some embodiments at least one (optionally two or three or more) peptide fragment(s) comprises a full antibody epitope of a coronavirus S protein, while at least one (optionally two, three, four, or more) other peptide fragment(s) comprises a partial sequence of said antibody epitope.

In some embodiments, some or all of these epitopes are neutralising antibody epitopes. In some embodiments, the polypeptide comprises at least one neutralising antibody epitope. In some embodiments, the polypeptide comprises at least two neutralising antibody epitopes. In some embodiments, the polypeptide comprises three or more neutralising antibody epitopes. Preferably, the neutralising antibody epitopes are linear neutralising antibody epitopes of a coronavirus S protein, optionally the SARS-CoV-2 S protein. A neutralising antibody epitope sequence is a sequence against which neutralising antibodies are generated. The RBD and NTD of S1, the fusion peptide of S2, the areas connecting and/or immediately adjacent HR1 and HR2, and the S2' cleavage site have been documented as hot spots for neutralising antibody stimulation. Some linear epitope regions are beginning to be defined from patient cohorts, for example residues S562-579, S818-835, S1148-1159, S455–469, S556–570, S16–30, S205–219, S455–469, S475–499, S556–570, S721–733, S793–812, S1106–1120, S1205–1222 of the SARS-CoV-2 (where S denotes S protein) (Poh, C.M., et al. (2020); Shuai Lu, et al. (2021); Liu, L., et al. (2020)) and for S2 in e.g. CN111647053. The polypeptide of the invention may comprise one, two, three, at least three or more of these epitopes. In some embodiments the polypeptide comprises at least one, at least two, at least three, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or 16 of these epitopes. In some embodiments the polypeptide comprises at least one, at least two, at least three, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or 16 peptide fragments each comprising one or more of these epitopes, in full or in part. It will be appreciated that the boundaries of these epitopes as defined by residue numbers can vary slightly. Therefore, in some embodiments, a peptide fragment may comprise one or more of the above recited epitopes or variants thereof.

In some embodiments, neutralising antibodies sera of patients receiving the polypeptide of the invention demonstrates IC_{50} (half maximal inhibitory concentration) values of at least 100, as measured by pseudotyped lentivirus, for example as per Poh, C.M., et al. (2020). In some embodiments, sera of patients receiving the polypeptide of the invention has IC_{50} of at least 200. In some embodiments, sera of patients receiving the polypeptide of the invention has

IC_{50} of at least 300. In some embodiments, sera of patients receiving the polypeptide of the invention has IC_{50} of at least 500. In some embodiments, sera of patients receiving the polypeptide of the invention has IC_{50} of at least 600. In some embodiments, sera of patients receiving the polypeptide of the invention has IC_{50} of at least 800. In some embodiments, sera of patients receiving the polypeptide of the invention has IC_{50} between 900 and 2000. In some embodiments, this neutralisation is stimulated after a single administration. In some embodiments, this neutralisation is stimulated after one or two booster administrations.

In some embodiments, mean neutralising titres are between 1- and 5-fold those seen in COVID-19 human convalescent sera ('HCS'). In some embodiments, mean neutralising titres are at least those seen in COVID-19 HCS, at least 2-fold those seen in COVID-19 HCS, or at least 3-fold.

In some embodiments, the neutralising antibody titre may be further increased by co-administering the polypeptide of the present invention with a native amino acid sequence derived from a coronavirus S protein against which the polypeptide of the present invention is targeted (by means of the sequences comprised within the polypeptide's peptide fragments). For example, in some embodiments the polypeptide of the present invention comprises peptide fragments comprising sequences derived from a coronavirus RBD (and therefore can be said to be targeting the RBD, by stimulating anti-RBD antibody production). Such a polypeptide may be co-administered with the RBD of said coronavirus, and this can increase the neutralising ability of resulting antibodies. In some embodiments the polypeptide of the present invention comprises one or more peptide fragments comprising sequences derived from a coronavirus RBM. Such a polypeptide may be co-administered with the RBM of said coronavirus, and this can increase the neutralising ability of resulting antibodies (see Example 3.4 and Figure 6 and 7). In some embodiments, the polypeptide of the present invention comprises one or more peptide fragments comprising sequences derived from a coronavirus HR1 and/or HR2. Such a polypeptide may be co-administered with the whole or a portion (containing HR1 and/or HR2 as appropriate) of the S2 protein of said coronavirus, and this can increase the neutralising ability of resulting antibodies. In some embodiments, mean IC_{50} of antibodies produced upon co-administration of a polypeptide of the invention and a native amino acid sequence (e.g. the RBM of a coronavirus, optionally SARS-CoV-2) is higher than the mean IC_{50} of antibodies produced upon administration of polypeptide of the invention alone.

The native amino acid sequence to be co-administered may comprise the wildtype sequence of the whole or a portion of a coronavirus protein (optionally a SARS-CoV-2 protein), or a variant thereof.

In some embodiments, sera of patients receiving a combination of polypeptide of the invention and a native coronavirus amino acid sequence (eg the RBD, optionally the RBM of a coronavirus optionally SARS-CoV-2) has IC_{50} of at least 100. In some embodiments, sera of patients receiving a combination of polypeptide of the invention and a native coronavirus amino acid sequence (eg the RBD, optionally the RBM of a coronavirus optionally SARS-CoV-2) has IC_{50} of at least 200. In some embodiments, sera of patients receiving a combination of polypeptide of the invention and a native coronavirus amino acid sequence (eg the RBD, optionally the RBM of a coronavirus optionally SARS-CoV-2) has IC_{50} of at least 300. In some embodiments, sera of patients receiving a combination of polypeptide of the invention and a native coronavirus amino acid sequence (eg the RBD, optionally the RBM of a coronavirus optionally SARS-CoV-2) has IC_{50} of at least 500. In some embodiments, sera of patients receiving a combination of polypeptide of the invention and a native coronavirus amino acid sequence (eg the RBD, optionally the RBM of a coronavirus optionally SARS-CoV-2) has IC_{50} of at least 600. In some embodiments, sera of patients receiving a combination of polypeptide of the invention and a native coronavirus amino acid sequence (eg the RBD, optionally the RBM of a coronavirus optionally SARS-CoV-2) has IC_{50} of at least 700. In some embodiments, sera of patients receiving a combination of polypeptide of the invention and a native coronavirus amino acid sequence (eg the RBD, optionally the RBM of a coronavirus optionally SARS-CoV-2) has IC_{50} of between 800 and 3000. In some embodiments, this neutralisation is stimulated after a single co-administration. In some embodiments, this neutralisation is stimulated after one or two booster co-administrations.

In some embodiments, mean neutralising titres are between 1- and 5-fold those seen in COVID-19 human convalescent sera. In some embodiments, mean neutralising titres are at least those seen in COVID-19 HCS, at least 2-fold those seen in COVID-19 HCS, or at least 3-fold.

In some embodiments, the polypeptide of the present invention and the native coronavirus amino acid sequence are co-administered simultaneously. In some embodiments, the polypeptide and the native coronavirus amino acid sequence are co-administered simultaneously as a single formulation. In some embodiments, the polypeptide and the native coronavirus amino acids are produced separately and mixed to form a single formulation for administration to a subject. In other embodiments, the polypeptide and the native coronavirus

amino acid sequence are produced together e.g. expressed from a single DNA vector or construct, and co-purified. In some embodiments, the polypeptide and native coronavirus amino acid sequence are co-administered sequentially or in a staggered regimen.

In some embodiments, the polypeptide of the present invention and the native coronavirus amino acid sequence are co-administered with an adjuvant (e.g. MPL) which may be administered as a formulation with both of the polypeptide of the present invention and the native coronavirus amino acid sequence, or with either one of the polypeptide of the present invention and the native coronavirus amino acid sequence (the other being co-administered as a separate formulation).

In some embodiments, one or more exogenous protease cleavage site sequences are located between each peptide fragment of the polypeptide of the present invention. In preferred embodiments, the cleavage sequence itself is not immunogenic and does not form a target for antibody or T cell responses either pre- or post-cleavage.

In an embodiment, the one or more exogenous protease cleavage site sequences are cleavage site sequences of a protease present in the subject (also known as the 'host') to whom the polypeptide is administered, such that the polypeptide may be cleaved within the host into its peptide fragments. Said protease may act extracellularly or, more preferably, intracellularly. Said protease may be a non-host protease delivered in combination with the polypeptide or its encoding nucleic acid. More preferably, said protease is a host protease. A host protease may be constitutively present, present only upon induction, or otherwise.

In some embodiments, the polypeptide comprises an intracellularly cleavable arrangement of peptide fragments.

In some embodiments, said one or more exogenous protease cleavage site sequence(s) is a cleavage site sequence of a furin, a trypsin, or a cell surface transmembrane protease/serine ('TMPRSS') protease. In some embodiments, said one or more exogenous protease cleavage site sequence(s) is a cleavage site sequence of a serine endopeptidase (for example Factor X or Factor Xa, optionally wherein the cleavage site is Ile-Glu-Gly-Arg, wherein cleavage occurs after the Arg). In some embodiments, said one or more exogenous protease cleavage site sequence(s) is a cleavage site sequence of HRV 3 C protease, optionally wherein the cleavage site is Leu-Glu-Val-Leu-Phe-Gln/Gly-Pro, wherein cleavage occurs between the glutamyl and glycyl residues. In some embodiments, said one or more exogenous protease cleavage site sequence(s) may be a cleavage site sequence of HIV protease. In some

embodiments, said one or more exogenous protease cleavage site sequence(s) may be a cleavage site sequence of metalloproteinases. In some embodiments, said one or more exogenous protease cleavage site sequence(s) may be a cleavage site sequence of tryptases. In some embodiments, said one or more exogenous protease cleavage site sequence(s) may be a cleavage site sequence of other proteases such as CD13 (human aminopeptidase N). In a preferred embodiment, the one or more exogenous protease cleavage site sequence(s) is a cleavage site sequence of a cathepsin, optionally a cysteine cathepsin. Cysteine cathepsins include, but are not limited to, cathepsin B, cathepsin K, cathepsin L, cathepsin S, and cathepsin X.

In a preferred embodiment, the one or more exogenous protease cleavage site sequence(s) is a cleavage site sequence of cathepsin S. Cathepsin S recognises and cleaves at a number of amino acid sequences, any of which could be used in the present invention, including but not limited to Arg-Cys-Gly-Leu, Thr-Val-Gly-Leu, Thr-Val-Gln-Leu, X-Asn-Leu-Arg, X-Pro-Leu-Arg, X-Ile-Val-Gln, and X-Arg-Met-Lys, wherein X is any amino acid. In some embodiments, the one or more exogenous protease cleavage sites may be any combination of those mentioned above, in any number.

In one embodiment of the present invention, X-Arg-Met-Lys is the exogenous protease cleavage site sequence. In embodiments exemplified herein, Leu-Arg-Met-Lys ('LRMK', SEQ ID NO: 34) is the exogenous protease cleavage site sequence.

In one embodiment, the Leu-Arg-Met-Lys 'LRMK' cleavage site sequence of cathepsin S is the one or more exogenous protease cleavage site sequence(s), having amino acid sequence: LRMK **(SEQ ID NO: 34)**

$CD8^+$ T cells (also 'Cytotoxic T Lymphocytes', 'CTLs') target and lyse diseased and/or infected cells. T cell responses are important in SARS-CoV-2 clearance and immunity (Jagannathan, P., Wang, T.T. (2021). Traditionally, MHC class I molecules are understood to present non-self fragments of intracellular origin for $CD8^+$ T cell recognition and activation; for example, a virally infected cell presents the fragmented products of proteasomal digestion of intracellular viral proteins on MHC class I molecules. $CD4^+$ T cells ('T helper cells', 'T_H') assist in the activation and expansion of other immune cells, including T cells and B cells. Traditionally, MHC class II molecules are understood to present non-self fragments of molecules of extracellular origin for $CD4^+$ T cell recognition and activation, wherein the non-self fragments have been internalised by antigen-presenting cells and degraded in lysosomes. More recently, cross-presentation has been shown known to occur in professional antigen-precenting cells

at low levels in addition to these traditional pathways, whereby internalised extracellular non-self fragments may be presented on MHC class I molecules e.g. for priming.

The Applicant has previously shown that CD4+ and CD8+ T cell responses can be generated by administration of pools of peptides having overlapping sequences and comprising T cell epitopes, as described in WO2007125371A2, or by administration of a long protein comprising multiple T cell epitopes fused together in a protease-cleavable manner, as described in WO2016095812A1 and Cai, L. *et al.*, (2017).

According to the present invention, the peptide fragments of the present invention, having been cleaved by a protease, may be processed and presented, for example via MHC class I and class II molecules, to cells of the immune system. Amino acid sequences deriving from the peptide fragments of the present invention stimulate $CD8^+$ and $CD4^+$ T cells via their presentation via MHC class I and class II molecules, respectively.

A number of sequences within coronavirus proteins are known to be, or are predicted to be, T cell epitopes. For example, residues 865-873 of S protein, and residues 1208-1216 of S protein, as well as sequences from the M protein and from non-structural/accessory proteins (Nelde, A., et al. (2021)). In some embodiments, the polypeptide of the invention comprises a peptide fragment comprising SEQ ID NO: 37. In some embodiments, the polypeptide of the invention comprises a peptide fragment comprising SEQ ID NO: 38.

It is likely that many more remain to be identified. In one embodiment of the invention, one or more peptide fragments comprise at least one known or suspected $CD8^+$ T cell epitope (optionally from a coronavirus protein, optionally the S protein, optionally the S protein of SARS-CoV-2) and/or at least one known or suspected $CD4^+$ T cell epitope (optionally from a coronavirus protein, optionally the S protein, optionally the S protein of SARS-CoV-2). In another embodiment, the polypeptide comprises one, two, three, four, five, at least 6, 7, 8, 9, 10, 11, or at least 12 $CD8^+$ epitopes (optionally from a coronavirus protein, optionally the S protein, optionally the S protein of SARS-CoV-2). In another embodiment, the polypeptide comprises one, two, three, four, five, at least 6, 7, 8, 9, 10, 11, or at least $CD4^+$ epitopes (optionally from a coronavirus protein, optionally the S protein, optionally the S protein of SARS-CoV-2).

In some embodiments, at least one peptide fragment of the polypeptide of the invention comprises one or more CD4+ T cell epitopes of a coronavirus protein, optionally a coronavirus S protein, optionally a SARS-CoV-2 S protein. In some embodiments, at least one peptide

fragment comprises a CD4+ T cell epitope, optionally the full sequence of the epitope. In some embodiments, at least two peptide fragments comprise a CD4+ T cell epitope, optionally the full sequence of the epitope. In some embodiments, at least three peptide fragments comprise a CD4+ T cell epitope, optionally the full sequence of the epitope. Since in some embodiments the peptide fragments comprise sequences which overlap, in some embodiments at least one (optionally two or three) peptide fragment(s) comprises a full CD4+ T cell epitope of a coronavirus protein, while at least one (optionally two, three, four, or more) other peptide fragment(s) comprises a partial sequence of said T cell epitope.

In some embodiments, at least one peptide fragment of the polypeptide of the invention comprises one or more CD8+ T cell epitopes of a coronavirus protein, optionally a coronavirus S protein, optionally a SARS-CoV-2 S protein. In some embodiments, at least one peptide fragment comprises a CD8+ T cell epitope, optionally the full sequence of the epitope. In some embodiments, at least two peptide fragments comprise a CD8+ T cell epitope, optionally the full sequence of the epitope. In some embodiments, at least three peptide fragments comprise a CD8+ T cell epitope, optionally the full sequence of the epitope. Since in some embodiments the peptide fragments comprise sequences which overlap, in some embodiments at least one (optionally two or three) peptide fragment(s) comprises a full CD8+ T cell epitope of a coronavirus protein, while at least one (optionally two, three, four, or more) other peptide fragment(s) comprises a partial sequence of said T cell epitope.

In some embodiments, at least one peptide fragment comprises some or all of a CD4+ T cell epitope of a coronavirus protein and some or all of a CD8+ T cell epitope of a coronavirus protein. In some embodiments, at least one T cell epitope comprised within the polypeptide or a peptide fragment thereof is a cross-reactive T cell epitope.

In some embodiments, the polypeptide of the invention stimulates the T cell response. In some embodiments, the polypeptide of the invention stimulates the T cell response in a manner specific to the coronavirus protein(s) from which the polypeptide is derived. In some embodiments, the polypeptide stimulates the $CD8^+$ T cell response. In some embodiments, the polypeptide stimulates the $CD4^+$ T cell response. In some embodiments, the polypeptide of the invention stimulates both the $CD8^+$ and $CD4^+$ T cell response.

In some embodiments, the polypeptide of the invention comprises overlapping peptide fragments, which further strengthens the T cell response (Zhang *et al.*, 2009). Further, the use of overlapping peptides more comprehensively represents the range of potential T cell epitopes.

In some embodiments, T cell responses are specific to the RBD, optionally the SARS-CoV-2 RBD. In some embodiments, T cell responses are specific to the HR1 and/or HR2, optionally the SARS-CoV-2 HR1 and/or HR2. T cell stimulation can be quantified via direct *ex vivo* ELISpot assay with peripheral blood mononuclear cells (PBMCs). In some embodiments, the polypeptide of the present invention stimulates CD8+ and CD4+ T cell responses of between at least 50, at least 70, or at least 100 IFNγ spot forming units (SFU)/10^6 PBMC after overnight stimulation. In some embodiments, the polypeptide of the present invention stimulates median CD8+ and CD4+ T cell responses of at least 200 SFU/10^6 PBMC after overnight stimulation. In some embodiments, the polypeptide of the present invention stimulates median CD8+ and CD4+ T cell responses of at least 500 SFU/10^6 PBMC after overnight stimulation. In some embodiments, the polypeptide of the present invention stimulates median CD8+ and CD4+ T cell responses of at least 700 SFU/10^6 PBMC after overnight stimulation. In some embodiments, the polypeptide of the present invention stimulates median CD8+ and CD4+ T cell responses of at least 800, at least 900, at least 1000, or at least 1200 SFU/10^6 PBMC after overnight stimulation. In some embodiments, the polypeptide of the present invention stimulates median CD8+ and CD4+ T cell responses of between 50 and 1200 SFU/1x10^6 PBMC. In some embodiments, the polypeptide of the present invention stimulates median CD8+ and CD4+ T cell responses of between 50 and 1200 SFU/2.5x10^5 PBMC.

In some embodiments, mean T cell responses (e.g. as measured by spot forming units (SFU)/10^6 PBMC) are between 1- and 5-fold those seen in PBMCs derived from COVID-19 convalescent patients. In some embodiments, mean T cell responses are at least those seen in PBMCs derived from COVID-19 convalescent patients, at least 2-fold those seen in PBMCs derived from COVID-19 convalescent patients, or at least 3-fold.

The presence of multiple, optionally overlapping, peptide fragments in polypeptides of the present invention means said polypeptides have broad applicability across various HLA types within any given population, and reduces the need for HLA typing, via the ability to tile, or provide greater coverage of, one or more epitopes and by providing alternative options for immune recognition.

In a preferred embodiment, the polypeptide comprises 8 to 18 peptide fragments, 9 to 17 peptide fragments, 10 peptide fragments, 11 peptide fragments, 12 peptide fragments, 13 peptide fragments, 14, 15, or 16 peptide fragments, some of which comprise overlapping sequences derived from the S1 subunit, preferably the RBD and/or RBM, of the S protein of a severe acute respiratory syndrome-related coronavirus, preferably SARS-CoV-2, and the

remainder of which comprise overlapping sequences derived from the S2 subunit, preferably the HR1 and/or HR2 region, of the S protein of a severe acute respiratory syndrome-related coronavirus, preferably SARS-CoV-2, and wherein each peptide fragment is separated in amino acid sequence from its adjacent peptide fragment(s) by means of a LRMK cleavage site sequence of cathepsin S, such that each peptide fragment can be intracellularly liberated from the polypeptide. In said preferred embodiment, the polypeptide raises neutralising antibodies which, in the presence of infectious viral particles, block viral binding to host receptors and block viral entry and, further, the polypeptide stimulates T cell responses against the peptide fragments of the polypeptide and/or T cell epitopes contained therein. The skilled person will appreciate that 'derived from' carries the meaning outlined above.

In some embodiments, the polypeptide of the invention is immunostimulatory. In some embodiments, one or more of the peptide fragments of the polypeptide of the invention are immunostimulatory. In some embodiments, one or more of the sequences comprised within the peptide fragments of the polypeptide of the invention are immunostimulatory. 'Immunostimulatory' as referred to herein means stimulates, motivates, causes, and/or produces an immune response when administered to a subject. In preferred embodiments, said immune response comprises an adaptive immune response against the sequences comprised within the polypeptide and derived from coronavirus. In some embodiments, said adaptive immune response comprises the generation of antibodies against the polypeptide and/or against one or more peptide fragments and/or sequences comprised therein. In other embodiments, said adaptive immune response comprises the activation and/or proliferation of CD8+ and/or CD4+ T cells. In some embodiments, said adaptive immune response comprises the generation of antibodies against sequences of the peptide fragments of the polypeptide of the invention and, further, the activation and/or proliferation of CD8+ and/or CD4+ T cells specific to sequences comprised within the peptide fragments of the polypeptide of the invention. The skilled person would appreciate that immunostimulation can be measured by the detection of increased antibody production, increased T cell production, increased cytokine production, increased chemokine production. The skilled person will also appreciate that such detection may be through biochemical analysis, molecular analysis and/or medical examination as appropriate. For example, epitope-specific T cells may be quantified by interferon-γ enzyme-linked immunospot ('ELISpot') assay.

In some embodiments, each peptide fragment is represented only once within the polypeptide of the present invention. In other embodiments, the polypeptide of the present invention comprises multiple versions of a given peptide fragment. These versions may be identical in

sequence or may have some degree of sequence variation, for example in order to represent various SARS-CoV-2 variant strains or various coronaviruses.

It is commonly understood that antibodies are generated against discontinuous, conformational epitopes and that, for this reason, the native 3D structure of the protein should be retained – at least in the local epitope area – for the effective generation of neutralising antibodies. Linear epitopes, representing two-dimensional amino acid sequences in the absence of conformational context, are not expected to generate neutralising antibodies. However, surprisingly, the polypeptide of the present invention generates neutralising antibodies against coronavirus S protein (see Example 3.4 and Figures 6 and 7). The mechanism by which this is achieved is the subject of ongoing investigation.

The polypeptide of the present invention is able to selectively present key functional regions (e.g. the RBD, further optionally the RBM, and/or HR1 and/or HR2 of the S protein) to the immune system. Thus the polypeptide of the present invention comprises a higher proportion of key functional regions than the native or wildtype protein from which its fragments derive. Furthermore, in some embodiments these key functional regions are represented multiple times and/or in multiple (e.g. overlapping) forms, further increasing the proportion of key functional regions contained within the polypeptide. This increases the likelihood that neutralising antibodies will be raised i.e. against these functional regions. The polypeptide of the present invention preferentially presents key functional regions to the immune system for B cell recognition and antibody generation. Neutralising antibody titres can be further increased by co-administration with a native amino acid sequence from the whole or a part of a coronavirus protein, e.g. the S protein or a part thereof (e.g. the RBD, further optionally the RBM, and/or HR1 and/or HR2 of the S protein), optionally of SARS-CoV-2.

There is a risk that antibodies generated from a prior delivery of coronavirus proteins into a host (from prior infection, vaccination, or otherwise) motivates, upon subsequent exposure of the host to infectious coronavirus particles, antibody dependent enhancement ('ADE') of disease. It is suspected that the generation of high titres of neutralising antibodies and/or reduced presentation of non-neutralising epitopes will lower the risk of ADE (Lee *et al.,* 2020). In some embodiments, the polypeptide of the invention preferentially provides sequences derived from functionally key regions, aka neutralising epitopes, of a coronavirus protein, preferably multiple sequences which are derived from one or more functionally key regions, further preferably multiple sequences which overlap and which are derived from one or more functionally key regions. Thus, in some embodiments the polypeptide increases the probability

that high-affinity neutralising antibodies are produced in high titres and reduces the risk of ADE.

On a population-wide level, antibodies produced in response to viral challenge exhibit variation in epitope specificity and/or immunodominant epitopes. As described above, in some embodiments the polypeptide of the invention provides multiple peptide fragments each representing different portions of key functional regions of said viral protein (e.g. the RBD, optionally further the RBM, of the S1 subunit, and/or the HR1 and/or the HR2 regions of the S2 subunit of the S protein), optionally with sequence overlap. This means that a polypeptide of the present invention can compensate population-wide epitope variation via a tiling, or greater coverage of, key functional regions such as the RBD, optionally the RBM, and/or the HR1 and/or HR2 regions. By offering alternative epitopes for immune recognition, polypeptides of the invention are effective at generating neutralising antibodies in a larger proportion of the vaccinated population than proteins containing only single representations of epitope sequences.

In some embodiments, the polypeptide of the invention and/or its component peptide fragments has an immunogenicity equivalent to the ChAdOx1 nCoV-19 vaccine candidate (van Doremalen *et al* (2020)). In some embodiments, the polypeptide of the invention and/or its component peptide fragments stimulates neutralising antibodies in 100% of subjects, in 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, at least 85%, at least 80%, at least 75%, or at least 70% of subjects after a single dose. In some embodiments, the polypeptide of the invention and/or its component peptide fragments stimulates neutralising antibodies in 100% of subjects, in 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, at least 85%, at least 80% at least 75%, or at least 70% of subjects after two doses. In some embodiments, the polypeptide of the invention and/or its component peptide fragments stimulates neutralising antibodies in 100% of subjects, in 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, at least 85%, at least 80%, at least 75%, or at least 70% of subjects after three doses.

In some embodiments, the polypeptide of the invention and/or its component peptide fragments stimulates coronavirus protein-specific T cell responses in 100% of subjects, in 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, at least 85%, at least 80%, at least 75%, or at least 70% of subjects after a single dose. In some embodiments, the polypeptide of the invention and/or its component peptide fragments stimulates coronavirus protein-specific T cell responses in 100% of subjects, in 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, at least 85%, at least 80%, at least 75%, or at least 70% of subjects after two

doses. In some embodiments, the polypeptide of the invention and/or its component peptide fragments stimulates coronavirus protein-specific T cell responses in 100% of subjects, in 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, at least 85%, at least 80%, at least 75%, or at least 70% of subjects after three doses.

In one embodiment, any one of the two or more peptide fragments of the present invention may comprise any one of the sequences SEQ ID NOs 1 to 12, as detailed below, or a variant thereof. In another embodiment, any one of the three or more peptide fragments of the present invention may comprise any one of the sequences SEQ ID NOs 1 to 12, as detailed below, or a variant thereof. In another embodiment, the polypeptide comprises one or more, two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, nine or more, ten or more, eleven or more, or twelve of the sequences SEQ ID NOs 1 to 12, as detailed below, or a variant thereof:

PF1 (30 aa):	SVLYNSASFSTFKCYGVSPTKLNDLCFTNV	**(SEQ ID NO: 1)**
PF2 (30 aa):	GVSPTKLNDLCFTNVYADSFVIRGDEVRQI	**(SEQ ID NO: 2)**
PF3 (30 aa):	YADSFVIRGDEVRQIAPGQTGKIADYNYKL	**(SEQ ID NO: 3)**
PF4 (30 aa):	APGQTGKIADYNYKLPDDFTGCVIAWNSNN	**(SEQ ID NO: 4)**
PF5 (30 aa):	PDDFTGCVIAWNSNNLDSKVGGNYNYLYRL	**(SEQ ID NO: 5)**
PF6 (30 aa):	LDSKVGGNYNYLYRLFRKSNLKPFERDIST	**(SEQ ID NO: 6)**
PF7 (30 aa):	FRKSNLKPFERDISTEIYQAGSTPCNGVEG	**(SEQ ID NO: 7)**
PF8 (30 aa):	EIYQAGSTPCNGVEGFNCYFPLQSYGFQPT	**(SEQ ID NO: 8)**
PF9 (31 aa):	FNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE	**(SEQ ID NO: 9)**
PF10 (36 aa):	DISGINASVVNIQKEIDRLNVAKNLNESLIDLQELG	**(SEQ ID NO: 10)**
PF11 (30 aa):	EIDRLNEVAKNLNESLIDLQELGKYEQYIKW	**(SEQ ID NO: 11)**
PF12 (30 aa):	IDLQELGKYEQYIKWPWYIWLGFIAGLIAIV	**(SEQ ID NO: 12)**

Any one of SEQ ID NOs 1 to 12 may be present within the polypeptide in any order. In some embodiments, any one of SEQ ID NOs 1 to 12 be present more than once.

As discussed previously, the polypeptide of the invention comprises one or more overlapping sequence(s). As a further illustrative example, if the polypeptide comprised a first peptide fragment having a first sequence, the first sequence being SEQ ID NO: 1, and the polypeptide also comprised a second peptide fragment having a second sequence, the second sequence being SEQ ID NO: 2, then the polypeptide would comprise one overlapping sequence (which is the same in each fragment). In this instance, the overlapping sequence would have the amino acid sequence: GVSPTKLNDLCFTNV

SEQ ID NOs 1 to 12 derive from the S protein of SARS-CoV-2 and correspond to the original Wuhan L strain, which has S protein sequence SEQ ID NO: 13 (Uniprot accession number P0DTC2). Numerous mutations have arisen in the S protein during the course of the SARS-CoV-2 pandemic, giving rise to subsequent viral variants (e.g. the D614G mutation has become prevalent in circulation; emerging variants continue to arise, e.g. B.1.1.7, B 1.351, P.1, B.1.427, B.1.429, B.1.617.2). Further mutations will continue to emerge. Variants of SEQ ID NOs 1 to 12, and 13 to 33 as follows, comprising mutated sequences corresponding to these subsequent strains can be readily envisaged. In some embodiments the polypeptide of the invention comprises peptide fragments comprising one or more sequences SEQ ID NOs 1 to 12 wherein one or more amino acids within said SEQ ID NOs 1 to 12 is altered to correspond to mutations within one or more Variants of Concern. Such mutations characteristic of current Variants of Concern may be found in Table A. In some embodiments, the polypeptide of the invention comprises peptide fragments comprising one or more sequences SEQ ID NOs 1 to 12 comprising the N501Y mutation. In some embodiments, the polypeptide of the invention comprises peptide fragments comprising one or more sequences SEQ ID NOs 1 to 12 comprising the E484K mutation. In some embodiments, the polypeptide of the invention comprises peptide fragments comprising one or more sequences SEQ ID NOs 1 to 12 comprising the L452R mutation.

In one embodiment, the S protein has the amino acid sequence SEQ ID NO: 13 (Uniprot accession number P0DTC2) or a variant thereof. SEQ ID NO: 13 is the sequence of the original L strain from Wuhan, which emerged in 2019. In some embodiments, the S protein sequence comprises one or more mutations characteristic of circulating strains, Variants of Interest, Variants Under Investigation, and/or Variants of Concern. Variant-defining mutations of the S protein of Variants of Concern (as defined by Public Health England at time of filing) are set out in Table A. In some embodiments, the S protein sequence comprises the D614G mutation. In some embodiments, the S protein sequence comprises the N501Y mutation. In some embodiments, the S protein sequence comprises mutation L452R. In some embodiments, the S protein sequence comprises mutation E484K. In some embodiments, the S protein sequence is the sequence of a Variant of Concern. In some embodiments, the S protein sequence is the B.1.1.7 S protein sequence. In some embodiments, the S protein sequence is the B.1.351 S protein sequence. In some embodiments, the S protein sequence is the P.1 S protein sequence. In some embodiments, the S protein sequence is the B.1.1.7 with E484K S protein sequence. In some embodiments, the S protein sequence is the B.1.617.2 S protein sequence. In some embodiments, the S protein sequence comprises one or more of the mutations provided in Table A.

Variant of Concern (according to UK PHE Technical briefing 10, 7 May 2021)	**Variant-defining spike protein mutations**
B.1.1.7	Δ69/70, Δ144, (E484K*), (S494P*), N501Y, A570D, D614G, P681H, T716I, S982A, D1118H, (K1191N*)
B.1.351	D80A, D215G, (Δ241/242/243*), (R246I*) K417N, E484K, N501Y, (D614G*), A701V
P.1	L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I
B.1.1.7 with E484K	Δ69/70, Δ144, E484K, (S494P*), N501Y, A570D, D614G, P681H, T716I, S982A, D1118H, (K1191N*)
B.1.617.2	T19R, 156-158del, L452R, T478K, D614G, P681R, D950N

Table A: mutations of Variants of Concern (at time of filing). (*) = detected in some sequences but not all sequences.

Further S protein sequences, and variants thereof, that result from emerging viral variants and emerging mutations will be apparent to the skilled person. A repository with up-to-date genomic definitions for all Variants of Concern and Variants Under Investigation, curated by Public Health England, can be accessed on GitHub (phe-genomics/variant_definitions).

SEQ ID NOs 1 to 9 comprise sequences derived from the RBD of the S1 subunit of the S protein of SARS-CoV-2. In some embodiments, the S1 subunit has the amino acid sequence SEQ ID NO: 14 or a variant thereof. SEQ ID NO: 14 consists of residues 13-685 of SEQ ID NO: 13. SEQ ID NO: 14 is the sequence of the original L strain. In some embodiments, the S1 sequence comprises one or more mutations characteristic of circulating strains, Variants of Interest, Variant Under Investigation, and/or Variants of Concern. In some embodiments, the S1 sequence comprises the D614G mutation. In some embodiments, the S1 sequence comprises the N501Y mutation. In some embodiments, the S1 sequence comprises mutation L452R. In some embodiments, the S1 sequence comprises mutation E484K. In some embodiments, the S1 sequence is a S1 sequence of a Variant of Concern. In some embodiments, the S1 sequence is a B.1.1.7 S1 sequence. In some embodiments, the S1 sequence is a B.1.351 S1 sequence. In some embodiments, the S1 sequence is a P.1 S1 sequence. In some embodiments, the S1 sequence is a B.1.1.7 with E484K S1 sequence. In

some embodiments, the S1 sequence is a B.1.617.2 S1 sequence. In some embodiments, the S1 sequence comprises one or more of the mutations provided in Table A.

In some embodiments, the RBD has the amino acid sequence SEQ ID NO: 15, or a variant thereof. SEQ ID NO: 15 is as defined in Fig 4 of Lan, J., *et al.* (2020) and consists of residues 336-516 of SEQ ID NO: 13. In some embodiments, the RBD has the amino acid sequence SEQ ID NO: 16, or a variant thereof. SEQ ID NO: 16 is as defined in Extended Data Fig.1 of Shang, J., *et al.* (2020) and consists of residues 319-529 of SEQ ID NO: 13.

SEQ ID NO: 15 and 16 are sequences of the original L strain. In some embodiments, the RBD sequence comprises one or more mutations characteristic of circulating strains, Variants of Interest, Variant Under Investigation, and/or Variants of Concern. In some embodiments, the RBD sequence comprises the N501Y mutation. In some embodiments, the RBD sequence comprises mutation L452R. In some embodiments, the RBD sequence comprises mutation E484K. In some embodiments, the RBD sequence is the RBD sequence of a Variant of Concern. In some embodiments, the RBD sequence is the B.1.1.7 RBD sequence. In some embodiments, the RBD sequence is the B.1.351 RBD sequence. In some embodiments, the RBD sequence is the P.1 RBD sequence. In some embodiments, the RBD sequence is the B.1.1.7 with E484K RBD sequence. In some embodiments, the RBD sequence is the B.1.617.2 RBD sequence. In some embodiments, the RBD sequence comprises one or more of the mutations provided in Table A as appropriate.

Sequences contained within SEQ ID NOs 5 to 9 derive, at least in part, from the RBM of the S1 subunit of the S protein of SARS-CoV-2. In some embodiments, the RBM has the amino acid sequence SEQ ID NO: 17, or a variant thereof. SEQ ID NO: 17 is as defined in Lan, J., *et al.* (2020) and consists of residues 438-506 of SEQ ID NO: 13.

SEQ ID NO: 17 is a sequence of the original L strain. In some embodiments, the RBM sequence comprises one or more mutations characteristic of circulating strains, Variants of Interest, Variant Under Investigation, and/or Variants of Concern. In some embodiments, the RBM sequence comprises the N501Y mutation. In some embodiments, the RBM sequence comprises mutation L452R. In some embodiments, the RBM sequence comprises mutation E484K. In some embodiments, the RBM sequence is a RBD sequence of a Variant of Concern. In some embodiments, the RBM sequence is a B.1.1.7 RBM sequence. In some embodiments, the RBM sequence is a B.1.351 RBM sequence. In some embodiments, the RBM sequence is a P.1 RBM sequence. In some embodiments, the RBM sequence is a B.1.1.7 with E484K RBM sequence. In some embodiments, the RBM sequence is a B.1.617.2

RBM sequence. In some embodiments, the RBM sequence comprises one or more of the mutations provided in Table A as appropriate. In some embodiments the RBM sequence comprises one or more of mutations E484K, S494P, N501Y, K417T, T478K, L452R.

In some embodiments, the RBM has amino acid sequence SEQ ID NO: 36 (comprising E484K, N501Y, and L452R) or a variant thereof.

SEQ ID NOs: 10 to 12 comprise sequences derived from the S2 subunit of the S protein of SARS-CoV-2. In some embodiments, the S2 subunit has the amino acid sequence SEQ ID NO: 18, or a variant thereof. SEQ ID NO: 18 consists of residues 686-1273 of SEQ ID NO: 13. SEQ ID NO: 18 is a sequence of the original L strain. In some embodiments, the S2 sequence comprises one or more mutations characteristic of circulating strains, Variants of Interest, Variant Under Investigation, and/or Variants of Concern. In some embodiments, the S2 sequence is a S2 sequence of a Variant of Concern. In some embodiments, the S2 sequence is a B.1.1.7 S2 sequence. In some embodiments, the S2 sequence is a B.1.351 S2 sequence. In some embodiments, the S2 sequence is a P.1 S2 sequence. In some embodiments, the S2 sequence is a B.1.1.7 with E484K S2 sequence. In some embodiments, the S2 sequence is a B.1.617.2 S2 sequence. In some embodiments, the S2 sequence comprises one or more of the mutations provided in Table A as appropriate.

SEQ ID NOs: 10 to 12 comprise sequences derived, at least in part, from the HR2 region of the S2 subunit of the S protein of SARS-CoV-2. In some embodiments, the HR2 region has the amino acid sequence SEQ ID NO: 19, or a variant thereof. SEQ ID NO: 19 consists of residues 1159-1211 of SEQ ID NO: 13. SEQ ID NO: 19 is a sequence of the original L strain. In some embodiments, the HR2 sequence comprises one or more mutations characteristic of circulating strains, Variants of Interest, Variant Under Investigation, and/or Variants of Concern. In some embodiments, the HR2 sequence is a HR2 sequence of a Variant of Concern. In some embodiments, the HR2 sequence is a B.1.1.7 HR2 sequence. In some embodiments, the HR2 sequence is a B.1.351 HR2 sequence. In some embodiments, the HR2 sequence is a P.1 HR2 sequence. In some embodiments, the HR2 sequence is a B.1.1.7 with E484K HR2 sequence. In some embodiments, the HR2 sequence is a B.1.617.2 HR2 sequence. In some embodiments, the HR2 sequence comprises one or more of the mutations provided in Table A as appropriate.

In some embodiments, the polypeptide comprises peptide fragments comprising sequences derived, at least in part, from the HR1 region of the S2 subunit of the S protein of SARS-CoV-2. In some embodiments, the HR1 region has amino acid sequence SEQ ID NO: 35, or a

variant thereof. SEQ ID NO: 35 consists of residues 910–988 of SEQ ID NO: 13. SEQ ID NO: 35 is a sequence of the original L strain. In some embodiments, the HR1 sequence comprises one or more mutations characteristic of circulating strains, Variants of Interest, Variant Under Investigation, and/or Variants of Concern. In some embodiments, the HR1 sequence is a HR1 sequence of a Variant of Concern. In some embodiments, the HR1 sequence is a B.1.1.7 HR1 sequence. In some embodiments, the HR1 sequence is a B.1.351 HR1 sequence. In some embodiments, the HR1 sequence is a P.1 HR1 sequence. In some embodiments, the HR1 sequence is a B.1.1.7 with E484K HR1 sequence. In some embodiments, the HR1 sequence is a B.1.617.2 HR1 sequence. In some embodiments, the HR1 sequence comprises one or more of the mutations provided in Table A as appropriate.

In another embodiment, any one of the two or more peptide fragments of the present invention may comprise a sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or at least 99% identity, optionally 75-100%, 80-100%, 85-100%, 90-100%, 91-100%, 92-100%, 93-100%, 94-100%, 95-100%, 96-100%, 97-100%, 98-100% identity to one or more of SEQ ID NOs 1 to 12.

In one embodiment, the polypeptide comprises any one or more of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, and/or SEQ ID NO:12, wherein a protease cleavage site sequence is located between each sequence, optionally wherein the protease cleavage site sequence is LRMK SEQ ID NO: 34.

In one embodiment, the polypeptide comprises SEQ ID NO: 1, and SEQ ID NO: 2, and SEQ ID NO: 3, and SEQ ID NO:4, and SEQ ID NO:5, and SEQ ID NO:6, and SEQ ID NO:7, and SEQ ID NO:8, and SEQ ID NO:9, and SEQ ID NO:10, and SEQ ID NO:11, and SEQ ID NO:12, wherein the LRMK SEQ ID NO: 34 is located between each of SEQ ID NO: 1, and SEQ ID NO: 2, and SEQ ID NO: 3, and SEQ ID NO:4, and SEQ ID NO:5, and SEQ ID NO:6, and SEQ ID NO:7, and SEQ ID NO:8, and SEQ ID NO:9, and SEQ ID NO:10, and SEQ ID NO:11, and SEQ ID NO:12.

Accordingly, in one exemplary embodiment of the invention, a polypeptide is provided having the amino acid sequence SEQ ID NO: 21, or a variant thereof. In another exemplary embodiments, a polypeptide is provided having amino acid sequence SEQ ID NO: 20 (which comprises SEQ ID NO: 21 with a N-terminal His6 tag), or a variant thereof.

In one exemplary embodiment, a polypeptide is provided having the nucleic acid sequence SEQ ID NO: 22 or a variant thereof.

The skilled person will understand that sequences comprised within peptide fragments of the present invention may be optimised, optionally for MHC presentation for example by (but not limited to) introduction of one or more hydrophobic, hydrophilic, acidic, and/or basic amino acids at any location in the peptide fragment(s) such that binding to the MHC class I groove, the MHC class II grove, TAP, or a chaperone molecule is increased. The skilled person will understand that additional amino acid sequences may be appended to the polypeptide of the present invention for purposes of detection, isolation, secretion, purification, or similar. Accordingly, in some embodiments the polypeptide may comprise a tag, for example (but not limited to) a His tag, a fluorescent protein, biotin, OmpA, GST, MBP, CBP, Myc, HA, FLAG. The skilled person will understand that additional amino acid sequences may be appended to the polypeptide of the present invention for purposes of function, efficacy or similar. Accordingly, in some embodiments the polypeptide may comprise a signal sequence, a membrane-penetrating sequence (for example but not limited to CPP), and/or an immunostimulatory and/or adjuvant sequence.

In one embodiment, any one of the two or more peptide fragments of the present invention may comprise a sequence comprising one or more of the epitopes defined by SEQ ID NOs 23 to 33, or any variant thereof:

SASFSTFKCYGVSPTKL	**(SEQ ID NO: 23)**
GVSPTKL	**(SEQ ID NO: 24)**
KLPDDFTGCV	**(SEQ ID NO: 25)**
PDDFTGCV	**(SEQ ID NO: 26)**
NLDSKVGGNYNYLYRL	**(SEQ ID NO: 27)**
LDSKVGGNYNYLYRLFRKSNLKPFERDIST	**(SEQ ID NO: 28)**
FRKSNLKPFERDISTEIYQ	**(SEQ ID NO: 29)**
SYGFQPT	**(SEQ ID NO: 30)**
SYGFQPTNGVGYQPY	**(SEQ ID NO: 31)**
DISGINASVVNIQKEIDRLN	**(SEQ ID NO: 32)**
IDLQEL	**(SEQ ID NO: 33)**

SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 28, and SEQ ID NO: 30 are linear antibody epitopes. In some embodiments, the polypeptide of the present invention may comprise peptide fragments comprising one or more of antibody epitopes selected from SEQ ID NOs:

23, 25, 28, and 30. In some embodiments, the polypeptide may comprise one or more peptide fragments comprising the full sequence of one or more of SEQ ID NOs: 23, 25, 28 or 30, and/or may comprise one or more peptide fragments comprising partial sequences of one or more of SEQ ID NOs: 23, 25, 28 or 30. For example, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 27, and SEQ ID NO: 29 are partial sequences of antibody SEQ ID NOs: 23, 25, 28, and 30. In some embodiments, the polypeptide may comprise one or more peptide fragments comprising one or more of SEQ ID NOs: 23, 25, 28 or 30, and/or may comprise one or more peptide fragments comprising one or more of SEQ ID NOs: 24, 26, 27, or 29.

In another embodiment, any one of the two or more peptide fragments of the present invention may comprise one or more sequence(s) derived from one or more of SEQ ID NOs 23 to 33. In further embodiments, any one of the two or more peptide fragments of the present invention may comprise one or more sequence(s) having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or at least 99% identity, optionally 75-100%, 80-100%, 85-100%, 90-100%, 91-100%, 92-100%, 93-100%, 94-100%, 95-100%, 96-100%, 97-100%, 98-100% identity to one or more of SEQ ID NOs 23 to 33.

In an aspect, a polynucleotide encoding any of the polypeptides described herein is provided. In one embodiment, the polynucleotide encodes a polypeptide comprising between two and 30 peptide fragments derived from one or more structural protein(s) encoded by coronavirus sub-genomic RNA, wherein at least two of the two or more peptide fragments comprise sequences which derive from the same protein and which overlap, and wherein the polypeptide further one or more protease cleavage site sequences located between each of the two or more peptide fragments.

In another embodiment, the polynucleotide encodes a polypeptide which comprises three or more peptide fragments, wherein at least two of said three or more peptide fragments comprise sequences which overlap, and further wherein at least one of said three or more peptide fragments comprises a sequence derived from the S1 subunit of the S protein of a severe acute respiratory syndrome-related coronavirus, optionally SARS-CoV-2, and at least one of said three or more peptide fragments comprises a sequence derived from the S2 subunit of the S protein of a severe acute respiratory syndrome-related coronavirus, optionally SARS-CoV-2.

In a preferred embodiment, the polynucleotide encodes a polypeptide which comprises 8 to 18 peptide fragments, 9 to 17 peptide fragments, 10 peptide fragments, 11 peptide fragments, 12 peptide fragments, 13 peptide fragments, 14, 15, or 16 peptide fragments, some of which

comprise overlapping sequences derived from the S1 subunit, preferably the RBD and/or RBM, of the S protein of a severe acute respiratory syndrome-related coronavirus, preferably SARS-CoV-2, the rest of which comprise overlapping sequences derived from the S2 subunit, preferably the HR1 and/or HR2 region, of the S protein of a severe acute respiratory syndrome-related coronavirus, preferably SARS-CoV-2, and wherein each peptide fragment is separated in amino acid sequence from its adjacent peptide fragment(s) by means of a LRMK cleavage site sequence of cathepsin S, such that, once inside a cell, each peptide fragment can be liberated from the polypeptide.

In some embodiments, the polynucleotide of the invention comprises DNA and/or RNA nucleotide sequences derived from the RNA sequences which encode for structural proteins of a coronavirus, optionally the S protein of a severe acute respiratory syndrome-related virus, optionally SARS-CoV-2. In some embodiments, said nucleic acid sequences will have 60 % sequence identity, 65 % sequence identity, 70 %, 75 %, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or at least 99% identity, optionally 75-100%, 80-100%, 85-100%, 90-100%, 91-100%, 92-100%, 93-100%, 94-100%, 95-100%, 96-100%, 97-100%, 98-100% identity to the sub-genomic RNA sequences from which they derive.

The skilled person will understand that the polynucleotide of the invention may comprise various enhancer elements and/or promoter sequences. Suitable promoters, whether constitutive or inducible, will be immediately apparent to the skilled person, and include but are not limited to T7, lac, Sp6, araBad, trp, Ptac, CMV, EF1a, CAG, PGK, TRE, U6, UAS.

One aspect of the invention provides polypeptides as defined above in a composition or pharmaceutical composition. In one embodiment, a composition or pharmaceutical composition comprises a polynucleotide encoding a polypeptide of the invention. The compositions may comprise, consist essentially of, or consist of the polypeptide and/or polynucleotide of the invention.

In one embodiment, a pharmaceutical composition comprising the polypeptide and/or the polynucleotide of the invention is provided. The pharmaceutical compositions may be for human or animal usage in human and veterinary medicine. The pharmaceutical composition may be formulated according to route of administration. In one embodiment, the pharmaceutical composition is formulated for oral, nasal, ocular, buccal, vaginal, rectal, transdermal, intravenous, intramuscular or subcutaneous administration. In a preferred route of administration, the pharmaceutical composition is formulated for administration by injection. In some embodiments, delivery by injection can be by subcutaneous, intravenous,

intramuscular, intraperitoneal, or intradermal injection. In another preferred route of administration, the pharmaceutical composition is formulated for administration by oral ingestion and/or inhalation, for example.

In one embodiment, the composition or pharmaceutical composition additionally comprises a pharmaceutically acceptable delivery vehicle. The polypeptide of the invention and/or the polynucleotide of the invention may be administered to a subject by means of said delivery vehicle. In one embodiment, the pharmaceutically acceptable delivery vehicle is a viral vector, optionally an adenovirus, an adeno-associated virus, MVA, HSV. In another embodiment, the pharmaceutically acceptable delivery vehicle is a bacterial vector. In another embodiment, the pharmaceutically acceptable delivery vehicle is a plasmid, a nanoparticle, a lipoparticle, a polymeric particle, or a virus-like particle.

In one embodiment, the polypeptide does not require vector delivery or delivery by means of a delivery vehicle. In one aspect of the invention is provided a formulation comprising one or more polypeptide(s) as defined above. In some embodiments, the formulation may be for administration to a subject as a vaccine or a therapy. In some embodiments the formulation is a vaccine formulation. In some embodiments, a first dose formulation is the same or different from a booster formulation. In some embodiments the formulation additionally comprises a native amino acid sequence derived from a coronavirus S protein against which the polypeptide of the present invention is targeted (by means of the sequences comprised within the polypeptide's peptide fragments). The native amino acid sequence may comprise the wildtype sequence of the whole or a portion of a coronavirus protein (optionally a SARS-CoV-2 protein), or a variant thereof. For example, in some embodiments the formulation comprises a polypeptide of the present invention comprising peptide fragments comprising sequences derived from a coronavirus RBD (and therefore can be said to be targeting the RBD, by stimulating anti-RBD antibody production) and, additionally, the RBD of said coronavirus. In some embodiments the formulation comprises a polypeptide of the present invention comprising one or more peptide fragments comprising sequences derived from a coronavirus RBM and, additionally, the RBM of said coronavirus. In some embodiments, the formulation comprises polypeptide of the present invention comprising one or more peptide fragments comprising sequences derived from a coronavirus HR1 and/or HR2 and, additionally, the S2 protein of said coronavirus. In some embodiments, the formulation comprises polypeptide of the present invention comprising one or more peptide fragments comprising sequences derived from a coronavirus HR1 and, additionally, the HR1 of said coronavirus. In some embodiments, the formulation comprises polypeptide of the present invention comprising one

or more peptide fragments comprising sequences derived from a coronavirus HR2 and, additionally, the HR2 of said coronavirus.

In some embodiments, the formulation comprises a polypeptide of the present invention and the whole or part of a coronavirus S protein. In some embodiments, the formulation comprises a polypeptide of the present invention and the whole or part of a coronavirus S1 subunit. In some embodiments, the formulation comprises a polypeptide of the present invention and the whole or part of a coronavirus S2 subunit. In some embodiments, the formulation comprises a polypeptide of the present invention and a coronavirus RBD. In some embodiments, the formulation comprises a polypeptide of the present invention and a coronavirus RBM. In some embodiments, the formulation comprises a polypeptide of the present invention and a coronavirus HR1. In some embodiments, the formulation comprises a polypeptide of the present invention and a coronavirus HR2.

In some embodiments, the formulation additionally comprises one or more pharmaceutically acceptable adjuvants, for example but not limited to MPL, Alum, AS501, montanide.

In one embodiment, the composition or pharmaceutical composition or formulation optionally comprises one or more pharmaceutically acceptable carriers or one or more pharmaceutically acceptable excipients. In one embodiment, the composition or pharmaceutical composition or formulation optionally comprises one or more pharmaceutically acceptable adjuvants, for example but not limited to MPL, Alum, AS501, montanide. In one embodiment, the composition or pharmaceutical composition or formulation is optionally admixed with one or more pharmaceutically acceptable diluents, excipients or carriers. Examples of such suitable excipients for the different forms of pharmaceutical compositions described herein may be found in the "Handbook of Pharmaceutical Excipients", 2nd Edition, (1994), Edited by A Wade and PJ Weller.

The composition or pharmaceutical composition or formulation may comprise one or more additional components. In one embodiment, the composition or pharmaceutical composition additionally comprises a pharmaceutically acceptable carrier. In one embodiment, the carrier is suitable for injectable delivery. In another embodiment, the carrier is suitable for pulmonary delivery. In another embodiment, the carrier is suitable for oral delivery. In one embodiment, the composition or pharmaceutical composition additionally comprises a therapeutically active agent.

In one embodiment, a vaccine comprising a polypeptide of the invention is provided. In one embodiment, a vaccine comprising a polynucleotide encoding a polypeptide of the invention is provided.

In one embodiment, a vaccine comprising a polypeptide of the invention and a native amino acid sequence of the whole or a part of a coronavirus protein, optionally the S protein, S1 protein, S2 protein, RBD, RBM, HR1 and/or HR2, is provided. In some embodiments, the native amino acid sequence is the RBD or the RBM of a coronavirus, optionally SARS-CoV-2. In one embodiment, a vaccine comprising a polynucleotide encoding a polypeptide of the invention and additionally comprising a second polypeptide encoding a native amino acid sequence of the whole or a part of a coronavirus protein, optionally the S protein, S1 protein, S2 protein, RBD, RBM, HR1 and/or HR2, is provided. In some embodiments, the native amino acid sequence is the RBD or the RBM of a coronavirus, optionally SARS-CoV-2.

In one embodiment, in human usage, the composition or pharmaceutical composition or vaccine is administered to a human subject at a dose which lies between 5 $\mu g.kg^{-1}$ and 5 $mg.kg^{-1}$. In some embodiments, this may be administered to a human at approximately 50 $\mu g.kg^{-1}$. In some embodiments, this may be administered to a human at approximately 5 $\mu g.kg^{-1}$ to 100 $\mu g.kg^{-1}$. In some embodiments, this may be administered to a human at approximately 5 $\mu g.kg^{-1}$ to 30 $\mu g.kg^{-1}$. In some embodiments, this may be administered to a human at approximately 25 $\mu g.kg^{-1}$ to 75 $\mu g.kg^{-1}$. In some embodiments, this may be administered to a human at approximately 50 $\mu g.kg^{-1}$ to 100 $\mu g.kg^{-1}$. In some embodiments, this may be administered to a human at approximately 75 $\mu g.kg^{-1}$ to 100 $\mu g.kg^{-1}$.

The compositions, pharmaceutical compositions and vaccines of the invention can elicit an immune response in a subject, preferably an immune response to a coronavirus, more preferably a betacoronavirus, even more preferably a severe acute respiratory syndrome-related coronavirus, most preferably SARS-CoV-2. In some embodiments, the immune response is a protective immune response. In some embodiments, the immune response reduces the symptoms or severity of associated diseases, optionally COVID-19, in a subject.

In one embodiment, a method for the manufacture of a coronavirus vaccine is provided, comprising expressing a polynucleotide of the invention *in vitro*, to produce a polypeptide of the invention, and isolating and/or purifying said polypeptide. Examples of cell lines, expression systems, and purification techniques suitable for the *in vitro* expression of said polynucleotide described herein may be found in the “Protein expression handbook;

Recombinant protein expression and purification technologies", (2015) ThemoFisher Scientific gibco education series.

In one embodiment, a method for the manufacture of a coronavirus vaccine is provided, comprising constructing, isolating, amplifying and/or purifying a polynucleotide of the invention. Suitable techniques for the construction, isolation, amplification and/or purification of a polynucleotide will be apparent to the skilled person, and may also be found in, for example, "DNA Purification and Protocols Applications Guide", (2012) Promega.

In one aspect, a method for the prophylactic and/or post-exposure treatment of a coronavirus in a subject is provided, comprising administering to a subject a therapeutically effective amount of a polypeptide of the invention. In one embodiment, the coronavirus is a betacoronavirus, preferably a severe acute respiratory syndrome-related virus. In some embodiments, the coronavirus is SARS-CoV or SARS-CoV-2. The method may comprise administering one dose of therapeutically effective amount. The method may comprise administering a first dose and, subsequently, a booster dose.

In one embodiment, the method for the prophylactic and/or post-exposure treatment of a coronavirus in a subject additionally comprises administering to a therapeutically effective amount of a native amino acid sequence of the whole or a part of a coronavirus protein, optionally the RBD, RBM, HR1 and/or HR2. The native amino acid sequence to be co-administered may comprise the wildtype sequence of the whole or a portion of a coronavirus protein (optionally a SARS-CoV-2 protein), or a variant thereof.

In one aspect, a method for the prophylactic and/or post-exposure treatment of a coronavirus in a subject is provided, comprising co-administering to a subject a therapeutically effective amount of a polypeptide of the invention and a native amino acid sequence of the whole or a part of a coronavirus protein, optionally the RBD, RBM, HR1 and/or HR2. In some embodiments, the polypeptide of the present invention and the native coronavirus amino acid sequence are co-administered simultaneously. In some embodiments, the polypeptide and the native coronavirus amino acid sequence are co-administered simultaneously as a single formulation. In some embodiments, the polypeptide and the native coronavirus amino acids are produced separately and mixed to form a single formulation for administration to a subject. In other embodiments, the polypeptide and the native coronavirus amino acid sequence are produced together e.g. expressed from a single DNA vector or construct, and co-purified. In some embodiments, the formulation is administered to a subject once. In another embodiment, the formulation is administered to a subject in a first administration, and

subsequently the formulation is administered to the subject again in a booster administration. In some embodiments, the polypeptide and native coronavirus amino acid sequence are co-administered sequentially or in a staggered regimen.

In one aspect, a method for the production of anti-coronavirus antibodies in a subject is provided, optionally for the production of neutralising anti-coronavirus antibodies, said method comprising administering to a subject a therapeutically effective amount of a polypeptide of the invention. In one embodiment, the coronavirus is a betacoronavirus, preferably a severe acute respiratory syndrome-related virus. In some embodiments, the coronavirus is SARS-CoV or SARS-CoV-2. In one embodiment, the method for the prophylactic and/or post-exposure treatment of a coronavirus in a subject additionally comprises administering to a therapeutically effective amount of a native amino acid sequence of the whole or a part of a coronavirus protein, optionally the RBD, RBM, HR1 and/or HR2. The native amino acid sequence to be co-administered may comprise the wildtype sequence of the whole or a portion of a coronavirus protein (optionally a SARS-CoV-2 protein), or a variant thereof. Co-administration may occur as above.

In one embodiment, the polypeptide and, optionally, also the native coronavirus amino acid sequence, is administered by means of its encoding polynucleotide(s). In one embodiment, the polypeptide and, optionally, also the native coronavirus amino acid sequence is administered as a composition, pharmaceutical composition or vaccine described above. In one embodiment, the polypeptide and, optionally, also the native coronavirus amino acid sequence is administered in a delivery vehicle as outlined above. In one embodiment, the polypeptide and, optionally, also the native coronavirus amino acid sequence is administered with a pharmaceutically acceptable carrier as outlined above. In one embodiment the subject is a mammal, preferably a human.

In one aspect, there is provided a polypeptide, or polynucleotide encoding said polypeptide, of the invention for use as a medicament.

In one aspect, there is provided a polypeptide (or polynucleotide encoding said polypeptide) for use in the treatment of a coronavirus, preferably a betacoronavirus, preferably a severe acute respiratory syndrome-related coronavirus, optionally SARS-CoV and most preferably SARS-CoV-2.

In one aspect, there is provided a polypeptide (or polynucleotide encoding said polypeptide) for use in the vaccination against a coronavirus, preferably a betacoronavirus, preferably a

severe acute respiratory syndrome-related coronavirus, optionally SARS-CoV and most preferably SARS-CoV-2.

The polypeptide (or polynucleotide) can be used to invoke an immune response, optionally an antibody response and/or a T cell response. The polypeptide (or polynucleotide) can be used to invoke an immune response, optionally an antibody response and/or a T cell response, specific to the coronavirus protein(s) from which the sequences comprised within the polypeptide (or polynucleotide) are derived.

In one aspect, there is provided a use of a polypeptide of the invention, or polynucleotide encoding said polypeptide, in the manufacture of a medicament for treating a coronavirus infection, preferably a betacoronavirus infection, preferably a severe acute respiratory syndrome-related coronavirus infection, optionally SARS-CoV and most preferably SARS-CoV-2.

In one aspect, there is provided a use of a polypeptide of the invention, or polynucleotide encoding said polypeptide, in the manufacture of a vaccine for preventing a coronavirus infection, preferably a betacoronavirus infection, preferably a severe acute respiratory syndrome-related coronavirus infection, optionally SARS-CoV and most preferably SARS-CoV-2.

In one embodiment, diagnostic methods for the detection of a coronavirus infection, optionally a SARS-CoV-2 infection are provided. In one embodiment, the infection is active. In other embodiments, the infection has occurred prior. In another embodiment, diagnostic methods for the detection of cellular immune memory to a previous coronavirus, optionally SARS-CoV-2, infection are provided. In another embodiment, a method of detecting active infection is provided. In one embodiment, a method for the diagnosis of an active or prior coronavirus infection in a subject or patient is provided, wherein the method comprises the steps of (a) delivering a polypeptide of the invention, wherein the polypeptide comprises at least two peptide fragments comprising sequences which overlap, and wherein at least one of said peptide fragments comprises a CD4+ and/or CD8+ T cell epitope, further wherein a protease cleavage site is located between each peptide fragment, to a subject or patient, optionally into the skin of a subject and (b) detecting T cell stimulation. In some embodiments, the coronavirus is a betacoronavirus. In some embodiments, the coronavirus is a severe acute respiratory syndrome-related virus. In some embodiments, the coronavirus is SARS-CoV-2. In some embodiments, detecting T cell stimulation comprises detecting the production of immune cytokines, preferably T cell cytokines, wherein T cell cytokines are cytokines

produced by activated and/or stimulated T cells, such as gamma interferon. In some embodiments, the step of detecting the production of said immune cytokines, preferably T cell cytokines, comprises medical examination, e.g. comprising detecting redness and/or inflammation at the site of administration. In other embodiments, the step of detecting the production of immune cytokines, preferably T cell cytokines, comprises molecular or biochemical analysis, for example (but not limited to) enzyme linked immunospot ('ELISpot') assay. Detecting pronounced T cell stimulation, step (b), in response to administration of the polypeptide, step (a), overnight or within 16 hours of administration provides a positive indication that the subject is, or has previously been, infected with coronavirus. In other words, a rapid T cell response upon administration of the polypeptide of the invention to a subject is indicative of primed T cell responses. In some embodiments, the coronavirus is SARS-CoV-2. In some embodiments, the patient is a mammal, optionally a human.

In another embodiment, a method for the diagnosis of an active or prior coronavirus, optionally SARS-CoV-2, infection in a sample is provided, wherein the method comprises the steps of (a) delivering a polypeptide of the invention to a sample and (b) detecting T cell stimulation. In some embodiments, the sample is a blood sample derived from a patient, optionally a PBMC sample. The step of detecting the production of immune cytokines, preferably T cell cytokines, may comprise molecular or biochemical analysis, for example (but not limited to) enzyme linked immunospot ('ELISpot') assay. Detecting T cell stimulation, step (b), optionally with net spot of at least 10 SFU/2.5x10^5 cells or above, wherein 'net spot' is calculated as the difference between antigen stimulated sample and sample without antigen stimulation (net spot = antigen stimulated - no antigen) in response to administration of the polypeptide, step (a), within 16 hours of administration provides a positive indication that the subject is, or has previously been, infected with coronavirus, optionally SARS-CoV-2. The sample may be an isolated sample (i.e. previously obtained from a subject or patient).

The invention is described by reference to the following non-limiting Examples.

EXAMPLE 1

A polypeptide vaccine ('ROP-COVS', SEQ ID NO: 20) was designed towards the SARS-CoV-2 protein domains which are most actively involved in viral entry into a host cell. This ROP-COVS is a recombinant polypeptide comprising 12 peptide fragments ('PF's), each linked to the next via a LRMK cleavage sequence (SEQ ID NO: 34) of cathepsin S, such that the PFs can be liberated intracellularly upon digestion by cathepsin S. Each PF is numbered 1 to 12

according to sequential amino acid position within the ROP, with PF1 being the OSP most proximate to the N-terminus and PF12 the most proximate to the C-terminus. The sequences of the PFs are as follows:

PF1 (30 aa): SVLYNSASFSTFKCYGVSPTKLNDLCFTNV **(SEQ ID NO: 1)**
PF2 (30 aa): GVSPTKLNDLCFTNVYADSFVIRGDEVRQI **(SEQ ID NO: 2)**
PF3 (30 aa): YADSFVIRGDEVRQIAPGQTGKIADYNYKL **(SEQ ID NO: 3)**
PF4 (30 aa): APGQTGKIADYNYKLPDDFTGCVIAWNSNN **(SEQ ID NO: 4)**
PF5 (30 aa): PDDFTGCVIAWNSNNLDSKVGGNYNYLYRL **(SEQ ID NO: 5)**
PF6 (30 aa): LDSKVGGNYNYLYRLFRKSNLKPFERDIST **(SEQ ID NO: 6)**
PF7 (30 aa): FRKSNLKPFERDISTEIYQAGSTPCNGVEG **(SEQ ID NO: 7)**
PF8 (30 aa): EIYQAGSTPCNGVEGFNCYFPLQSYGFQPT **(SEQ ID NO: 8)**
PF9 (31 aa): FNCYFPLQSYGFQPTNGVGYQPYRVVVLSFE **(SEQ ID NO: 9)**
PF10 (36 aa): DISGINASVVNIQKEIDRLNVAKNLNESLIDLQELG **(SEQ ID NO: 10)**
PF11 (30 aa): EIDRLNEVAKNLNESLIDLQELGKYEQYIKW **(SEQ ID NO: 11)**
PF12 (30 aa): IDLQELGKYEQYIKWPWYIWLGFIAGLIAIV **(SEQ ID NO: 12)**

Each PF shares a portion of its sequence (aka 'overlaps') with at least one other. For example, amino acids 1 to 15 of PF2 comprise amino acids 16 to 30 of PF1 i.e. a so-called overlap; amino acids 1 to 15 of PF3 comprise amino acids 16 to 30 of PF2 i.e. another so-called overlap.

1. Gene sequence design

PFs 1 to 9 were selected to tile the SARS-CoV-2 S1 receptor binding domain ('RBD') (SEQ ID NO: 15 or 16) and to comprise a number of whole or partial antibody and T cell epitopes of the RBD. PFs 10 to 12 were selected to tile the C-terminal end of the SARS-CoV-2 S2 HR2 region (SEQ ID NO: 19) and the proximal region of S2 (amino acids 483 to 543 of SEQ ID NO: 18), and to comprise whole or partial antibody and T cell epitopes thereof. Each PF, or 'peptide fragment', is linked to the next by a LRMK sequence (SEQ ID NO: 34). The resulting designed ROP-COVS had amino acid sequence SEQ ID NO: 20. A N-terminal His_6 tag was added for purification of the ROP-COVS (SEQ ID NO: 21).

The *E. coli* codon-optimized gene sequence encoding the resultant His-tagged ROP-COVS protein is presented as SEQ ID NO: 22. This gene sequence was cloned through clonal amplification in DH5α *E.coli,* identified via colony PCR and inserted into a pET30a vector (forming plasmid Y0028023-1, Figure 1).

2. Protein production

2.1 Expression

Plasmid Y0028023-1 was transformed into BL21 (DE3) *E.coli*. Electrophoretic analysis confirmed that the ROP-COVS gene inserted successfully (Figure 2). Flasks (250 mL) containing 50 mL of LB medium (containing 50 ug/ml kanamycin sulfate) were used for cultivation. The strain was inoculated at a proportion of 1:500. The bacteria were incubated with rotary shaking (150 rpm) at 37 °C overnight. The cell culture was transferred into 1.2 L 2YT medium (containing 50 ug/ml kanamycin sulfate) at a proportion of 1:100. Once the OD600 value reached 0.8, IPTG was added to a final concentration of 0.2 mM to induce the expression of ROP-COVS. The bacteria were incubated with rotary shaking of 200 rpm at 37 °C.

Bacteria were harvested by centrifugation at 4500 rpm for 30 min. SDS-PAGE was used to analyze the induction rate and demonstrated successful induction after a 4-hour incubation (Figure 3A).

2.2 Collection of inclusion bodies

The harvested wet bacteria were resuspended and washed once with 0.9% NaCl, with a washing ratio of 10 ml/g and centrifugal conditions of 4500 rpm, 4 °C, 30 min. After washing, the wet bacteria were dissolved with lysis buffer (20 mM Tris-HCl, 300 mM NaCl, 20 mM Imidazole, 1% Triton X-100, 1 mM DTT, 1 mM PMSF, pH 8.0) in 10 ml/g, and the pellets were lysed by sonication for 60 cycles (3s on and 5s off).

After lysis, the soluble and insoluble fractions were analysed by SDS-PAGE, which indicated that the target protein ROP-COVS was expressed as inclusion bodies in cells. The inclusion bodies were collected by centrifugation at 9500 rpm for 30 min. The supernatant was discarded. The inclusion bodies were then washed twice with washing buffer 1 (20 mM Tris-HCl, 300 mM NaCl, 1% Triton X-100, 2 mM EDTA, 5 mM DTT, pH 8.0) and once with washing buffer 2 (20 mM Tris-HCl, pH8.0).

2.3 Purification

The inclusion bodies were dissolved in Buffer A (20 mM Tris-HCl, 300mM NaCl, 8 M Urea, pH 8.0) and magnetically mixed over night at 4 °C. The suspension was subjected to centrifugation (18000 rpm, 30 min, 4 °C) to remove the undissolved fractions. The supernatant was loaded into the Ni-NTA column (Smart-Lifesciences) pre-equilibrated with Buffer A. Fractions containing target protein were eluted with a 0-300 mM Imidazole gradient in 20 mM

Tris-HCl buffer containing 300 mM NaCl (pH 8.0). SDS-PAGE was used to analyze the result of purification (Figure 3B).

2.4 Refolding

Fractions with over 95% purity (those in Lanes 6-8 of Figure 3B) were collected. Dialysis was used for refolding: the fractions were first dialyzed to refolding buffer 1 (1×PBS, 4 mM GSH, 0.4 mM GSSG, 0.4 M L-Arginine, 1 M Urea, 5% Glycerol, 0.5% Sarkosyl, pH 7,4) and then to dialyzed to refolding buffer 2 (1x PBS, 5% glycerol). SDS-PAGE was used to analyze the refolding result (Figure 3C).

2.5 Production of SARS-CoV-2 RBM

In some embodiments, RBM is co-administered with ROP-COVS. SARS-CoV-2 RBM (SEQ ID NO: 17) was expressed from *E.coli* according to standard procedures and purified by affinity chromatography according to standard procedures.

3. Vaccination and demonstration of immune response

3.1 Preparation of BALB/c mice

40 mice (SPF grade, 5-6 weeks, female) were bought from Changzhou Cavens Co., Ltd. To allow them to adapt to the new environment, the mice were fed for one week before vaccination.

3.2 Vaccination

Regime 1:

The mice were divided into two groups and vaccinated according to Table 1, below. The mice were vaccinated on day 0, day 14 and day 21. Each mouse was injected subcutaneously with 100 ul mixture of Antigens (or PBS as a control) and MPL. The dose of MPL was followed as per the instruction. At day 24, 3 days after the final vaccination, all mice were sacrificed.

	Day	Vaccination (V) Group			Control (C) Group		
		No.	Antigen	Adjuvant	No.	Antigen	Adjuvant
1st Vaccination	0	10	50 ug ROP-COVS	MPL	10	PBS	MPL
2nd Vaccination	14		25 ug ROP-COVS	MPL		PBS	MPL
3rd Vaccination	21		25 ug ROP-COVS	MPL		PBS	MPL

Table 1. ROP-COV vaccination regime 1

Regime 2:

In some embodiments, RBM is co-administered with ROP-COVS. The mice were divided into three groups and vaccinated according to Table 2, below. The mice were vaccinated on day 0, day 14 and day 21. Each mouse was injected subcutaneously with 100 ul total mixture of Antigens (or S protein as a control) and MPL. The dose of MPL was followed as per the instruction. At day 24, 3 days after the final vaccination, all mice were sacrificed.

Dose	Day	Vaccination (V) Group 1			V Group 2			Control (C) Group		
		No.	Antigen	Adjuvant	No.	Antigen	Adjvnt	No.	Antigen	Adjvnt
1st	0	5	50 ug ROP-COVS	MPL	5	50 ug ROP-COVS + 50 ug RBM	MPL	5	50 ug S protein	MPL
2nd	14		25 ug ROP-COVS	MPL		25 ug ROP-COVS + 25 ug RBM	MPL		50 ug S protein	MPL
3rd	21		25 ug ROP-COVS	MPL		25 ug ROP-COVS + 25 ug RBM	MPL		50 ug S protein	MPL

Table 2. ROP-COV vaccination regime 2

3.3 Antibody titration (ELISA)

The whole blood was extracted from the sacrificed mice and the serum separated by centrifugation (12000 rpm, 3 min). ELISA-based antibody titration was carried out on sera derived from mice immunized with Regime 1: ROP-COVS and control (PBS). A 96-well ELISA plate was coated with 2 ug/ml SARS-COV-2 RBD protein (Sino Biological, 40592-V08B) and incubated at 37 °C for 3 hours. The plate was washed by PBST (PBS containing 0.05% Tween-20) for 5 times and then blocked with Pierce Protein-Free Blocking Buffer (Thermo Fisher). After blocking, the plate was washed for 5 times and the serum was added into the wells and incubated for 1 hour at 37 °C. The plate was then washed, and Goat Anti Mouse

(Abcam, ab6789) added. Antibody titres are markedly and significantly higher in vaccinated mice than control mice at all dilutions trialed (Figure 4). Thus, results demonstrate that vaccination with ROP-COVS leads to pronounced anti-RBD antibody production.

3.4 Surrogate neutralization assay (ELISA)

Mouse sera was extracted and separated as per 3.3 from mice vaccinated according to Regime 2. An ELISA-based surrogate neutralization assay based on Ig competition with the hACE2-RBD binding interaction was carried out according to the following protocol:

1. Plate coated with 20 ug/ml RBD, 100 ul/well, incubated at 4 °C overnight.
2. Plate washed with PBST. Plate blocked with 2.5% BSA, 200 ul/well, incubated at 37°C for 1 h.
3. Plate washed with PBST. The mice sera were diluted to 1:100, 1:200, 1:400, 1:800 and 1:1600. The sera with different dilution were added into the plate, 50 ul/well. 20 ug/ml ACE2-hFc were then added into the plate, 50 ul/well. The plate was incubated at 37 °C for 30 min.
4. Plate washed with PBST. 100 ul/well of anti-hFc antibodies (HRP conjugated) were added. The plate was incubated at room temperature for 30 min.
5. Plate washed with PBST. 100 ul/well TMB substrate added, and allowed to stand for 5-10 min.
6. 50 ul/well of 15% H2SO4 was added into the plate. The OD value was read at 450 nm.

Alternative ELISA-based surrogate neutralization assays are also suitable, for example as described in Tan, C.W., *et al.* (2020).

Results are shown in Figure 6. Vaccination with ROP-COVS stimulates higher neutralizing antibody titres than vaccination with S protein (shown by lower absorbance). Vaccination with both ROP-COVS and RBM produces the highest neutralizing antibody titres.

This surrogate neutralization assay was repeated using IgG purified from mice sera (vaccinated according to Regime 2 and obtained as per 3.3) according to standard protocol. Results are shown in Figure 7. At concentrations of 100 ug/l and below, neutralizing titres in response to ROP-COVS are higher than to S protein. Vaccination with a combination of ROP-COVS and RBM again stimulates highest neutralizing antibody titres at all IgG concentrations.

3.5 Neutralization assay

Mice will be vaccinated according to Regime 1 and/or 2 and sera extracted and separated as per 3.3. Neutralization assays will be performed with pseudotyped or chimeric SARS-CoV-2 virus particles according to standard protocols for example as described in Nie, J., *et al.* or

Schmidt F, *et al*. More preferably, neutralization assays will be carried out using replication-competent SARS-CoV-2 at BSL-3 using standard protocol as described in Amanat, F., *et al.* Results will demonstrate that antibodies produced by mice vaccinated with ROP-COVS and with a combination of ROP-COVS + RBM block viral entry and/or replication. Results will demonstrate that the combination of ROP-COVS + RBM is more potent for generation of neutralizing antibodies than ROP-COVS alone.

3.6 ELISPOT analysis of T cell response

Spleens will be extracted from sacrificed mice (having been vaccinated according to Regime 1 and/or 2) and will be strained through a mesh, loaded to murine splenocyte separation medium (Solarbio), and centrifuged at 1000g for 22 minutes before transferring the layered lymphocytes to a new tube with cell culture medium. The cells will be washed twice by RPMI 1640. 2.5 x 105 splenocytes per well will be used for stimulation in ELISPOT assays. CD4+ or CD8+ T cells will be purified by negative or positive selection using microbeads kit (Miltenyi, Germany) as per the manufacturer's instructions. Assays will be performed using ELISPOT kits (Mabtech, Sweden).

Briefly, splenocytes will be restimulated overnight with 5 μg/well SARS-CoV-2 S protein or ROP-COVS in anti-5 IFN-γ-Ab precoated plates (Millipore). Cells will be discarded, and biotinylated anti-IFN-γ antibody will be added for two hours at room temperature, followed by another one hour of incubation at room temperature with alkaline phosphatase (ALP) conjugated streptavidin. After color develops, the reaction will be stopped by washing plates with tap water and plates will be air-dried. Spots will be counted with an ELISPOT reader (CTL). Results will demonstrate that ROP-COVS can stimulate pronounced CD4+ and CD8+ T cell responses.

3.7 Pre-clinical trials

In vivo pre-clinical trials can be conducted following standard protocols (see e.g. Muñoz-Fontela, C., *et al*.). Neutralizing antibody titres and ELISpot assay PMBC T cell responses will be measured. Results will demonstrate that vaccination with ROP-COVS alone or in combination with RBM generates protective anti-SARS-CoV-2 immune responses.

References

Jaimes, J. A., André, N. M., Chappie, J. S., Millet, J. K., & Whittaker, G. R. (2020). Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop. Journal of molecular biology, 432(10), 3309–3325. https://doi.org/10.1016/j.jmb.2020.04.009

de Groot, R. J. et al. in 'Virus Taxonomy, Ninth Report of the International Committee on Taxonomy of Viruses' (eds King, A. M. Q. et al.) 806–828 (Elsevier Academic Press, 2012).

Lee, W.S., Wheatley, A.K., Kent, S.J. *et al.* (2020) Antibody-dependent enhancement and SARS-CoV-2 vaccines and therapies. *Nat Microbiol* 5, 1185–1191. https://doi.org/10.1038/s41564-020-00789-5

Thanh Le, T., Cramer, J.P., Chen, R., Mayhew, S. (2020) Evolution of the COVID-19 vaccine development landscape. *Nat. Rev. Drug Discov.* https://www.nature.com/articles/d41573-020-00151-8

Coronaviridae Study Group of the International Committee on Taxonomy of Viruses (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature microbiology*, *5*(4), 536–544. https://doi.org/10.1038/s41564-020-0695-z

Alba Grifoni, Daniela Weiskopf, Sydney I. Ramirez, Jose Mateus, Jennifer M. Dan, Carolyn Rydyznski Moderbacher, Stephen A. Rawlings, Aaron Sutherland, Lakshmanane Premkumar, Ramesh S. Jadi, Daniel Marrama, Aravinda M. de Silva, April Frazier, Aaron F. Carlin, Jason A. Greenbaum, Bjoern Peters, Florian Krammer, Davey M. Smith, Shane Crotty, Alessandro Sette, (2020) Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals, *Cell*, 181(7), 1489-1501.e15, https://doi.org/10.1016/j.cell.2020.05.015.

Daniel Altmann, Rosemary Boyton (2020) SARS-CoV-2 T cell immunity: Specificity, function, durability, and role in protection, *Science Immunology*, 5(49), eabd6160, https://doi.org/10.1126/sciimmunol.abd6160

Zhang H, Hong H, Li D, Ma S, Di Y, Stoten A, Haig N, Di Gleria K, Yu Z, Xu XN, McMichael A, Jiang S. (2009) Comparing pooled peptides with intact protein for accessing cross-presentation pathways for protective CD8+ and CD4+ T cells. *The Journal of Biological Chemistry.* 284(14) 9184–9191. https://doi.org/10.1074/jbc.m809456200

Watson, C. T., Glanville, J., & Marasco, W. A. (2017). The Individual and Population Genetics of Antibody Immunity. *Trends in immunology*, *38*(7), 459–470. https://doi.org/10.1016/j.it.2017.04.003

van Doremalen, N., Lambe, T., Spencer, A. et al. (2020) ChAdOx1 nCoV-19 vaccine prevents SARS-CoV-2 pneumonia in rhesus macaques. *Nature* https://doi.org/10.1038/s41586-020-2608-y

Tan, C.W., Chia, W.N., Qin, X. *et al*. A SARS-CoV-2 surrogate virus neutralization test based on antibody-mediated blockage of ACE2–spike protein–protein interaction. *Nat Biotechnol* 38, 1073–1078 (2020). https://doi.org/10.1038/s41587-020-0631-z

Nie, J., Li, Q., Wu, J. *et al*. Quantification of SARS-CoV-2 neutralizing antibody by a pseudotyped virus-based assay. *Nat Protoc* 15, 3699–3715 (2020). https://doi.org/10.1038/s41596-020-0394-5

Schmidt F, Weisblum Y, Muecksch F, Hoffmann HH, Michailidis E, Lorenzi JCC, Mendoza P, Rutkowska M, Bednarski E, Gaebler C, Agudelo M, Cho A, Wang Z, Gazumyan A, Cipolla M, Caskey M, Robbiani DF, Nussenzweig MC, Rice CM, Hatziioannou T, Bieniasz PD. Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and chimeric viruses. *J Exp Med*. 2020 Nov 2;217(11):e20201181. doi: 10.1084/jem.20201181

Amanat, F., White, K. M., Miorin, L., Strohmeier, S., McMahon, M., Meade, P., Liu, W.-C., Albrecht, R. A., Simon, V., Martinez-Sobrido, L., Moran, T., García-Sastre, A., & Krammer, F. (2020). An in vitro microneutralization assay for SARS-CoV-2 serology and drug screening. *Current Protocols in Microbiology*, 58, e108. doi: 10.1002/cpmc.108.

Muñoz-Fontela, C., Dowling, W.E., Funnell, S.G.P. et al. Animal models for COVID-19. *Nature* 586, 509–515 (2020). https://doi.org/10.1038/s41586-020-2787-6

Tai, W., He, L., Zhang, X. *et al*. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell Mol Immunol* 17, 613–620 (2020). https://doi.org/10.1038/s41423-020-0400-4

Yi, C., Sun, X., Ye, J. et al. Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies. Cell Mol Immunol 17, 621–630 (2020). https://doi.org/10.1038/s41423-020-0458-z

Shang, J., Ye, G., Shi, K. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* 581, 221–224 (2020). https://doi.org/10.1038/s41586-020-2179-y

Lan, J., Ge, J., Yu, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 215–220 (2020). https://doi.org/10.1038/s41586-020-2180-5

Xia, S., Liu, M., Wang, C. et al. Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. Cell Res 30, 343–355 (2020). https://doi.org/10.1038/s41422-020-0305-x

Poh, C.M., Carissimo, G., Wang, B. et al. Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. Nat Commun 11, 2806 (2020). https://doi.org/10.1038/s41467-020-16638-2

Shuai Lu, Xi-xiu Xie, Lei Zhao, Bin Wang, Jie Zhu, Ting-rui Yang, Guang-wen Yang, Mei Ji, Cui-ping Lv, Jian Xue, Er-hei Dai, Xi-ming Fu, Dong-qun Liu, Lun Zhang, Sheng-jie Hou, Xiao-lin Yu, Yu-ling Wang, Hui-xia Gao, Xue-han Shi, Chang-wen Ke, Bi-xia Ke, Chun-guo Jiang, Rui-tian Liu, The immunodominant and neutralization linear epitopes for SARS-CoV-2, Cell Reports, Volume 34, Issue 4, 2021, https://doi.org/10.1016/j.celrep.2020.108666

Liu, L., Wang, P., Nair, M.S. et al. Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. Nature 584, 450–456 (2020). https://doi.org/10.1038/s41586-020-2571-7

Cai L, Zhang J, Zhu R, Shi W, Xia X, Edwards M, Finch W, Coombs A, Gao J, Chen K, Owen S, Jiang S, Lu W. Protective cellular immunity generated by cross-presenting recombinant overlapping peptide proteins. *Oncotarget.* 2017 Aug 24;8(44):76516-76524. doi: 10.18632/oncotarget.20407

Nelde, A., Bilich, T., Heitmann, J.S. *et al.* SARS-CoV-2-derived peptides define heterologous and COVID-19-induced T cell recognition. *Nat Immunol* 22, 74–85 (2021). https://doi.org/10.1038/s41590-020-00808-x

Jagannathan, P., Wang, T.T. Immunity after SARS-CoV-2 infections. *Nat Immunol* 22, 539–540 (2021). https://doi.org/10.1038/s41590-021-00923-3

CLAIMS

1. A polypeptide for vaccination against and/or treatment of a coronavirus infection comprising two or more peptide fragments comprising sequences derived from coronavirus S protein, and one or more exogenous protease cleavage site sequences located between each of the two or more peptide fragments.

2. The polypeptide of any one of the preceding claims, wherein at least two of the two or more peptide fragments comprise at least one overlapping sequence.

3. The polypeptide of any one of the preceding claims, wherein at least two of the two or more peptide fragments comprise sequences derived from the S1 and/or S2 subunit of the S protein.

4. The polypeptide of any one of the preceding claims, wherein the coronavirus is a betacoronavirus, optionally a severe acute respiratory syndrome-related coronavirus, optionally SARS-CoV-2.

5. The polypeptide of any one of the preceding claims, wherein the coronavirus is a human coronavirus.

6. The polypeptide of any one of the preceding claims, wherein at least one of the two or more peptide fragments comprises a sequence derived from the receptor binding domain (RBD), optionally the receptor binding motif (RBM), of the S1 subunit.

7. The polypeptide of any one of the preceding claims, wherein at least one of the two or more peptide fragments comprises a sequence derived from the HR2 and/or HR1 domain of the S2 subunit.

8. The polypeptide of any one of the preceding claims, wherein the polypeptide comprises three or more peptide fragments each comprising sequences derived from the S protein of the coronavirus, wherein at least two of the three or more peptide fragments comprise sequences derived from the same S protein subunit, and wherein at least two of the three or more peptide fragments comprise at least one overlapping sequence.

9. The polypeptide of claim 8, wherein the three or more peptide fragments comprise at least one peptide fragment comprising a sequence derived from the S1 subunit, and at least one peptide fragment comprising a sequence derived from the S2 subunit.

10. The polypeptide of any one of the preceding claims, wherein at least one peptide fragment comprises one or more antibody epitope(s) of a coronavirus S protein, optionally a SARS-CoV-2 S protein.

11. The polypeptide of claim 10, wherein the antibody epitope is a linear antibody epitope, optionally a neutralising linear antibody epitope.

12. The polypeptide of any one of the preceding claims, wherein the polypeptide, when administered to a subject, stimulates production of anti-coronavirus antibodies in said subject, optionally neutralising antibodies.

13. The polypeptide of any one of the preceding claims, wherein at least one peptide fragment comprises one or more CD4+ T cell epitope(s) and/or one or more CD8+ T cell epitope(s) of a coronavirus S protein, optionally a SARS-CoV-2 S protein.

14. The polypeptide of any one of the preceding claims, wherein the polypeptide, when administered to a subject or cell sample, stimulates T cell responses specific to the coronavirus in said subject or cell sample.

15. The polypeptide of any one of the preceding claims, wherein the polypeptide has five or more peptide fragments, more preferably ten or more peptide fragments.

16. The polypeptide of any one of the preceding claims, wherein the one or more protease cleavage site sequence is a cathepsin protease cleavage sequence, preferably cathepsin S.

17. The polypeptide of any one of the preceding claims, wherein the one or more protease cleavage site sequence is LRMK (SEQ ID NO.: 34).

18. The polypeptide of any one of the preceding claims, wherein the peptide fragments comprise at least one of the following epitopes:

SEQ ID NO: 23; SEQ ID NO: 24; SEQ ID NO: 25; SEQ ID NO: 26; SEQ ID NO: 27; SEQ ID NO: 28; SEQ ID NO: 29; SEQ ID NO: 30; SEQ ID NO: 31; SEQ ID NO: 32; SEQ ID NO: 33.

19. The polypeptide of any one of the preceding claims, wherein at least one peptide fragment comprises a sequence with at least 90% sequence identity to a sequence selected from the group:
SEQ ID NO: 1; SEQ ID NO: 2; SEQ ID NO: 3; SEQ ID NO: 4; SEQ ID NO: 5; SEQ ID NO: 6; SEQ ID NO: 7; SEQ ID NO: 8; SEQ ID NO: 9; SEQ ID NO: 10; SEQ ID NO: 11; SEQ ID NO: 12,
and, optionally, wherein the polypeptide is immunostimulatory.

20. A formulation, optionally a SARS-CoV-2 vaccine formulation, for administration to a subject comprising the polypeptide of any one of the preceding claims.

21. A formulation for administration to a subject comprising the polypeptide of any one of claims 1-19 and the RBD, RBM, HR1 and/or HR2 of a coronavirus, optionally SARS-CoV-2.

22. A polynucleotide encoding the polypeptide of any one of claims 1-19.

23. A composition comprising a delivery vehicle and the polynucleotide of claim 22, optionally further comprising one or more polynucleotide(s) encoding the RBM, RBD, HR1, and/or HR2 of a coronavirus, optionally SARS-CoV-2.

24. The composition of claim 23, additionally comprising a pharmaceutically acceptable carrier.

25. A method for manufacturing a vaccine comprising:
expressing the polynucleotide of claim 22 in one or more cells *in vitro* to produce a polypeptide according to any one of claims 1 to 19; and
isolating the polypeptide.

26. The method of claim 25, additionally comprising the steps of:
expressing a polynucleotide encoding at least the RBM of a coronavirus, optionally SARS-CoV-2, in one or more cells *in vitro* to produce a protein sequence comprising at least the RBM;

isolating the protein sequence; and

combining the polypeptide and the protein sequence into a formulation for administration.

27. A method for vaccination against and/or treatment of a coronavirus infection comprising administering, to a subject in need thereof, the polypeptide of any one of claims 1 to 19.

28. The method of claim 27, additionally comprising administering, to the subject, the RBM, RBD, HR1, and/or HR2 of a coronavirus, optionally SARS-CoV-2.

29. A method for vaccination against and/or treatment of a coronavirus infection comprising administering, to a subject in need thereof, the polynucleotide of claim 20 or the composition of claim 23 or claim 24.

30. The method of any one of claims 27 to 29, wherein the polypeptide or polynucleotide is administered to a human at any dose between 5 $\mu g.kg^{-1}$ and 5 $mg.kg^{-1}$.

31. The method of any one of claims 27 to 30, wherein the coronavirus is a betacoronavirus, optionally a severe acute respiratory syndrome-related coronavirus, further optionally SARS-CoV-2.

32. A method for the diagnosis of a coronavirus infection comprising the steps of:
(a) delivering to a patient the polypeptide of any one of claims 1 to 19; and
(b) detecting T cell stimulation.

33. The method of claim 32 wherein the step of detecting T cell stimulation comprises detecting the production of cytokines.